Responses to the Reviews of "Characterization of Organic Aerosol across the Global Remote Troposphere: A comparison of ATom measurements and global chemistry models"

Anonymous Referee #1

This is a well-written paper. I recommend accepting it, but clarifying as noted below:

We thank the reviewer for interesting suggestions. We have modified the manuscript to address all of his/her concerns.

R1.1) Line 46-47: highest levels measured at what altitudes?

A1.1) The highest levels were measured in the lower troposphere (below 4km). This is now explained in the revised manuscript:

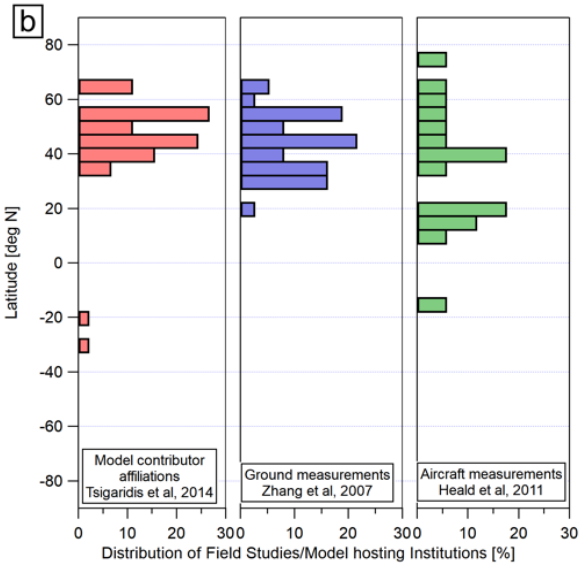
"*OA concentrations have a strong seasonal and zonal variability, with the highest levels measured in the lower troposphere in the summer and over the regions influenced by the biomass burning from Africa (up to 10 ug sm^{-3}).*"

R1.2) Line 75: You might consider adding Zhu et al., 2019 to the list of references here or in line 79. Zhu, J., Penner, J. E., Yu, F., Sillman, S., Andreae, M., and Coe, H., 2019: Organic aerosol nucleation, climate and land use change: Decrease in radiative forcing, Nature Communications, 10, Article No. 423, https://www.nature.com/articles/s41467-019-08407-7.

A1.2) The suggested reference has been added.

R1.3) Fig 1b: lines 1402-1407: Is the distribution shown for the AEROCOM models at the ground or at altitudes sampled by aircraft? What is meant by "distribution of studies" when referring to the models? (explain in caption, please, not just text)

A1.3) We have revised the figure axis label, legends, and figure caption to better describe the data shown on the figure. For the models, what we are showing is the geographical distribution of the institutions hosting/running the GCMs that participated in the AEROCOM-II comparison, which, very much like the measurements, have a very strong bias towards the Northern Hemisphere (NH). While GCMs certainly try to cover the global troposphere, both the bias in constraining measurements and funding sources will lead to more optimization of these models for the mid-latitude NH. The updated figure and caption are shown below:



Caption Fig. 2b: *"(b, right) Geographical distribution of institutions at which the AeroCom-II models were ran/developed (based on the affiliation of all authors) and of the field measurements included in two major literature overview studies (Zhang et al., 2007; Heald et al., 2011) for the OA ground and aircraft AMS as a function of latitude. For the aircraft campaigns, the average latitude for the full deployment was taken."*

R1.4) Line 136: Is there something that distinguishes "ATom models" from other models? Strange terminology.

A1.4) The term "Atom models" is shorthand used to refer to current models (current as of beginning of 2019) that have been ran for the ATom field project as explained in the text:

"*ATom measurements were compared with results of eight global models that simulated the time period of the ATom-1 and 2 campaigns (August 2016 and February 2017), using reanalysis meteorology (and a spin-up time of at least six to twelve months). These are referred hereafter as ATom models [..]*"

We have updated the title of sections 2.1 and 2.2 to read "*ATom model simulations*" and "*AeroCom-II model climatology*".

R1.5) Line 178-179: what fraction of hydrophilic organic material is incorporated into precipitation in GOCART? i.e. what is the Kappa value used in this model?

A1.5) The GOCART model emits 50% of POA in hydrophilic and 50% in hydrophobic mode. The model allows a conversion from a hydrophobic to hydrophilic mode with an e-folding time of 2.5 days. All SOA from biogenic, anthropogenic, and biomass burning sources are treated as hydrophilic. The hydrophilic OA is removed by large-scale and convective warm clouds, while the hydrophobic OA is removed by ice clouds. The hydrophilic particles undergo hygroscopic growth according to the equilibrium parameterization of Gerber (1985). This is now explained in the manuscript as:

"*The primary emitted OC and SOA are separated into hydrophobic (50%) and hydrophilic (50%) species, with a 2.5 days e-folding time conversion from hydrophobic to hydrophilic organic particles. All SOAs from biogenic, anthropogenic, and biomass burning sources are treated as hydrophilic particles. Both types of organic particles are dry deposited. The hydrophilic OA is removed by large-scale and convective warm clouds, while hydrophobic OA is removed by ice clouds. The hydrophilic particles undergo hygroscopic growth according to the equilibrium parameterization of Gerber (1985).*"

Gerber, H. E.: Relative-humidity parameterization of the Navy Aerosol Model (NAM), Tech. Rep. NRL Report 8956, Naval Research Laboratory, 1985.

R1.6) Line 237: add reference for CMIP6 global inventory.

A1.6) The reference has been added:

"*The two simulations with the GEOS-Chem 12.0.1 global chemistry model (Bey et al., 2001) use emissions based on CMIP6 global inventory (Hoesly et al., 2018; Feng et al., 2019) with regional improvements for anthropogenic sources,..*"

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Feng, L., Smith, S. J., Braun, C., Crippa, M., Gidden, M. J., Hoesly, R., Klimont, Z., van Marle, M., van den Berg, M., and van der Werf, G. R.: Gridded Emissions for CMIP6, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-195, in review, 2019.

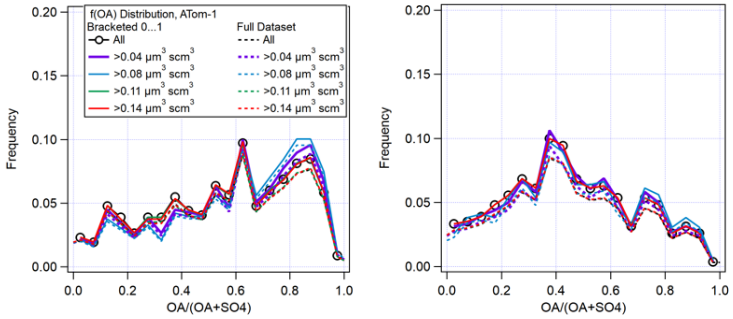
R1.7) Line 426-429: The averaging procedure you used is not clear. If the values are < 3 sigma detection limit, shouldn't you replace the value by zero (so as not to bias the average high)?

A1.7) No! This is an important misconception among some modelers. The data have the correct statistical behavior, i.e. the average of a period of zero concentrations is near zero, as is verified frequently by measuring filtered ambient air in flight. Thus both negative and positive values below DL need to be retained in the dataset. Concentrations cannot be negative, but measurements can be thought of as the sum of concentrations and statistical noise, and can be negative when the real concentrations are zero or very low. A bias would be created if we removed measurements below <0 (or below DL), which we are not doing and generally caution against. This is already explained in the manuscript (L428-430 of the ACPD version), but we have expanded this discussion to make it clearer:

"*Note that a large fraction of the 1-minute OA values in the remote free troposphere were below the local 3σ detection limit. The data of periods of zero concentration (sampling ambient air through a particle filter) do average to zero. Some negative measurements are present, and this is normal for measurements of very low concentrations in the presence of instrumental noise. Averaging of longer periods, as done for the figures in this paper, reduces the detection limit. We therefore caution future data users that the reported data should be averaged as needed, as replacing below-detection limit (or negative) values by other values introduces biases on averages.*"

We have also included an additional figure into the SI that evaluates possible biases in the fractional data by filtering the data based on an independent measurement (the NOAA aerosol volume measurement on ATom, Brock et al 2019) and included some additional discussion in the main text:

"For fractional ratio analysis, measurements were averaged to 5-minute time resolution to reduce the noise in the ratios due to noise in the denominator. The results are not very sensitive to the 5-minute averaging (compared to 1-minute) as shown in Figure S12 for OA to sulfate ratios. The same figure also illustrates that excluding ratios affected by negative concentrations (the non-bracketed case, overall these are about 15% of the dataset) does not really affect the fractional distribution, with the variance between the two cases diminishing as the averaging interval increases. To further confirm that there is no inherent bias in the fractional products regardless of the treatment of low concentration values, an additional sensitivity analysis was performed where data was filtered by an independent measurement proxy for aerosol mass, the aerosol volume measured in ATom (Brock et al, 2019). Using a range of value that encompasses the regime where the AMS calculated volume to aerosol measured volume exhibited increased noise (Guo et al, 2019), no systematic bias was found (Figure S13), with variations of about 10% in fractional volume for different filtering conditions."



Caption Fig. S13*: Exploring the impact of thresholding the 5-min averaged data by a minimum detectable aerosol volume in the PM1 range (from the NOAA SD product, see Guo et al 2019 for details) when computing the the OA/(OA+SO4) distributions in Figure 9.*

R1.8) Line 550-551: Other than the reduction in spread of the AerocomII-sub models compared to full AeroCom II ensemble, this statement is not supported by comparing Fig S2 with Fig 3.

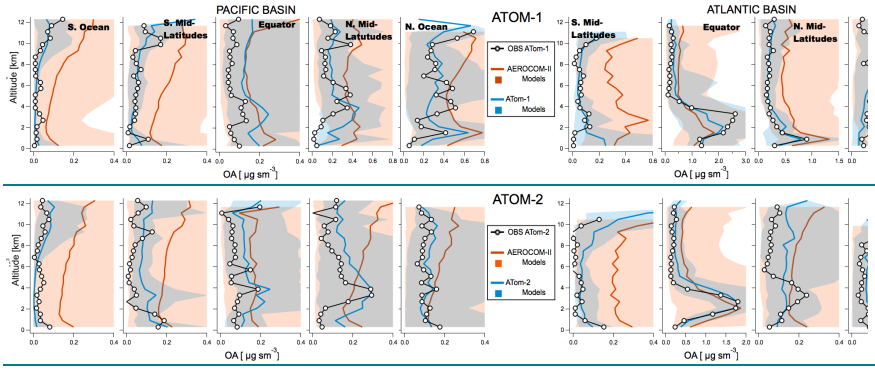
A1.8) We have revised this text to make the point clearer:

"*This reduction in the ensemble spread may partially be explained by a smaller size of the ATom model ensemble (see Fig. S2), which also includes models with a more up-to-date OA representation. In order to explore this point further, results for a subset of AeroCom II models (using earlier versions of models in the ATom ensemble) show only a slight reduction (~10%) in the model spread, with however some regional differences i.e. an improved agreement with observations in the MBL, but an increase in the model bias and spread in the LS (Figure S2). Thus, model improvement for the more recent models appears to be the main reason for the reduced spread."*

R1.9) Line 558-559: you should plot these profiles on a linear scale. It is hard to judge how different the models are using a logarithmic scale.

A1.9) Given that the modeled and observed values span more than one order of magnitude we have used the log scale to visually facilitate the model/obs comparisons but also to allow us to keep the same x-axis span for various regions. We have also added a new supplementary figure (similar to figure 4) using a linear scale in the updated manuscript, as Figure S5 (shown below). This is now explained in the revised manuscript:

"Note that the use of a wide logarithmic scale (to be able to span all the observations) may make the observed differences appear small, although they often reach factors of 2-10 and larger (Figure S5 shows the same results on a linear scale)."

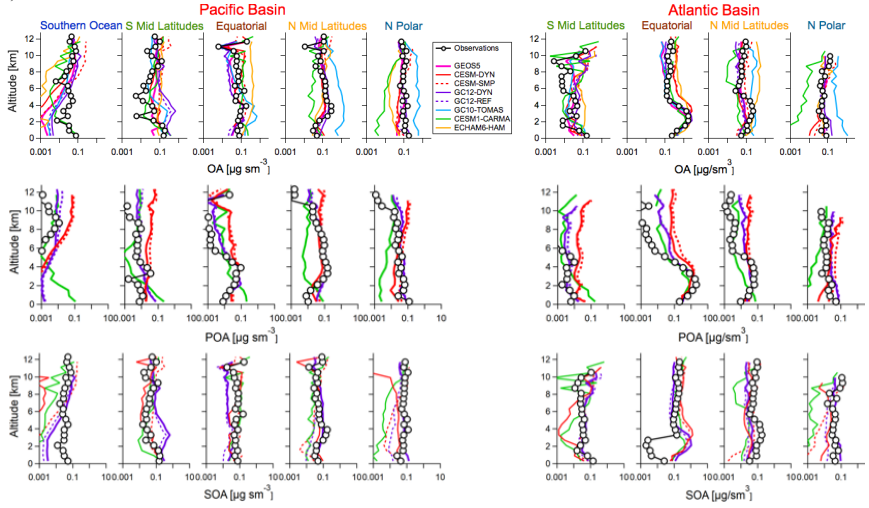


***Caption Fig. S5**: Similar to figure 4 but on a linear scale.*

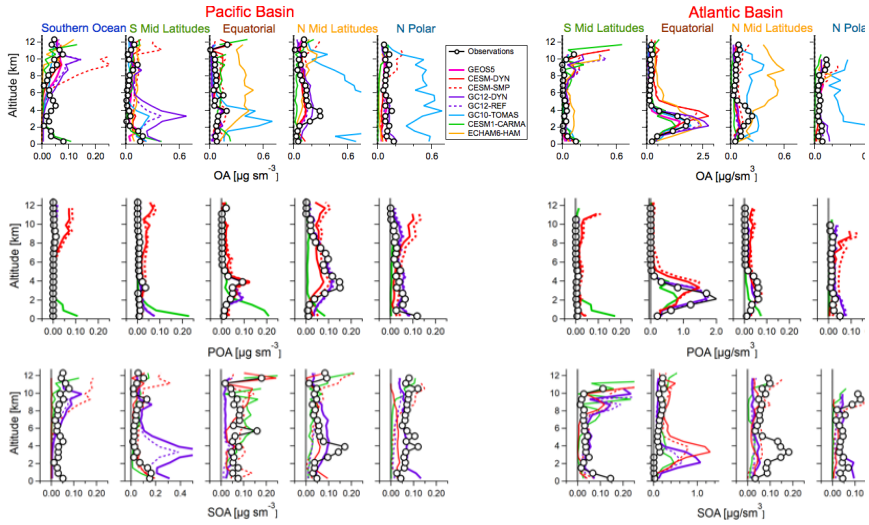
R1.10) Line 587-588: I would reference Fig S6 here, since it is on a linear scale. And you should change S7 to linear scale.

A1.10) We have revised the paper to include both the log (Figure S7a) and linear scale figures (Figure S7b) as part of Figure S7.

a)



b)



Caption Fig. S7*: As Figure S6 for ATom-2 shown both on a logarithmic (a) and linear (b) scales.*

R1.11) Line 766-769: what is meant by POA/OA being shifted rightward? Makes no sense to me.

A1.11) The predicted POA/OA ratio in GC12-REF is overpredicted compared to measurements in Figure 7 which is consistent with the results shown in Figure 8 for GC12-REF that have the right amount of POA and underpredict total OA.

This is now clarified in the manuscript as:

"*It should be noted that these results are consistent with the POA/OA frequency distribution shown in Figure 7 (the POA/OA ratio predicted by GC12-REF is larger than the measured ratio, which is consistent with the fact that POA is about the right amount, and OA is underpredicted in Figure 8*)."

Anonymous Referee #2

General Comments) This is an interesting study that makes comprehensive use of a unique dataset (ATom) to evaluate a series of models. The multi-model approach is particularly valuable for pinpointing model deficiencies in these remote environments. The authors present a thorough series of comparisons, however the conclusions are not well supported. This is primarily due to the reliance on an analysis to separate POA from SOA in the measurements which is not very well justified. More work is needed to expand this analysis (see below for suggestions), or remove it and alter the text accordingly, before the manuscript would be acceptable for publication.

We thank the reviewer for valuable suggestions. We hope that we have addressed all the concerns in a satisfactory manner. In particular, we have improved the POA analysis and associated discussions in the revised manuscript. Additional simulations have been performed with the GEOS-Chem model to document the sensitivity of our results to the simulated non-volatile vs. semi-volatile properties of POA.

R2.1) Figure 1: This figure is unclear and not sufficiently discussed in the main text. What does "distribution of studies" used as the x-axis of Figure 1b mean? – a more exact definition of what is plotted should be provided. In addition, the quantitative discussion of these AeroCom results in the abstract is unclear (line 37) – what does "factor of 400-1000" imply – that the spread of the means is of this range? This could more clearly be given as a percentage of the mean or median model, or as phrased in lines 100-103 as "model dispersion" in orders of magnitude. The manuscript does not fully discuss what is shown in Figure 1b.

A2.1) We already addressed the points about Figure 1b in A1.3 since the first reviewer had a similar comment, and refer reviewer 2 to the changes discussed there.

The factor of 400-1000 refers to the results in Figure 1a, which are described in L103-107 of the ACPD version, as well as in the caption of Figure 1. We have reworded the main text for clarity as:

"Our own analyses of the AeroCom-II results shown in Figure 1a indicate that model dispersion (quantified as the ratio of the average concentration of the highest model to that of the lowest one, in each region) increases not only with altitude but also with distance from the northern mid-latitude source (and data-rich) regions. The model spread is a factor of 10-20 in the free troposphere between the equator and northern mid-latitudes, and increases to a factor of 200-800 over the Southern Ocean and near the tropopause."

The caption of Figure 1a has been revised to read:

"Figure 1: (a, left) The ratio between the average OA concentrations of the highest to the lowest models (for each region) as predicted among 28 global chemistry transport models participating in the AeroCom phase II intercomparison study (Tsigaridis et al. 2014)."

And the abstract has been revised to read:

"OA predictions from AeroCom Phase II global models span two to three orders-of-magnitude"

R2.2) Section 2.1 would benefit from a bit more discussion of the methodology in selecting these models and the differences in their configurations. Are they all standard configurations (i.e. as downloaded), including emissions used, if not why were different parameters chosen? The level of detail in the description of the various models is quite uneven –the authors should ensure that the same information is provided for all models. Finally, are simulations performed and sampled to match the spatial location of the ATom aircraft (with emissions and meteorology matched to the year of the measurements)?

A2.2) The considered models span a range of complexity in terms of aerosols parameterizations, and some of the models have several OA schemes or aerosol modules (like CESM or GEOS-Chem). For each model we have referenced the publication that describes the baseline configuration, and the modifications that have been used in the runs included here. In the revised manuscript, we have made clear when a standard configuration is being used e.g. for the GEOS-Chem GC12-REF configuration:

"*Note that this GEOS-Chem REF simulation is similar to the version 12 default "complex option" which includes non-volatile POA and semi-volatile SOA (semi-volatile POA is an optional switch within this version used in Pai et al. 2020).*"

Pai, S. J., Heald, C. L., Pierce, J. R., Farina, S. C., Marais, E. A., Jimenez, J. L., Campuzano-Jost, P., Nault, B. A., Middlebrook, A. M., Coe, H., Shilling, J. E., Bahreini, R., Dingle, J. H., and Vu, K.: An evaluation of global organic aerosol schemes using airborne observations, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-331, in review, 2019.

We have also provided more details for some models and made sure that the description includes information on the emissions, aerosol module (composition, size representation), OA formation and removals.

The following description has been added for ECHAM6-HAM:

"*Aerosol particles are removed by dry and wet deposition. The wet deposition includes the below cloud scavenging by rain and in-cloud cloud scavenging for large-scale and convective systems (Croft et al., 2010).*"

Croft, B., Lohmann, U., Martin, R. V., Stier, P., Wurzler, S., Feichter, J., Hoose, C., Heikkilä, U., van Donkelaar, A., and Ferrachat, S.: Influences of in-cloud aerosol scavenging parameterizations on aerosol concentrations and wet deposition in ECHAM5-HAM, Atmos. Chem. Phys., 10, 1511–1543, https://doi.org/10.5194/acp-10-1511-2010, 2010.

Removal has been better described for GEOS-Chem:

"*The removal of gases and aerosols are treated similar to the GEOS-Chem 12.0.1 model (GC12-REF, see above).*"

The following was added for CESM2:

"*Simulations based on the CESM2.0 Earth system model use the standard version of the Whole Atmosphere Community Climate Model (WACCM6, Gettelman et al., 2019, Emmons et al., 2019).*"

ATom model simulations were performed with the emissions and meteorology matching the year of the measurements. This is now better explained in the manuscript:

"*ATom measurements were compared with results of eight global models that simulated the time period of the ATom-1 and 2 campaigns (August 2016 and February 2017), using the emissions and reanalysis meteorology corresponding to this period (and a spin-up time of at least six to twelve months).*"

In addition, a column has been added to Table 1 specifying the meteorological reanalysis used for each model.

R2.3) The manuscript is missing any discussion of the role of POA treatment in these comparisons. It's not 100% clear from Section 2.1 (e.g. no info provided on POA for ECHAM-HAM, GC10-TOMAS, or any of the CESM configurations), but it appears that all of these simulations use non-volatile POA. A number of modeling studies have implemented a semi-volatile treatment of POA since Robinson et al. (2007). It seems like a major weakness to draw general conclusions on OA model performance when using a

series of models which do not represent the semi-volatile nature of POA. It would be nice to see the authors add such a simulation to their suite, but if this proves impractical at this stage of the work, the manuscript should be altered considerably to acknowledge the gaps in the POA treatment and how this may have a substantial impact on the comparisons and conclusions drawn here. The lack of discussion of the simulation (and emissions) of POA also somewhat undercuts the discussion of Section 4.3. It's clear that models are underestimating the observed OM:OC during ATom, but if the models are over-estimating the POA to begin with (perhaps because it's all assumed to be non-volatile?) then this could be a compensating bias related solely to how POA is treated.

A2.3) As suggested by the reviewer we have clarified in the revised manuscript that models used in this study only include non-volatile POA parameterizations. Please see section 2.1:

"*In all models POA is treated as a non-volatile directly emitted species. In most models (see below) the primary emitted organic aerosol is artificially aged to transition between hydrophobic to hydrophilic POA.*"

This non-volatile treatment of POA in the models is consistent with the way the estimated POA has been derived from the ATom measurements. Indeed, the estimated POA is calculated from the POA/BC ratios representative of the ambient air values close to the emission sources, after most evaporation has occurred, but before substantial chemistry of POA has taken place. As a consequence, the estimated POA can be approximately considered to be non-volatile. As discussed in response A2.24, the model and measurement emission ratios are not significantly different. Therefore the comparison with the non-volatile POA representation from models is more appropriate than a comparison with a semi-volatile POA representation. This is now more clearly explained in the manuscript in section 4.4:

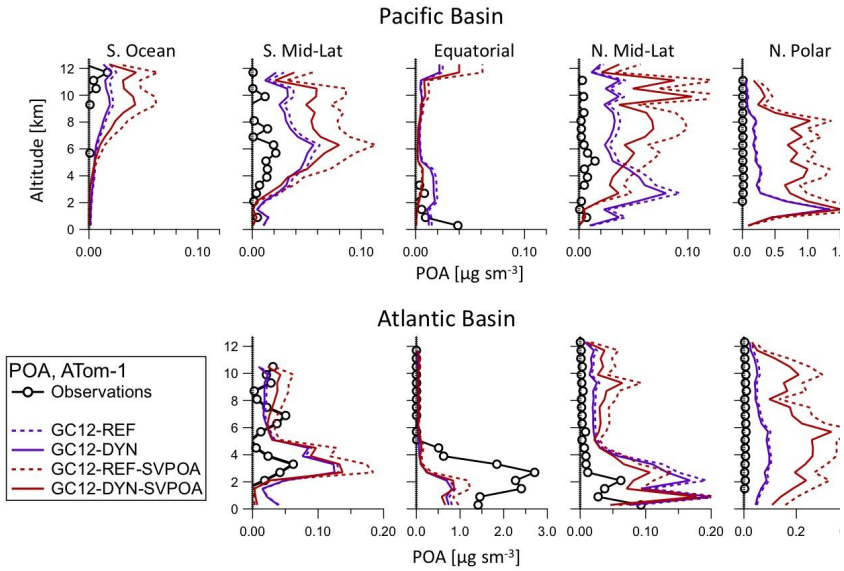
"*POA concentrations were estimated from the BC measurements by using an emission ratio appropriate to the airmass origin (biomass burning vs. anthropogenic), and using the f(BB) mass fraction from the PALMS single particle instrument (see Section 3.2). By using the POA/BC ratio at the source regions after most evaporation, but before POA chemical degradation has taken place, we implicitly assume POA to be chemically inert, while in reality it can slowly be lost to the gas-phase by heterogeneous chemistry (e.g. George and Abbatt, 2010; Palm et al., 2018). Thus, the observation-based method provides an upper limit to the fraction of POA. The model/measurement comparison is only shown for the CESM and GEOS-Chem model variants, as other participating models do not separate or did not report their POA and SOA fractions. In all simulations, POA was treated as a chemically inert directly emitted primary aerosol species that only undergoes transport, transformation from hydrophobic to hydrophilic state with ageing (1-2 days typically),*

coagulation, and dry and wet deposition. Importantly, the treatment of POA as non-volatile (rather than semi-volatile) in models is fully consistent with the assumptions for POA estimation from the measurements."

And in the conclusion:

"*The non-volatile POA treatment in models is consistent with the assumption of inert POA particles used to estimate POA from the measurements, and cannot explain the model bias. Indeed, sensitivity simulations with semi-volatile POA lead to a much larger model bias for OA in the upper troposphere and remote regions.* "

In addition, we have performed sensitivity simulations to estimate the effect of the non-volatile vs. semi-volatile POA assumption in the models on POA predictions. We have performed an additional simulation (GC12-REF-SVPOA) for ATom-1 based on GC12-REF, in which the non-volatile treatment of POA has been replaced by the semi-volatile POA parameterization based on Pye and Seinfeld, 2010 and using a two-product reversible partitioning model. This is a similar model configuration as used in Pai et al. (2020) under "the complex scheme" (though different emissions were used between their study and here). The comparison of POA vertical profiles between GC12-REF (non-volatile) and GC12-REF-SVPOA (semi-volatile) over various regions is shown in the figure below. The comparison indicates that the POA concentrations are larger in most regions when the semi-volatile POA parameterization is used.



Caption Fig. S16*: Sensitivity simulations to estimate the importance of the non-volatile vs. semi-volatile POA treatment in GEOS-Chem. The semi-volatile POA in GC12-REF-SVPOA (GC12-DYN-SVPOA) model configuration should be directly compared with the corresponding GC12-REF (GC12-DYN) non-volatile POA.*

This is now discussed in the revised manuscript:

"*Finally, we have examined whether the non-volatile treatment of POA in models could lead to these unrealistically high POA fractions in the remote regions. Figure S16 shows a comparison of POA vertical profiles as predicted by the GC12-REF simulations that use non-volatile POA and a sensitivity simulation GC12-REF-SVPOA that uses semi-volatile POA similar to the standard treatment in GEOS-Chem as described in Pai et al. (2020). Note, however, that Pai et al. (2020) included marine POA emissions, used different reanalysis meteorology, and a different model version (12.1.1 rather than 12.0.1 here), so their resulting comparisons to ATom measurements are somewhat different than found here for GC12-REF-SVPOA. The comparison indicates that the POA concentrations increase substantially in most regions when the semi-volatile POA parameterization is used. These results suggest that non-volatile treatment of POA is not responsible of the model bias.*"

R2.4) Some information on model configurations is missing that would be important for comparing model performance (could potentially be added to Table 1): what is the assumed OM:OC ratio, what are the global emission totals for key precursors (isoprene, monoterpenes, POA, etc.)?

A2.4) Information on OA/OC ratios was already provided in the ACPD manuscript. Please see the description: "*OA/OC of 1.4 is used in ECHAM6-HAM, whereas 1.8 is used in GEOS5 and GC10-TOMAS simulations for both POA and SOA. Other models calculated directly SOA concentrations without applying this conversion (CESM1-CARMA, CESM2-SMP, CESM2-DYN, GC12-REF and GC12-DYN), but for POA used the ratio of 1.8 (CESM1-CARMA, CESM2-DYN) and 2.1 (GC12-REF and GC12-DYN). Most of the AeroCom-II models used the ratio of 1.4 for all primary and secondary OA (Tsigaridis et al., 2014).*" This information is also shown again in Figure 5 when comparing with the measurements.

As suggested by the reviewer we have added the OA/OC ratios also to Table 1.

We do not have the total amount of precursors saved for all models, so that information has not been added. However, we reference the emission inventories that are used for each model.

R2.5) The estimation of the POA fraction in Section 3.2 is not well supported. First, the manuscript is missing a discussion of the uncertainty on the fBB from PALMS (lines 340-342). Second, the numbers in Table S1 do not support the averages used in the text, for example EFs for urban sources range over an order of magnitude (0.16-15.4) and the authors appear to have simply averaged these values, which seems highlight inappropriate. The example provided by the authors of using a single ratio of BB from Andraea (2019) leading to a POA fraction of > 100% in African plumes also illustrates the inappropriate application of a single number. EFs range significantly with fuel type, combustion conditions, and location; use of any single value is likely to lead to uncertainties that would vastly outweigh the value of the analysis. A more appropriate approach might be to take a lower limit set of EFs and an upper limit set of EFs, and bracket the POA estimation using first one and then the other. Absent such an analysis, this POA estimate seems unreliable and the results of Section 4 are highly questionable. The analysis of Figure S9 seems to go in this direction, but the range in EFs in this Figure do not represent the full range of values shown in Table S1. Given that all the conclusions in Section 4.4 hinge on this analysis, perhaps the authors could expand this discussion: describe the range in fBB values, and then the calculated POA contributions (from FF and BB separately) estimated for all the ATom data.

In order to explore the uncertainties in their methodology, the authors could also apply the same analysis to the model output of [BC] and assumed EFs (use first the same EFs as used in the measurement analysis and then the EFs used in the model) to see how an estimated POA_model would compare to the simulated POA. This could pinpoint whether flaws in methodology for estimating POA or flaws in the model simulation of POA dominate.

A2.5) The range quoted by the reviewer for urban sources is not correct. The ratio of 15.4 is for rural agricultural biomass burning, not for urban sources. In addition, we only used in our average the ratios for mixed urban air, as discussed in response A2.3, while the ratios for emission sources (e.g. individual cars) were only shown to support their consistency with the mixed urban air ratios. We have clarified Table S1 (shown below) to make clear which values are used in our averages (marked now in **bold**) and which are presented only for reference, and which apply to urban vs. BB sources (shown now in *italic*). In reality, the range of measured ratios for urban pollution is 0.5-2.4, and the uncertainty due to this effect is minor. In fact Figure S9 in the ACPD version already showed a sensitivity study with the urban ratio varying between 0.5 and 3, and showed that the effect of this ratio on the plots is minuscule, especially when compared to the model-measurement disagreement.

We have now clarified in the text how the averages for urban sources were calculated:

"Based on Table S1 data, we assume POA to be co-emitted with BC for anthropogenic fossil fuel / urban region POA (herein called FF_{ratio} for simplicity, even though much of it is non-fossil, Zotter et al., 2014; Hayes et al., 2015) at a ratio of 1.5 (average of all urban ambient air studies that report POA and BC for best intercomparability to the ATom dataset; including all urban studies results in a very similar number, 1.48)."

Furthermore, upon revisiting Andreae (2019) review for these responses, we noticed that using an OA/OC ratio of 1.8 for his data as we have done for all other studies compiled in Table S1 was incorrect, since he based his review on a universal value of OA/OC of 1.6 for biomass burning sources (see Section 2.1 in that review), which results in a small correction to the BB_{ratio} to 11.8 (instead of 13.5). Hence we have updated Figure 7 as well as Figure S8-S10 (all shown below) to reflect this change (which has minimal impact on f(POA)), and have modified the text in the manuscript accordingly:

"For biomass burning sources, we use a value of POA/BC = 11.8 (BB_{ratio}), based on the average of the recent review by Andreae (2019), which included over 200 previous determinations for a variety of fuels and burning conditions (since Andreae (2019) used and OA/OC ratio of 1.6 in his work, we have used that value to calculate POA/BC; we note that this is different from the 1.8 OA/OC ratio used for other studies listed in Table S1)."

