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 #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.*"

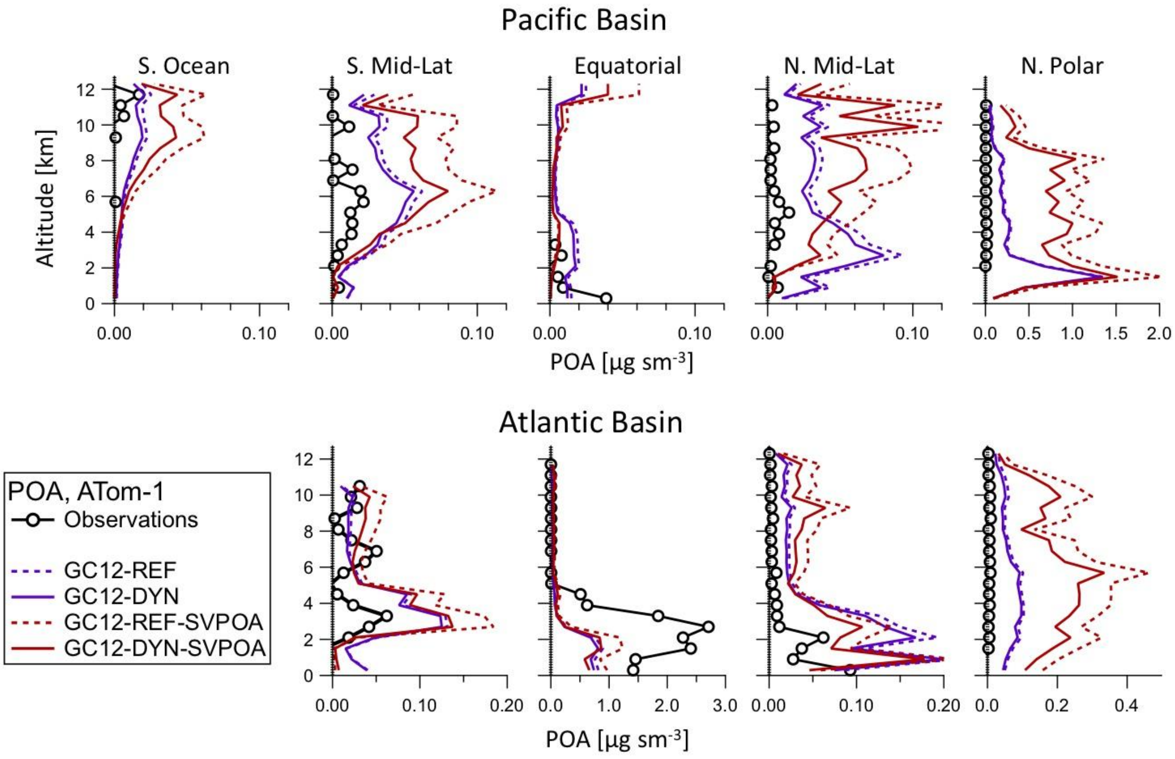
This non-volatile treatment of POA in the models is consistent with the way the estimated POA has been derived from the ATom measurements. Indeed, the estimated POA is calculated from the POA/BC ratios representative of the ambient air values close to the emission sources, after most evaporation has 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."*

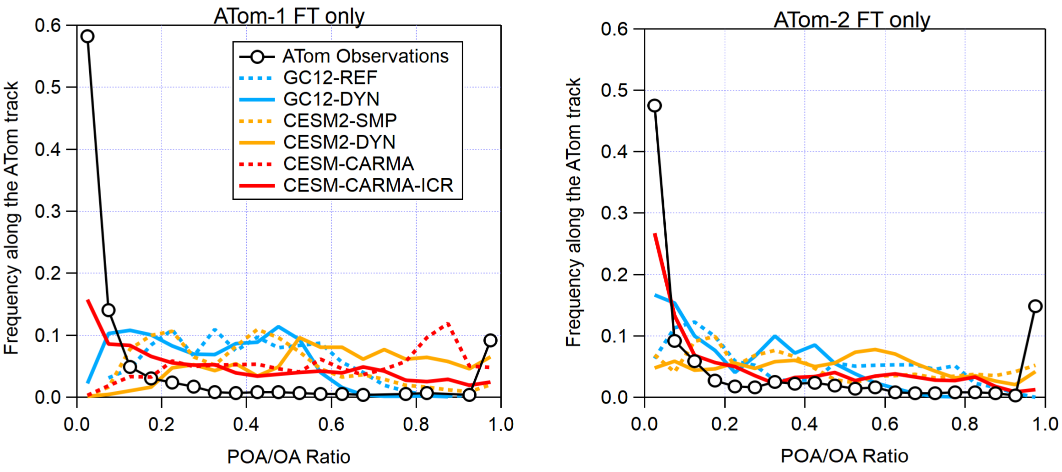
It is not possible to apply the measurement methodology to model outputs, as none of the models track separately BC from various emission sectors. In any case, we have now clearly shown that the measurement-based estimates are very robust against a wide range of assumptions.