We have also slightly revised the range of FF_{ratio} and BB_{ratio} that we consider in the sensitivity analysis shown in Figure S9. We cover a range of 0.5-2.4 for FF_{ratio}, consistent with the discussion above (and add one additional scenario). For the range of BB_{ratio} , we are using the lower and upper uncertainty ranges (in both OC and EC emissions) from Andreae (2019) for the major contributors to global BB (2-60), which also covers all the suggested averages for the individual sources (except peat) as well as the range of BB emissions used in the models (Table S2). Aerosol emissions from peat are a clear outlier, but their global contribution is small (about 5%) and, as a recent analysis shows, the peat sources with very large BB_{ratio}, are very localized (Watson et al, 2019), so they mostly contribute during the height of the South East Asian Fire season (September to October, Reddington et al, 2014), hence outside the sampling period for ATom-1 and 2.

Watson, J. G., Cao, J., Chen, L. W. A., Wang, Q., Tian, J., Wang, X., Gronstal, S., Ho, S. S. H., Watts, A. C. and Chow, J. C.: Gaseous, PM2.5 Mass, and Speciated Emission Factors from Laboratory Chamber Peat Combustion, Atmos. Chem. Phys. Discuss., 1–39, doi:10.5194/acp-2019-456, 2019.

It should be clear now that the sensitivity study in Figure S9 *does represent the full range of the literature emission ratios* shown in Table S1. And that illustrates the robustness of the POA results: even with the most extreme assumptions for the emission ratios, the POA/OA distribution changes little. The key is that BC is very low in most of the remote troposphere, and thus there are no realistic ratios of POA/BC that could possibly produce POA concentrations similar to those in most models. We have added Figure S11a to the SI (shown below) to illustrate the skewness of the BC/OA distribution.

Regarding the uncertainty in $f(BB)_{PALMS}$, while it should be clear that any uncertainty in this factor will have only a limited impact on f(POA), we have conducted an extra sensitivity study with the uncertainty estimated by the PALMS team (+/-5%), and have added Figure S11b to the SI.

We have also revised Figure S10, which explores the impact of very low OA values on the f(POA) distribution. In addition to showing the sensitivity of f(POA) to the choice of averaging interval (which reduces the percentage of points below detection limit) we also explore the impact of capping POA to OA (e.g. not allowing the estimated POA to be larger than OA). This new analysis shows that not capping POA results in very similar f(POA) profiles, with the exception of f(POA)=1. The 10-20% fractions calculated for the standard, capped case are actually a combination of data close to sources where POA estimated from the measurements was indeed larger than OA (and which in Figure S10 would show up at values >1) and cases where BC and hence POA was close to zero (BC<0.1 ng sm^{-3}) but OA was negative due to noise. As expected, this effect is somewhat less apparent at longer

averaging times (and more apparent for ATom-2, where there was a higher fraction of very low OA values). Since the non-capped case underestimates f(POA)=1, by not including the data close to sources, using the capped data is clearly better. However, due to the limitations in our ability to estimate POA when both BC and OA are very low our analysis likely overrepresents the amount of POA found in ATom. We have modified the discussion in Section 4.4 to reflect this:

"The differences are so large that they are pretty insensitive to details of the POA estimation method from the measurements, mostly because for the vast majority of the ATom track BC/OA ratios were extremely low and hence the exact magnitude of the multiplicative factor is secondary to the estimation of POA (Figure S11). As Figure S9 illustrates, the choice of FFratio has very little impact on the overall distribution of POA. On the other hand, while the BBratio does impact the overall distribution of POA, it mostly affects the points in the vicinity of the large Atlantic plumes. Since the POA/BC ratio in those plumes is fairly low, (see Section 3.2), using a very large BBratio mostly leads to an increase of the fraction of the points where POA > 100%. While the large range of published BB_{ratio} for different sources precludes a more accurate estimation by our method, for the purposes of the comparison with the model results we emphasize that even using the largest BB_{ratio}, f(SOA) is still significantly larger in the ATom dataset that in any of the models.

Additional sensitivity tests were performed to investigate the impact of noisy data and uncertainties of f(BB) on the estimation of POA. Figure S11 clearly shows that the impact of a misattribution of the aerosol type by the stated PALMS uncertainty (Froyd et al, 2019) is completely negligible. Figure S10 details how the choice of averaging interval (with longer averaging times reducing both the fraction of OA measurements under the DL and below zero) impact the distribution of POA. Overall, no large changes are observed for averaging times >5 min, and hence a 5 min averaging interval was used for the analysis in Figure 7. Figure S10 also illustrates how capping the histogram impacts the POA distribution. To capture the most realistic f(POA) distribution, the data in Fig 7 was capped at the extremes (so $f_{(POA)}<0$ is taken as $f_{(POA)}=0$, and $f_{(POA)}>1$ is taken as $f_{(POA)}=1$). As Fig S10 shows, data with $f_{(POA)}<0$ is almost exclusively due to very small (and always positive, since BC cannot go negative) POA values being divided by small, negative noise in total OA, and hence treating that fraction of the histogram as essentially $f_{(POA)}\sim0$ is justified. On the other end of the distribution, data where POA is larger than OA is mostly due to our average BB_{ratio} being larger than the one encountered in most of the BB plumes in ATom. Choosing a lower BB_{ratio}, as Fig S9b and S9d illustrate, leads to f(POA)>1 basically trending to zero, confirming our interpretation. This is a limitation of the dataset, and it does not seem appropriate to remove these points, since some fraction are likely dominated by POA. However, it shows that the POA estimation, especially for this part of the distribution likely overstates the importance of POA."

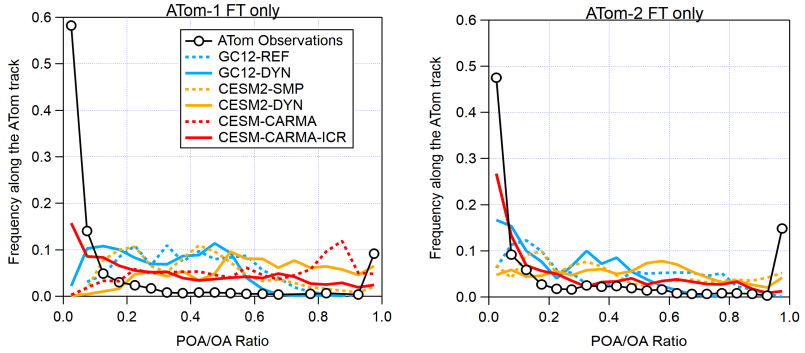
It is not possible to apply the measurement methodology to model outputs, as none of the models track separately BC from various emission sectors. In any case, we have now clearly shown that the measurement-based estimates are very robust against a wide range of assumptions.

Table S1: POA/BC ratios determined in previous field and laboratory emission studies. Studies that reported well constrained urban non-BB POA based on AMS PMF determinations (highlighted in bold) were averaged to determine the value used for $(POA/BC)_{anthro}$ *. Studies that reported* $(POA/BC)_{BB}$ *are shown in italics. For the average of* $(POA/BC)_{BB}$ *the weighted average reported by Andreae, 2019 was used.*

Source	**Technique**	**Type of emissions**	**POA/BC ratio (OA measured)**	**POA/BC ratio (OC measured, OA/OC of 1.8 used)**
Zhang et al. 2005	AMS PCA for POA EC from TOCA	Urban background	**1.41**	
Szidat et al. 2006	14C source apportionment for EC and OC	Urban mobile sources Residential burning		2.65 *11.3*
Ban-Weiss et al. 2008	OC: Filters (TOA) Aethalometer and filters for BC	Mobile sources: Light Duty Vehicles Diesel		 2.5 1.3
Aiken et al. 2009	AMS PMF for POA, SP2 for BC	Urban background	**0.8**	
Christian et al. 2010	TOT EC/OC analyzer	Cooking Stoves Trash Burning Brick Klinn Charcoal Klinn AG Burn		*6.3* *7.75* *0.27* *78* *200*
Chirico et al. 2010	AMS PMF for POA SP2 for BC	Tailpipe emissions, gas vehicle	0.16-0.3	
Minguillon et al. 2011	14C source apportionment for EC and OC, combined with AMS PMF	Urban backg. Rural backg. Biomass burning	 *15.4*	1.7 4
Huang et al. 2013	AMS PMF for POA, SP2 for BC	Urban backg. winter Urban backg. summer	**0.82** **1.27**	

Hayes et al. 2013	AMS PMF for POA, SP2 for BC	Urban background	**1.82** (average) 1.51 (more diesel influenced)	
Crippa et al. 2013	AMS PMF for POA, Aethalometer for BC	Urban mobile sources Cooking aerosol Residential burning	**0.5 (ave)** **0.5 (ave)** 3.4 (ave)	
Huang et al. 2015	Offline AMS and TOT OC/EC analyzer, ME2 analysis	Traffic Cooking BB	**0.5** **2.5** 11	
Zhang et al. 2015	14C source apportionment for EC and OC	Fossil fuel, coal burning Residential burning		1.6 *8.5*
Hu et al. 2016	AMS PMF for POA, SP2 for BC	Urban Background	**1.4**	
Kim et al. 2018	AMS PMF for POA, SP2 for BC	Urban background (70% HOA, 30% COA)	**2.2**	
Whatore et al. 2017	TOT EC/OC analyzer	African traditional stoves		*4.8*
Nault et al. 2018	AMS PMF for POA, SP2 for BC	Urban background	**2.38**	
Chen et al. 2018	AMS PMF for POA, SP2 for BC	BB urban BB rural	*6.25* *5*	
Chirico et al. 2011	AMS OA SP2 for BC	Tunnel mobile emissions	0.4	
Kim et al. 2017	AMS PMF for POA, SP2 for BC	Total urban POA (40% BB, 27% HOA, 33% COA)	3.2	
Andreae, 2019	Review (OA/OC of 1.6 used per the methodology of the review)	Savanna Tropical forest Temperate forest Boreal forest		*9.1* *13.8* *31.7* *22*

		Peat	*227*
		AG	*18.7*
		Dung	*9.9*
		Biofuel	*52.6*
		Charcoal	*13*
		Average (this work)	***11.8***



Caption Fig. 7: *Frequency distribution of observed and simulated ratio of POA to total OA in the free troposphere during ATom-1 and ATom-2 as computed by the GC12-, CESM2-, and CESM1-CARMA models.*

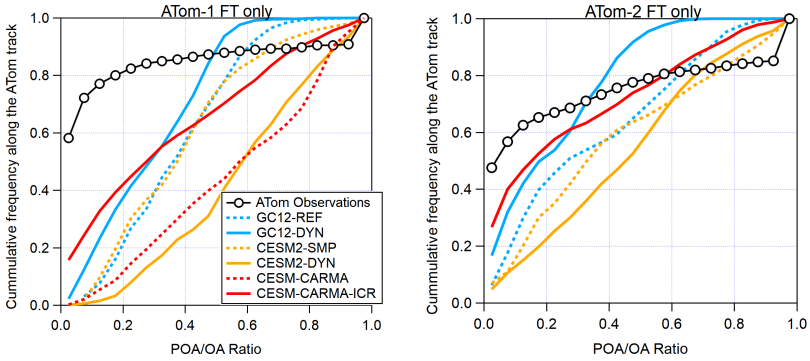
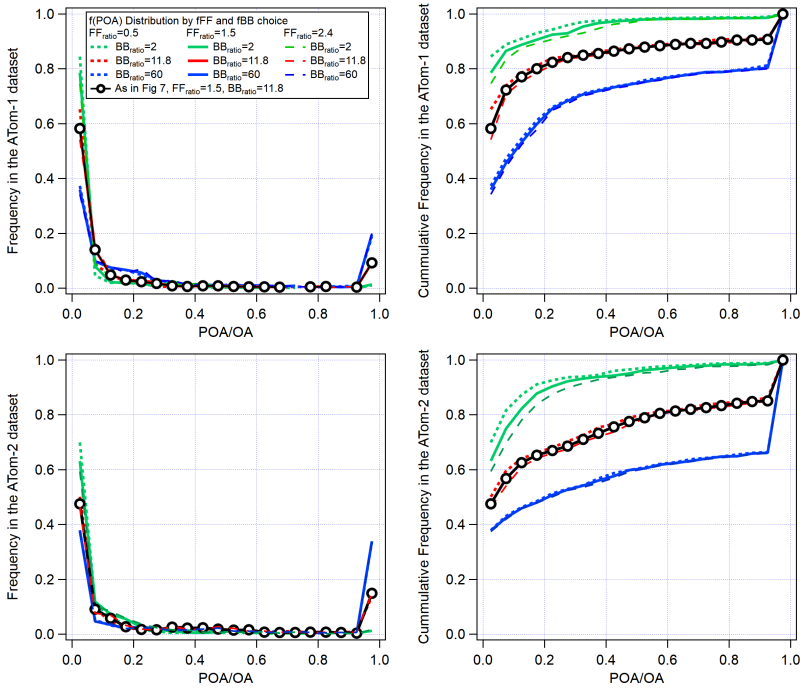


Figure S8: POA/OA distributions (free troposphere only) from Figure 7 shown as cumulative distributions (CDF). Note that for the OA/BC ratios observed for ATom specifically, the green curves in Fig S9b and S9d (BB_{ratio}=2) are closer to the real distribution.

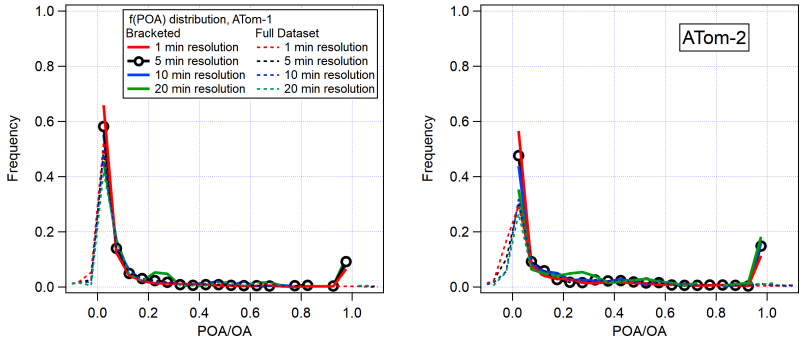


598

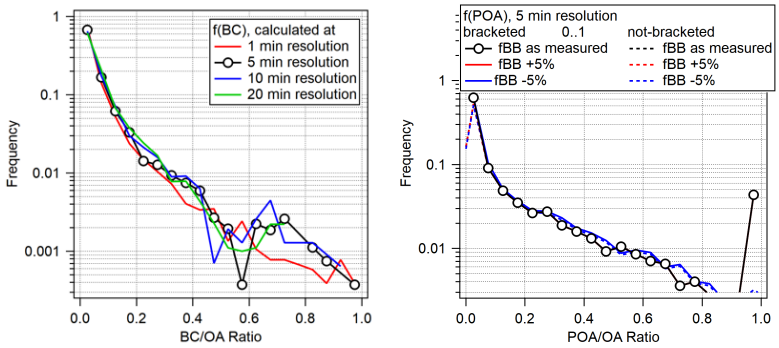
599

600

601 ***Caption Fig. S9:*** *Sensitivity of the overall measured POA/OA distribution to different*
602 *estimates of POA/BC ratios for both urban and BB sources covering the range of values*
603 *shown in Table S1 and S2, both for the frequency and cumulative frequency distribution*
604 *(left/right) and ATom-1 and 2 (top/bottom). Note that for the choice of BBratio ranges, we*
605 *used the range (within uncertainties) for the main global BB contributors and excluded*
606 *one clear outlier, peat. This is justified since peat is a small source, mostly localized to SE*
607 *Asia, and the main emissions of peat BB aerosol are outside the sampling periods of ATom-*
608 *1 and 2 (Reddington et al, 2014).*



Caption Fig. S10*: Exploring the impact of OA data below detection limit (DL) by increasing the averaging interval on the POA/OA distributions in Figure 7 for ATom 1 and 2 (a 5 min averaging interval was used throughout the analysis discussed in Section 4.4). Also shown is the comparison of a capped (so $f_{(POA)}=0$ includes $f_{(POA)}<0$, and $f_{(POA)}=1$ includes $f_{(POA)}>1$) vs. an unconstrained histogram, for the same set of averaging intervals. In the manuscript, 5-minute averaging (capped) is used*



Caption Fig. S11: *(left) Distribution of BC/OA ratios that are used as the basis of the estimation of $f_{(POA)}$ for all ATom deployments, shown using different averaging intervals (right) Effect of the 5% uncertainty in the f(BB) reported by the PALMS instrument on the estimation of $f_{(POA)}$, using both bracketed and not bracketed data (cf. Figure S10).*

Added SI reference:

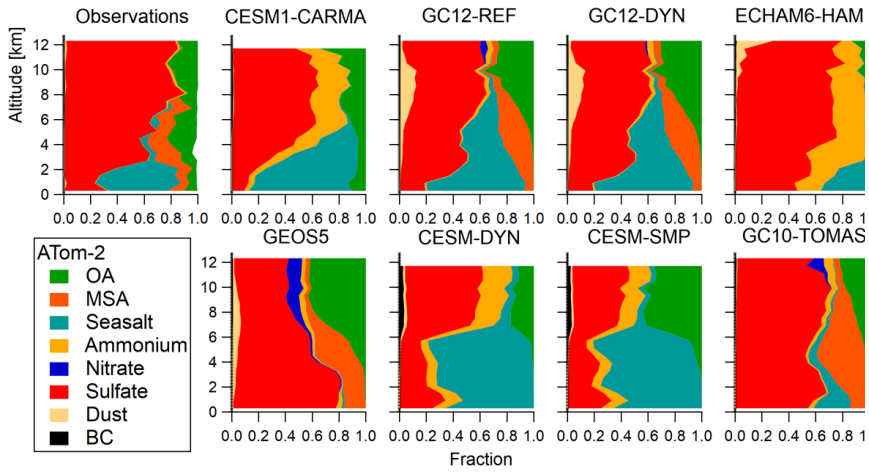
Reddington, C. L., Yoshioka, M., Balasubramanian, R., Ridley, D., Toh, Y. Y., Arnold, S. R. and Spracklen, D. V.: Contribution of vegetation and peat fires to particulate air pollution in Southeast Asia, Environ. Res. Lett., 9(9), doi:10.1088/1748-9326/9/9/094006, 2014.

R2.6) Lines 813-821: Figure 10 seems interesting, but it feels like an aside. The details of how these models treat inorganics (including nitrate, ammonium, sea salt, and dust) and the relevant emissions, which would be necessary to understand these differences are not included in the manuscript. Thus, the authors should either eliminate this text in favor of a more focused discussion of the OA results (as suggested in point #5 above), or substantially enhance the model description section to include the relevant details.

A2.6) While this paper focuses on OA, it is still of interest to document the relative importance of OA and other species, and how these vary substantially across different models. Several papers from our groups and others have been published that address some of those components, and others are in preparation. We believe it is still of broad interest to keep this figure to provide context for the OA results. We have added further explanations to the text with suitable references for ATom analyses and modeling of the other chemical components:

"The discrepancies between the observed and predicted composition of submicron aerosol over remote regions can be quite large for other constituents as well. Figure 10 shows the comparison of measured and predicted composition of the submicron aerosol over the Southern Ocean (during the NH winter) where the disagreement in simulated sea salt, nitrates, ammonium, and MSA often exceeds the contribution of OA. While the observations show a more uniform distribution of non-marine aerosol with higher values in the mid and upper troposphere, respectively, most models tend to simulate highest fractions of OA (and sulfate) towards the tropopause. This may also be explained by the uncertainties in modeled wet removal of aerosol that has been discussed above. Specific studies have discussed and continue to investigate the ATom measurements and simulations of different components in more detail, including black carbon (Katich et al., 2018; Ditas et al., 2019), MSA (Hodshire et al., 2019), sulfate-nitrate-ammonium (Nault et al., 2019), and sea salt (Yu et al, 2019; Bian et al., 2019; Murphy et al., 2019)."

For consistency with the treatment in Figure 2b, we have also included both the modeled and measured submicron dust to Figure 10. The measurements only reflect the low end of the dust distribution (< 500 nm), and do not fully match the size range of the model-reported submicron dust (as shown in Table 1). Hence it is expected that observations will have lower dust concentrations than the models.



We have updated the Figure caption to read:

Caption Fig. 10: *"[..] Note that while the modeled and measured submicron sea-salt size ranges agree fairly well (Table 1), this is not quite the case for dust. Given that the accumulation mode dust in the models presented contains larger sizes than the AMS range (< 500 nm), it is expected for the modeled dust concentration to be larger than measured."*

Bian, H., et al. (2019), Observationally constrained analysis of sea salt aerosol in the marine atmosphere 3, Atmos. Chem. Phys., doi:10.5194/acp-2019-18.

Ditas, J., et al. (2018), Strong impact of wildfires on the abundance and aging of black carbon in the lowermost stratosphere, Proc. Natl. Acad. Sci., 811595-11603, doi:10.1073/pnas.1806868115.

Hodshire, A., et al. (2019), The potential role of methanesulfonic acid (MSA) in aerosol formation and growth and the associated radiative forcings, Atmos. Chem. Phys., 19, 3137-3160, doi:10.5194/acp-19-3137-2019.

Katich, J., et al. (2018), Strong Contrast in Remote Black Carbon Aerosol Loadings Between the Atlantic and Pacific Basins, J. Geophys. Res., 123, 13,386-13,395, doi:10.1029/2018JD029206.

Nault, B., et al. (2019), Global Observations of Ammonium Balance and pH Indicate More Acidic Conditions and More Liquid Aerosols than Current Models Predict, Abstract A52C-08, presented at 2019 Fall Meeting, AGU, San Francisco, CA, 9-13 Dec.

Murphy, D. M., Froyd, K. D., Bian, H., Brock, C. A., Dibb, J. E., DiGangi, J. P., Diskin, G., Dollner, M., Kupc, A., Scheuer, E. M., Schill, G. P., Weinzierl, B., Williamson, C. J., and Yu, P.: The distribution of sea-salt aerosol in the global troposphere, Atmos. Chem. Phys., 19, 4093-4104, https://doi.org/10.5194/acp-19-4093-2019, 2019.

Yu, P., Froyd, K. D., Portmann, R. W., Toon, O. B., Freitas, S. R., Bardeen, C. G., Brock, C., Fan, T., Gao, R.-S., Katich, J. M., Kupc, A., Liu, S., Maloney, C., Murphy, D. M., Rosenlof, K. H., Schill, G., Schwarz, J. P. and Williamson, C.: Efficient In-Cloud Removal of Aerosols by Deep Convection, Geophys. Res. Lett., 46(2), 1061–1069, doi:10.1029/2018GL080544, 2019.

Minor Comments

R2.7) The mixed capitalization in the title is a bit odd.

A2.7) The mixed capitalization has been removed: "*Characterization of organic aerosol across the global remote troposphere: A comparison of ATom measurements and global chemistry models*".

The mixed capitalization for the mission name (ATom, Atmospheric Tomography mission) is in accordance with the official mission acronym and description: https://espo.nasa.gov/atom

R2.8) Line 69: The authors might consider rephrasing. The word "major" implies a larger role in RF than OC contributes in the AR5 assessment cited (i.e. GHG dominate the RF, and even amongst aerosols, the effect of OC is considerably less than the inorganics or BC).

A2.8) We agree with the reviewer. The sentence has been changed to read:

"*They are associated with adverse health effects (Mauderly and Chow, 2008, Shiraiwa et al., 2017) and contribute radiative forcing in the climate system (Boucher et al., 2013).*"

R2.9) Line 92: Hodzic et al (2016) do not use the "same field campaigns" – rather they use a subset of those previously analyzed by Heald et al. (2011) with some additional campaigns.

A2.9) We agree with the reviewer's comment, and have updated the text to read:

"*For a subset of 9 recent aircraft campaigns, Hodzic et al. (2016) showed that OA is likely a more dynamic system than represented in chemistry-climate models, with both stronger production and stronger removals.*"

R2.10) Lines 87-91: Pai et al., ACPD, 2019 provides a more recent evaluation of the standard GEOS-Chem model configurations (including comparisons with ATom) which should be discussed here and perhaps elsewhere in the manuscript, particularly as they do not see the same bias away from source that was highlighted in previous studies (Heald et al., 2011; Hodzic et al., 2016).

A2.10) This paper has not yet been accepted as of this writing, so we refrained from discussing it in detail, based on previous guidance from journal editors and reviewers. Now that it is accepted, we have added a reference in the revised paper, see response to R2.3.

R2.11) Table 1: Why are dust and seasalt sizes included here, and why are they listed in the sub-micron only? Dust and sea salt go well into the 10's of um in model simulations.

A2.11) We agree that sea-salt and dust are mostly present in the coarse mode, but they do have a tail in the submicron mode. It is their contribution to submicron aerosols only that is included in Figures 2 and 10, to provide a complete representation of all the chemical components present in submicron particles. Figure 10 in the submitted manuscript did not include dust, as explained in the response A2.6 we have added it in the revised version. We have also adjusted the caption accordingly, as documented in that response.

R2.12) Lines 198-199 and 756-757: Marais et al. (2016) replace the isoprene VBS with their mechanism for isoprene SOA. Please clarify whether isoprene SOA in your simulations follows this or whether it includes both that from the VBS of Pye et al. (2010) as well as that produced using the mechanism of Marais et al. (2016), which might lead to double-counting of isoprene SOA.

A2.12) For isoprene, there is no double-counting as the VBS has been replaced by the parameterization from Marais et al. (2016).

This is now more clearly explained:

"*The first configuration (called hereafter GC12-REF) includes the default (http://wiki.seas.harvard.edu/geos-chem/index.php) representation of SOA formation based on Marais et al. (2016) for isoprene-derived SOA, and on the volatility basis set (VBS) of Pye et al. (2010) for all other precursors.*"

This is also clarified lines 756-757:

"*It should be noted that in both cases, isoprene-SOA is formed in aqueous aerosols following Marais et al. (2016).*"

R2.13) Line 201: does "with the exception of the treatment of isoprene SOA" imply that photolytic removal does not apply to isoprene SOA in GC12-DYN?

A2.13) That is correct. This has been clarified in the revised manuscript: "*As in Hodzic et al. (2016) the GC12-DYN model version includes updated VBS SOA parameterization, updated dry and wet removal of organic vapors, and photolytic removal of SOA (except for isoprene-SOA).*"

R2.14) Line 249-250: does this imply that CESM2-DYN uses the same SOA yield parameters, photolytic loss, and updated Henry's law constants as GC12-DYN? If not, please clarify which differ.

A2.14) The treatment is similar in both models for the most part following the parameterization of Hodzic et al., 2016, at the exception of i) the isoprene-SOA formation (GC12-DYN used Marais et al., 2016); ii) the low-NOx yields (in CESM2-DYN only low-NOx yields are used). This is now more clearly explained in the manuscript:

"*This is a similar SOA scheme as used in GC12-DYN (with differences in the treatment of isoprene-SOA based on Marais et al. 2016 in GC12-DYN, and the use of both low- and high-NOx VBS yields in GC12-DYN).*"

R2.15) Line 275: "in a climatological way" is not defined here. Suggest remove as the later text describes how the model is sampled.

A2.15) We have removed this text as suggested by the reviewer.

R2.16) Figure S1 should be included in the main text given that it shows a central comparison of ATom-2 with the models.

A2.16) We respectfully disagree. Figure S1 shows that the trends discussed for ATom-1 hold for ATom-2 as well.

R2.17) Section 3.1: The measurement description section should include the detection limits and uncertainties on the AMS data during ATom and how this might impact the comparisons. I noted that some of this is given in lines 415-423, but it seems like this belongs earlier in the measurement description section, or at least that the authors could refer the reader to this later discussion in their manuscript, so that they know it will be addressed.

A2.17) These items are discussed at length in the references provided (Schroeder et al, 2018; Nault et al, 2018; Jimenez et al., 2019), but we agree with the reviewer that a brief summary and referral to Section 3.3 would improve the readability of the manuscript. Hence we have added the following to Section 3.1:

"AMS data was acquired at 1 Hz time resolution and independently processed and reported at both 1 s and 60 s time resolutions (Jimenez et al., 2019a). The later product, with more robust peak fitting at low concentrations was exclusively used as the primary dataset in this work. Detection limits at different time resolutions/geographical bins relevant to this study are discussed in Section 3.3. The overall 2σ accuracies of the AMS measurement (38% for OA, 34% for sulfate and other inorganics) are discussed in Bahreini et al. (2008) and Jimenez et al. (2019b)."

Bahreini, R., Ervens, B., Middlebrook, A. M., Warneke, C., de Gouw, J. A., DeCarlo, P. F., Jimenez, J. L., Brock, C. A., Neuman, J. A., Ryerson, T. B., Stark, H., Atlas, E., Brioude, J., Fried, A., Holloway, J. S., Peischl, J., Richter, D., Walega, J., Weibring, P., Wollny, A. G. and Fehsenfeld, F. C.: Organic aerosol formation in urban and industrial plumes near Houston and Dallas, Texas, J. Geophys. Res., 114, D00F16, doi:10.1029/2008JD011493, 2009.

Jimenez, J.L., P. Campuzano-Jost, D.A. Day, B.A. Nault, D.J. Price, and J.C. Schroder. ATom: L2 Measurements from CU High-Resolution Aerosol Mass Spectrometer (HR-AMS). ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1716, 2019a.

Jimenez, J.L., et al.: Evaluating the Consistency of All Submicron Aerosol Mass Measurements (Total and Speciated) in the Atmospheric Tomography Mission (ATom), Abstract A31A-08, presented at 2019 Fall Meeting, AGU, San Francisco, CA, 9-13 Dec., 2019b.

R2.18) Line 329: could the authors be more explicit? Does this imply that biomass burning OA from Africa is larger in size than typical?

A2.18) Not at all, this just refers to the fact that a linear regression is quite sensitive to high points, and that on average the African BB plumes have 10x higher concentrations than the data outside of them. As discussed in Brock et al. (2019) and Jimenezet al. (2019), for the measurements there is no systematic bias apparent in the comparisons with the particle sizing instruments in this range.

For the models discussed, both GC10-TOMAS and CESM1-CARMA do show about 15%-20% contribution of coarse aerosols contribution to OA in the BB plumes, and removing those improves the correlation for GC10-TOMAS (0.97). This is not the case for the standard version of CESM1-CARMA, since without the convective fix it also shows a substantial contribution of large aerosols in the UT.

We have modified the text to clarify this:

"(Slopes for ATom-1 linear regressions: CESM-1CARMA:0.91, GC10-TOMAS: 0.94, ECHAM6-HAM 1.00) mostly influenced by the high concentration points in the biomass plumes off Africa that have a large effect on the linear regressions, since they are about 10 times larger than the bulk of the dataset)"

R2.19) Lines 330-339: what is the size range of the aerosols detected by the PALMS instrument?

A2.19) The PALMS instrument reports mass products in the range 100-3000 nm geometric based on the NOAA size distribution data (Brock et al, 2019, Froyd et al, 2019). All the PALMS data included in this work has been computed to match the size range of the AMS (so D_{aero} 40...1250 nm, see Knote et al., 2011, and Jimenez et al., 2019) using the measured density. Hence the PALMS data reported here is consistent with the AMS data, with the possible exception of (less frequent) particle growth events in the upper troposphere where a significant mass fraction is below the optical detection limit of the PALMS (roughly 100-150 nm D_{geo}, see Froyd et al, 2019). The text in the manuscript has been modified to explain this more clearly:

"For all PALMS data used in this work (biomass burning fraction and dust) the AMS transmission function was applied to ensure that both instruments were characterizing approximately the same particle size range."

R2.20) Line 339: unclear. Why is the AMS transmission function applied to the PALMS data?

A2.20) See the response to the previous comment (A2.19).

R2.21) Line 343: Given that fBB from PALMS is a derived quantity and not a direct measurement the statement that the PALMS fBB "is more useful as a particle tracer" is a bit bold and requires a citation. Or the language should be softened to "may be more useful".

A2.21) We have revised this text to further explain what we meant and why this is the best choice for our analyses, with the available dataset. Note that the next sentence in the manuscript (L343-345 in the ACPD version, not referred to by the reviewer, but very important for this choice) provides an additional, and likely more important reason for the usefulness of this parameter for our purposes. The revised text reads:

"*This parameter correlates quite well with other gas-phase BB tracers, and is more useful as a particle tracer since its lifetime follows that of the particles. Importantly, it is not impacted by the long lifetimes of the gas-phase tracers (e.g. 9 months for CH_3CN) and unrelated removal processes (e.g. ocean uptake for CH_3CN and HCN) that result in highly variable backgrounds. Hence $f(BB)_{PALMS}$ has a much higher contrast ratio and linearity for particle BB impacts, compared to the available gas-phase tracers in the ATom dataset. An airmass was classified as non-BB influenced when $f(BB)_{PALMS}$ was lower than 0.30 (Hudson et al, 2004) as shown in Figure 2b. $f(BB)_{PALMS}$ was also used to assess the impact of POA on the total OA burden (next section); note that no thresholding was applied in that case.*"

R2.22) Lines 346-349: These sentences seem to conflate primary and biomass burning, which are not necessarily the same thing. If the implication is that the analysis assumes no SOA from biomass burning (such as suggested by Hodshire et al., 2019), that assumption should be stated explicitly here.

A2.22) The order of these 3 sentences was confusing and we have reorder them to first explain how we separate BB and non-BB airmasses, and then how we calculate the POA fraction in OA. See the revised manuscript:

"*An airmass was classified as non-BB influenced when $f(BB)_{PALMS}$ was lower than 0.30 (Hudson et al, 2004) as shown in Figure 2b. For both ATom-1 and 2, about 76% of measurements were classified as not influenced by biomass burning. $f(BB)_{PALMS}$ was also used to assess the impact of POA on the total OA burden (next section); note that no thresholding was applied in that case.*"

Furthermore, $f(BB)_{PALMS}$ is an important variable on that estimation process, as explained in the next section. At this point in the text, this has no bearing on the SOA formation ability of BB sources.

Later, in the next section, we do state (L375-377): *"We note the measured total OA/BC of ~3.5 (conservatively assuming that all OA is POA) observed on both ATom missions for the large African sourced BB plumes over the Equatorial Atlantic"* Indeed we are assuming here that **on those strong African BB plumes measured near the source region** all OA is POA. However, there is no explicit assumption applied that all BB OA is POA in our POA estimation method. Depending on the plumes encountered in the global atmosphere and their OA/BC ratio, some of their OA can be classified as SOA by our method.

The key point is that, since **the main result is that POA is surprisingly low compared to models, we are trying to make conservative assumptions that maximize POA**. In this way the measurement-based estimate cannot be criticized as being biased low and providing too low POA.

R2.23) Line 382: what are the units on the POA? In units of carbon or was an OM:OC applied?

A2.23) We only use OA (Organic Aerosol), POA (Primary Organic Aerosol) and SOA (Secondary Organic Aerosol) in units of μg sm^{-3}, as stated in Section 3.1, in this manuscript. By definition these include carbon and any other elements that are part of the organic molecules constituting OA. As described in Section 4 and Table 1, some of the older models still use OC, but this is a less useful metric (mostly left over from a time in which only OC could be measured) that we have tried to avoid for the discussion of concentrations in this work.

Importantly, as described above, to derive the FF_{ratio} from Table S1 we have relied exclusively on studies that actually report OA, and not OC, since the uncertainty in those determinations is substantially larger and also less applicable to the instrument payload on ATom. We have modified the text to clarify this point:

"Based on Table S1 data, we assume POA to be co-emitted with BC for anthropogenic fossil fuel / urban region POA (herein called FF_{ratio} for simplicity, even though much of it is non-fossil, Zotter et al., 2014; Hayes et al., 2015) at a ratio of 1.55 (average of all studies that report POA and BC)"

R2.24) Line 391: EFs range orders of magnitude and these ranges in both the model and measurements are being compared. It's not clear that "no significant bias is apparent", they

could easily differ by a factor of two on average – perhaps the authors rather mean something like "the ranges in values are consistent".

A2.24) We have calculated the averages for measurements (3.5-4 for residential, and 1-1.8 for traffic) and the emission inventories (4.6-5.9 for residential, and 1.1-1.4 for traffic) to confirm that they are similar. We have reworded this text to address this comment as:

"The averages and ranges of the measurement and model ratios are similar, and thus no significant model bias on the ratios is apparent."

R2.25) Section 3.3: The authors have focused on the means for their model-measurement comparisons. It would be useful to examine whether this is an appropriate metric – are the distributions skewed? Can the models capture the shape of the distribution? Might a comparison of medians in Section 4 provide different results?

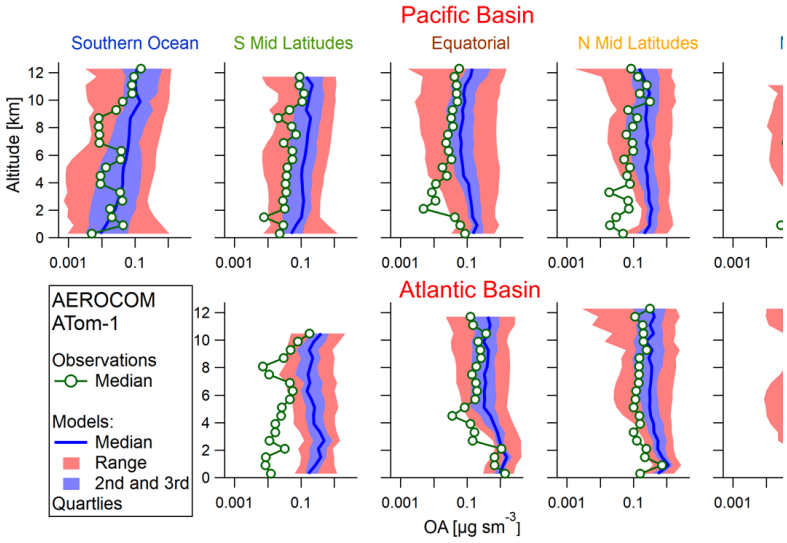
A2.25) For the model-measurement comparisons of the AeroCom-II and ATom model ensembles, we have compared both the box plots for various regions using medians (Figure 3) and the vertical profiles using the means of each ensemble (Figure 4). The results of those analyses are consistent, and show a factor of 2-3 overestimation by the AeroCom-II model ensemble of the measured OA in various regions.

The plot below shows the comparison of medians for the observed and predicted OA in various regions for (a) the AeroCom-II model ensemble and for (b) the ATom model ensemble. These plots are to be compared with Figure 5 in the manuscript that showed OA mean concentrations. Here again, the comparison suggests that using medians results in a slightly lower values for all datasets (as expected), but does not change the conclusions of the model-measurement comparisons. We have added the plots below to the SI as Figure S18.

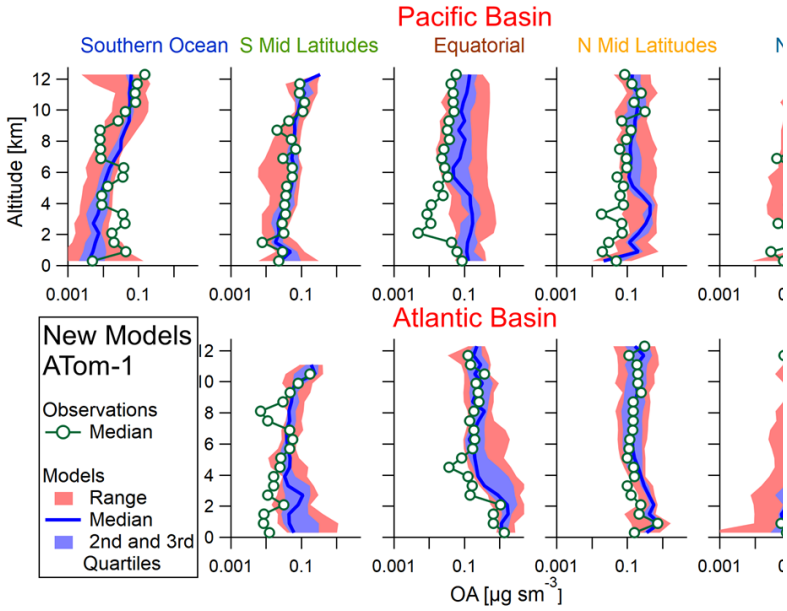
This is now explained in the revised manuscript:

"*We note that using the ensemble median OA profiles instead of ensemble mean OA profiles (as shown in Figure 5 and S7) results in a slightly lower values of OA but does not change the conclusions of the model-measurement comparisons. (Figure S18).*"

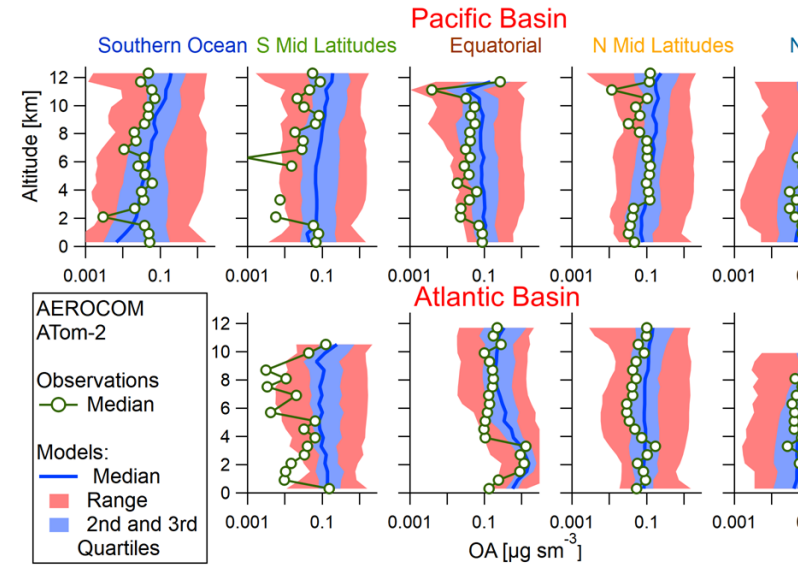
For the evaluation of the individual models the statistics are shown in Table 2. As suggested by the reviewer we have in addition compared distribution plots of OA mass concentrations for the observations and various models.


Pacific Basin
Southern Ocean
S Mid Latitudes
Equatorial
N Mid Latitudes
Altitude [km]
Atlantic Basin
AEROCOM
ATom-1
Observations
Median
Models:
Median
Range
2nd and 3rd
Quartlies
OA [µg sm-3]

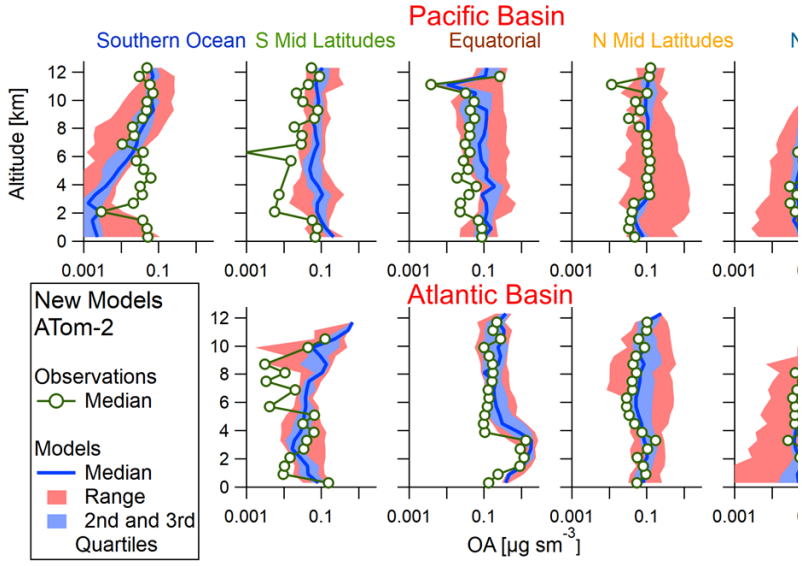
a)


Pacific Basin
Southern Ocean
S Mid Latitudes
Equatorial
N Mid Latitudes
Altitude [km]
Atlantic Basin
New Models
ATom-1
Observations
Median
Models
Range
Median
2nd and 3rd
Quartiles
OA [µg sm-3]

b)



c)

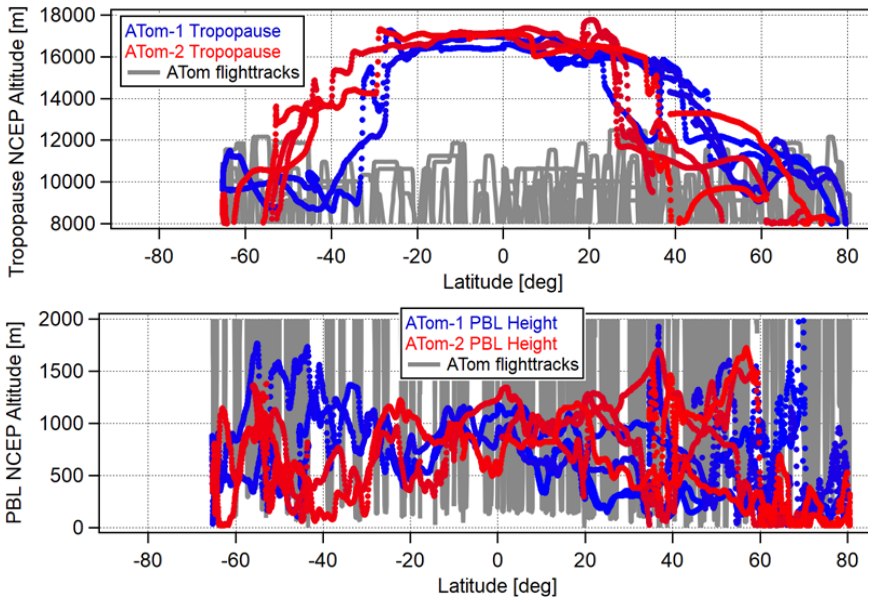


d)

Caption Fig. S18*: Comparison of OA median vertical profiles as measured during ATom-1 and predicted by the (a) AeroCom-II model ensemble and (b) ATom model ensemble . Panels (c) and (d) show the same for ATom-2, respectively (similar to figure 5 in the paper that compares OA average profiles).*

R2.26) Lines 435-436: could you provide the range of MBL heights and tropopause heights along the flight tracks used here?

A2.26) We have added the NCEP reanalysis values of the PBL and tropopause heights to the Supp. Info. as Figure S17. GEOS5 values are very similar:



Caption Fig. S17: *(top) tropopause heights from the NCEP reanalysis at each Lat/Long flown for ATom-1 (blue) and ATom-2 (red). (bottom) Planetary boundary layer (PBL) heights obtained in the same way. Values from the GEOS-5 model are very similar to these. ATom-1 and 2 flight tracks are included in grey for context.*

The DC-8 ceiling is about 13 km (42000 ft, in practice 39000-41000 ft was the maximum altitude for most flights), which means that we only sampled the stratosphere at latitudes higher than 30 degrees. Based on these data, we modified the manuscript to document this as:

"The tropopause height varied during ATom between 8 and 16.5 km; given the DC-8 ceiling (42 kft, 12.8 km) the stratosphere was only sampled at latitudes higher than 30 degrees in both hemispheres. The MBL height varied between up to 1.5 km in the mid-latitudes, ~1 km in the tropics, and sometimes <150 m (lowest DC-8 altitude) for some of the sampling in the polar troposphere."

R2.27) Figure 2b: please provide either total number of points or percentages of total dataset for the categories should here.

A2.27) As requested by the reviewer we have calculated the percentages of data in each category. This information has been added to the caption of Figure 2b:

"In ATom-1, BB-only represents 24% of the data, clean MBL 8%, clean FT 57% and clean UT 12%, whereas in ATom-2 BB-only represents 24%, clean MBL 9%, clean FT 53%, clean UT 15%."

R2.28) Line 472: suggest inserting the word "likely" to "less polluted than ATom-1, likely due to a" since you haven't definitely compared emissions or source contributions.

A2.28) We have modified the text as requested.

R2.29) Line 497-498: The statement "It should be noted.." is surprising. The authors haven't shown any analysis for this and Figures 2a and S1 clearly show elevated OA in the North Pacific which seems likely associated with Asian source. Could the authors explain?

A2.29) This may have been unclear as originally written. We were trying to say that we did not see large extended plumes. But we do agree that the elevated OA in the North Pacific is likely associated with the Asian outflow. To clarify this point, we have modified the sentence and referenced the corresponding figures in the revised manuscript:

"*It should be noted that Asian pollution was likely an important contributor to the North Pacific Basin, especially between 2 and 6 km, in both ATom deployments (see figures 2a and S1).*"

R2.30) Line 537: The NMB in Table 2 for CESM1-CARMA is given as -33.2%, so the -20-30% range seems incorrect.

A2.30) The range has been corrected. See the updated text in the response A2.31 below.

R2.31) Table 2 indicates that the NMB for all models is positive for ATom-2. I didn't see this surprising result discussed in the text.

A2.31) See response in A2.32.

R2.32) Line 539-541: This statement is incorrect as it only appears to apply to CESM. According to Table 2, while GC12-DYN is slightly less biased than GC12-REF for ATom1, the reverse is true for ATom-2.

A2.32) To address reviewer's comments (2.31 and 2.32) we have separated the discussion into ATom-1 and ATom-2 (NH summer and NH winter) periods. The revised manuscript has been updated to read:

"*During the NH summer (ATom-1), models using the VBS parameterization from Pye et al. (2010) tend to underpredict the OA concentrations by 43% for GC12-REF and 33% for CESM1-CARMA for ATom-1, most likely due to the excessive evaporation of the formed SOA in remote regions and low yields for anthropogenic SOA (Schroder et al., 2018; Shah et al., 2019). Models using the VBS parameterization from Hodzic et al. (2016) (CESM2-DYN and GC12-DYN) where OA is less volatile and also OA yields are corrected for wall losses show an improved agreement with observations especially for CESM2-DYN (with NMB of ~5%), and to a lesser extent for GC12-DYN (NMB of ~33%). During the NH winter (ATom-2) characterized by a lower production of SOA, both VBS approaches lead to an overestimation of the predicted OA. This is likely caused by excessively high levels of primary emitted OA as discussed in section 4.4.*"

R2.33) Section 4.4: The POA to OA ratio is derived, not directly measured (e.g. line 685).

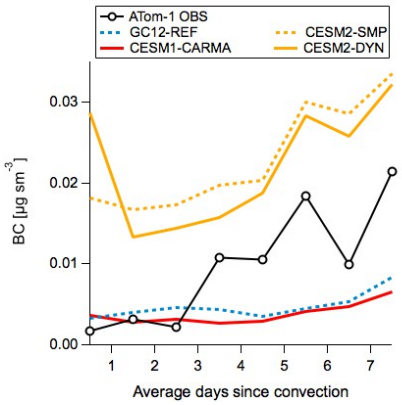
A2.33) We believe that this is already very clear to a reader of this section, since the method is described in detail. However, to reduce possible confusion we have changed the text at this location to read:

"Most models fail to reproduce the overwhelming dominance of SOA that is inferred from the measurements during ATom-1, while the discrepancies are less severe during NH winter (ATom-2)."

R2.34) Lines 726-728, and 737-738: given these statements why does Figure 8 not include a comparison of BC with and without in-cloud removal?

A2.34) The reason the BC was not shown is because we do not have CESM1-CARMA results for BC for both simulations with and without in-cloud removal improvements.

Figure for BC is shown below and includes only CESM1-CARMA simulations with in-cloud removal improvements (as described in Table 1 of the manuscript). Thus, this figure has not been included in the main section of the paper, but we have added it to the SI (Figure S15).



Caption Fig. S15: *Measured and predicted BC concentrations during ATom-1 as a function of the number of days since the air mass was processed through convection.*

Anonymous Referee #3

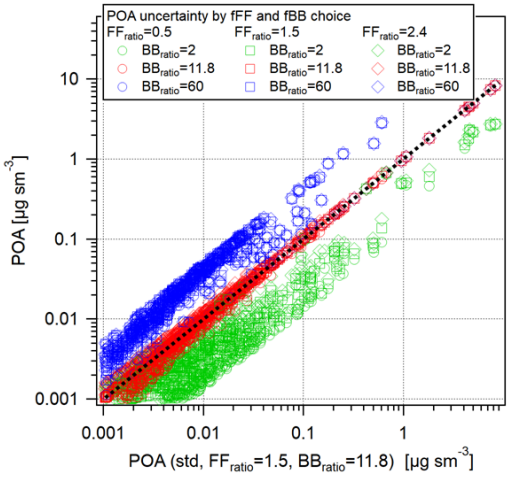
Hodzic presents a comparison of a large set of global models and observations from the free troposphere with a focus on organic aerosol. I applaud the authors on this large undertaking in terms of number of models and synthesis of observations. The manuscript contains a large amount of information. My major comments are regarding the POA estimation and better illustrating the utility of DYN configuration (which may just be a bit buried).

We thank the reviewer for the encouraging evaluation. We have modified the manuscript to address his/her concerns on the POA and DYN model configurations and other topics.

Major comments:

R3.1). Section 3.2 and POA estimation: Can the uncertainty in the POA estimates be quantified, ideally with error bars (e.g. in Fig 6 vertical profiles or Fig 8 POA)? Some ideas to consider regarding the POA estimation:

The uncertainty on the POA estimation is dominated by the choice of the biomass burning POA/BC emission ratio. We have documented both in the SI and in A2.5 at length that except for the pure BB points the uncertainty in the distribution of f(POA) is rather small (and for the purer BB points it's biased high toward more POA). Regarding OA, the following figure shows the uncertainty range from Figure S9 in concentration space:



So in absolute terms, the uncertainty in the estimation of total POA is about 20. But this does barely affect the actual fractions shown in Figure 6, since only a very small number of points actually contributes to the larger concentrations. For Figure 8 the uncertainty is likely larger, but will mostly affect the absolute numbers, not the trends with convection, which is the point of that figure.

R3.1-a) There are anthropogenic POA sources that do not have significant amounts of BC. See, for example, Figure 2 of Reff et al. (https://pubs.acs.org/doi/10.1021/es802930x) which indicates large emissions of dust associated with anthropogenic activity (road dust, construction dust) that have high OC relative to BC. While fossil fuel-BC may be a general proxy for anthropogenic activity, is the ratio 1.55 reflective of that general behavior? What does the EPA NEI indicate as FFratio including all anthropogenic PM sources (e.g. Fig 2 of Reff et al.)?

A3.1-a) We thank the reviewer for providing this additional reference. We have included values for the EPA NEI inventory in the supplementary materials (see Table S2). The associated ratios for the traffic and residential combustion sources in the NEI inventory (for traffic = 1.8, and residential = 8.2) are similar to those reported for other inventories and already discussed in the paper (section 3.2). Our ratios do not include emissions associated with fugitive dust from road, tire and construction, which is typically found in larger particles (D_{aero} > 1 μm, Zhao et al., 2017).

This is now explained in the manuscript:

"*It should be noted that urban model ratios do not include emissions associated with fugitive dust from road, tire and construction, as those are typically found in larger particles than those studied here (Zhao et al., 2017).*"

Zhao, G., Chen, Y., Hopke, P.K., Holsen, T.M., Dhaniyala, S.: Characteristics of traffic-induced fugitive dust from unpaved roads, Aerosol Science and Technology, 51:11, 1324-1331, DOI: 10.1080/02786826.2017.1347251, 2017.

Also see response to R3.1-d. Because we are determining the ratios from urban ambient air measurements, sources that do not emit BC (but do contribute on the submicron range) are also implicitly included.

R3.1-b) Can you plot data (obs or model) as a function of f(BB) influenced to get an intercept at f(BB)=0 and f(BB)=1 for comparison to FFratio and BBratio specified in text?

A3.1-b) Below are the requested quantile plots. On the left we are showing the ratio of the mean OA to mean BC quantiles, while on the right we are using medians. As already shown in Figure S11a and emphasized in A2.5, it highlights for the vast majority of cases along the ATom track, BC was a very small contributor to particulate carbon. And it also shows, as discussed in the manuscript, that the measured OA/BC in the BB plumes encountered during ATom were actually lower than the global average BB_{ratio} we use in our estimation. These plots have been added to the paper as Fig. S19.

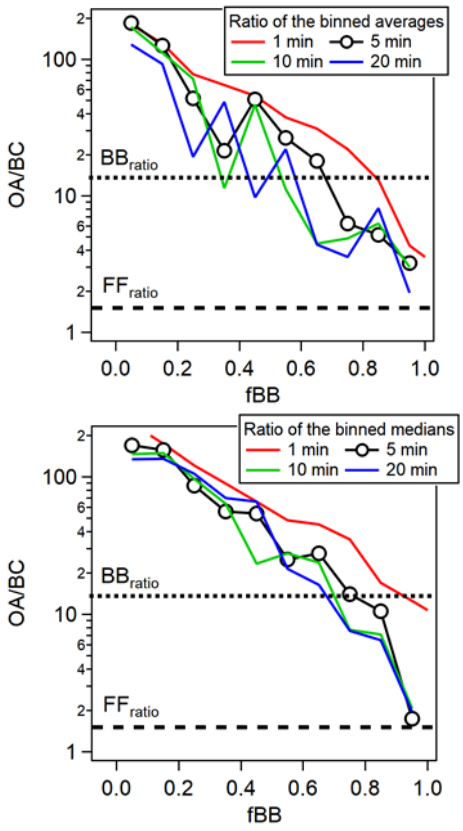
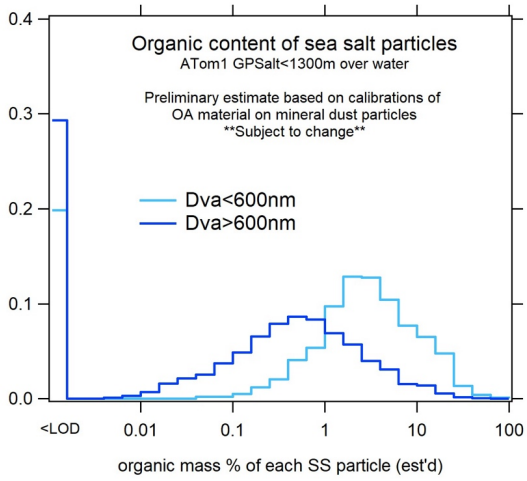


Figure S19: Distribution of the OA/BC ratio as a function of the fraction of BB influence measured by $f(BB)_{PALMS}$, calculated both as binned averages (left) and binned medians (right) for AToOm-1. Also shown are the OA/BC ratios that we currently assume based on the literature review for both anthropogenic (FF_{ratio}) and biomass burning sources (BB_{ratio}).

R3.1-c) For the calculation of observed POA, should there be small amounts of OA associated with sea spray and dust? You have sea salt and dust observations in Figure 2b and elsewhere that could be used to estimate that POA.

A3.1-c) We have added the following text to address this point:

"The contribution of POA from sea spray is difficult to constrain. As an order-of-magnitude estimate, marine POA is roughly calculated based on preliminary calibrations of OA on mineral dust particles from the PALMS instrument (personal communication K. Froyd). Using this calibration, the average OA by mass on sea salt was <10% for the large majority of MBL sampling (>85%). Since sea salt contributed 4% (11%) of mass in the AMS size range for ATom-1(2) (Figure 2), we estimate that marine POA is on the order of ~1% of aerosol mass in the AMS size range, and possibly much lower. Thus we think that it is reasonable to neglect the contribution of marine POA to this dataset. Future studies will refine this estimate."



POA associated with sea-salt particles in the marine boundary layer (<1300m) as reported by the PALMS instrument during ATom-1.

R3.1-d) Would it be better to label the estimate from equation 1 as "combustion POA"? I would assume models are also more specifically combustion POA?

A3.1-d) It would be inaccurate to label the estimated POA as combustion in the current model outputs as we cannot separate the combustion-emitted POA from other emission sources. This would require adding additional tracers and redoing the model simulations.

For the measurements, it would again be inaccurate to consider the estimated POA as only due to combustion. We have added the following text to the manuscript to clarify this point:

"The studies used to derive the emission ratio used ambient data in urban air, where all sources mix together and impact the POA/BC ratio, and thus the ratios include the impact of POA sources that may not emit BC."

R3.1-e) Is non-differential removal of BC and POA appropriate if POA is semivolatile? Consider that near a fire, concentrations could be high enough for IVOCs to be partitioned into the particle. As the airmass is diluted, POA will decrease more rapidly than a conserved tracer. This may explain why the larger BBratio gives POA >100% (line 378).

A3.1-e) This would be true if we were using the directly measured POA. But we are deriving POA from BC, and our method to estimate POA implicitly assumes that POA is non-volatile, and that it does not evaporate (and that also it is not lost to other processes such as photolysis or heterogeneous oxidation, even though those processes are known to be active in the atmosphere). This is done to obtain an upper limit for the measured POA, so that the key result that POA is too high in the models is reinforced. Therefore, POA > 100% in African plumes cannot be due to POA evaporation (and is instead likely due to uncertainties associated with the use of the global BB average ratios for those specific plumes; in fact the measured OA/BC value is within the combined uncertainty range for tropical forest given by Andreae (2019) 2.9..24.7).

The sensitivity of our results to the semi-volatile nature of POA has already been discussed in A2.3.

R3.1-f) For model estimates of POA, is hydrophilic OC considered POA or SOA? Can models just label the hydrophilic OC as "SOA" and get the right properties for endpoints (health, climate) of interest?

A3.1-f) Primary emitted hydrophilic OC is considered as POA. Indeed, the accepted definition is that for carbon emitted in the particle phase, even if it reacts in the particle phase, remains POA. For SOA to be produced from POA, POA needs to evaporate, and the evaporated organics gases to undergo gas-phase oxidation and then recondence into the particle phase. This is not happening in the model for the hydrophilic POA. This is consistent with earlier studies e.g. Pye and Seinfeld, 2010; Tsigaridis et al., 2014; Pai et al., 2019.

R3.1-g) Did you consider using a CTM to verify the method in equation 1 or back calculate the model effective FFratio? Does it reproduce the model POA (hydrophobic OC)?

A3.1-g) As already explained in the response A2.5, we did not apply this methodology to the model outputs as models do not track separately BC for various sources.

R3.1-h) BC is chemically aged. Do measurement techniques measure BC effectively at all atmospheric lifetimes?

A3.1-h) There is no evidence (that we are aware of, nor supplied by the reviewer) in the literature or in our measurements of refractory BC (such as measured in ATom) undergoing chemical loss to the gas-phase in the atmosphere. However, there are well-known aging effects (including coagulation, condensation, cloud processing) that do change the microphysical arrangement of BC by causing its physical shape to change and associating it with increasing amounts of internally mixed materials. Extensive testing as published in the literature (for example, Cross et al., (2010)) has shown that the measurement technique used in ATom to measure BC concentrations is insensitive to these aging effects. We have added the following text to the manuscript to address this point:

"Note that BC can physically age but it is not lost in any significant amount to the gas-phase due to chemical processes in the atmosphere."

Cross et al., (2010) Soot Particle Studies—Instrument Inter-Comparison—Project Overview, Aerosol Science and Technology, 44:8, 592-611, DOI: 10.1080/02786826.2010.482113.

R3.1-i) Line 650: The observation-method is not necessarily an upper-bound limit on the fraction of POA as it does not consider OA emitted in sea spray, dust, and may not consider all anthropogenic forms of POA. Consider rephrasing and/or demonstrating it is a limit by adding error bars by using more conservative (higher) FFratios.

A3.1-i) We respectfully disagree with the reviewer's statement. As explained in response A3.1-d above, our method implicitly accounts for all pollution sources of POA, even if they do not emit BC. As documented in response A3.1-c above, the estimated impact of marine POA in the submicron range during ATom is very small, and does not affect our results. Any amount of POA present in dust would be even smaller, as the dust concentrations in the submicron range (Fig. 2) are even smaller than those of sea salt, and the fraction of OA in dust is also very low.

R3.2) Line 355: Can you elaborate on why the PMF doesn't work to separate SOA and POA? When (what timescale, location, altitude, other factor) does the PMF stop working?