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 Peat AG Dung Biofuel Charcoal Average (this work)		*9.1* *13.8* *31.7* *22* *227* *18.7* *9.9* *52.6* *13* ***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.*

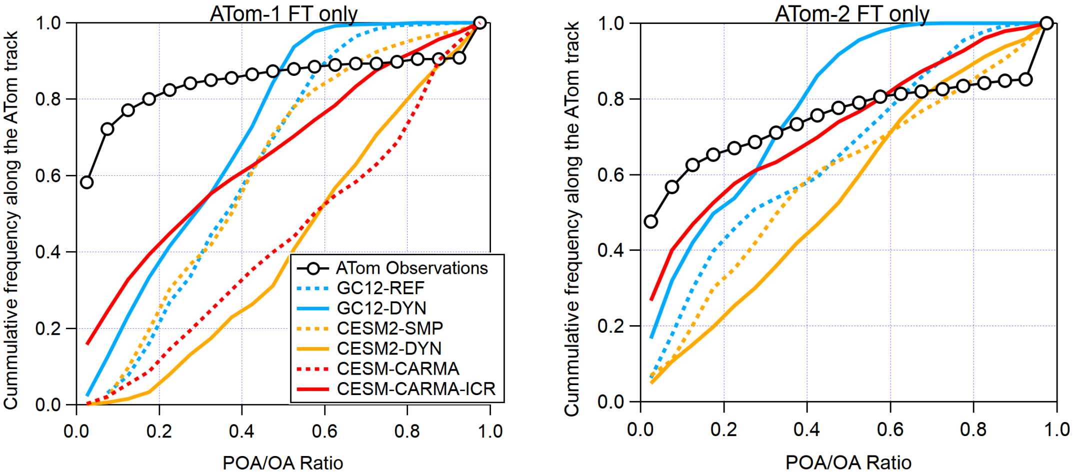
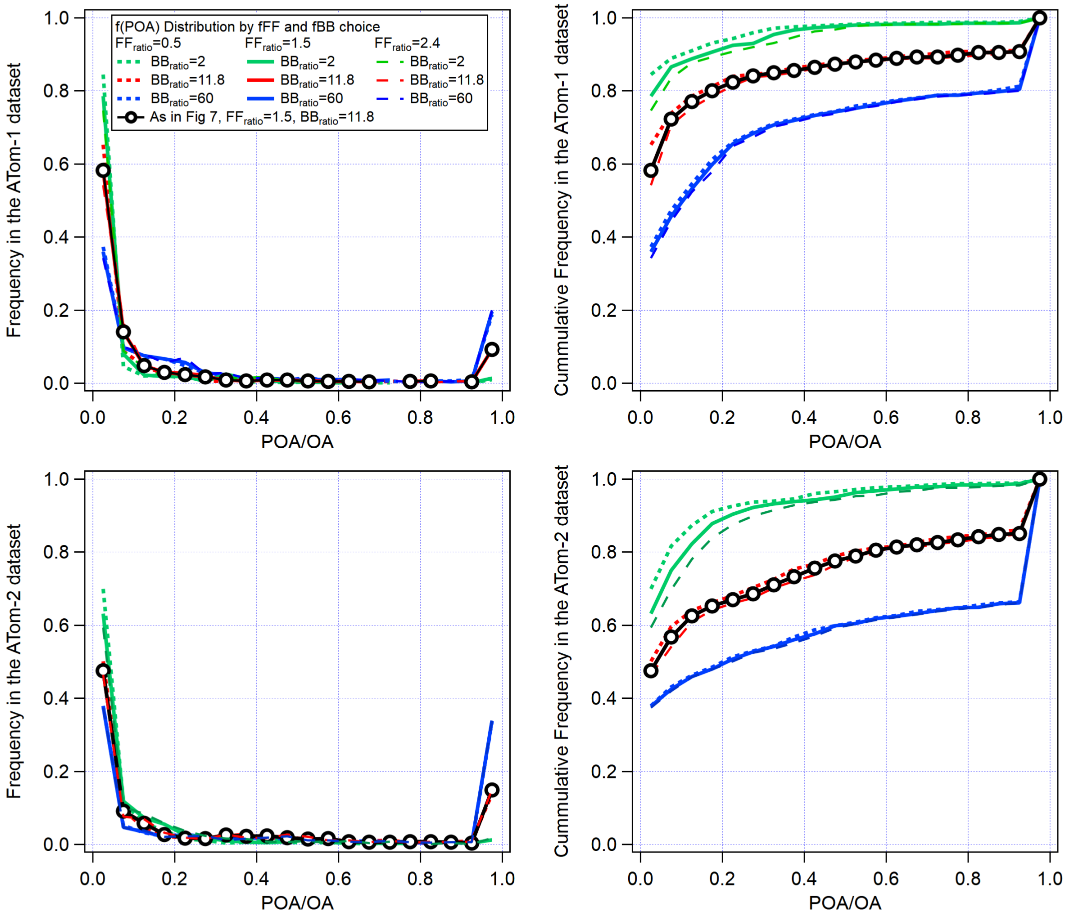
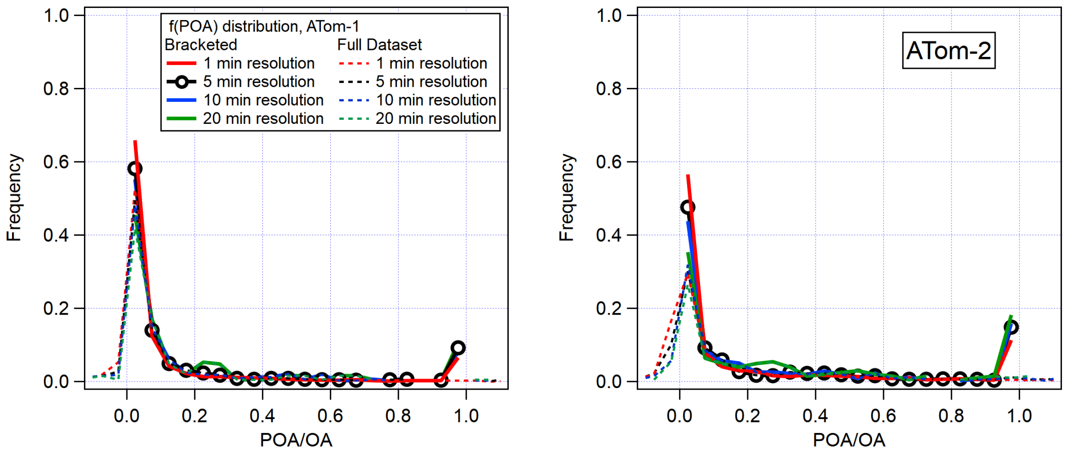