A3.2) Extracting linear factors by PMF always "works" in a technical sense, the question is whether it can provide the information that one seeks. This depends on the information content of the dataset and the questions asked. We have updated the text at this location to document this issue in more detail (since we do get this question relatively frequently), as:

"This approach is not suitable for ATom. To accurately resolve a minor factor such as POA in an AMS dataset, there needs to be a combination of: (a) Sufficient OA mass concentration, so that the signal-to-noise of the spectra is sufficient; (b) Enough fractional mass for the factor to be resolved (>5% in urban areas per Ulbrich et al. (2009), probably a larger fraction at low concentrations such as in ATom); (c) Sufficient spatio-temporal variability ("contrast") in the relative contributions of different factors, since that is part of what PMF uses to extract the factors; (d) Sufficient difference in the spectra of the different factors (for the same reason as (c)), and (e) relatively invariant spectra for each factor across the dataset (as this is a key assumption of the PMF algorithm). As an example of a near ideal case, in Hodshire et al (2019) we extracted MSA by PMF from the ATom-1 data, and were able to match that factor with our independently calibrated MSA species. A very distinct and nearly invariant mass spectrum was measured repeatedly near sources (MBL) (and was mostly absent elsewhere, thus providing strong spatio-temporal contrast) and accounted for about 6% of the fractional mass and 15% of the variance in time. Thus all the conditions were met. For POA, on the other hand, the air sampled in ATom and coming from e.g. Asia has POA and SOA very well mixed, with little change on their relative mass fractions vs. time (as the aircraft flies through that airmass). POA is very low, as documented later in this paper. Atmospheric aging makes the spectra from all OA sources more and more similar as measured by AMS spectra (Jimenez et al., 2009). Thus most of the conditions above are not satisfied for extracting POA by PMF analysis of this dataset."

Ulbrich, I.M., M.R. Canagaratna, Q. Zhang, D.R. Worsnop, and J.L. Jimenez. Interpretation of Organic Components from Positive Matrix Factorization of Aerosol Mass Spectrometric Data. Atmospheric Chemistry and Physics , 9, 2891-2918, 2009.

Hodshire, A.L., P. Campuzano-Jost, J.K. Kodros, B. Croft, B.A. Nault, J.C. Schroder, J.L. Jimenez, J.R. Pierce. The potential role of methanesulfonic acid (MSA) in aerosol formation and growth and the associated radiative forcings. Atmos. Chem. Phys., 19, 3137-3160, 2019, https://doi.org/10.5194/acp-19-3137-2019

Jimenez, J.L., M.R. Canagaratna, N.M. Donahue, A.S.H. Prevot, Q. Zhang, J.H. Kroll, P.F. DeCarlo, J.D. Allan, H. Coe, N.L. Ng, A.C. Aiken, K.D. Docherty, I.M. Ulbrich, A.P. Grieshop, A.L. Robinson, J. Duplissy, J. D. Smith, K.R. Wilson, V.A. Lanz, C. Hueglin, Y.L. Sun, J. Tian, A. Laaksonen, T. Raatikainen, J. Rautiainen, P. Vaattovaara, M. Ehn, M. Kulmala, J.M. Tomlinson, D.R. Collins, M.J. Cubison , E.J. Dunlea, J.A. Huffman, T.B. Onasch, M.R. Alfarra, P.I. Williams, K. Bower, Y. Kondo, J. Schneider, F. Drewnick, S. Borrmann, S. Weimer, K. Demerjian, D. Salcedo, L. Cottrell, R. Griffin, A. Takami, T. Miyoshi, S. Hatakeyama, A. Shimono, J.Y Sun, Y.M. Zhang, K. Dzepina, J.R. Kimmel, D. Sueper, J.T. Jayne, S.C. Herndon, A.M. Trimborn, L.R. Williams, E.C. Wood, C.E. Kolb, A.M. Middlebrook, U. Baltensperger, and D.R. Worsnop. Evolution of Organic Aerosols in the Atmosphere. Science, 326, 1525-1529, 2009. doi: 10.1126/science.1180353.

R3.3) Perhaps the strengths of DYN could be better isolated/highlighted (e.g. what life-times, OA regimes, POA levels, etc does DYN perform better for?). The abstract statement: "concept of a more dynamic OA. . .with enhanced removal of primary OA, and a stronger production of secondary OA in global models needed to provide a better agreement with observations" could use more support.
Evaluation seems mixed: Figure 8, indicates GC12 and CESM1-CARMA have reasonable POA (while CESM2-DYN has overestimated POA). Line 803 indicates DYN evaluates better, but that effect seems marginal or secondary to other issues in Figure 9 b at least for GC. RMSE in Table 2 also tends to increase when going from base to DYN treatment in CESM2 and GC12. Figure 7 improvements seem mixed.

A3.3) We agree with the reviewer that given the extreme complexity of the dataset and the models, it is difficult to identify the model configuration that works the best all the time. We do not conclude that DYN is providing better results than other models under all conditions (e.g. CESM2-DYN still has a problem with the POA removal in convection as the fix implemented in CESM-CARMA has not yet been implemented in this version). We have reworded the text in section 4.1 to better describe the model behavior in the NH summer and winter, and indicate more clearly when and where the stronger SOA production provides better results:

"*During the NH summer (ATom-1), models using the VBS parameterization from Pye et al. (2010) tend to underpredict the OA concentrations by 43% for GC12-REF and 33% for CESM1-CARMA for ATom-1, most likely due to the excessive evaporation of the formed SOA in remote regions and low yields for anthropogenic SOA (Schroder et al., 2018; Shah et al., 2019). Models using the VBS parameterization from Hodzic et al. (2016) (CESM2-DYN and GC12-DYN) where OA is less volatile and also OA yields are corrected for wall losses show an improved agreement with observations especially for CESM2-DYN (with NMB of ~5%), and to a lesser extent for GC12-DYN (NMB of ~33%). During the NH winter*

(ATom-2) characterized by a lower production of SOA, both VBS approaches lead to an overestimation of the predicted OA. This is likely caused by excessively high levels of primary emitted OA as discussed in section 4.4."

More importantly (not specific to DYN), in this study we have shown as stated in the conclusions (and abstract) that "*the OA system seems to be more dynamic with a need for an enhanced removal of primary OA, and a stronger production of secondary OA in global models to provide a better agreement with observations.*" And that is supported by the fact that models that have improved in-cloud removal of POA tend to perform better with regard to POA concentrations and vertical profiles in the upper troposphere than those that don't have it (e.g. CESM1-CARMA with improved in-cloud removal vs. without improved in-cloud removal in Figure 8 and S6; GC12 in Figure 8 and S6 compared to CESM2). We have also shown that making POA semi-volatile instead of non-volatile in model simulations aggravates the model bias in the upper troposphere, and that removal by deep convective clouds and possibly photolysis are needed to address model bias.

The need for a stronger SOA production in models is supported by the fact that models that have correct POA (CESM1-CARMA or GC12) need a stronger production of SOA to match the measured concentrations. The comparison of GC12-REF with two SOA formation mechanisms (using the same removals) in Figure 8 illustrates that a stronger production of SOA would lead to an improved agreement with measurements.

Minor comments:

R3.4) Abstract states "OA predictions from AeroCom Phase II...span a factor of 400- 1000. . ." should that be the inter-model variability spans a factor of 400-1000 or the concentrations predicted span that range?

A3.4) The first one. This has already been clarified in response A2.1 above.

R3.5) Did all models assume nonvolatile POA?

A3.5) Yes, see response A2.3.

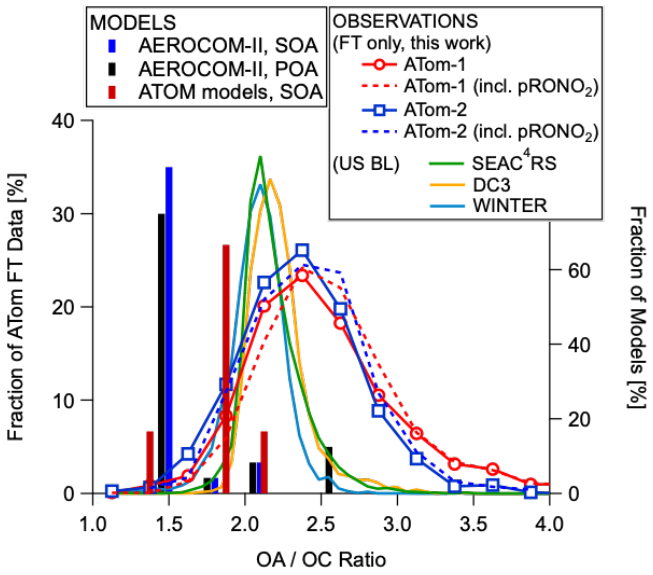
R3.6) Do OA/OC ratios include consideration of S and N in the form of organosulfates and organonitrates?

A3.6) The following text has been added to the manuscript to address this point:

"Note that for organosulfates (R-O-SO_2H and organonitrates (R-O-NO_2, $pRONO_2$ in the following) only one oxygen is included in the reported OA/OC, as the fragments of these species are typically the same as for inorganic species in the AMS (Farmer et al., 2010). However in ATom organosulfates are estimated to account for ~1% of the total sulfate (based on PALMS data, see Liao et al., 2015 for the methodology). Since sulfate and OA concentrations are comparable, organosulfates would only increase the OA/OC by ~1% on average. Organonitrates are reported from the AMS for ATom. Their impact on OA/OC is not propagated for the default values, to maintain consistency with a large set of OA/OC measurements by AMS in the literature, and since they would increase OA/OC on average by only 4.5% (ATom-1) and 2.2% (ATom-2), which is smaller than the uncertainty of this measurement. However, we show the results with both methods in Fig. 5 to fully document this topic.

Farmer, D.K., A. Matsunaga, K.S. Docherty, J.D. Surratt, J.H. Seinfeld, P.J. Ziemann, and J.L. Jimenez. Response of an Aerosol Mass Spectrometer to Organonitrates and Organosulfates and implications for Atmospheric Chemistry. Proceedings of the National Academy of Sciences of the USA, 107, 6670-6675, doi: 10.1073/pnas.0912340107, 2010.

Liao, J., K.D. Froyd, D.M. Murphy, F.N. Keutsch, G. Yu, P.O.Wennberg, J.St. Clair, J.D. Crounse, A. Wisthaler, T. Mikoviny, T.B. Ryerson, I.B. Pollack, J. Peischl, J.L. Jimenez, P. Campuzano Jost, D.A. Day, B.E. Anderson, L.D. Ziemba, D.R. Blake, S. Meinardi, G. Diskin. Airborne organosulfates measurements over the continental US. Journal of Geophysical Research-Atmospheres, 120, 2990–3005, doi:10.1002/2014JD022378, 2015.



Caption Fig. 5*: Distribution of the OA / OC ratio as measured during ATom-1 and -2. Also included (in dashed lines) are the distributions of OA/OC values that included the contribution of organic nitrates (pRONO₂). Values for the recent aircraft campaigns (SEAC4RS, DC3 and WINTER) that took place over continental US regions closer to continental source regions are also shown (Schroder et al., 2018). The bars (right axis) show the OA/OC used for SOA and POA by the models included in the AeroCom and ATom ensemble, with OA/OC=1.4 being the modal value for the former and 1.8 for the latter.*

R3.7) At times, it was a bit confusing if the AeroCom-II ensemble referred to circa 2014 models or the actual AeroCom II results paired with measurements. On line 103, the "AeroCom results" could be clarified as "AeroCom models" or "models used in AeroCom." Line 277 might be better as "monthly average results of 28 global models" instead of "results of 28 global models."

A3.7) We have modified the text as suggested by the reviewer: "*Our own analyses of the AeroCom-II models..*" and "*We consider the monthly average results of 28 global models,..*".

R3.8) Figure 1: Is "author affiliations" supposed to be in the caption?

A3.8) We have added *"Model contributor affiliations"* to the figure label and caption. See also response A1.3.

R3.9) Line 163: While marine production of OA may be smaller than continental production, it's possible marine production contributes more to concentrations over the oceans than it's global production rate would suggest. I recommend rewording sentence. What concentrations of SOA are predicted for oceanic isoprene sources?

A3.9) SOA production from oceanic isoprene sources is not included as already indicated in the manuscript.

As suggested by the reviewer, we have modified the sentence to read: "*None of the models includes the marine production of OA which is estimated to be ~3 orders of magnitude smaller than the continental production of OA from both isoprene and monoterpene precursors (Kim et al., 2017), but could be important in the MBL.*"

See also the response to A3.1-c above.

R3.10) How does averaging data allow below detection values to be used?

A3.10) This seems to be a very widespread issue with modelers' understanding of our measurements. We have included the following text to section 3.3. to clarify and document this point:

"Per standard statistics, the precision of a measurement decreases (i.e., gets better) with the square root of the number of points (or time interval) sampled. I.e. the precision of an average can be approximated by the standard error of the mean (σ/sqrt(n), where n is the number of measurements averaged), and it is better than the precision of the individual data points (σ). This also applies to the detection limit, since it is just 3 times the precision. Note that a detection limit is not meaningful unless the averaging time is specified. For example, let's assume that the detection limit is 20 ng m^{-3} (1-second), and the data points over 60 consecutive seconds are all 10 ng m^{-3}. All 1-second measurements are below the 1-second DL. However the average (10 ng m^{-3}) is now above the DL for 1-minute averages, which is 20/sqrt(60) = 2.6 ng m^{-3}."

R3.11) Line 605: qualify the OA/OC of 1.4 as "fossil fuel" combustion as biomass combustion tends to have OA/OC of 1.7.

A3.11) We have added this qualifier as requested: "*A low OA/OC ratio is indicative of freshly emitted OA from fossil fuel combustion (typically ~1.4),..*"

1514

Characterization of organic aerosol across the global remote troposphere: A comparison of ATom measurements and global chemistry models

Alma Hodzic[1], Pedro Campuzano-Jost[2,3], Huisheng Bian[4], Mian Chin[4], Peter R. Colarco[4], Douglas A. Day[2,3], Karl D. Froyd[2,8], Bernd Heinold[6], Duseong S. Jo[2,3], Joseph M. Katich[2,8], John K. Kodros[5], Benjamin A. Nault[2,3], Jeffrey R. Pierce[5], Eric Ray[2,8], Jacob Schacht[6], Gregory P. Schill[2,8], Jason C. Schroder[2,3], Joshua P. Schwarz[2,8], Donna T. Sueper[2,3], Ina Tegen[6], Simone Tilmes[1], Kostas Tsigaridis[7,9], Pengfei Yu[8,10], Jose L. Jimenez[2,3]

[1]*National Center for Atmospheric Research, Boulder, CO, USA*

[2]*Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA*

[3]*Department of Chemistry, University of Colorado, Boulder, CO, USA*

[4]*NASA Goddard Space Flight Center, Greenbelt, MD, USA*

[5]*Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA*

[6]*Leibniz Institute for Tropospheric Research, Leipzig, Germany*

[7]*Center for Climate Systems Research, Columbia University, New York, NY, USA*

[8]*NOAA Earth System Research Laboratory (ESRL), Chemical Sciences Division, Boulder, CO, USA*

[9]*NASA Goddard Institute for Space Studies, New York, NY, USA*

[10]*Institute for Environmental and Climate Research, Jinan University, Guangzhou, Guangdong, China*

Key words: organic aerosol, remote atmosphere, ATom field campaign.

Abstract.

The spatial distribution and properties of submicron organic aerosols (OA) are among the key sources of uncertainty in our understanding of aerosol effects on climate. Uncertainties are particularly large over remote regions of the free troposphere and Southern Ocean, where very little data has been available, and where OA predictions from AeroCom Phase II global models span two to three orders-of-magnitude, greatly exceeding the model spread over source regions. The (nearly) pole-to-pole vertical distribution of non-refractory aerosols was measured with an aerosol mass spectrometer onboard the NASA DC8 aircraft as part of the Atmospheric Tomography (ATom) mission during the northern hemisphere summer (August 2016) and winter (February 2017). This study presents the first extensive characterization of OA mass concentrations and their level of oxidation in the remote atmosphere. OA and sulfate are the major contributors by mass to submicron aerosols in the remote troposphere, together with sea salt in the marine boundary layer. Sulfate was dominant in the lower stratosphere. OA concentrations have a strong seasonal and zonal variability, with the highest levels measured in the lower troposphere in the summer and over the regions influenced by the biomass burning from Africa (up to 10 $\mu g\ sm^{-3}$). Lower concentrations (~0.1-0.3 $\mu g\ sm^{-3}$) are observed in the northern mid- and high- latitudes and very low concentrations (< 0.1 $\mu g\ sm^{-3}$) in the southern mid- and high- latitudes. The ATom dataset is used to evaluate predictions of eight current global chemistry models that implement a variety of commonly used representations of OA sources and chemistry, as well as of the AeroCom-II ensemble. The current model ensemble captures the average vertical and spatial distribution of measured OA concentrations, and the spread of the individual models remains within a factor of 5. These results are significantly improved over the AeroCom-II model ensemble, which shows large overestimations over these regions. However, some of the improved agreement with observations occurs for the wrong reasons, as models have the tendency to greatly overestimate the primary OA fraction, and underestimate the secondary fraction. Measured OA in the remote free troposphere are highly oxygenated with organic aerosol to organic carbon (OA/OC) ratios of ~ 2.2-2.8 and are 30-60% more oxygenated than in current models, which can lead to significant errors in OA concentrations. The model/measurement comparisons presented here support the concept of a more dynamic OA system as proposed by Hodzic et al. (2016), with enhanced removal of primary OA, and a stronger production of secondary OA in global models needed to provide a better agreement with observations.

1 Introduction

Organic aerosols (OA) are a complex mixture of directly emitted primary OA (POA) and chemically produced secondary OA (SOA) from anthropogenic and biogenic emission sources. They are associated with adverse health effects (Mauderly and Chow, 2008, Shiraiwa et al., 2017) and contribute radiative forcing in the climate system (Boucher et al., 2013). The currently limited understanding of processes involved in the formation, ageing, and removal of organic compounds results in large uncertainties in (i) the predicted global OA burden, (ii) relative contributions of emissions vs. chemistry to OA formation, (iii) spatial distribution, and (iv) impacts on radiation and clouds (Kanakidou et al., 2005, Hallquist et al., 2009, Heald et al., 2011, Spracklen et al., 2011, Tsigaridis et al., 2014, Hodzic et al., 2016, Shrivastava et al., 2017, Tsigaridis and Kanakidou, 2018, Zhu et al., 2019). The uncertainties are particularly large in the estimated global burden of SOA that range from 12 to 450 Tg y^{-1} (see Fig. 9 of Hodzic et al., 2016), and in their direct and indirect radiative forcing that range from -0.08 to -0.33 W m^{-2}, and -0.60 to −0.77 W m^{-2}, respectively (Spracklen et al., 2011, Myhre et al., 2013, Scott et al., 2014, Hodzic et al., 2016, Tsigaridis and Kanakidou, 2018). Reducing these uncertainties is becoming more important as OA is on a path to becoming the dominant fraction of the submicron anthropogenic aerosol mass globally due to the ongoing efforts to reduce SO_2 emissions and associated sulfate aerosols.

Model performance has been especially poor in the remote regions of the atmosphere where OA measurements available for model evaluation have been sparse (especially aloft). Using data from 17 aircraft campaigns mostly located in the northern hemisphere Heald et al. (2011) showed that the skill of the global GEOS-Chem model in predicting the vertical distribution of OA was significantly decreased in remote regions compared to polluted near-source regions. The study pointed out the limitations of commonly used SOA formation mechanisms that are based on chamber data; these have the tendency to underpredict OA in source regions and overpredict OA in the remote troposphere. For a subset of 9 recent aircraft campaigns, Hodzic et al. (2016) showed that OA is likely a more dynamic system than represented in chemistry-climate models, with both stronger production and stronger removals. These authors suggested that additional removal mechanisms via e.g. photolytic or heterogeneous reactions of OA particles are needed to explain low OA concentrations observed in the upper troposphere where direct cloud scavenging is less efficient. The recent global multi-model comparison study (Tsigaridis

Deleted: a major
Deleted: ion
Deleted: to

et al., 2014) within the AeroCom Phase II project illustrates well the amplitude of model uncertainties simulating OA mass concentrations and the contrast in model performance between near-source and remote regions. The results indicate that model dispersion (the spread between the models with the lowest and highest predicted OA concentrations) increases with altitude from roughly 1 order of magnitude near the surface to 2-3 orders of magnitude in the upper troposphere. Our own analyses of the AeroCom-II models shown in Figure 1a indicate that model dispersion (quantified as the ratio of the average concentration of the highest model to that of the lowest one, in each region) increases not only with altitude but also with distance from the northern mid-latitude source (and data-rich) regions. The model spread is a factor of 10-20 in the free troposphere between the equator and northern mid-latitudes, and increases to a factor of 200-800 over the Southern Ocean and near the tropopause. It is not surprising that model spread is lower closer to source regions where it is mostly driven by uncertainties in emissions and SOA production yields. Spread is expected to be larger in remote regions where models are also impacted by uncertainties in transport, chemical ageing and removal. The lowest model dispersion also coincides with the regions of the northern hemisphere (NH) or the African biomass burning outflow where models have been evaluated the most (Figure 1b), emphasizing the need for further model/observation comparison studies in remote regions (of the southern hemisphere (SH) in particular).

Here, we present a unique data set of airborne aerosol mass spectrometer measurements of OA mass concentrations collected onboard the NASA DC-8 as part of the Atmospheric Tomography (ATom) mission. The aircraft sampled the vertical structure of the atmosphere from near-surface (0.2 km) to the lower-stratosphere (LS) regions (12 km altitude) over both the Pacific and Atlantic basins (to limit the influence of source regions) with a quasi-global spatial coverage from 82°N to 67°S. This dataset is used to perform the first systematic global-scale multi-model evaluation of the chemistry-climate models focusing on OA in the remote troposphere over the remote oceans. We focus on the NH summer (August 2016, ATom-1) and NH winter (February 2017, ATom-2) deployments. Overall these ATom missions sampled the marine boundary layer (MBL) for 10% of the flight tracks, 12% of the time the remote lower stratosphere, and the rest the free troposphere. The model-observation comparisons are aimed at identifying discrepancies in terms of OA mass concentrations and vertical distribution, their fractional contribution to submicron aerosols, and their oxidation level in global models.

The modeling framework is described in Section 2. Section 3 describes the ATom dataset and the spatial and vertical distributions of OA over the Atlantic and Pacific regions. Section 4 presents the comparisons of ATom-1 and -2 data to multi-model predictions from both the AeroCom-II models, and the ensemble of eight current model simulations of the ATom campaign. Section 5 presents the conclusions of the study and discusses its implications.

2 Modeling framework

2.1 ATom model simulations

ATom measurements were compared with results of eight global models that simulated the time period of the ATom-1 and 2 campaigns (August 2016 and February 2017), using the emissions and reanalysis meteorology corresponding to this period (and a spin-up time of at least six to twelve months). These are referred hereafter as ATom models and include the NASA global Earth system model GEOS5, the aerosol-climate model ECHAM6-HAM, three versions of the NCAR Community Earth System Model (CESM), and three versions of the global chemistry GEOS-Chem model. Simulations were performed at various horizontal resolutions ranging from relatively high ~50km (GEOS5) and ~100km (CESM2 models) resolutions to somewhat coarser grids of ~200km (CESM1-CARMA, GEOS-Chem) and ~400km for GC10-TOMAS. The advantage of using the same host model (in the cases of variants of CESM2 and GEOS-Chem) is that the dynamics and emissions remain comparable. Models differ greatly in their treatment of emissions, gas-phase chemistry, aerosol chemistry and physical processes, and aerosol coupling with radiation and clouds, among others. Table 1 describes the configuration of various models (e.g. meteorology, emissions), and their treatment of OA. In this section we only summarize the main features and parameters directly impacting the OA simulations. Some models do not include SOA chemistry and instead assume that SOA is directly emitted proportional to the emissions of its precursors (ECHAM6-HAM, CESM2-SMP, GEOS5, GC10-TOMAS), while others have more complex treatments of organic compounds, their chemistry, and partitioning into particles (GC12-REF, GC12-DYN, GC10-TOMAS, CESM1-CARMA, CESM2-DYN). It should be noted that models that directly emit SOA assume that SOA is a non-volatile species that remains irreversibly in the particle phase. In all models POA is treated as a non-volatile directly emitted species. In most models (see below) the primary emitted organic aerosol is artificially aged to transition between hydrophobic to hydrophilic POA. There are some commonalities between simulations for

Deleted: -GOCART

Deleted: -GOCART

Deleted: -GOCART

the treatment of biogenic emissions, which are based in all models on the Model of Emissions of Gases and Aerosols from Nature (MEGAN, Guenther et al., 2012) to generate meteorology-dependent emissions of volatile organic compounds. None of the models includes the marine production of OA which is estimated to be ~3 orders of magnitude smaller than the continental production of OA from both isoprene and monoterpene precursors (Kim et al., 2017), but could be important in the MBL. This contribution could however be larger for sea-spray biological material from phytoplankton with predicted contributions of 0.01 to 0.1 $\mu g\ m^{-3}$ to surface submicron aerosol over remote oceanic regions (Vergara-Temprado et al., 2017, Middlebrook et al., 1998). Below we only provide a brief description of most important processes that influence OA for each model.

GEOS5 was run in a configuration similar to Bian et al. (2019) using the anthropogenic emissions from HTAP v2 (Janssens-Maenhout et al., 2015) and biomass burning emissions from the Quick Fire Emission Dataset (QFED v2.54). Aerosols are simulated within the GOCART bulk aerosol module and include externally mixed particles of black carbon (BC), organic carbon (OC), sulfate, ammonium, nitrate, dust and sea salt (Colarco et al., 2010, Bian et al., 2017). The formation of SOA is based on a prescribed 10% formation yield from the monoterpene emissions. The primary emitted OC and SOA are separated into hydrophobic (50%) and hydrophilic (50%) species, with a 2.5 days e-folding time conversion from hydrophobic to hydrophilic organic particles. All SOAs from biogenic, anthropogenic, and biomass burning sources are treated as hydrophilic particles. Both types of organic particles are dry deposited. The hydrophilic OA is removed by large-scale and convective warm clouds, while hydrophobic OA is removed by ice clouds. The hydrophilic particles undergo hygroscopic growth according to the equilibrium parameterization of Gerber (1985).

Deleted: -GOCART

The ECHAM6.3-HAM2.3 standard version (Tegen et al., 2019) was run using updated anthropogenic emissions (Schacht et al, 2019) combining the ECLIPSE (Klimont et al., 2017) emissions, with the Russian anthropogenic BC emissions from Huang et al. (2015). For biomass burning the Global Fire Assimilation System (GFAS, Kaiser et al., 2012) biomass burning emissions are used, however, without the scaling factor of 3.4 suggested by Kaiser et al. (2012). Aerosol composition and processes are simulated using the Hamburg Aerosol Model (HAM2, Zhang et al., 2012), that considers an aerosol internal mixture of sulfate, BC, OC, sea salt, and mineral dust. The aerosol population and their microphysical interactions are simulated using seven log-normal modes, including the

nucleation mode, soluble and insoluble Aitken, accumulation and coarse modes. In the model configuration used in this publication the formation of SOA is based on a prescribed 15% mass yield from monoterpene emissions only (Dentener et al., 2006). Aerosol particles are removed by dry and wet deposition. The wet deposition includes the below cloud scavenging by rain and in-cloud cloud scavenging for large-scale and convective systems (Croft et al., 2010).

The two simulations with the GEOS-Chem 12.0.1 global chemistry model (Bey et al., 2001) use emissions based on CMIP6 global inventory (CEDS historical emissions up to 2014 and future emissions based on climate scenarios, Hoesly et al., 2018; Feng et al., 2019) with regional improvements for anthropogenic sources, and on GFED v.4 for biomass burning emissions (Giglio et al., 2013). Both simulations use the bulk aerosol representation and differ only in the treatment of SOA formation and removal. The first configuration (called hereafter GC12-REF) includes the default (http://wiki.seas.harvard.edu/geos-chem/index.php) representation of SOA formation based on Marais et al. (2016) for isoprene-derived SOA, and on the volatility basis set (VBS) of Pye et al. (2010) for all other precursors. Note that this GEOS-Chem REF simulation is similar to the version 12 default "complex option" which includes non-volatile POA and semi-volatile SOA (semi-volatile POA is an optional switch within this version used in Pai et al. 2020). The second configuration (referred to as GC12-DYN) includes a more dynamic representation of the SOA lifecycle based on Hodzic et al. (2016), with the exception of the treatment of isoprene SOA that is formed in the aqueous aerosols as in Marais et al. (2016). As in Hodzic et al. (2016) the GC12-DYN model version includes updated VBS SOA parameterization, updated dry and wet removal of organic vapors, and photolytic removal of SOA (except for isoprene-SOA). SOA formation is based on wall-corrected chamber yields (Zhang et al., 2014) for the traditional precursors (isoprene, monoterpenes, sesquiterpenes, benzene, toluene, xylene) and on yields derived from an explicit chemical mechanism for higher molecular weight n-alkanes and n-alkenes species (Hodzic et al., 2016). The removal of gas-phase oxidized volatile organics uses updated Henry's law solubility coefficients from Hodzic et al. (2014), and photolytic removal of SOA (Hodzic et al., 2015). In addition to OA, the model includes BC and dust, and simulates the chemistry and gas-particle partitioning of inorganic compounds such as sulfate, ammonium, nitrate and sea salt using the ISORROPIA II thermodynamic model (Fountoukis and Nenes, 2007). In both GEOS-Chem configurations, BC and primary OC are simulated with a hydrophobic and hydrophilic fraction for each. At the time of emission,

80% of BC and 50% of primary OC are considered as hydrophobic. Hydrophobic aerosols are converted to hydrophilic aerosols with an e-folding lifetime of 1.15 days. An OA/OC ratio of 2.1 is assumed to convert POC to POA, and SOA is simulated as OA mass (i.e. no OA/OC ratio assumption is needed for SOA, except for comparison with OC measurements). Soluble gases and aerosols are removed by both dry and wet deposition. Wet deposition includes scavenging in convective updrafts, and in-cloud and below-cloud scavenging from large-scale precipitation (Liu et al., 2001). Hydrophobic aerosols (BC and POA) are scavenged in convective updrafts following Wang et al. (2014).

GC10-TOMAS is based on the GEOS-Chem version 10.01 coupled with TwO Moment Aerosol Sectional microphysics scheme (TOMAS) and ran in a similar configuration to that described in Kodros et al. (2016). The model computes the evolution of sulfate, sea salt, primary and secondary OA, BC, and dust aerosols described by 15 internally mixed size bins (of which six were analyzed for these comparisons, cf. Table 1). Anthropogenic emissions are based on the EDGAR v4 global inventory with regional improvements, while the biomass burning emissions are from GFED v3. SOA are irreversibly made from the emitted parent precursor, considering a 10% mass yield from monoterpene emissions, and an emission flux of 0.2 Tg of SOA per Tg of CO for the anthropogenic CO emissions. The removal of gases and aerosols are treated similar to the GEOS-Chem 12.0.1 model (GC12-REF, see above).

Simulations based on the CESM2.0 Earth system model use the standard version of the Whole Atmosphere Community Climate Model (WACCM6, Gettelman et al., 2019, Emmons et al., 2019). Details on the specific of the model configurations are described in detail in Tilmes et al. (2019) i.e. CESM2-SMP and CESM2-DYN correspond to the specified dynamics WACCM6-SOAG and WACCM6-VBSext simulations described in that work, respectively. Emissions are based on the CMIP6 global inventory for the year 2014 for anthropogenic sources, and on the QFED version 2.4 for the wildfires inventory. Aerosols are represented with the modal aerosol scheme (MAM4, Liu et al., 2012) that includes BC, primary and secondary OA, sulfate, dust and sea salt. Four modes are considered including Aitken, accumulation and coarse size modes, and an additional primary carbon mode. Only the accumulation mode was used in this work. The CESM2-SMP and CESM2-DYN simulations differ in their treatment of OA. CESM2-SMP forms OA directly using fixed mass yields from primary emitted precursors (isoprene, monoterpenes, aromatics) without explicitly simulating their oxidation and partitioning. These mass yields

are increased by a factor of 1.5 to match the anthropogenic aerosol indirect forcing (Liu et al., 2012). The second configuration (referred to as CESM2-DYN) includes the formation and removal parameterizations of organics of Hodzic et al. (2016), as implemented into CESM2 by Tilmes et al. (2019) for all species based on low-NOx VBS yields only. This is a similar SOA scheme as used in GC12-DYN (with differences in the treatment of isoprene-SOA based on Marais et al. 2016 in GC12-DYN, and the use of both low- and high-NOx VBS yields in GC12-DYN). Organic gases and aerosols undergo dry and wet deposition as described in Liu et al. (2012). It should be noted that CESM2-SMP does not include deposition of intermediate organic vapors. Aerosol wet scavenging considers in-cloud scavenging (the removal of cloud-borne particles that were activated at the cloud base) and below-cloud scavenging for both convective and grid-scale clouds.

CESM1-CARMA simulations use the configuration described in Yu et al. (2019) which is based on CESM1 and the sectional Community Aerosol and Radiation Model for Atmospheres (CARMA v3.0). Anthropogenic emissions are those from the Greenhouse gas-Air pollution Interactions and Synergies (GAINS) model, and biomass burning emissions are from the Global Fire Emission Database (GFED v3, van der Werf et al., 2010). In CARMA, 20 size bins are used for both pure sulfate particles (bins from 0.2 nm to 1.3 μm in radius, only used up to 500 nm) and mixed aerosols composed of BC, primary and secondary OC, dust, sea salt, and sea-spray sulfate (bins from 0.05–8.7 μm in radius, again, only analyzed up to 500 nm). SOA formation is based on the VBS approach from Pye et al. (2010). The removal of OA occurs only by dry and wet deposition. Compared to the CESM2-SMP and CESM2-DYN simulations, the convective removal of aerosols uses the modified scheme described in Yu et al. (2019) which accounts for aerosol secondary activation from the entrained air above the cloud base, and the scavenging of activated aerosols in convective updrafts. The default CESM can transport aerosols from the cloud base to the top of the cloud in strong convective updrafts in one time step without scavenging them, while the new scheme allows for a more efficient removal off all aerosols inside convective clouds. A sensitivity simulation is performed for ATom-1 to quantify the effect of this improved removal on OA concentrations (Section 4.5).

2.2 AeroCom-II model climatology

The ATom measurements are also compared to the global model OA predictions generated within the Phase II Aerosol Comparisons between Observations and Models (AeroCom-II) project (Schulz et al., 2009). We consider the monthly average results of 28

global models, which is a subset of those presented in Tsigaridis et al. (2014), based on the availability of model results. It should be noted that the meteorological forcing used in these models is mostly based on the year 2006, while the anthropogenic and biomass burning emissions are mostly representative of the year 2000. For comparison purposes, the monthly mean model outputs for the months of August (ATom-1) and February (ATom-2) are interpolated along the flight path (latitude, longitude, and altitude), and averaged the same way as the measurements (see section 3.2).

3 Description of ATom measurements

3.1 Submicron aerosol data

The measurements of non-refractory submicron aerosols were performed onboard the NASA DC8 aircraft as part of the ATom field study (Wofsy et al., 2018) using the University of Colorado Aerodyne High-Resolution Time-of-Flight Aerosol Mass Spectrometer (AMS in the following, Canagaratna et al., 2007, DeCarlo et al., 2006).

We use measurements from both the NH summer (August 2016, ATom-1) and winter (February 2017, ATom-2) deployments. Figure 2a shows the flight path and the vertical extent of the ATom-1 dataset colored by OA mass concentrations (see Figure S1 for ATom-2). The aircraft performed systematic vertical sampling with ~140 vertical profiles per campaign throughout the troposphere from the near surface ~0.2 km to the upper troposphere/lower stratosphere region at ~13 km altitude. Details on the operation of the CU AMS on board the DC-8 are reported in Schroder et al. (2018), Nault et al. (2018), and Jimenez et al. (2019b). AMS data was acquired at 1 Hz time resolution and independently processed and reported at both 1 s and 60 s time resolutions (Jimenez et al., 2019a). The later product, with more robust peak fitting at low concentrations was exclusively used as the primary dataset in this work. Detection limits at different time resolutions/geographical bins relevant to this study are discussed in Section 3.3. The overall 2σ accuracies of the AMS measurement (38% for OA, 34% for sulfate and other inorganics) are discussed in Bahreini et al. (2008) and Jimenez et al. (2019b).

For ATom, the AMS reported the standard non-refractory aerosol species OA, sulfate, nitrate, ammonium, and chloride, with the response for all the nominally inorganic species characterized by in-field calibrations. In addition, it also reported methanesulfonic acid (MSA, Hodshire et al., 2019a describes the AMS MSA methods and calibrations for ATom) and sea salt for for D_{geo}<450 nm (based on the method of Ovadnevaite et al., 2012). Both

of these species were important to achieve closure with the volume calculated from the on-board sizing instruments in the marine boundary layer (Jimenez et al., 2019b). Another important refractory submicron species not captured by the AMS measurements is BC. This was measured on ATom with the NOAA SP2 instrument (Katich et al., 2018). It should be noted that aircraft measurements of aerosol mass concentrations are given in $\mu g\ sm^{-3}$ (i.e., under standard conditions of 1 atm and 273.15 K).

For ATom the AMS measured particles with geometric diameters (based on the campaign-wide average density of 1640 kg m^{-3}, Jimenez et al., 2019b) of between D_{geo}~60 and 295 nm with ~100% efficiency (and between 35 and 460 nm with 50% efficiency). Here we denote the AMS aerosol data as "submicron" mass (based on the more usual definition using aerodynamic diameter, which is larger than the geometric diameter; DeCarlo et al., 2004), with the assumption that non-refractory aerosol are small contributors to mass above the AMS size range. As shown in Brock et al., 2019, the accumulation mode for the ATom sampling environment only extended up to 500 nm, and hence, as expected for a background tropospheric environment, this approximation is appropriate. Very good agreement was observed with the integrated volume calculated from the number size distributions for ATom (Brock et al., 2019). A low bias compared to a typical submicron definition can occur in thick biomass burning plumes and in the lower stratosphere at times (Jimenez et al., 2019b). As detailed in Table 1, the accumulation mode for the bulk models discussed in this study overlaps with the size range of the AMS, and for the sectional models (CESM1-CARMA, GEOS-Chem-TOMAS, ECHAM6-HAM) only the bins that match the AMS size range were used. As expected based on the previous discussion, however, a comparison of the total OA calculated by these sectional models with the modeled OA inside the AMS size-range showed small differences (Slopes for ATom-1 linear regressions: CESM1-CARMA:0.91, GC10-TOMAS: 0.94, ECHAM6-HAM 1.00) mostly influenced by the high concentration points in the biomass plumes off Africa that have a large effect on the regression since they are about 10 times larger than the bulk of the dataset).

Refractory and non-refractory aerosol composition was also measured using the NOAA Particle Analysis by Laser Mass Spectrometry (PALMS) instrument. PALMS classifies individual aerosol particles into compositional classes including biomass burning (Hudson et al., 2004), sea salt (Murphy et al., 2019), mineral dust (Froyd et al., 2019), and others. Mass concentrations for these particles types are derived by combining PALMS

composition data with aerosol size distribution measurements (Froyd et al., 2019). Good agreement overall was found for OA, sulfate and seasalt between the two particle mass spectrometers during ATom once the AMS and PALMS instrument transmissions were accounted for (Jimenez et al., 2019b). For all PALMS data used in this work (biomass burning fraction and dust) the AMS transmission function was applied to ensure that both instruments were characterizing approximately the same particle range.

For a particular airmass, the mass fraction of biomass burning (BB) aerosol reported by the PALMS instrument $f(BB)_{PALMS}$ (Thompson and Murphy, 2000; Froyd et al., 2019) was then used to evaluate the degree of BB influence. This parameter correlates quite well with other gas-phase BB tracers, and is more useful as a particle tracer since its lifetime follows that of the particles. Importantly, it is not impacted by the long lifetimes of the gas-phase tracers (e.g. 9 months for CH_3CN) and unrelated removal processes (e.g. ocean uptake for CH_3CN and HCN) that result in highly variable backgrounds. Hence $f(BB)_{PALMS}$ has a much higher contrast ratio and linearity for particle BB impacts, compared to the available gas-phase tracers in the ATom dataset. An airmass was classified as non-BB influenced when $f(BB)_{PALMS}$ was lower than 0.30 (Hudson et al, 2004) as shown in Figure 2b. For both ATom-1 and 2, about 74% of measurements were classified as not influenced by biomass burning. $f(BB)_{PALMS}$ was also used to assess the impact of POA on the total OA burden (next section); note that no thresholding was applied in that case.

Deleted: This parameter correlates quite well with other gas-phase BB tracers, and is more useful as a particle tracer. Importantly, it is not impacted by the long lifetimes of the gas-phase tracers (e.g. 9 months for CH_3CN) and has hence a much higher contrast ratio in the ATom dataset. An airmass was classified as non-BB influenced when $f(BB)_{PALMS}$ was lower than 0.30 (Hudson et al, 2004) as shown in Figure 2b. It was also used to assess the impact of POA on the total OA burden (next section).

3.2 Estimation of the POA fraction for the ATom dataset

For model evaluation purposes, it is important to know whether the source of OA is primary or secondary. For ground studies close to sources (e.g. Jimenez et al., 2009) Positive Matrix Factorization of AMS mass spectra (PMF, Ulbrich et al., 2009) can be used to estimate the contribution of primary sources (mostly from transportation, heating, cooking, and biomass burning) to total OA. This approach is not suitable for ATom. To accurately resolve a minor factor such as POA in an AMS dataset, there needs to be a combination of: (a) Sufficient OA mass concentration, so that the signal-to-noise of the spectra is sufficient; (b) Enough fractional mass for the factor to be resolved (>5% in urban areas per Ulbrich et al. (2009), probably a larger fraction at low concentrations such as in ATom); (c) Sufficient spatio-temporal variability ("contrast") in the relative contributions of different factors, since that is part of what PMF uses to extract the factors; (d) Sufficient difference in the spectra of the different factors (for the same reason as (c)), and (e) relatively invariant spectra for each factor across the dataset (as this is a key assumption of the

PMF algorithm). As an example of a near ideal case, in Hodshire et al (2019) we extracted MSA by PMF from the ATom-1 data, and were able to match that factor with our independently calibrated MSA species. A very distinct and nearly invariant mass spectrum was measured repeatedly near sources (MBL) (and was mostly absent elsewhere, thus providing strong spatio-temporal contrast) and accounted for about 6% of the fractional mass and 15% of the variance in time. Thus all the conditions were met. For POA, on the other hand, the air sampled in ATom and coming from e.g. Asia has POA and SOA very well mixed, with little change on their relative mass fractions vs. time (as the aircraft flies through that airmass). POA is very low, as documented later in this paper. Atmospheric aging makes the spectra from all OA sources more and more similar as measured by AMS spectra (Jimenez et al., 2009). Thus most of the conditions above are not satisfied for extracting POA by PMF analysis of this dataset.

Instead, in this work we have estimated POA based on the fact that it is co-emitted with BC as part of the combustion processes releasing both species in source regions, and that BC is not impacted by chemical aging processes over the lifetime of the airmass. Note that BC can physically age but it is not lost in any significant amount to the gas-phase due to chemical processes in the atmosphere. We assume non-differential removal (and transport) of the BC fraction relative to the rest of the POA (the two are generally internally mixed, Lee et al., 2015). Table S1 summarizes recent POA/BC and POC/EC emission ratio determinations for urban background sites, which best represent real mixes of pollution sources, and for individual sources of POA (from mobile sources – commonly referred as HOA – and cooking aerosol – COA). Based on Table S1 data, we assume POA to be co-emitted with BC for anthropogenic fossil fuel / urban region POA (herein called FF_{ratio} for simplicity, even though much of it is non-fossil, Zotter et al., 2014; Hayes et al., 2015) at a ratio of 1.5 (average of all urban ambient air studies that report POA and BC for best intercomparability to the ATom dataset; including all urban studies results in a very similar number, 1.48). Mobile source measurements in general exhibit lower ratios (POA/OA ratio 0.5-1.5) while COA determination typically ranges from 2 to 3. Hence, the ratio used here is a good estimate for a diverse mix of urban sources as appropriate for ATom. The studies used to derive the emission ratio used ambient data in urban air, where all sources mix together and impact the POA/BC ratio, and thus the ratios include the impact of POA sources that may not emit BC. It should be noted that urban model ratios do not include emissions associated with fugitive dust from road, tire and construction, as those are typically found in larger particles than those studied here (Zhao et al., 2017).

Deleted: This approach is not suitable for ATom, since the aerosol is too aged to be effectively classified into POA and SOA by PMF. However, PMF analyses can still be useful for other purposes, such as identifying MSA, as previously observed in remote environments (Hodshire et al., 2019a). ¶

For biomass burning sources, we use a value of POA/BC = 11.8 (BB_{ratio}), based on the average of the recent review by Andreae (2019), which included over 200 previous determinations for a variety of fuels and burning conditions (since Andreae (2019) used and OA/OC ratio of 1.6 in his work, we have used that value to calculate POA/BC; we note that this is different from the 1.8 OA/OC ratio used for other studies listed in Table S1). We note the measured total OA/BC of ~3.5 (conservatively assuming that all OA is POA) observed on both ATom missions for the large African-sourced BB plumes over the Equatorial Atlantic. We note that using the larger BB_{ratio} from Andreae (2019) leads to a POA fraction >> 100% in the ATom African plumes. We also perform sensitivity studies with values of both FF_{ratio} and BB_{ratio} within the literature range.

The PALMS determined mass fraction of biomass impacted aerosol ($f(BB)_{PALMS}$) can then be used to determine a total POA contribution from both types of sources:

$$POA = [BC]*(FF_{ratio}+(BB_{ratio}- FF_{ratio})*f(BB)_{PALMS}) \quad \text{(Eq. 1)}$$

Further detail is provided in Table S2, which summarizes the POA/BC ratios used in the emission inventories implemented in current models. Overall, there is reasonable agreement with the measurements in Table S1, with FF_{ratio} ranging from ~0.5 for diesel fuels, to >2 for energy production and ~5 for residential emissions (which include some BB). On the other hand, for biomass burning, the emission inventories ratios range from ~5 for crop, to ~15 for forest, and up to ~50 for peatland. While generally consistent with the values discussed by Andreae (2019), they are on the lower end of the ranges discussed in that work. The averages and ranges of the measurement and model ratios are similar, and thus no significant model bias on the ratios is apparent.

PALMS detection efficiency increases with size across the accumulation mode, and therefore the f(BB) number fraction is weighted to the larger size end of the accumulation mode. In very clean regions of the upper troposphere (typically <0.15 μg sm^{-3} submicron mass) particles below the PALMS size range can contribute significantly to aerosol mass (Williamson et al., 2019; Jimenez et al., 2019b). If BB particles are not evenly distributed across the entire accumulation mode (due to preferential removal in convective updrafts of primary aerosol, cf. Yu et al., 2019 and Section 4.5; and preferential condensation of SOA on smaller particles), then the f(BB) reported by PALMS will be an overestimation. For the final analysis these periods where left in the dataset, and therefore for the LS the reported POA is likely overestimated for these regions, although their impact on the mass-weighted campaign average is negligible.

The contribution of POA from sea spray is difficult to constrain. As an order-of-magnitude estimate, marine POA is roughly calculated based on preliminary calibrations of OA on mineral dust particles from the PALMS instrument (personal communication K. Froyd). Using this calibration, the average OA by mass on sea salt was <10% for the large majority of MBL sampling (>85%). Since sea salt contributed 4% (11%) of mass in the AMS size range for ATom-1(2) (Figure 2), we estimate that marine POA is on the order of ~1% of aerosol mass in the AMS size range, and possibly much lower. Thus we think that it is reasonable to neglect the contribution of marine POA to this dataset. Future studies will refine this estimate.

3.3 Data processing for comparisons

For the comparisons between the measurements and the various global models, data were averaged both vertically and zonally to minimize the impact of smaller plumes or vertical gradients in aerosol concentrations that might not be captured by coarse resolution models. For the same reason, all data near airports was removed from the datasets prior to analysis (up to about 3 km on the climb in/out). In order to restrict this analysis to the remote troposphere, the last leg of the ATom-1 mission (over the continental US) was taken out of the dataset as well. Data was binned into 5 large latitude regions as shown in Figure 2a including southern polar (55-80°S, "S.Polar"), southern mid-latitudes (25-55°S, "S.Mid"), equatorial (25°S-25°N, "Equatorial"), northern mid-latitudes (25-55°N, "N.Mid"), northern polar (55-80°N, "N.Polar") and analyzed separately for the Pacific and Atlantic basins. For data in each of these latitude regions, altitude profiles were calculated with a constant 600 m altitude resolution. According to both variability in the cleanest air and statistical analysis of the organic background subtraction (Drewnick et al. 2009), the 1σ precision at low concentrations for one-minute data ranged between 20 and 50 ng sm^{-3}, or a 3σ detection limit between 60 and 150 ng sm^{-3} for the one-minute data (confirmed by frequent filter blanks). Per standard statistics, the precision of a measurement decreases (i.e., gets better) with the square root of the number of points (or time interval) sampled. I.e. the precision of an average can be approximated by the standard error of the mean (σ/sqrt(*n*), where *n* is the number of measurements averaged), and it is better than the precision of the individual data points (σ). This also applies to the detection limit, since it is just 3 times the precision. Note that a detection limit is not meaningful unless the averaging time is specified. For example let's assume that the detection limit is 20 ng m^{-3} (1-second), and the data points over 60 consecutive seconds are all 10 ng m^{-3}. All 1-

second measurements are below the 1-second DL. However the average (10 ng m^{-3}) is now above the DL for 1-minute averages, which is 20/sqrt(60) = 2.6 ng m^{-3}. On average, each individual point in the profiles represents the average of about 25 min of ATom flight data. At that time resolution, the OA 1σ precision was about 10 ng sm^{-3}. Hence with very few exceptions (10 points for both missions combined), the OA concentrations in the averaged profiles reported are well above the instrumental detection limit in those regions. For model-measurement comparisons along flight tracks, model outputs and measurements were considered at 1-minute time resolution, which corresponds to ~0-700 m vertical resolution and ~0.05-0.15 degrees horizontal resolution. Note that a large fraction of the 1-minute OA values in the remote free troposphere were below the local 3σ detection limit. The data of periods of zero concentration (sampling ambient air through a particle filter) do average to zero. Some negative measurements are present, and this is normal for measurements of very low concentrations in the presence of instrumental noise. Averaging of longer periods, as done for the figures in this paper, reduces the detection limit. We therefore caution future data users that the reported data should be averaged as needed, as replacing below-detection limit (or negative) values by other values introduces biases on averages. For fractional ratio analysis, measurements were averaged to 5-minute time resolution to reduce the noise in the ratios due to noise in the denominator. The results are not very sensitive to the 5-minute averaging (compared to 1-minute) as shown in Figure S12 for OA to sulfate ratios. The same figure also illustrates that excluding ratios affected by negative concentrations (the non-bracketed case, overall these are about 15% of the dataset) does not really affect the fractional distribution, with the variance between the two cases diminishing as the averaging interval increases. To further confirm that there is no inherent bias in the fractional products regardless of the treatment of low concentration values, an additional sensitivity analysis was performed where data was filtered by an independent measurement proxy for aerosol mass, the aerosol volume measured in ATom (Brock et al., 2019). Using a range of value that encompasses the regime where the AMS calculated volume to aerosol measured volume exhibited increased noise (Jimenez et al., 2019b), no systematic bias was found (Figure S13), with variations of about 10% in fractional volume for different filtering conditions.

Some of the performed analysis required separating the dataset into vertical subsets. In this manuscript, we define the marine boundary layer (MBL) as the region below 1.5 times the calculated boundary layer height in the NCEP global model reanalysis. The free troposphere (FT) includes all data points between the top of MBL and the NCEP

tropopause height, and the LS region includes all points above the NCEP tropopause height. The tropopause height varied during ATom between 8 and 16.5 km; given the DC-8 ceiling (42 kft, 12.8 km) the stratosphere was only sampled at latitudes higher than 30 degrees in both hemispheres. The MBL height varied between up to 1.5 km in the mid-latitudes, ~1 km in the tropics, and sometimes <150 m (lowest DC-8 altitude) for some of the sampling in the polar troposphere.

3.4 Submicron aerosol composition

Figure 2b shows that during both NH summer and winter ATom deployments, OA is one of the three dominant components of the measured submicron aerosol in the remote troposphere, together with sulfate and sea salt. During ATom-1, average submicron aerosol concentrations were close to 0.8 μg sm^{-3} in the marine boundary layer and biomass burning outflow regions, and ~2 times lower in the free troposphere and lower stratosphere regions. ATom-2 had overall lower average concentrations below 0.4 μg sm^{-3} (vs. 0.5 μg sm^{-3} for ATom-1). As expected, sulfate (sulfuric acid in the lower stratosphere) is the dominant constituent in the MBL (~50%) and LS (50-70%), while the OA contribution is generally below 10% and 40%, respectively in those regions. A large fraction of sea salt aerosol is found in the MBL especially during the NH winter deployment (~30%, see Murphy et al., 2019).

OA is found to be a major constituent (~50%) of submicron aerosol in the clean (non-BB influenced) free troposphere. The contribution of OA is 1.4 times larger than that of sulfate during the NH summer, and 1.2 times lower than that of sulfate during the NH winter, which is likely due to a large contribution of the NH sources to SOA production in the NH summer. Biomass-burning events increase the OA contribution relative to that of sulfate, and lead to a higher contribution of OA to total during the ATom-1 mission (stronger BB influence).

3.5 Spatial and vertical distribution of OA

Figure 2a (and Fig. S1) shows the spatial and vertical distribution of OA mass concentrations measured during ATom-1 (and ATom-2) campaigns. Most data were taken over remote oceanic regions (and a few remote continental regions, primarily over the Arctic). The measured OA varies between extremely clean conditions (< 0.1 μg sm^{-3}) encountered mostly in the Pacific and Southern Ocean regions and moderately polluted conditions (> 2 μg sm^{-3}) in the biomass burning outflow regions. During ATom-1 (August

2016), a strong BB influence is observed in the lower troposphere (below 6 km) over the Atlantic basin off the African coast and over California with OA concentrations exceeding 10 μg sm^{-3}. OA associated with biomass burning is also present in the upper troposphere over equatorial regions and over Alaska, associated with the deep convective transport of biomass burning aerosols. The biomass burning contribution to carbonaceous aerosols in those regions during ATom-1 was also apparent in the black carbon measurements (Katich et al., 2019). ATom-2 was generally less polluted than ATom-1, likely due to a more limited global influence of biomass burning emissions during that period, and also to a less active photochemistry during winter months in the NH.

The measured OA is characterized by a strong latitudinal gradient. Figure 2c shows the average vertical profiles of measured OA over the selected latitudinal bands during August 2016. The cleanest airmasses are observed over the remote oceanic regions of the Southern Hemisphere (SH, 25-80°S) with OA mass concentrations below 0.06 μg sm^{-3}. These extremely low OA concentrations can be explained by the very low influence from continental emission sources, and presumably low marine POA and SOA precursor emissions. This is consistent with low concentrations of gas-phase pollutants (e.g. CO, ethane, propane). An enhancement can be noticed above 10 km in the lower stratosphere. In some cases, this could be related to the long-range transport of biomass burning aerosols from the tropics. By comparison, the Arctic region is more polluted with an order of magnitude higher OA levels compared to its analog of the SH (i.e. OA loadings ranging from 0.1 to 0.5 μg sm^{-3}). These concentrations are comparable to FT levels measured in the extratropical regions (25-55°N) of the NH. The equatorial marine regions (25°S-25°N) display the highest OA concentrations with a strong gradient between lower and upper troposphere. In the lower troposphere OA, concentrations are close to 1 μg sm^{-3}, and decrease down to 0.1 μg sm^{-3} at altitudes above 4km. The highest OA levels are associated with the African outflow over the southeastern Atlantic Ocean, which results from the transport of the biomass burning smoke from the sub-Saharan regions and increasing urban and industrial air pollution in southern West Africa (Flamant et al., 2018). Figure 2d shows that the Atlantic basin is often more polluted than the Pacific basin, not only because of the African biomass burning influence but also due to the contribution of anthropogenic pollution in the lower troposphere of the NH. It should be noted that Asian pollution was likely an important contributor to the North Pacific Basin, especially between 2 and 6 km, in both ATom deployments (see figures 2a and S1). Several-fold higher OA

concentrations are found near the surface (below 1km) over the southern Pacific compared to that same location in the southern Atlantic, which could be indicative of the stronger emission of marine OA in the Pacific basin.

In addition to spatial gradients, a strong summer-to-winter contrast is observed in OA concentrations. Figure 2e shows the ratio between OA vertical profiles measured in the NH summer ATom-1 vs. in the NH winter ATom-2. The NH is more polluted during the NH summer due to the photochemical production of SOA, as well as biomass burning emissions, leading to the tripling of OA concentrations in the extratropical regions (25-80°N) on average regardless of altitude. The doubling of OA loading in the lower troposphere at the equator (25°S-25°N) in the NH summer (August, ATom-1) is strongly influenced by the biomass burning activity in the sub-Saharan African region as already mentioned above. Likewise, OA concentrations are found to be generally higher in the SH during the SH summer. These zonal trends are broadly similar to the ones described in Katich et al (2018) for BC.

4 Model-measurement comparisons

4.1 Evaluation of predicted OA concentrations

Prior to evaluating model performance in simulating OA, we have assessed the ATom models' ability to simulate sulfate aerosols. According to the model evaluation shown in Table S3, the predicted sulfate concentrations are generally within 40% of the measured values, which is comparable to the AMS measurement uncertainties. The only exception is found for the ECHAM6-HAM model, which overestimates sulfate aerosols by a factor of two. These results imply that most ATom models capture relatively well the overall sulfate burden. However, large root mean square error (RMSE > 0.4 μg sm^{-3} for ATom-1 and > 0.2 μg sm^{-3} for ATom-2) is indicative of their limited skill in reproducing the observed variability in sulfate concentrations.

For OA, model evaluation metrics for the entire ATom-1 and ATom-2 campaigns are given in Table 2 for the eight ATom models and their ensemble, as well as the AeroCom-II ensemble. The results show that the normalized mean bias is substantially lower for the ATom model ensemble compared to AeroCom-II decreasing from 74% to 4% for ATom-1 and from 137% to 23% to ATom-2, which is within the measurement uncertainty range. The mean temporal correlations are substantially improved from 0.31 (0.38) for AeroCom-II to 0.66 (0.48) for ATom model ensemble during ATom-1 (ATom-2). However, results

vary strongly among ATom models. Models using prescribed emissions of non-volatile SOA have the tendency to overestimate the OA concentrations during both NH summer and winter deployments (with ~35-60% overestimation for CESM2-SMP, ~70-100% for ECHAM6-HAM, and up to 150% for GC10-TOMAS during ATom-2), with the exception of the GEOS5 model that on the contrary underestimates OA concentrations by 5-25%. During the NH summer (ATom-1), models using the VBS parameterization from Pye et al. (2010) tend to underpredict the OA concentrations by 43% for GC12-REF and 33% for CESM1-CARMA for ATom-1, most likely due to the excessive evaporation of the formed SOA in remote regions and low yields for anthropogenic SOA (Schroder et al., 2018; Shah et al., 2019). Models using the VBS parameterization from Hodzic et al. (2016) (CESM2-DYN and GC12-DYN) where OA is less volatile and also OA yields are corrected for wall losses show an improved agreement with observations especially for CESM2-DYN (with NMB of ~5%), and to a lesser extent for GC12-DYN (NMB of ~33%). During the NH winter (ATom-2) characterized by a lower production of SOA, both VBS approaches lead to an overestimation of the predicted OA. This is likely caused by excessively high levels of primary emitted OA as discussed in section 4.4.

Deleted: -GOCART

Figure 3 compares the average median ratios between modeled and observed OA concentrations for the ATom and AeroCom-II model ensembles for different regions (BB, MBL, FT, LS). The results show that the median ratio for the ATom model ensemble is close to unity in all regions. This is at least a factor of two improvement compared to AeroCom-II models, which were almost always biased high for the remote regions sampled in ATom. The model spread has also been reduced by a factor of 2-3 in all regions. This reduction in the ensemble spread may partially be explained by a smaller size of the ATom model ensemble (see Fig. S2), which also includes models with a more up-to-date OA representation. In order to explore this point further, results for a subset of AeroCom-II models (using earlier versions of models in the ATom ensemble) show only a slight reduction (~10%) in the model spread, with however some regional differences i.e. an improved agreement with observations in the MBL, but an increase in the model bias and spread in the LS (Figure S2). Thus, model improvement for the more recent models appears to be the main reason for the reduced spread.

4.2 Evaluation of predicted OA vertical distribution

Figure 4 compares the mean vertical profiles of OA measured during ATom-1 and -2 with the predictions of the model ensemble average based on the eight ATom models (Table

1) and 28 AeroCom-II models for the different latitudinal regions of the Pacific and Atlantic basins. Note that the use of a wide logarithmic scale (to be able to span all the observations) may make the observed differences appear small, although they often reach factors of 2-10 and larger (Figure S5 shows the results on a linear scale). For AeroCom-II, large latitudinal differences exist in the results with a better performance closer to source regions and large disagreement in the lower stratosphere and remote regions, as already suggested by the mission medians shown in Figure 3. The best AeroCom-II model performance is found over the equator in both basins, where the model ensemble captures within a factor of 2 the observed OA concentrations throughout the troposphere in the Pacific basin, and matches remarkably well the observations in the lower troposphere of the Atlantic basin that is heavily influenced by biomass burning emissions. Reasonable agreement is found for the OA vertical distribution over the NH Atlantic and Pacific oceans, especially in the lower troposphere (< 4 km). The largest model discrepancies (1-2 orders of magnitude) are found in the remote regions of the Southern Ocean and SH mid-latitudes during both seasons and basins. The model overestimation is also large over the NH mid-latitude Pacific basin in the upper troposphere. A spread of 2-3 orders of magnitude is observed around the ensemble average indicating a very large variability in individual model predictions. This evaluation of AeroCom-II models in remote regions is an extension of that performed at the surface for urban and remote stations by Tsigaridis et al. (2014) (as in that previous study, the data and model simulations compared are not synchronous in time). The tendency of the model ensemble to overpredict OA concentrations by a factor of 2 on average in the remote regions is consistent with the transition from the large underprediction in OA near the source region to a slight overprediction of OA in remote continental sites that was reported for most AeroCom-II models (Tsigaridis et al., 2014), and also observed for default parameterizations in other studies (Heald et al., 2011; Hodzic et al., 2016).

By comparison, the results of the ATom model ensemble show a much better agreement with observations. The model spread is still substantial, but mostly below a factor of 5. Figures S6 and S7 show OA vertical profiles for individual ATom models and the spread in their results. In most regions, the ATom model ensemble captures reasonably well both the absolute concentrations as well as the shape of the vertical profiles. In the biomass burning outflow and NH mid-latitude regions, the ATom ensemble average better captures the higher OA concentrations in the boundary layer and lower OA concentrations in the lower stratosphere than the AeroCom-II ensemble. We note that using the ensemble

median OA profiles instead of ensemble mean OA profiles (as shown in Figure 5 and S7) results in a slightly lower values of OA but does not change the conclusions of the model-measurement comparisons (Figure S18).

4.3 Oxidation level of organic aerosols (OA/OC ratios)

In addition to OA mass concentrations, we also evaluate the model's ability to simulate their degree of oxygenation, an indicator of their oxidation and aging (Aiken et al., 2008; Kroll et al., 2011). Ambient measurements of the oxidation level of organic particles are limited (Aiken et al., 2008, Canagaratna et al., 2015), and the ATom dataset provides the first global distribution of O/C and OA/OC ratios for the remote aerosol. The OA/OC ratio is an estimate of the average molecular weight of organic matter per carbon weight, and it mostly depends on the oxygen content (i.e. the O/C ratio), in the absence of significant concentrations of organonitrates and -sulfates. It is needed to compare measurements of organic aerosol mass (from e.g. AMS) with organic carbon measurements (from e.g. thermooptical methods). It is also needed to compare the various types of measurements to model concentrations, which are sometimes carried internally as OA and sometimes as OC. A low OA/OC ratio is indicative of freshly emitted OA from fossil fuel combustion (typically ~1.4), and its value increases with increased processing of organics in the atmosphere. Figure 5 shows that in the remote regions the bulk of measured OA/OC ratios during ATom-1 and -2 range between 2.2 and 2.5, and are larger than values of 2.1 ± 0.2 found in the polluted US continental outflow regions that were sampled during SEAC4RS, WINTER and DC3 field campaigns (Schroder et al., 2018). These values indicate that remote OA is highly oxidized and chemically processed.

Note that for organosulfates (R-O-SO_2H and organonitrates (R-O-NO_2, $pRONO_2$ in the following) only one oxygen is included in the reported OA/OC, as the fragments of these species are typically the same as for inorganic species in the AMS (Farmer et al., 2010). However in ATom organosulfates are estimated to account for ~1% of the total sulfate (based on PALMS data, see Liao et al., 2015 for the methodology). Since sulfate and OA concentrations are comparable, organosulfates would only increase the OA/OC by ~1% on average. Organonitrates are reported from the AMS for ATom. Their impact on OA/OC is not propagated for the default values, to maintain consistency with a large set of OA/OC measurements by AMS in the literature, and since they would increase OA/OC on average by only 4.5% (ATom-1) and 2.2% (ATom-2), which is smaller than the uncertainty of this

measurement. However, we show the results with both methods in Fig. 5 to fully document this topic.

Importantly, this ratio is also used to calculate the total OA mass concentration for models that provided their outputs in terms of organic carbon concentrations ($[OA]_i = [OC]_i \times OA/OC_{ratio}$). Most Models use a constant OA/OC ratio, but the value used varies substantially. OA/OC of 1.4 is used in ECHAM6-HAM, whereas 1.8 is used in GEOS5 and GC10-TOMAS simulations for both POA and SOA. Other models calculated directly SOA concentrations without applying this conversion (CESM1-CARMA, CESM2-SMP, CESM2-DYN, GC12-REF and GC12-DYN), but for POA used the ratio of 1.8 (CESM1-CARMA, CESM2-SMP, CESM2-DYN) and 2.1 (GC12-REF and GC12-DYN). Most of the AeroCom-II models used the ratio of 1.4 for all primary and secondary OA (Tsigaridis et al., 2014). The comparison with measurements shows that the measured values are ~40% larger than those assumed in some of the ATom models, and 60-80% larger than used in AeroCom-II models. The comparison between the observed and predicted OA/OC vertical profiles (Fig. S3) shows that AeroCom-II models tend to generally underpredict this ratio, and do not capture its increase in remote regions. As a result, this underestimation of OA/OC ratios and the use of a constant value could substantially impact the comparisons of OA mass concentrations for several models considered in this study (ECHAM6-HAM, GEOS5, CESM1-CARMA and GC10-TOMAS). If we correct for the underestimated OA/OC ratio using the ATom measured values of 2.2 (to be conservative) and compare to previously discussed biases in Table 2, the overprediction of the ECHAM6-HAM model is increased to ~110-160%, and that of GC10-TOMAS to 180% during ATom-2 while having ~15% bias in ATom-1, whereas GEOS5 results now overestimate up to 30% during ATom1, and perform much better during ATom-2.

Deleted: -GOCART

Deleted: -GOCART

Deleted: -GOCART

These results demonstrate that current global chemistry-climate models use unrealistically low OA/OC ratios, which results in a large underestimate of the degree of oxidation of OA in remote regions. Inaccurate prediction of OA oxidation as it ages could impact not only the calculations of OA burden, but also its optical properties as the absorption of OA changes with its degree of oxidation (through the formation and destruction of brown carbon, Laskin et al., 2015, Forrister et al., 2015). However, models used in this study did not include these effects.

4.4 Contribution of primary vs. secondary OA

We further assess whether global models can adequately predict the relative contributions of primary and secondary OA. We strive to quantify these fractions with the most straightforward methods (with the fewest assumptions) for both models and measurements. POA concentrations were estimated from the BC measurements by using an emission ratio appropriate to the airmass origin (biomass burning vs. anthropogenic), and using the f(BB) mass fraction from the PALMS single particle instrument (see Section 3.2). By using the POA/BC ratio at the source regions after most evaporation, but before POA chemical degradation evaporation has taken place, we implicitly assume POA to be chemically inert, while in reality it can slowly be lost to the gas-phase by heterogeneous chemistry (e.g. George and Abbatt, 2010; Palm et al., 2018). Thus, the observation-based method provides an upper limit to the fraction of POA. The model/measurement comparison is only shown for the CESM and GEOS-Chem model variants, as other participating models do not separate or did not report their POA and SOA fractions. In all simulations, POA was treated as a chemically inert directly emitted primary aerosol species that only undergoes transport, transformation from hydrophobic to hydrophilic state with ageing (1-2 days typically), coagulation, and dry and wet deposition. Importantly, the treatment of POA as non-volatile (rather than semi-volatile) in models is fully consistent with the assumptions for POA estimation from the measurements.

Figure 6 compares the vertical profiles of measurement-derived POA during ATom-1 and predicted by the CESM2-DYN model over clean remote regions of the Pacific basin and northern polar Atlantic that are not influenced by biomass burning. Comparisons for other models are similar (not shown). Observations show that POA is extremely small in remote regions, whereas the model predicts that about half of the OA is made of POA in those areas. Although the model reproduces quite well the measured total OA, it tends to severely overpredict the amount of POA and underpredict that of SOA over clean remote regions (with the two errors canceling each other when it comes to total OA). Over the biomass burning regions (not shown here) it can be difficult to directly quantify POA and SOA with this method, as total OA remains about constant, while POA decreases with aging and SOA increases (Cubison et al., 2011; Jolleys et al., 2015; Hodshire et al., 2019b). However, given this evolution the method used here would lead to an overestimate of POA for this reason.

A more general comparison is made in Figure 7, using the frequency distributions of the measured and simulated fraction of POA/OA, for the free troposphere only (Figure S8 shows the corresponding cumulative distributions). Observations indicate that most remote FT airmasses contain less than 10% POA, except for biomass burning plumes that are considered mostly primary. A slightly higher proportion of POA is seen in ATom-2, which is consistent with a slower photochemical production of SOA during NH winter. These results indicate that the remote OA is consistently dominated by SOA regardless of the season and location. The comparison with models reveals a very large discrepancy in the predicted vs. measured POA vs. SOA contributions. Models have a general tendency to severely overpredict the fraction of POA and underpredict that of SOA, displaying a much wider frequency distribution than the measurements (as also shown for POA and SOA vertical profiles for individual models on Figures S6 and S7). In GC12-REF, CESM2-DYN and CESM1-CARMA (without improved in-cloud removal) predictions for ATom-1, more than a half of the remote OA is POA, while that is very rarely observed in the free troposphere (possibly only during strong biomass burning events). Most models fail to reproduce the overwhelming dominance of SOA that is inferred from the measurements during ATom-1, while the discrepancies are less severe during NH winter (ATom-2). These seasonal differences suggest that model errors could be partially due to inefficient production of SOA, although removal errors also probably play a major role (see next section).

The differences are so large that they are pretty insensitive to details of the POA estimation method from the measurements, mostly because for the vast majority of the ATom track BC/OA ratios were extremely low and hence the exact magnitude of the multiplicative factor is secondary to the estimation of POA (Figure S11). As Figure S9 illustrates, the choice of FF_{ratio} has very little impact on the overall distribution of POA. On the other hand, while the BB_{ratio} does impact the overall distribution of POA, it mostly affects the points in the vicinity of the large Atlantic plumes. Since the POA/BC ratio in those plumes is fairly low, (see Section 3.2), using a very large BB_{ratio} mostly leads to an increase of the fraction of the points where POA > 100%. While the large range of published BB_{ratio} for different sources precludes a more accurate estimation by our method, for the purposes of the comparison with the model results we emphasize that even using the largest BB_{ratio}, fraction of SOA is still significantly larger in the ATom dataset that in any of the models.

Additional sensitivity tests were performed to investigate the impact of noisy data and uncertainties of f(BB) on the estimation of POA. Figure S11 clearly shows that the impact of a misattribution of the aerosol type by the stated PALMS uncertainty (Froyd et al., 2019) is completely negligible. Figure S10 details how the choice of averaging interval (with longer averaging times reducing both the fraction of OA measurements under the DL and below zero) impact the distribution of POA. Overall, no large changes are observed for averaging times >5 min, and hence a 5 min averaging interval was used for the analysis in Figure 7. Figure S10 also illustrates how capping the histogram impacts the POA distribution. To capture the most realistic $f_{(POA)}$ distribution, the data in Fig 7 was capped at the extremes (so $f_{(POA)}<0$ is taken as $f_{(POA)}=0$, and $f_{(POA)}>1$ is taken as $f_{(POA)}=1$). As Figure S10 shows, data with $f_{(POA)}<0$ is almost exclusively due to very small (and always positive, since BC cannot go negative) POA values being divided by small, negative noise in total OA, and hence treating that fraction of the histogram as essentially fPOA~0 is justified. On the other end of the distribution, data where POA is larger than OA is mostly due to our average BB_{ratio} being larger than the one encountered in most of the BB plumes in ATom. Choosing a lower BB_{ratio}, as Figures S9b and S9d illustrate, leads to $f_{(POA)}>1$ basically trending to zero, confirming our interpretation. This is a limitation of the dataset, and it does not seem appropriate to remove these points, since some fraction are likely dominated by POA. However, it shows that the POA estimation, especially for this part of the distribution likely overstates the importance of POA.

A comparison between simulations that have the same treatment of POA, and only differ in their chemistry and removal of SOA (e.g. CESM2-SMP vs. CESM2-DYN; GC12-REF vs. GC12-DYN) indicate that a more complex SOA treatment does not always result in a more accurate simulation of the primary/secondary character of OA, a result that was also found in the AeroCom-II multi-model intercomparison (Tsigaridis et al., 2014).

Finally, we have examined whether the non-volatile treatment of POA in models could lead to these unrealistically high POA fractions in the remote regions. Figure S16 shows a comparison of POA vertical profiles as predicted by the GC12-REF simulations that use non-volatile POA and a sensitivity simulation GC12-REF-SVPOA that uses semi-volatile POA similar to the standard treatment in GEOS-Chem as described in Pai et al. (2020). Note, however, that Pai et al. (2020) included marine POA emissions, used different reanalysis meteorology, and a different model version (12.1.1 rather than 12.0.1 here), so their resulting comparisons to ATom measurements are somewhat different than found

here for GC12-REF-SVPOA. The comparison indicates that the POA concentrations increase substantially in most regions when the semi-volatile POA parameterization is used. These results suggest that non-volatile treatment of POA is not responsible of the model bias.

4.5 Sensitivity to OA formation and removal

In this section, we further investigate some of the possible reasons for the incorrect model predictions of the relative contributions of POA and SOA in remote regions. Given the tendency of models to underestimate OA close to anthropogenic source regions and overestimate OA downwind in past studies (e.g. Heald et al., 2011; Tsigaridis et al., 2014; Hodzic et al., 2016), in this section we investigate the sensitivity of OA to increasing sources and increasing removals. We have performed two additional model simulations to test the sensitivity of the POA/SOA fractions to uncertainties in the representation of (i) wet scavenging, based on the CESM1-CARMA simulations in which we have removed the improvements in the aerosol removal by the convective updrafts (Yu et al., 2019); and of (ii) SOA formation based on the GC12-REF simulations in which we have replaced the SOA formation VBS mechanism (Pye et al., 2010) by an updated VBS mechanism that uses chamber wall-loss corrected SOA yields (Hodzic et al., 2016, same formation scheme that is used in GC12-DYN and CESM2-DYN runs, but with removals kept identical to GC12-REF). The results of these two sensitivity simulations are displayed on Figure 8, which shows measured and predicted mass concentrations of OA, POA, SOA and sulfate for ATom-1 as a function of the number of days since the air mass was processed through convection. One should keep in mind that this is an averaged plot that included airmasses from various regions and altitudes, and not a Lagrangian plot following the same airmass.

Sensitivity to in-cloud scavenging in convective clouds. Inefficient wet removal of primary OA could contribute to the POA overprediction in global models, especially in the tropics. Previous global model studies have reported two to three orders of magnitude overestimation of primary carbonaceous species such as BC in the free troposphere when the removal in the convective updrafts was not included (e.g. Schwarz et al., 2013, Yu et al., 2019). A strong reduction due to convective removal is also expected for POA concentrations, as POA is a primary species co-emitted with BC at the surface and internally mixed with it (Lee et al., 2015), and that is typically coated by secondary inorganics and organics over short timescales (Petters et al., 2006; Jiang et al., 2010; Wang et al., 2010). Figures 7a and 8 compare the simulations of CESM1-CARMA with

and without improved convective in-cloud scavenging during ATom-1. The improved in-cloud scavenging scheme considers aerosol activation into cloud droplets from entrained air above the cloud base, which is more realistic and results in a more efficient removal of aerosols in the upper troposphere by convection. E.g. a two order of magnitude reduction in BC in the upper FT was reported by Yu et al. (2019), resulting in much improved agreement with observations. Similar results were observed for sea salt aerosols in Murphy et al. (2019). Figure 8 shows that all submicron aerosol species simulated in CESM1-CARMA are strongly impacted by the in-cloud removal above the cloud base. POA concentrations are reduced by an order of magnitude while sulfate is reduced by 30% leading in both cases to a much-improved agreement with observations. SOA is reduced by ~30% as well, which leads to an underprediction of measured SOA concentrations. The overall impact on OA concentrations is a significant reduction, which leads to ~20% underestimation of OA in the aged remote air during ATom-1.

For the CESM2-DYN model that does not have improved in-cloud removal, the reasonable agreement (within 20%) with the observed OA concentrations thus results from coincidental error compensation between the overpredicted POA and underpredicted SOA. The prescribed SOA formation and the artificial 50% adjustment of SOA emissions based on Liu et al. (2012) in CESM2-SMP leads to an overestimation of observed SOA in aged remote airmasses.

Sensitivity to SOA formation. In addition, we have also tested the sensitivity of the OA composition to the choice of the SOA formation mechanism. Figure 8 compares the results of the GC12-REF model that uses SOA formation yields derived from traditional chamber experiments (Pye et al., 2010) and those corrected for loses of organic vapors onto chamber walls as proposed in Hodzic et al. (2016). Previous studies have reported that chamber wall losses could lead up to a factor of 4 underprediction of formed SOA (Zhang et al., 2014; Krechmer et al., 2016). It should be noted that in both cases, isoprene-SOA is formed in aqueous aerosols following Marais et al. (2016). The comparison shows a factor of 3 increase in SOA concentrations when the updated SOA formation is considered leading to a much better agreement with the observed SOA as well as the observed total OA. GC12-REF predicts well the amount of POA and overpredicts somewhat the amount of sulfate aerosols, which is expected as it already includes the improved aerosol removal in convective updrafts (Wang et al., 2014). Figure S6 also shows that POA vertical distribution is well captured in GEOS-Chem in most regions, except over the polar north

Pacific. It should be noted that these results are consistent with the POA/OA frequency distribution shown in Figure 7 (the POA/OA ratio predicted by GC12-REF is larger than the measured ratio, which is consistent with the fact that POA is about the right amount, and OA is underpredicted in Figure 8).

These sensitivity simulations suggest that a stronger convective removal of POA and a stronger production of SOA might be needed to correctly represent not only the total OA concentrations but also its primary and secondary nature in remote free troposphere and remote ocean regions. Accurate predictions of the OA concentration, composition, and source contributions for the right reasons are key for accurately predicting their lifecycle and radiative impacts. Only when there is confidence that the sources are accurately predicted, we can have confidence in OA predictions for pre-industrial and future conditions, as well as to evaluate PM mitigation strategies.

4.6 OA and sulfate relative contributions in FT

Finally, we assess the model ability to predict relative amounts of OA and sulfate in the free troposphere where they are the two major constituents of the submicron aerosol (Figure 2b). Accurate predictions of their relative contributions are crucial to determine the hygroscopicity of the submicron aerosol, and its ability to serve as a cloud condensation nuclei (CCN) in the remote free troposphere (Carslaw et al., 2013; Brock et al., 2016).

Figure 9a compares the average measured relative fractions of sulfate (36%) and carbonaceous aerosols (OA=59% and BC=5%) in the FT with those predicted by individual models during ATom-1. The CESM2 models best reproduce the observed relative contributions, with a slight underestimation of OA (57% instead of 59%) for CESM2-DYN, and a slight overestimation of OA (63% instead of 59%) for CESM2-SMP. GEOS5 has 15% more OA relative to sulfate than observed. All other models underestimate both OA and BC relative fractions. For instance, in GC12-REF and -DYN, both the BC and OA fractions are ~40% (relative) lower than observed.

Deleted: -GOCART

Figure 9b shows the frequency distribution of observed and predicted fractions of OA relative to sulfate during ATom-1 and -2 in the free troposphere. Most models fail to reproduce the relatively uniform nature of the observed distributions during ATom-1, with typically narrower model shapes around a preferred ratio. The NH summer measurements indicate that OA > sulfate in ~55% of the samples (consistent with Fig. 2b), while models generally tend to underestimate the relative OA contribution. In particular, GEOS-Chem

and ECHAM6-HAM tend to overestimate the relative contribution of sulfate. A better agreement is found for GEOS5, CESM1-CARMA and CESM2-DYN, which follow more closely the shape of the observed distribution. The comparisons also suggest that the more complex SOA treatment of SOA formation and removal proposed by Hodzic et al. (2016) in the same host model leads to an improved agreement with observations (e.g. CESM2-DYN vs. CESM2-SMP; GC12-DYN vs. GC12-REF). It should be noted that CESM2-SMP uses fixed SOA yields that were increased by 50% as suggested by Liu et al. (2012), leading to an overestimation of the relative contribution of OA compared to that of sulfate in the free troposphere. During the NH winter (ATom-2), measurements show a somewhat higher proportion of sulfate aerosols (vs. ATom-1), which is consistent with a slower production of SOA in the NH during winter and a reduced influence of biomass burning. Similar conclusions are found for the evaluation of different models. It is worth mentioning that the comparison performed for the whole ATom-1 and 2 dataset (not shown) leads to similar results with even slightly stronger overestimation of the sulfate relative contribution compared to OA.

The discrepancies between the observed and predicted composition of submicron aerosol over remote regions can be quite large for other constituents as well. Figure 10 shows the comparison of measured and predicted composition of the submicron aerosol over the Southern Ocean (during the NH winter) where the disagreement in simulated sea salt, nitrates, ammonium, and MSA often exceeds the contribution of OA. While the observations show a more uniform distribution of non-marine aerosol with higher values in the mid and upper troposphere, respectively, most models tend to simulate highest fractions of OA (and sulfate) towards the tropopause. This may also be explained by the uncertainties in modeled wet removal of aerosol that has been discussed above. Specific studies have discussed and continue to investigate the ATom measurements and simulations of different components in more detail, including particle number (Williamson et al., 2019), black carbon (Katich et al., 2018; Ditas et al., 2019), MSA (Hodshire et al., 2019), sulfate-nitrate-ammonium (Nault et al., 2019), and sea salt (Yu et al, 2019; Bian et al., 2019; Murphy et al., 2019).

Deleted: -GOCART

5 Conclusions and implications

Our understanding and representation in global models of the lifecycle of the OA remain highly uncertain, especially in remote regions where constraints from measurements have been very sparse. We have performed a systematic evaluation of the performance of eight global chemistry climate models and of 28 AeroCom-II models in simulating the latitudinal and vertical distribution of OA and its composition in the remote regions of the Atlantic and Pacific marine boundary layer, free troposphere and lower stratosphere, using the unique measurements from the ATom campaign. Our simulations are conducted for both ATom-1 and ATom-2 deployments that took place in August 2016 and February 2017, respectively. The main conclusions of the comparison are as follows:

- The AeroCom-II ensemble average tends to be biased high by a factor of 2-5 in comparison to measured vertical OA profiles in the remote atmosphere during both NH summer and NH winter. The ensemble spread increases from a factor of 40 in the NH source regions to a factor of 1000 in remote regions of the Southern Ocean. The evaluation of AeroCom-II models in the remote regions provides an extension of the previous evaluation with continental ground data by Tsigaridis et al. (2014). We note that the data from the AeroCom-II models were based on monthly mean values from a different simulated year than the ATom campaigns; however, the consistent model biases are strong enough that we would not expect our conclusions to change for a different modeled year.
- The results of the ATom model ensemble used in this work show a much better agreement with the OA observations in all regions and reduced model variability. However, some of the agreement is for the wrong reasons, as most models severely overestimate the contribution of POA and underestimate the contribution of SOA to total OA. Sensitivity simulations indicate that the POA overestimate in CESM could be due to an inadequate representation of primary aerosol removal by convective clouds, (additional convective removal per Yu et al. (2019) in CESM1-CARMA led to a better agreement with observations). Most models have insufficient production of SOA, and sensitivity studies indicate that a stronger production of SOA is needed to capture the measured concentrations. The photochemical ageing of POA which was not considered here (unlike for SOA) could also contribute to the model overestimation. The non-volatile POA treatment in models is consistent with the assumption of inert POA particles used to estimate POA from measurements, and cannot explain the

model bias. Indeed, sensitivity simulations with semi-volatile POA lead to a much larger model bias for OA in the upper troposphere and remote regions. The compensation between errors in POA and SOA in remote regions is however a recurring issue in OA modeling (de Gouw and Jimenez, 2009). For instance, it was found in the urban outflow regions such as Mexico City during MILAGRO 2006 field campaign (Fast et al., 2009; Hodzic et al., 2009); Paris during MEGAPOLI 2009 (Zhang et al., 2013); the Los Angeles area during CalNex-2010 (Baker et al., 2015; Woody et al., 2016); the NE US outflow during WINTER 2015 (Schroder et al., 2018; Shah et al., 2019).

- Additional errors in simulated OA concentrations can arise from the use of too low OA/OC ratios when model results (often calculated as OC) are converted to OA for comparison with measurements. We note that OA is the most atmospherically-relevant quantity, while OC is an operational quantity, partially a relic from a period in which only OC could be separately quantified (although also of some use for carbon budget studies). It should also be noted that most emission inventories still use OC as the primary variable, which is why the use of accurate OA/OC ratios is still key for all models. We show that the OA/OC ratio used in most models is too low compared to measured values that range mostly from 2.2 to 2.5, resulting in errors in OA mass of ~70% for AeroCom-II models and ~30% for current models that use organic carbon to track OA mass. Remote OA is thus highly oxidized and chemically processed. These results demonstrate that current global chemistry-climate models underestimate the degree of oxidation of OA in remote regions and need to consider further chemical ageing of OA, which could impact the calculations of its burden, and optical and hygroscopic properties.
- The results also show that in most models (except CESM2) the predicted OA contribution to the total submicron aerosol is underestimated relative to sulfate in the remote FT where OA and sulfate are the dominant submicron aerosols (important for climate). Accurate predictions of composition of submicron particles remains challenging in remote regions and should be the topic of future studies.

Key implications of our results are: (i) Model errors on the relative contribution of POA and SOA to OA reduce our confidence on the ability to simulate radiative forcing over time or OA health impacts; (ii) Model errors for the relative contributions of sulfate and organics to the submicron aerosol in the free troposphere could lead to errors in the predicted CCN or radiative forcing of aerosols as inorganics are more hygroscopic than OA; (iii) the OA

system seems to be more dynamic with a need for an enhanced removal of primary OA, and a stronger production of secondary OA in global models to provide a better agreement with observations.

Acknowledgements. The authors want to thank the ATom leadership team and the NASA logistics and flight crew for their contributions to the success of ATom. Authors acknowledge Dr. Rebecca Buchholz (NCAR) for providing the emissions used for the CESM2 simulations. The ATOM measurements and analyses were supported by NASA grants NNX15AH33A, NNX15AJ23G, and 80NSSC19K0124. AH was supported by the National Center for Atmospheric Research, which is operated by the University Corporation for Atmospheric Research on behalf of the National Science Foundation. JRP and JKK were supported by the US Department of Energy's Atmospheric System Research, an Office of Science, Office of Biological and Environmental Research program, under grant no. DE-SC0019000. This project has received support from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819169), and from EPA STAR grant 83587701-0. This manuscript has not been reviewed by EPA, and no endorsement should be inferred. We would like to acknowledge high-performance computing support from Cheyenne provided by NCAR's Computational and Information Systems Laboratory. We thank C. Brock for the aerosol volume data, and D. Murphy for useful discussions. We thank the ATom leadership team, science team and the NASA DC-8 crew for their contributions to the success of the ATom mission.

Code/Data availability: Data can be obtained from the ATom website: https://doi.org/10.3334/ORNLDAAC/1581.

L2 Measurements from CU High-Resolution Aerosol Mass Spectrometer (HR-AMS) can be obtained from the ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1716.

Author contribution: A. Hodzic, P. Campuzano-Jost and J.L. Jimenez performed the measurement / model comparisons, wrote and revised the manuscript. P. Campuzano-Jost, D.A. Day, B.N. Nault, J.C. Schroder, D.T. Sueper, and J. L. Jimenez performed and analyzed the AMS measurements. K.D. Froyd and G.P. Schill performed and analyzed the PALMS measurements. J.P. Schwarz and J.M. Katich performed the BC measurements. H. Bian, M. Chin, P.R. Colarco, B. Heinold, A. Hodzic, D.S. Jo, J.K.

Kodros, J.R. Pierce, E. Ray, J. Schacht, I. Tegen, S. Tilmes, K. Tsigaridis, and P. Yu provided model output. All authors provided comments on the manuscript.

References

Aiken, A.C., P.F. DeCarlo, J.H. Kroll, D.R. Worsnop, J.A. Huffman, K. Docherty, I.M. Ulbrich, C. Mohr, J.R. Kimmel, D. Sueper, Q. Zhang, Y. Sun, A. Trimborn, M. Northway, P.J. Ziemann, M.R. Canagaratna, T.B. Onasch, R. Alfarra, A.S.H. Prevot, J. Dommen, J. Duplissy, A. Metzger, U. Baltensperger, and J.L. Jimenez. O/C and OM/OC Ratios of Primary, Secondary, and Ambient Organic Aerosols with High Resolution Time-of-Flight Aerosol Mass Spectrometry Environmental Science & Technology, 42, 4478–4485, doi: 10.1021/es703009q, 2008.

Andreae, M. O.: Emission of trace gases and aerosols from biomass burning – An updated assessment, Atmos. Chem. Phys. Discuss., 1–27, doi:10.5194/acp-2019-303, 2019.

Bahreini, R., Ervens, B., Middlebrook, A. M., Warneke, C., de Gouw, J. A., DeCarlo, P. F., Jimenez, J. L., Brock, C. A., Neuman, J. A., Ryerson, T. B., Stark, H., Atlas, E., Brioude, J., Fried, A., Holloway, J. S., Peischl, J., Richter, D., Walega, J., Weibring, P., Wollny, A. G. and Fehsenfeld, F. C.: Organic aerosol formation in urban and industrial plumes near Houston and Dallas, Texas, J. Geophys. Res., 114, D00F16, doi:10.1029/2008JD011493, 2009.

Baker, K. R., Carlton, A. G., Kleindienst, T. E., Offenberg, J. H., Beaver, M. R., Gentner, D. R., Goldstein, A. H., Hayes, P. L., Jimenez, J. L., Gilman, J. B., de Gouw, J. A., Woody, M. C., Pye, H. O. T., Kelly, J. T., Lewandowski, M., Jaoui, M., Stevens, P. S., Brune, W. H., Lin, Y.-H., Rubitschun, C. L., and Surratt, J. D.: Gas and aerosol carbon in California: comparison of measurements and model predictions in Pasadena and Bakersfield, Atmos. Chem. Phys., 15, 5243-5258, https://doi.org/10.5194/acp-15-5243-2015, 2015.

Bey, I., Jacob, D. J., Yantosca, R. M., and Logan, J. A.: Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation, J Geophys. Res., 106, 23073–23095, 2011.

Bian, H., Froyd, K., Murphy, D. M., Dibb, J., Chin, M., Colarco, P. R., Darmenov, A., da Silva, A., Kucsera, T. L., Schill, G., Yu, H., Bui, P., Dollner, M., Weinzierl, B., and Smirnov, A.: Observationally constrained analysis of sea salt aerosol in the marine atmosphere, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-18, 2019.

2653 Bian, H., Chin, M., Hauglustaine, D. A., Schulz, M., Myhre, G., Bauer, S. E., Lund, M. T.,
2654 Karydis, V. A., Kucsera, T. L., Pan, X., Pozzer, A., Skeie, R. B., Steenrod, S. D., Sudo, K.,
2655 Tsigaridis, K., Tsimpidi, A. P., and Tsyro, S. G.: Investigation of global nitrate from the
2656 AeroCom Phase III experiment, Atmos. Chem. Phys., 17, 12911-12940,
2657 https://doi.org/10.5194/acp-17-12911-2017, 2017.

2658 Boucher, O., D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M.
2659 Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S.K. Satheesh, S. Sherwood, B.
2660 Stevens and X.Y. Zhang: Clouds and Aerosols. In: Climate Change 2013: The Physical
2661 Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the
2662 Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M.
2663 Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)].
2664 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

2665 Bowman, K. P.: Large-scale isentropic mixing properties of the Antarctic polar vortex from
2666 analyzed winds. J. Geophys. Res., 98 (D12), 23 013–23 027, 1993.

2667 Brock, C. A., Wagner, N. L., Anderson, B. E., Attwood, A. R., Beyersdorf, A., Campuzano-
2668 Jost, P., Carlton, A. G., Day, D. A., Diskin, G. S., Gordon, T. D., Jimenez, J. L., Lack, D.
2669 A., Liao, J., Markovic, M. Z., Middlebrook, A. M., Ng, N. L., Perring, A. E., Richardson, M.
2670 S., Schwarz, J. P., Washenfelder, R. A., Welti, A., Xu, L., Ziemba, L. D., and Murphy, D.
2671 M.: Aerosol optical properties in the southeastern United States in summer – Part 1:
2672 Hygroscopic growth, Atmos. Chem. Phys., 16, 4987-5007, https://doi.org/10.5194/acp-16-
2673 4987-2016, 2016.

2674 Brock, C. A., Williamson, C., Kupc, A., Froyd, K. D., Erdesz, F., Wagner, N., Richardson,
2675 M., Schwarz, J. P., Gao, R.-S., Katich, J. M., Campuzano-Jost, P., Nault, B. A., Schroder,
2676 J. C., Jimenez, J. L., Weinzierl, B., Dollner, M., Bui, T. and Murphy, D. M.: Aerosol size
2677 distributions during the Atmospheric Tomography Mission (ATom): methods,
2678 uncertainties, and data products, Atmos. Meas. Tech., 12(6), 3081–3099,
2679 doi:10.5194/amt-12-3081-2019, 2019.

2680 Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q.,
2681 Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A. M., Delia, A., Williams, L. R.,
2682 Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and Worsnop,
2683 D. R.: Chemical and microphysical characterization of ambient aerosols with the Aerodyne

Aerosol Mass Spectrometer, Mass Spectrom. Rev., 26, 185–222, https://doi.org/10.1002/mas.20115, 2007.

Canagaratna, M. R., Jimenez, J. L., Kroll, J. H., Chen, Q., Kessler, S. H., Massoli, P., Hildebrandt Ruiz, L., Fortner, E., Williams, L. R., Wilson, K. R., Surratt, J. D., Donahue, N. M., Jayne, J. T., and Worsnop, D. R.: Elemental ratio measurements of organic compounds using aerosol mass spectrometry: characterization, improved calibration, and implications, Atmos. Chem. Phys., 15, 253-272, https://doi.org/10.5194/acp-15-253-2015, 2015.

Carslaw, K. S., Lee, L. a, Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V, Woodhouse, M. T., Regayre, L. a and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing., Nature, 503(7474), 67–71, doi:10.1038/nature12674, 2013.

Colarco, P., da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground based aerosol optical depth, J. Geophys. Res., 115, D14207, doi:10.1029/2009JD012820, 2010.

Croft, B., Lohmann, U., Martin, R. V., Stier, P., Wurzler, S., Feichter, J., Hoose, C., Heikkilä, U., van Donkelaar, A., and Ferrachat, S.: Influences of in-cloud aerosol scavenging parameterizations on aerosol concentrations and wet deposition in ECHAM5-HAM, Atmos. Chem. Phys., 10, 1511–1543, https://doi.org/10.5194/acp-10-1511-2010, 2010.

Cubison, M. J., Ortega, A. M., Hayes, P. L., Farmer, D. K., Day, D., Lechner, M. J., Brune, W. H., Apel, E., Diskin, G. S., Fisher, J. A., Fuelberg, H. E., Hecobian, A., Knapp, D. J., Mikoviny, T., Riemer, D., Sachse, G. W., Sessions, W., Weber, R. J., Weinheimer, A. J., Wisthaler, A., and Jimenez, J. L.: Effects of aging on organic aerosol from open biomass burning smoke in aircraft and laboratory studies, Atmos. Chem. Phys., 11, 12049-12064, https://doi.org/10.5194/acp-11-12049-2011, 2011.

DeCarlo, P., Slowik, J., Worsnop, D., Davidovits, P., Jimenez, J., Stainken, K., Williams, L., Jayne, J., Kolb, C. and Rudich, Y.: Particle Morphology and Density Characterization by Combined Mobility and Aerodynamic Diameter Measurements. Part 1: Theory, Aerosol Sci. Technol., 38(12), 1185–1205, doi:10.1080/027868290903907, 2004.

DeCarlo, P. F., Kimmel, J. R., Trimborn, A., Northway, M. J., Jayne, J. T., Aiken, A. C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K. S., Worsnop, D. R., and Jimenez, J. L.:

Field-Deployable, High-Resolution, Time-of-Flight Aerosol Mass Spectrometer, Anal. Chem., 78, 8281–8289, https://doi.org/10.1021/ac061249n, 2006.

de Gouw, J., and J.L. Jimenez. Organic Aerosols in the Earth's Atmosphere. Environmental Science & Technology, 43, 7614–7618, 2009. DOI: 10.1021/es9006004

Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., Ginoux, P., Gong, S., Hoelzemann, J. J., Ito, A., Marelli, L., Penner, J. E., Putaud, J.-P., Textor, C., Schulz, M., van der Werf, G. R., and Wilson, J.: Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom, Atmospheric Chemistry and Physics, 6, 4321–4344, https://doi.org/10.5194/acp- 6-4321-2006, https://www.atmos-chem-phys.net/6/4321/2006/, 2006.

Ditas, J., Ma, N., Zhang, Y., Assmann, D., et al.: Strong impact of wildfires on the abundance and aging of black carbon in the lowermost stratosphere, Proc. Natl. Acad. Sci., 811595-11603, doi:10.1073/pnas.1806868115, 2018.

Drewnick, F., Hings, S. S., Alfarra, M. R., Prevot, A. S. H. and Borrmann, S.: Aerosol quantification with the Aerodyne Aerosol Mass Spectrometer: detection limits and ionizer background effects, Atmos. Meas. Tech., 2(1), 33–46, 2009.

Emmons, L. K., Orlando, J. J., Tyndall, G., Schwantes, R. H., Kinnison, D. E., Marsh, D. R., Mills, M. J., Tilmes, S., and Lamarque, J.-F.: The MOZART Chemistry Mechanism in the Community Earth System Model version 2 (CESM2), to be Submitted to J. Adv. Modeling Earth Systems, 2019.

Farmer, D.K., A. Matsunaga, K.S. Docherty, J.D. Surratt, J.H. Seinfeld, P.J. Ziemann, and J.L. Jimenez. Response of an Aerosol Mass Spectrometer to Organonitrates and Organosulfates and implications for Atmospheric Chemistry. Proceedings of the National Academy of Sciences of the USA, 107, 6670-6675, doi: 10.1073/pnas.0912340107, 2010.

Fast, J., Aiken, A. C., Allan, J., Alexander, L., Campos, T., Canagaratna, M. R., Chapman, E., DeCarlo, P. F., de Foy, B., Gaffney, J., de Gouw, J., Doran, J. C., Emmons, L., Hodzic, A., Herndon, S. C., Huey, G., Jayne, J. T., Jimenez, J. L., Kleinman, L., Kuster, W., Marley, N., Russell, L., Ochoa, C., Onasch, T. B., Pekour, M., Song, C., Ulbrich, I. M., Warneke, C., Welsh-Bon, D., Wiedinmyer, C., Worsnop, D. R., Yu, X.-Y., and Zaveri, R.: Evaluating simulated primary anthropogenic and biomass burning organic aerosols during MILAGRO: implications for assessing treatments of secondary organic aerosols, Atmos. Chem. Phys., 9, 6191-6215, https://doi.org/10.5194/acp-9-6191-2009, 2009.

Feng, L., Smith, S. J., Braun, C., Crippa, M., Gidden, M. J., Hoesly, R., Klimont, Z., van Marle, M., van den Berg, M., and van der Werf, G. R.: Gridded Emissions for CMIP6, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-195, in review, 2019.

Flamant, C., Knippertz, P., Fink, A. H., Akpo, A., Brooks, B., Chiu, C. J., Coe, H., Danuor, S., Evans, M., Jegede, O., Kalthoff, N., Konaré, A., Liousse, C., Lohou, F., Mari, C., Schlager, H., Schwarzenboeck, A., Adler, B., Amekudzi, L., Aryee, J., Ayoola, M., Batenburg, A. M., Bessardon, G., Borrmann, S., Brito, J., Bower, K., Burnet, F., Catoire, V., Colomb, A., Den- jean, C., Fosu-Amankwah, K., Hill, P. G., Lee, J., Lothon, M., Maranan, M., Marsham, J., Meynadier, R., Ngamini, J.-B., Rosenberg, P., Sauer, D., Smith, V., Stratmann, G., Taylor, J. W., Voigt, C., and Yoboué, V.: The Dynamics-Aerosol-Chemistry- Cloud Interactions in West Africa field campaign: Overview and research highlights, B. Am. Meteorol. Soc., 99, 83–104, https://doi.org/10.1175/BAMS-D-16-0256.1, 2018.

Forrister, H., Liu, J., Scheuer, E., Dibb, J., Ziemba, L., Thornhill, K. L., Anderson, B., Diskin, G., Perring, A. E., Schwarz, J. P., Campuzano-Jost, P., Day, D. A., Palm, B. B., Jimenez, J. L., Nenes, A. and Weber, R. J.: Evolution of brown carbon in wildfire plumes, Geophys. Res. Lett., 42(11), doi:10.1002/2015GL063897, 2015.

Fountoukis, C. and Nenes, A.: ISORROPIA II : a computationally efficient thermodynamic equilibrium model for K+-Ca2+-Mg2+-Na+-SO42--NO3--Cl--H2O aerosols, Atmos. Chem. Phys., 7, 4639–4659, 2007.

Froyd, K. D., Murphy, D. M., Brock, C. A., Campuzano-Jost, P., Dibb, J. E., Jimenez, J.-L., Kupc, A., Middlebrook, A. M., Schill, G. P., Thornhill, K. L., Williamson, C. J., Wilson, J. C., and Ziemba, L. D.: A new method to quantify mineral dust and other aerosol species from aircraft platforms using single particle mass spectrometry, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2019-165, in review, 2019.

George, I. J. and Abbatt, J. P. D.: Heterogeneous oxidation of atmospheric aerosol particles by gas-phase radicals, Nature Chemistry, 2, 713–722, 2010.

Gettelman, A., Mills, M. J., Kinnison, D. e., Garcia, R. R., Smith, A. K., Marsh, D. R., Tilmes, S., Vitt, F., Bardeen, C. G., McInerny, J., Liu, H.-L., Solomon, S. C., Polvani, L. M., Emmons, L. K., Lamarque, J.-F., Richter, J. H., Glanville, A. S., Bacmeister, J. T., Phillips, A. S., Neale, R. B., Simpson, I. R., DuVivier, A. K., Hodzic, A., Randel, W. J.: The

Whole Atmosphere Community Climate Model Version 6 (WACCM6), J. of Geophys. Res., in review, 2019.

Gerber, H. E.: Relative-humidity parameterization of the Navy Aerosol Model (NAM), Tech. Rep. NRL Report 8956, Naval Research Laboratory, 1985.

Giglio, L., Randerson, J. T. and Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), J. Geophys. Res. Biogeosciences, 118, 317–328, 2013.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, https://www.geosci-model-dev.net/5/1471/2012/, 2012.

Hayes, P. L., Carlton, A. G., Baker, K. R., Ahmadov, R., Washenfelder, R. A., Alvarez, S., Rappenglück, B., Gilman, J. B., Kuster, W. C., de Gouw, J. A., Zotter, P., Prévôt, A. S. H., Szidat, S., Kleindienst, T. E., Offenberg, J. H., Ma, P. K., and Jimenez, J. L.: Modeling the formation and aging of secondary organic aerosols in Los Angeles during CalNex 2010, Atmos. Chem. Phys., 15, 5773-5801, https://doi.org/10.5194/acp-15-5773-2015, 2015.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, Th. F., Monod, A., Prévôt, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmos. Chem. Phys., 9, 5155-5236, https://doi.org/10.5194/acp-9-5155-2009, 2009.

Heald, C. L., Coe, H., Jimenez, J. L., Weber, R. J., Bahreini, R., Middlebrook, A. M., Russell, L. M., Jolleys, M., Fu, T.-M., Al- Ian, J. D., Bower, K. N., Capes, G., Crosier, J., Morgan, W. T., Robinson, N. H., Williams, P. I., Cubison, M. J., DeCarlo, P. F., and Dunlea, E. J.: Exploring the vertical profile of atmospheric organic aerosol: comparing 17 aircraft field campaigns with a global model, Atmos. Chem. Phys., 11, 12673–12696, doi:10.5194/acp-11-12673-2011, 2011.

Hodshire, A. L., Campuzano-Jost, P., Kodros, J. K., Croft, B., Nault, B. A., Schroder, J. C., Jimenez, J. L. and Pierce, J. R.: The potential role of methanesulfonic acid (MSA) in

aerosol formation and growth and the associated radiative forcings, Atmos. Chem. Phys., 19(5), 3137–3160, doi:10.5194/acp-19-3137-2019, 2019a.

Hodshire, A., A. Akherati, M. Alvarado, B. Brown-Steiner, S. Jathar, J.L. Jimenez, S. Kreidenweis, C. Lonsdale, T. Onasch, A. Ortega, J. Pierce. Aging Effects on Biomass Burning Aerosol Mass and Composition: A Critical Review of Field and Laboratory Studies. Environ. Sci. Technol., submitted, 2019b.

Hodzic, A., Jimenez, J. L., Madronich, S., Aiken, A. C., Bessagnet, B., Curci, G., Fast, J., Lamarque, J.-F., Onasch, T. B., Roux, G., Schauer, J. J., Stone, E. A., and Ulbrich, I. M.: Modeling organic aerosols during MILAGRO: importance of biogenic secondary organic aerosols, Atmos. Chem. Phys., 9, 6949-6981, https://doi.org/10.5194/acp-9-6949-2009, 2009.

Hodzic, A., Aumont, B., Knote, C., Lee-Taylor, J., Madronich, S., and Tyndall, G.: Volatility dependence of Henry's law constants of condensable organics: Application to estimate depositional loss of secondary organic aerosols, Geophys. Res. Lett., 41, 4795–4804, doi:10.1002/2014GL060649, 2014.

Hodzic, A., Madronich, S., Kasibhatla, P. S., Tyndall, G., Aumont, B., Jimenez, J. L., Lee-Taylor, J., and Orlando, J.: Organic photolysis reactions in tropospheric aerosols: effect on secondary organic aerosol formation and lifetime, Atmos. Chem. Phys., 15, 9253-9269, https://doi.org/10.5194/acp-15-9253-2015, 2015.

Hodzic, A., Kasibhatla, P. S., Jo, D. S., Cappa, C. D., Jimenez, J. L., Madronich, S., and Park, R. J.: Rethinking the global secondary organic aerosol (SOA) budget: stronger production, faster removal, shorter lifetime, Atmos. Chem. Phys., 16, 7917-7941, https://doi.org/10.5194/acp-16-7917-2016, 2016.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Huang, K., Fu, J. S., Prikhodko, V. Y., Storey, J. M., Romanov, A., Hodson, E. L., Cresko, J., Morozova, I., Ignatieva, Y., and Cabaniss, J.: Russian anthropogenic black carbon:

Emission reconstruction and Arctic black carbon simulation, J. Geophys. Res. Atmos., 120, 11,306– 11,333, doi:10.1002/2015JD023358, 2015.

Hudson, P. K., Murphy, D. M., Cziczo, D. J., Thomson, D. S., de Gouw, J. A., Warneke, C., Holloway, J., Jost, H. J. and Hübler, G.: Biomass-burning particle measurements: Characteristics composition and chemical processing, J. Geophys. Res. D Atmos., 109(23), 1–11, doi:10.1029/2003JD004398, 2004.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmüller, R., Denier van der Gon, H., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411–11432, https://doi.org/10.5194/acp-15-11411-2015, 2015.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, a. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, a. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, a. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. a., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, a. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., Worsnop, D. R., Worsnop, D. R., Dunlea, J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, a. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., Worsnop, D. R., Dunlea, E. J., et al.: Evolution of Organic Aerosols in the Atmosphere, Science 80., 326(5959), 1525–1529, doi:10.1126/science.1180353, 2009.

Jimenez, J.L., P. Campuzano-Jost, D.A. Day, B.A. Nault, D.J. Price, and J.C. Schroder: ATom: L2 Measurements from CU High-Resolution Aerosol Mass Spectrometer (HR-

AMS). ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1716, 2019a.

Jimenez, J.L., et al.: Evaluating the Consistency of All Submicron Aerosol Mass Measurements (Total and Speciated) in the Atmospheric Tomography Mission (ATom), Abstract A31A-08, presented at 2019 Fall Meeting, AGU, San Francisco, CA, 9-13 Dec., 2019b.

Jolleys, M. D., Coe, H., McFiggans, G., Taylor, J. W., O'Shea, S. J., Le Breton, M., Bauguitte, S. J.-B., Moller, S., Di Carlo, P., Aruffo, E., Palmer, P. I., Lee, J. D., Percival, C. J. and Gallagher, M. W.: Properties and evolution of biomass burning organic aerosol from Canadian boreal forest fires, Atmos. Chem. Phys., 15(6), 3077–3095, doi:10.5194/acp-15-3077-2015, 2015.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527–554, https://doi.org/10.5194/bg-9-527-2012, https://www.biogeosciences.net/9/527/2012/, 2012.

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Vignati, E., Stephanou, E. G., and Wilson, J.: Organic aerosol and global climate modelling: a review, Atmos. Chem. Phys., 5, 1053-1123, https://doi.org/10.5194/acp-5-1053-2005, 2005.

Katich, J. M., Samset, B. H., Paul Bui, T., Dollner, M., Froyd, K.D., Campuzano-Jost, P., Nault, B. A., Schroder, J. C., Weinzierl, B., Schwarz, J. P.: Strong Contrast in Remote Black Carbon Aerosol Loadings Between the Atlantic and Pacific Basins, J. Geophys. Res. Atmos., 123 (23), pages 13,386-13,395, https://doi.org/10.1029/2018JD029206, 2018.

Kim, M. J., G. A. Novak, M. C. Zoerb,M. Yang, B. W. Blomquist, B. J. Huebert, C. D. Cappa, and T. H. Bertram: Air-Sea exchange of biogenic volatile organic compounds and the impact on aerosol particle size distributions, Geophys. Res. Lett.,44, 3887–3896, doi:10.1002/2017GL072975, 2017.

Klimont, Z., K. Kupiainen, C. Heyes, P. Purohit, J. Cofala, P. Rafaj, J. Borken-Kleefeld, and W. Schöpp,: Global anthropogenic emissions of particulate matter including black carbon, Atmos. Chem. Phys., 17(14), 8681-8723, doi: 10.5194/acp-17-8681-2017, 2017.

Kodros, J. K., Cucinotta, R., Ridley, D. A., Wiedinmyer, C., and Pierce, J. R.: The aerosol radiative effects of uncontrolled combustion of domestic waste, Atmos. Chem. Phys., 16, 6771-6784, https://doi.org/10.5194/acp-16-6771-2016, 2016.

Krechmer, J.E., D. Pagonis, P.J. Ziemann, and J.L. Jimenez. Quantification of gas-wall partitioning in Teflon environmental chambers using rapid bursts of low-volatility oxidized species generated in-situ. Environmental Science and Technology, 50, 5757–5765, doi:10.1021/acs.est.6b00606, 2016.

Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E. and Worsnop, D. R.: Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol., Nat. Chem., 3(2), 133–9, doi:10.1038/nchem.948, 2011.

Laskin, A., Laskin, J., and Nizkorodov, S.: Chemistry of Atmospheric Brown Carbon, Chem. Rev. 2015, 115, 10, 4335-4382, 2015.

Lee, A. K. Y., Willis, M. D., Healy, R. M., Onasch, T. B., and Abbatt, J. P. D.: Mixing state of carbonaceous aerosol in an urban environment: single particle characterization using the soot particle aerosol mass spectrometer (SP-AMS), Atmos. Chem. Phys., 15, 1823-1841, https://doi.org/10.5194/acp-15-1823-2015, 2015.

Liao, J., K.D. Froyd, D.M. Murphy, F.N. Keutsch, G. Yu, P.O.Wennberg, J.St. Clair, J.D. Crounse, A. Wisthaler, T. Mikoviny, T.B. Ryerson, I.B. Pollack, J. Peischl, J.L. Jimenez, P. Campuzano Jost, D.A. Day, B.E. Anderson, L.D. Ziemba, D.R. Blake, S. Meinardi, G. Diskin. Airborne organosulfates measurements over the continental US. Journal of Geophysical Research-Atmospheres, 120, 2990–3005, doi:10.1002/2014JD022378, 2015.

Liu, H. Y., D. J. Jacob, I. Bey, and R. M. Yantosca: Constraints from Pb-210 and Be-7 on wet deposition and transport in a global three-dimensional chemical tracer model driven by assimilated meteorological fields, J. Geophys. Res., 106(D11), 12,109–12,128, 2001.

Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A., Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A. M. L., Hess, P., Mahowald, N., Collins, W., Iacono, M. J., Bretherton, C. S., Flanner, M. G., and Mitchell, D.: Toward a minimal representation of aerosols in climate models: description and evaluation in the Community Atmosphere Model CAM5, Geosci. Model Dev., 5, 709-739, https://doi.org/10.5194/gmd-5-709-2012, 2012.

Marais, E. A., Jacob, D. J., Jimenez, J. L., Campuzano-Jost, P., Day, D. A., Hu, W., Krechmer, J., Zhu, L., Kim, P. S., Miller, C. C., Fisher, J. A., Travis, K., Yu, K., Hanisco, T. F., Wolfe, G. M., Arkinson, H. L., Pye, H. O. T., Froyd, K. D., Liao, J., and McNeill, V. F.: Aqueous-phase mechanism for secondary organic aerosol formation from isoprene: application to the southeast United States and co-benefit of SO2 emission controls, Atmos. Chem. Phys., 16, 1603-1618, https://doi.org/10.5194/acp-16-1603-2016, 2016.

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J.-H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, Atmos. Chem. Phys., 13, 1853-1877, https://doi.org/10.5194/acp-13-1853-2013, 2013.

Mauderly, J. L., & Chow, J. C: Health Effects of Organic Aerosols, Inhalation Toxicology, 20:3, 257-288, DOI: 10.1080/08958370701866008, 2008.

Middlebrook, A. M., Murphy, D. M. and Thomson, D. S.: Observations of organic material in individual marine particles at Cape Grim during the First Aerosol Characterization Experiment (ACE 1), J. Geophys. Res. Atmos., 103(D13), 16475–16483, doi:10.1029/97JD03719, 1998.

Murphy, D. M., Froyd, K. D., Bian, H., Brock, C. A., Dibb, J. E., DiGangi, J. P., Diskin, G., Dollner, M., Kupc, A., Scheuer, E. M., Schill, G. P., Weinzierl, B., Williamson, C. J., and Yu, P.: The distribution of sea-salt aerosol in the global troposphere, Atmos. Chem. Phys., 19, 4093-4104, https://doi.org/10.5194/acp-19-4093-2019, 2019.

Nault, B. A., Campuzano-Jost, P., Day, D. A., Schroder, J. C., Anderson, B., Beyersdorf, A. J., Blake, D. R., Brune, W. H., Choi, Y., Corr, C. A., de Gouw, J. A., Dibb, J., DiGangi,

J. P., Diskin, G. S., Fried, A., Huey, L. G., Kim, M. J., Knote, C. J., Lamb, K. D., Lee, T., Park, T., Pusede, S. E., Scheuer, E., Thornhill, K. L., Woo, J.-H., and Jimenez, J. L.: Secondary organic aerosol production from local emissions dominates the organic aerosol budget over Seoul, South Korea, during KORUS-AQ, Atmos. Chem. Phys., 18, 17769-17800, https://doi.org/10.5194/acp-18-17769-2018, 2018.

Ovadnevaite, J., Ceburnis, D., Canagaratna, M., Berresheim, H., Bialek, J., Martucci, G., Worsnop, D. R. and O'Dowd, C.: On the effect of wind speed on submicron sea salt mass concentrations and source fluxes, J. Geophys. Res. Atmos., 117(D16), 1–11, doi:10.1029/2011JD017379, 2012.

Pai, S. J., Heald, C. L., Pierce, J. R., Farina, S. C., Marais, E. A., Jimenez, J. L., Campuzano-Jost, P., Nault, B. A., Middlebrook, A. M., Coe, H., Shilling, J. E., Bahreini, R., Dingle, J. H., and Vu, K.: An evaluation of global organic aerosol schemes using airborne observations, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-331, in press, 2020.

Palm, B. B., de Sá, S. S., Day, D. A., Campuzano-Jost, P., Hu, W., Seco, R., Sjostedt, S. J., Park, J.-H., Guenther, A. B., Kim, S., Brito, J., Wurm, F., Artaxo, P., Thalman, R., Wang, J., Yee, L. D., Wernis, R., Isaacman-VanWertz, G., Goldstein, A. H., Liu, Y., Springston, S. R., Souza, R., Newburn, M. K., Alexander, M. L., Martin, S. T., and Jimenez, J. L.: Secondary organic aerosol formation from ambient air in an oxidation flow reactor in central Amazonia, Atmos. Chem. Phys., 18, 467–493, https://doi.org/10.5194/acp-18-467-2018, 2018.

Petters, M. D., Prenni, A. J., Kreidenweis, S. M., DeMott, P. J., Matsunaga, A., Lim, Y. B., and Ziemann, P. J. Chemical aging and the hydrophobic-to-hydrophilic conversion of carbonaceous aerosol, Geophys. Res. Lett., 33, L24806, doi:10.1029/2006GL027249, 2006.

Pye, H. O. T., Chan, A. W. H., Barkley, M. P., and Seinfeld, J. H.: Global modeling of organic aerosol: The importance of reactive nitrogen (NOx and NO_3), Atmos. Chem. Phys., 10, 11,261–11,276, doi:10.5194/acp-10-11261-2010, 2010.

Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., Lane, T. E., Pandis, S. N., and Pierce, J. R.: Rethinking organic aerosols: Semivolatile emissions and photochemical aging, Science, 315, 1259–1262, 2007.

2997 Schroder, J. C., Campuzano-Jost, P., Day, D. A., Shah, V., Larson, K., Sommers, J. M.,
2998 Sullivan, A. P., Campos, T., Reeves, J. M., Hills, A., Hornbrook, R. S., Blake, N. J.,
2999 Scheuer, E., Guo, H., Fibiger, D. L., McDuffie, E. E., Hayes, P. L., Weber, R. J., Dibb, J.
3000 E., Apel, E. C., Jaeglé, L., Brown, S. S., Thornton, J. A. and Jimenez, J. L.: Sources and
3001 Secondary Production of Organic Aerosols in the Northeastern United States during
3002 WINTER, J. Geophys. Res. Atmos., 123(14), 7771–7796, doi:10.1029/2018JD028475,
3003 2018.

3004 Schacht, J., Heinold, B., Quaas, J., Backman, J., Cherian, R., Ehrlich, A., Herber, A.,
3005 Huang, W. T. K., Kondo, Y. Massling, A, Sinha, P.R., Weinzierl, B. Zanatta, M., and Tegen,
3006 I.: The importance of the representation of air pollution emissions for the modeled
3007 distribution and radiative effects of black carbon in the Arctic, Atmos. Chem. Phys.
3008 Discuss., 2019, 1-39, doi: 10.5194/acp-2019-71, 2019.

3009 Schulz, M., Chin, M., and Kinne, S.: The aerosol model comparison project, AeroCom,
3010 phase II: clearing up diversity, IGAC newsletter, No. 41, 2–11, 2009.

3011 Scott, C. E., Rap, A., Spracklen, D. V., Forster, P. M., Carslaw, K. S., Mann, G. W., Pringle,
3012 K. J., Kivekäs, N., Kulmala, M., Lihavainen, H., and Tunved, P.: The direct and indirect
3013 radiative effects of biogenic secondary organic aerosol, Atmos. Chem. Phys., 14, 447-
3014 470, https://doi.org/10.5194/acp-14-447-2014, 2014.

3015 Shah, V., Jaeglé, L., Jimenez, J. L., Schroder, J. C., Campuzano-Jost, P., Campos, T. L.,
3016 et al.: Widespread pollution from secondary sources of organic aerosols during winter in
3017 the Northeastern United States. Geophysical Research Letters, 46, 2974– 2983.
3018 https://doi.org/10.1029/2018GL081530, 2019.

3019 Shiraiwa, M., Ueda, K., Pozzer, A., Lammel, G., Kampf, C. J.: Aerosol Health Effects from
3020 Molecular to Global Scales, Environ. Sci. Technol. 51 (23), pp 13545–13567, 2017.

3021 Shrivastava, M., Cappa, C.D., Fan, J., et al: Recent advances in understanding secondary
3022 organic aerosol: Implications for global climate forcing, Rev. Geophys., 55, 509–559,
3023 doi:10.1002/2016RG000540, 2017.

3024 Spracklen, D. V., Jimenez, J. L., Carslaw, K. S., Worsnop, D. R., Evans, M. J., Mann, G.
3025 W., Zhang, Q., Canagaratna, M. R., Allan, J., Coe, H., McFiggans, G., Rap, A., and
3026 Forster, P.: Aerosol mass spectrometer constraint on the global secondary organic aerosol
3027 budget, Atmos. Chem. Phys., 11, 12109–12136, doi:10.5194/acp-11-12109-2011, 2011.

Tegen, I., Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Bey, I., Schutgens, N., Stier, P., Watson-Parris, D., Stanelle, T., Schmidt, H., Rast, S., Kokkola, H., Schultz, M., Schroeder, S., Daskalakis, N., Barthel, S., Heinold, B., and Lohmann, U.: The global aerosol–climate model ECHAM6.3–HAM2.3 – Part 1: Aerosol evaluation, Geosci. Model Dev., 12, 1643-1677, https://doi.org/10.5194/gmd-12-1643-2019, 2019.

Thomson, D. S., Schein, M. E. and Murphy, D. M.: Particle Analysis by Laser Mass Spectrometry WB-57F Instrument Overview, Aerosol Sci. Technol., 33(1–2), 153–169, doi:10.1080/027868200410903, 2000.

Tilmes, S., Hodzic, A., Emmons, L.K., Mills, M.J., Gettelman, A., Kinnison, D.E., Park, M., Lamarque J.-F., Vitt, F., et al.: Climate forcing and trends of organic aerosols in the Community Earth System Model (CESM2), to be submitted to JAMES, 2019.

Tsigaridis, K., Daskalakis, N., Kanakidou, M., Adams, P. J., Artaxo, P., Bahadur, R., Balkanski, Y., Bauer, S. E., Bellouin, N., Benedetti, A., Bergman, T., Berntsen, T. K., Beukes, J. P., Bian, H., Carslaw, K. S., Chin, M., Curci, G., Diehl, T., Easter, R. C., Ghan, S. J., Gong, S. L., Hodzic, A., Hoyle, C. R., Iversen, T., Jathar, S., Jimenez, J. L., Kaiser, J. W., Kirkevåg, A., Koch, D., Kokkola, H., Lee, Y. H., Lin, G., Liu, X., Luo, G., Ma, X., Mann, G. W., Mihalopoulos, N., Morcrette, J.-J., Müller, J.-F., Myhre, G., Myriokefalitakis, S., Ng, N. L., O'Donnell, D., Penner, J. E., Pozzoli, L., Pringle, K. J., Russell, L. M., Schulz, M., Sciare, J., Seland, Ø., Shindell, D. T., Sillman, S., Skeie, R. B., Spracklen, D., Stavrakou, T., Steenrod, S. D., Takemura, T., Tiitta, P., Tilmes, S., Tost, H., van Noije, T., van Zyl, P. G., von Salzen, K., Yu, F., Wang, Z., Wang, Z., Zaveri, R. A., Zhang, H., Zhang, K., Zhang, Q., and Zhang, X.: The AeroCom evaluation and intercomparison of organic aerosol in global models, Atmos. Chem. Phys., 14, 10845–10895, doi:10.5194/acp-14-10845-2014, 2014.

Tsigaridis, K., and Kanakidou, M.: The Present and Future of Secondary Organic Aerosol Direct Forcing on Climate, Current Climate Change Reports, 2018, Volume 4, Issue 2, pp 84–98, 2018.

Turpin, B. J. and Lim, H. J.: Species contributions to PM2.5 mass concentrations: Revisiting common assumptions for estimating organic mass, Aerosol Sci. Tech., 35, 602–610, doi:10.1080/02786820152051454, 2001.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R. and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass

spectrometric data, Atmos. Chem. Phys., 9(9), 2891–2918, doi:10.5194/acp-9-2891-2009, 2009.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, https://doi.org/10.5194/acp-10-11707-2010, 2010.

Vergara-Temprado, J., Murray, B. J., Wilson, T. W., O'Sullivan, D., Browse, J., Pringle, K. J., Ardon-Dryer, K., Bertram, A. K., Burrows, S. M., Ceburnis, D., DeMott, P. J., Mason, R. H., O'Dowd, C. D., Rinaldi, M., and Carslaw, K. S.: Contribution of feldspar and marine organic aerosols to global ice nucleating particle concentrations, Atmos. Chem. Phys., 17, 3637-3658, https://doi.org/10.5194/acp-17-3637-2017, 2017.

Wang, J., Cubison, M. J., Aiken, A. C., Jimenez, J. L., and Collins, D. R.: The importance of aerosol mixing state and size-resolved composition on CCN concentration and the variation of the importance with atmospheric aging of aerosols, Atmos. Chem. Phys., 10, 7267-7283, https://doi.org/10.5194/acp-10-7267-2010, 2010.

Williamson, C. J., Kupc, A., Axisa, D., Bilsback, K.R., Bui, T., Campuzano-Jost, P., Dollner, M., Froyd, K., Hodshire, A. L., Jimenez, J. L., Kodros, J. K., Luo, G., Murphy, D. M., Nault, B. A., Ray, E. A., Weinzierl, B. B., Wilson, J. C., Yu, F., Yu, P., Pierce, J.F., Brock C. A.: A Large Source of Cloud Condensation Nuclei from New Particle Formation in the Tropics, Nature, 574, https://doi.org/10.1038/s41586-019-1638-9, 2019.

Wofsy, S.C., S. Afshar, H.M. Allen, E. Apel, E.C. Asher, B. Barletta, J. Bent, H. Bian, B.C. Biggs, D.R. Blake, N. Blake, I. Bourgeois, C.A. Brock, W.H. Brune, J.W. Budney, T.P. Bui, A. Butler, P. Campuzano-Jost, C.S. Chang, M. Chin, R. Commane, G. Correa, J.D. Crounse, P. D. Cullis, B.C. Daube, D.A. Day, J.M. Dean-Day, J.E. Dibb, J.P. DiGangi, G.S. Diskin, M. Dollner, J.W. Elkins, F. Erdesz, A.M. Fiore, C.M. Flynn, K. Froyd, D.W. Gesler, S.R. Hall, T.F. Hanisco, R.A. Hannun, A.J. Hills, E.J. Hintsa, A. Hoffman, R.S. Hornbrook, L.G. Huey, S. Hughes, J.L. Jimenez, B.J. Johnson, J.M. Katich, R.F. Keeling, M.J. Kim, A. Kupc, L.R. Lait, J.-F. Lamarque, J. Liu, K. McKain, R.J. Mclaughlin, S. Meinardi, D.O. Miller, S.A. Montzka, F.L. Moore, E.J. Morgan, D.M. Murphy, L.T. Murray, B.A. Nault, J.A. Neuman, P.A. Newman, J.M. Nicely, X. Pan, W. Paplawsky, J. Peischl, M.J. Prather, D.J. Price, E. Ray, J.M. Reeves, M. Richardson, A.W. Rollins, K.H. Rosenlof,

T.B. Ryerson, E. Scheuer, G.P. Schill, J.C. Schroder, J.P. Schwarz, J.M. St.Clair, S.D. Steenrod, B.B. Stephens, S.A. Strode, C. Sweeney, D. Tanner, A.P. Teng, A.B. Thames, C.R. Thompson, K. Ullmann, P.R. Veres, N. Vieznor, N.L. Wagner, A. Watt, R. Weber, B. Weinzierl, P. Wennberg, C.J. Williamson, J.C. Wilson, G.M. Wolfe, C.T. Woods, and L.H. Zeng. 2018. ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1581, 2018.

Woody, M. C., Baker, K. R., Hayes, P. L., Jimenez, J. L., Koo, B., and Pye, H. O. T.: Understanding sources of organic aerosol during CalNex-2010 using the CMAQ-VBS, Atmos. Chem. Phys., 16, 4081-4100, https://doi.org/10.5194/acp-16-4081-2016, 2016.

Yu, P., O. B. Toon, C. G. Bardeen, M. J. Mills, T. Fan, J. M. English, and R. R. Neely, Evaluations of tropospheric aerosol properties simulated by the community earth system model with a sectional aerosol microphysics scheme, J. Adv. Model. Earth Syst.,7, 865–914, doi:10.1002/2014MS000421, 2015.

Yu, P., Froyd, K. D., Portmann, R. W., Toon, O. B., Freitas, S. R., Bardeen, C. G., et al.: Efficient in-cloud removal of aerosols by deep convection. Geophysical Research Letters, 46. https://doi.org/10.1029/2018GL080544, 2019.

Zhang, K., O'Donnell, D., Kazil, J., Stier, P., Kinne, S., Lohmann, U., Ferrachat, S., Croft, B., Quaas, J., Wan, H., Rast, S., and Feichter, J.: The global aerosol-climate model ECHAM-HAM, version 2: sensitivity to improvements in process representations, Atmos. Chem. Phys., 12, 8911-8949, https://doi.org/10.5194/acp-12-8911-2012, 2012.

Zhang, Q. J., Beekmann, M., Drewnick, F., Freutel, F., Schneider, J., Crippa, M., Prevot, A. S. H., Baltensperger, U., Poulain, L., Wiedensohler, A., Sciare, J., Gros, V., Borbon, A., Colomb, A., Michoud, V., Doussin, J.-F., Denier van der Gon, H. A. C., Haeffelin, M., Dupont, J.-C., Siour, G., Petetin, H., Bessagnet, B., Pandis, S. N., Hodzic, A., Sanchez, O., Honoré, C., and Perrussel, O., 2013. Formation of organic aerosol in the Paris region during the MEGAPOLI summer campaign: evaluation of the volatility-basis-set approach within the CHIMERE model, Atmos. Chem. Phys., 13, 5767-5790, doi:10.5194/acp-13-5767-2013.

Zhang, X., Cappa, C. D., Jathar, S. H., McVay, R. C., Ensberg, J. J., Kleeman, M. J., and Seinfeld, J. H.: Influence of vapor wall loss in laboratory chambers on yields of secondary organic aerosol, P. Natl. Acad. Sci. USA, 111, 5802–5807, 2014.

Zhao, G., Chen, Y., Hopke, P.K., Holsen, T.M., Dhaniyala, S.: Characteristics of traffic-induced fugitive dust from unpaved roads, Aerosol Science and Technology, 51:11, 1324-1331, DOI: 10.1080/02786826.2017.1347251, 2017.

Zhu, J., Penner, J. E., Yu, F., Sillman, S., Andreae, M., and Coe, H.: Organic aerosol nucleation, climate and land use change: Decrease in radiative forcing, Nature Communications, 10, Article No. 423, https://www.nature.com/articles/s41467-019-08407-7, 2019.

Zotter, P., I. El-Haddad, Y. Zhang, P.L. Hayes, X. Zhang, Y.H. Lin, L. Wacker, J. Schnelle-Kreis, G. Abbaszade, R. Zimmermann, J.D. Surratt, R. Weber, J.L. Jimenez, S. Szidat, U. Baltensperger, A.S.H. Prévôt. Diurnal cycle of fossil and non-fossil carbon using radiocarbon analyses during CalNex. Journal of Geophysical Research-Atmospheres, 119, 6818–6835, doi:10.1002/2013JD021114, 2014.

3136 Tables

3137 Table 1: ATom global model configurations and their treatment of the most important processes affecting organic aerosols.

Models & horizontal res. & met. fields & config. reference	**Aerosol module**	**Submicron size[6] OA (dust/sea salt)**	**SOA precursors[1]**					**SOA production**	**Emission**	**POA/POC (SOA/SOC)**	**Removal**		
			ISO	**MT**	**SQ**	**ANT**	**$C_{>12}$**				**Standard[2]**	**Improved**	**Photolytic**
CESM1-CARMA (1.9°lon x 2.5°lat) MERRA-2 (Yu et al. 2019)	20 bins	< 500 nm (< 800 nm)	x	x		x		Semi-volatile using VBS (Pye et al. 2010)	GAIS and GFED v3	1.8 (N/A)	x	For convective updrafts (Yu et al. 2019)[3]	
CESM2-DYN (0.9° lon x 1.25° lat) GEOS5 (Tilmes et al. 2019)	4 modes	< 270 nm (< 800 nm)	x	x	x	x	x	Semi-volatile using VBS (Hodzic et al. 2016)	CMIP6 and QFED v2.4	1.8 (N/A)	x	Water solubility of organic gases per Hodzic et al. (2014)	For SOA (Hodzic et al. 2016)
CESM2-SMP GEOS5 (0.9° lon x 1.25° lat) (Tilmes et al. 2019)	4 modes	< 270 nm (< 800 nm)	x	x		x		Non-volatile with prescribed mass yields for all precursors[4]	CMIP6 and QFED v2.4	1.8 (N/A)	x		
ECHAM6-HAM ECHAM6 (1.87°lon x1.87°lat) (Tegen et al. 2019)	7 modes	< 500 nm (< 500 nm)		x				Non-volatile with 15% prescribed mass yields (Dentener et al. 2006)	ECLIPSE[5] and GFAS	1.4 (1.4)	x		
GC12-REF (2° lon x 2.5° lat) GEOS-FP (Bey et al. 2001)	Bulk	Bulk (< 500 nm)	x	x	x	x		Semi-volatile using VBS (Pye et al. 2010); non-volatile isoprene-	CMIP6 and GFED v4	2.1 (N/A)	x	For convective updrafts per Wang et al. 2014	

								SOA (Marais et al. 2016)					
GC12-DYN (2° lon x 2.5° lat) GEOS-FP (Bey et al. 2001)	Bulk	Bulk (< 500 nm)	x	x	x	x	x	Semi-volatile using VBS (Hodzic et al. 2016); non-volatile isoprene-SOA (Marais et al. 2016)	CMIP6 and GFED v4	2.1 (N/A)	x	For convective updrafts (Wang et al. 2014); Water solubility of organic gases (Hodzic et al. 2014)	For SOA (Hodzic et al. 2016)
GC10-TOMAS (5° lon x 4° lat) GEOS-FP (Kodros et al. 2016)	15 bins	< 316 nm (< 316 nm)		x		x		Non-Volatile using 10% mass yields for MT, 0.2 Tg SOA per Tg CO for anthropogenic emissions	EDGAR v4 and GFED v3	1.8 (1.8)	x	For convective updrafts (Wang et al. 2014)	
GEOS5 (0.5°lon x 0.625°lat) MERRA-2 (Bian et al. 2019)	Bulk	bulk (< 1 μm for dust, 500 nm for seasalt)	x	x		x		Non-Volatile, 10% mass yields for all precursors	HTAP and QFED v2.54	1.8 (1.8)	x		

(1) SOA precursors include isoprene (ISO), monoterpenes (MT), sesquiterpenes (SQ), anthropogenics (ANT) including aromatics such as benzene, toluene and xylene, as well as lumped shorter chain alkanes and alkenes; and higher molecular weight n-alkanes and n-alkenes (C>12).

(2) Standard removal includes dry deposition and sedimentation, as well as convective and large-scale scavenging of soluble organic gases and aerosols, and below-cloud scavenging of aerosols.

(3) A sensitivity simulation is performed with CESM1-CARMA without the improved scavenging in convective updrafts.

(4) 5% for lumped C<12 alkanes, 5% for lumped C<12 alkenes, 15% for aromatics, 4% for isoprene, 25% for monoterpenes.

(5) Anthropogenic BC emission are replaced in Russia with the dataset of Huang et al. (2015).

(6) Submicron size range (diameter) used in various models for comparison with the AMS data.

Table 2: Comparison of observed and simulated OA concentrations along ATom-1 and ATom-2 flights for eight global model simulations and their ensemble. The results of the model ensemble are also indicated. The statistical indicators are calculated as normalized mean bias $NMB(\%) = 100 \times \sum_i (M_i - O_i) / \sum_i O_i$; normalized mean error $NME(\%) = 100 \times \sum_i |(M_i - O_i)| / \sum_i O_i$; root mean square error $RMSE(\mu g\, m^{-3}) = \sqrt{(1/N) \sum_i (M_i - O_i)^2}$ and correlation coefficient (R^2) between modeled (M_i) and observed (O_i) data points. The mean of ATom-1 observations is ~0.23 μg m^{-3} and for ATom-2 is 0.11 μg m^{-3}. Figure S4 shows the normalized mean bias for all individual ATom model simulations for various latitudinal regions and for both the Atlantic and Pacific basins.

Organic aerosols	Avg.Mod. (μg m^{-3})	NMB (%)	NME (%)	RMSE (μg m^{-3})	R^2	Avg.Mod. (μg m^{-3})	NMB (%)	NME (%)	RMSE (μg m^{-3})	R^2
Model	*ATom-1 scores (August 2016)*					*ATom-2 scores (February 2017)*				
AeroCom-II Ens.	0.400	74.2	127.3	0.560	0.31	0.254	137	175	0.278	0.38
AeroCom-II Sub.(1)	0.335	47.0	111	0.557	0.28	0.242	127	178	0.290	0.27
ATom Ensemble	0.239	-4.5	64.6	0.372	0.66	0.139	23	92.6	0.224	0.48
CESM2-DYN	0.268	4.6	83.7	0.867	0.47	0.140	25.6	111.7	0.317	0.36
CESM2-SMP	0.349	36.3	94.3	0.556	0.51	0.175	57.2	125.4	0.299	0.31
CESM1-CARMA	0.155	-33.2	93.8	0.603	0.12	0.131	22.6	119.6	0.244	0.31
ECHAM6-HAM	0.400	73.6	143.6	0.714	0.24	0.214	100	184.0	0.363	0.23
GC12-DYN	0.142	-32.6	79.4	0.560	0.16	0.174	14.7	96.6	0.312	0.39
GC12-REF	0.122	-43.0	76.5	0.536	0.18	0.147	3.6	96.3	0.292	0.35
GC10-TOMAS	0.218	-14.4	86.5	0.644	0.16	0.313	150.0	223.7	0.537	0.12
GEOS5	0.242	-5.4	86.6	0.975	0.38	0.084	-24.9	86.4	0.268	0.29

(1) This is the subset of AeroCom-II model ensemble that includes only seven models that are similar to those that are included in the ATom ensemble (either the same model, or an older model version, or the same aerosol module). AeroCom-II Sub. incudes CAM5-MAM3, CCSM4-hem, ECHAM5-HAM2,

GEOSChem-APM 8.2, GEOSChem 9, GISS-TOMAS and GMI (see Tsigaridis et al., 2014 for their description).

Figures:

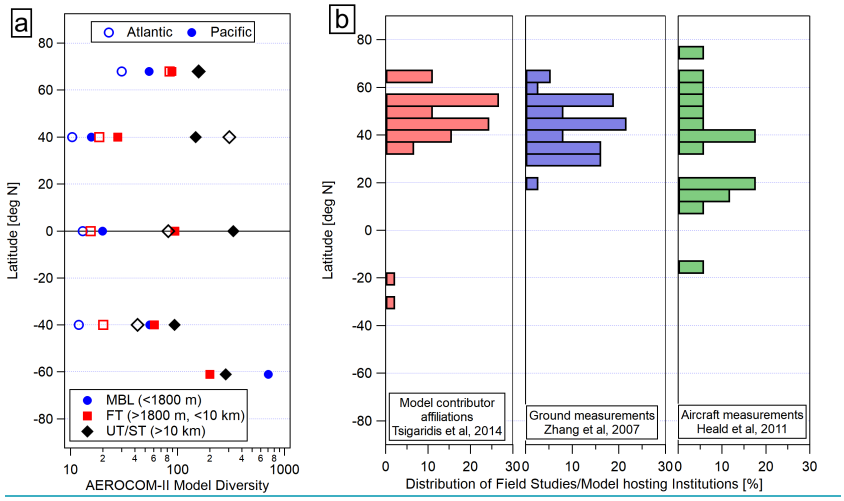
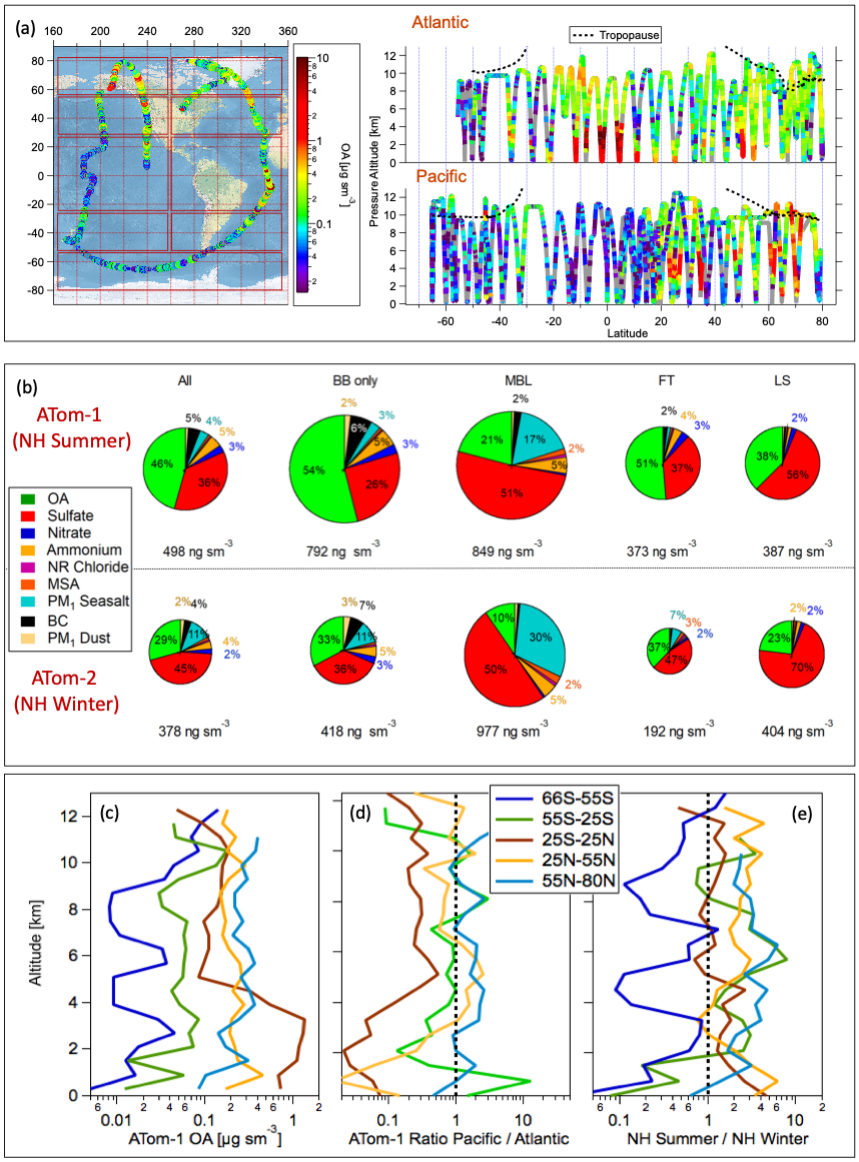


Figure 1: (a, left) The ratio between the average OA concentrations of the highest and the lowest models (for each region) as predicted among 28 global chemistry transport models participating in the AeroCom phase II intercomparison study (Tsigaridis et al. 2014); (b, right) Geographical distribution of institutions at which the AeroCom-II models were ran/developed (based on author affiliations) and of the field measurements included in two major literature overview studies (Zhang et al., 2007; Heald et al., 2011) for the OA ground and aircraft AMS as a function of latitude. For the aircraft campaigns, the average latitude for the full deployment was taken.


(a)
Atlantic
Pacific
Tropopause
Pressure Altitude [km]
Latitude
OA [µg sm^{-3}]
(b)
ATom-1
(NH Summer)
ATom-2
(NH Winter)
All
BB only
MBL
FT
LS
OA
Sulfate
Nitrate
Ammonium
NR Chloride
MSA
PM$_1$ Seasalt
BC
PM$_1$ Dust
498 ng sm^{-3}
792 ng sm^{-3}
849 ng sm^{-3}
373 ng sm^{-3}
387 ng sm^{-3}
378 ng sm^{-3}
418 ng sm^{-3}
977 ng sm^{-3}
192 ng sm^{-3}
404 ng sm^{-3}
(c)
(d)
(e)
66S-55S
55S-25S
25S-25N
25N-55N
55N-80N
Altitude [km]
ATom-1 OA [µg sm^{-3}]
ATom-1 Ratio Pacific / Atlantic
NH Summer / NH Winter

3177

Figure 2: (a, left) ATom-1 DC-8 flights during the August 2016 deployment. Red boxes indicate regions used for the latitude averaging of the model results. (a, right) Vertical distribution of OA concentrations ($\mu g\ sm^{-3}$) along ATom-1 flight tracks (b) Average submicron aerosol composition as measured in the biomass-burning influenced regions (BB only), and the non-BB influenced regions including the marine boundary layer (MBL), free troposphere (FT), and lower stratosphere (LS) for ATom-1 (upper plots) and ATom-2 (lower plots). The BB influenced airmasses were filtered using the PALMS data (see section 3.1). Contributions below 2% are shown but not labeled on the pie chart graph. In ATom-1, BB-only represents 24% of the data, clean MBL 8%, clean FT 57% and clean UT 12%, whereas in ATom-2 BB-only represents 3%, clean MBL 8%, clean FT 74%, clean UT 16%. (c) The average OA vertical profiles are shown for each latitude region as well as (d) the ratios between the Pacific and Atlantic Oceans in each region. (e) The seasonal contrast in OA concentrations as calculated as the ratio in OA concentrations between the NH summer (ATom-1) and NH winter (ATom-2) campaigns. The corresponding plots for ATom-2 can be found in Fig. S1.

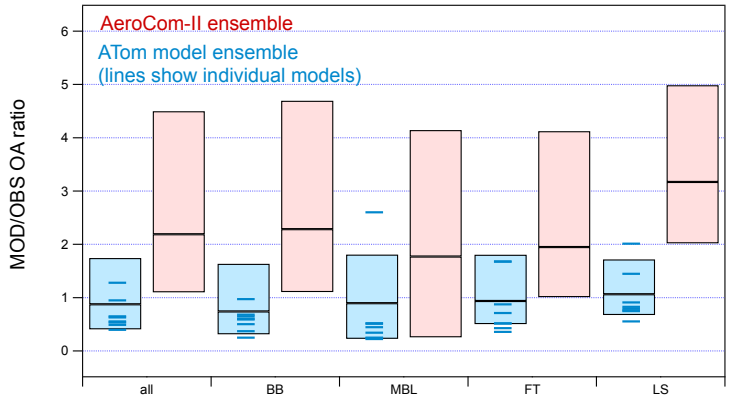


Figure 3: Ratios between predicted and observed OA concentrations for all ATom-1 flights as calculated for the ATom and AeroCom-II model ensembles in different regions ("BB" biomass burning influenced regions; "MBL" clean marine boundary layer; "FT" clean free troposphere' and "LS" lower stratosphere). Median of the ensemble ratio is shown as a horizontal line, while the boxes indicate 25th and 75th percentiles. Medians for the individual models included in the current ATom model ensemble are also shown as blue lines.

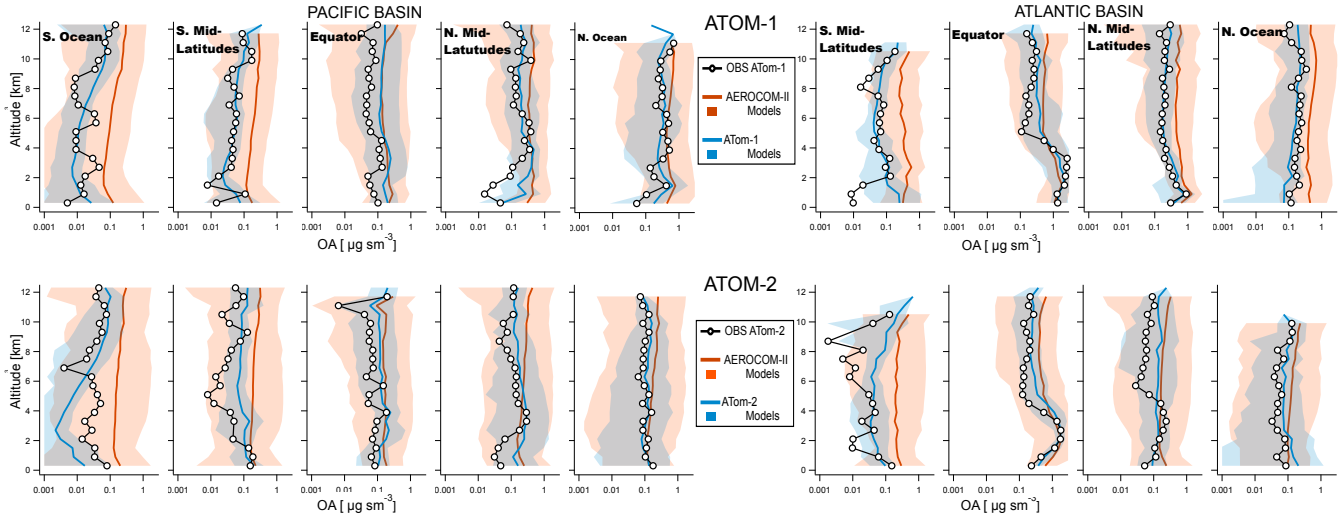


Figure 4: Comparison of latitude-averaged predicted OA vertical profiles with ATom-1 and -2 measurements taken over the Pacific (left side) and Atlantic (right side) basins. Results of the AeroCom-II model ensemble average are shown in red while those of the ATom model ensemble are shown in blue. Shaded areas indicate the variability (two standard deviations) within each model ensemble.

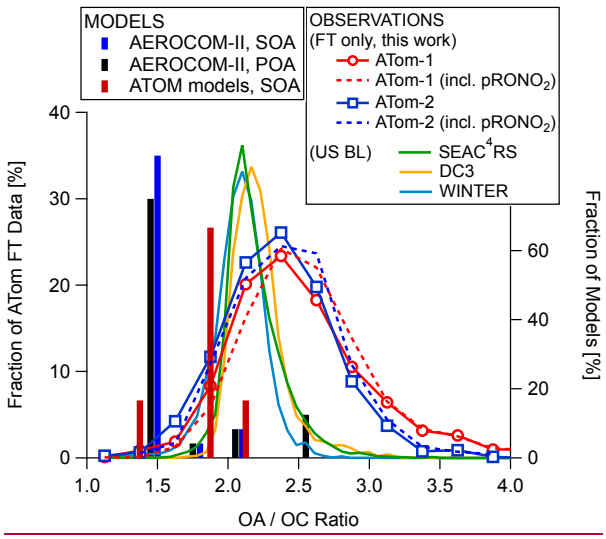


Figure 5: Distribution of the OA / OC ratio as measured during ATom-1 and -2. Values for the recent aircraft campaigns (SEAC4RS, DC3 and WINTER) that took place over continental US regions closer to continental source regions are also shown (Schroder et al., 2018). The bars (right axis) show the OA/OC used for SOA and POA by the models included in the AeroCom and ATom ensemble, with OA/OC=1.4 being the modal value for the former and 1.8 for the latter.

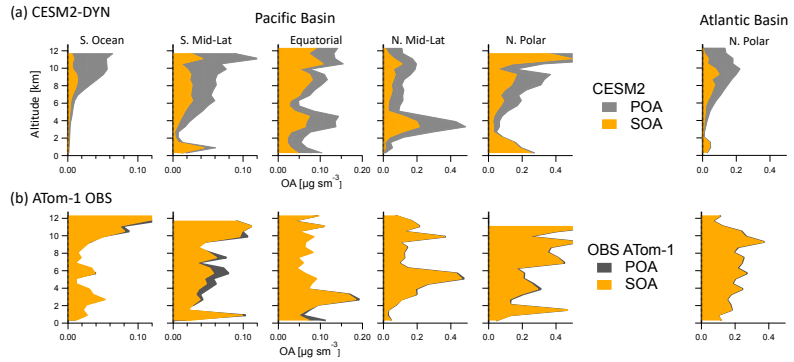


Figure 6: Comparison of averaged POA and SOA vertical profiles as observed during ATom and as predicted by the CESM2-DYN model over the non-BB influenced Pacific and Atlantic basins. The comparison is not shown for the strongly biomass burning influenced regions as all the OA is conservatively allocated to POA in those regions.

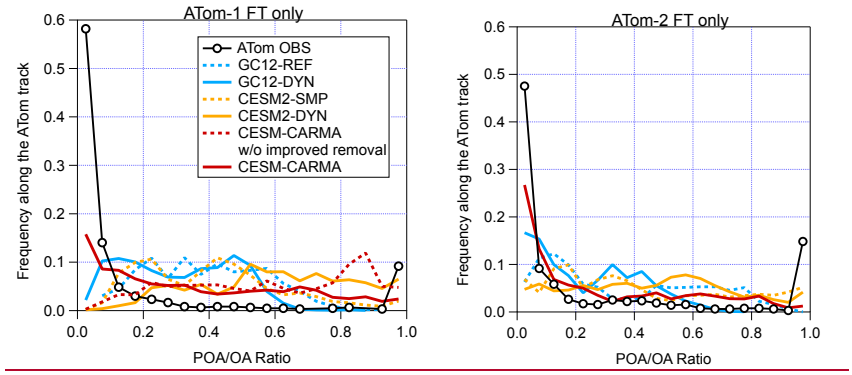


Figure 7: Frequency distribution of observed and simulated ratio of POA to total OA in the free troposphere during ATom-1 and ATom-2 as computed by the GC12-, CESM2-, and CESM1-CARMA models.

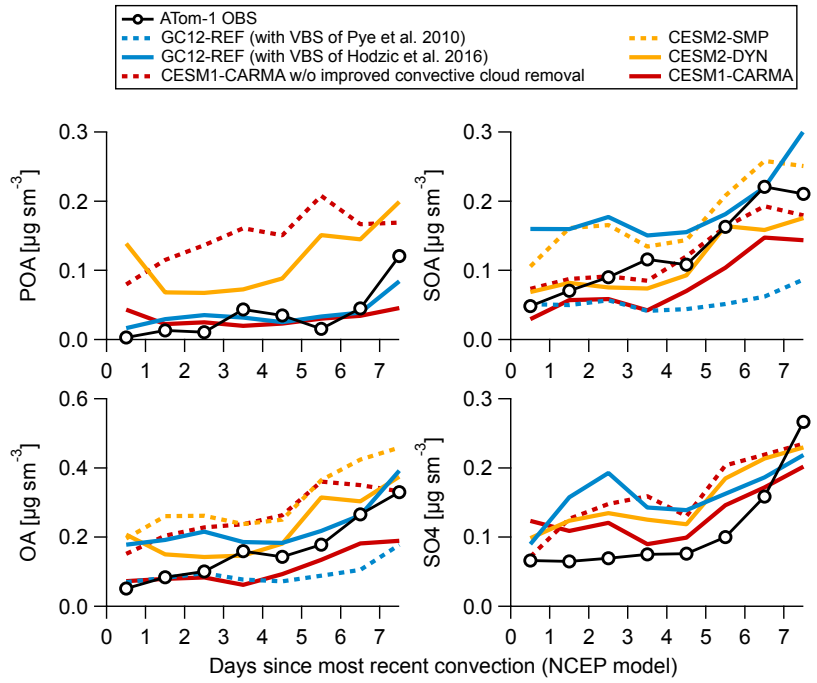
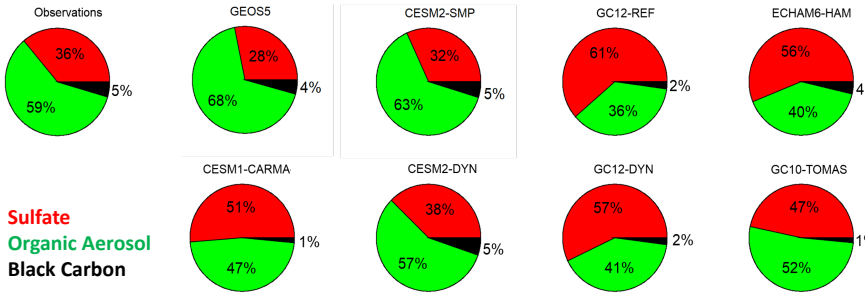


Figure 8: Measured and predicted mass concentrations of POA, SOA, OA and sulfate aerosols during ATom-1 as a function of the number of days since the air mass was processed through convection (based on a trajectory model from Bowman, 1993, and satellite cloud data from NASA Langley, https://clouds.larc.nasa.gov/). CESM2-SMP and CESM2-DYN have the same emissions and processing of POA and sulfate, and thus similar concentrations. The same is true for the two versions of GC12.

(a)



(b)

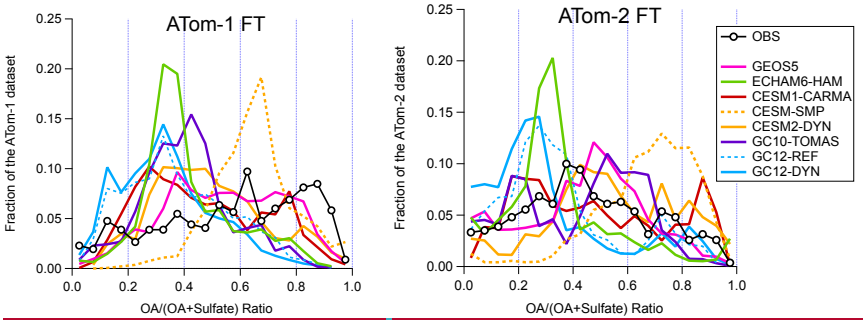


Figure 9: (a) Predicted and measured composition of submicron aerosols in the free troposphere as a function of the submicron aerosol mass concentrations during ATom-1. (b) Frequency distribution of observed and simulated ratio of organic to organic plus sulfate aerosols in the free troposphere during ATom-1 and -2.

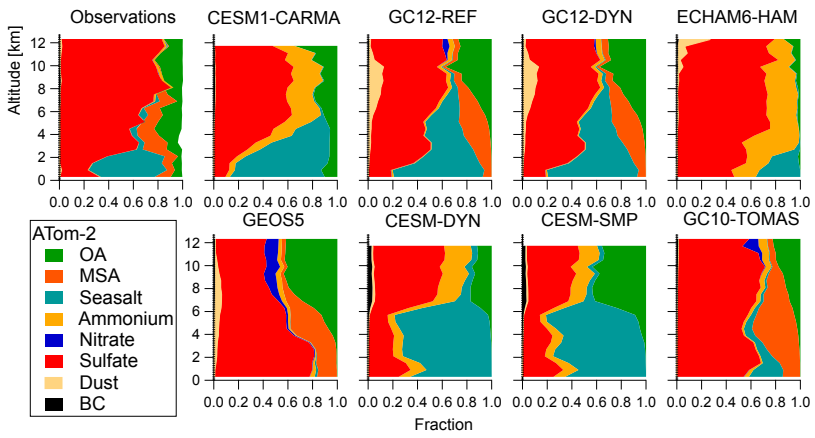


Figure 10: Comparison of measured and predicted composition of submicron aerosols as a function of altitude over the remote Southern Ocean region during NH Winter (ATom-2). For models that do not calculate ammonium in the aerosol (such as CESM1-CARMA, CESM2-SMP, CESM2-DYN and ECHAM6-HAM), ammonium was estimated from the sulfate mass assuming the formation of ammonium sulfate. Note that while the modeled and measured submicron sea salt size ranges agree fairly well (Table 1), this is not quite the case for dust. Given that the accumulation mode dust in the models presented contains larger sizes than the AMS range (< 500 nm), it is expected for the modeled dust concentration to be larger than measured.

3254