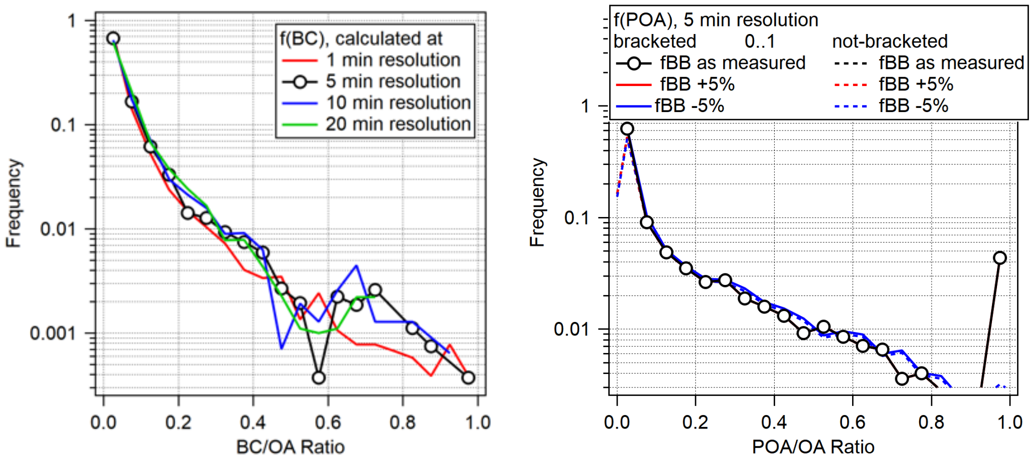
Figure S8: POA/OA distributions (free troposphere only) from Figure 7 shown as cumulative distributions (CDF). Note that for the OA/BC ratios observed for ATom specifically, the green curves in Fig S9b and S9d (BB_{ratio}=2) are closer to the real distribution.



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

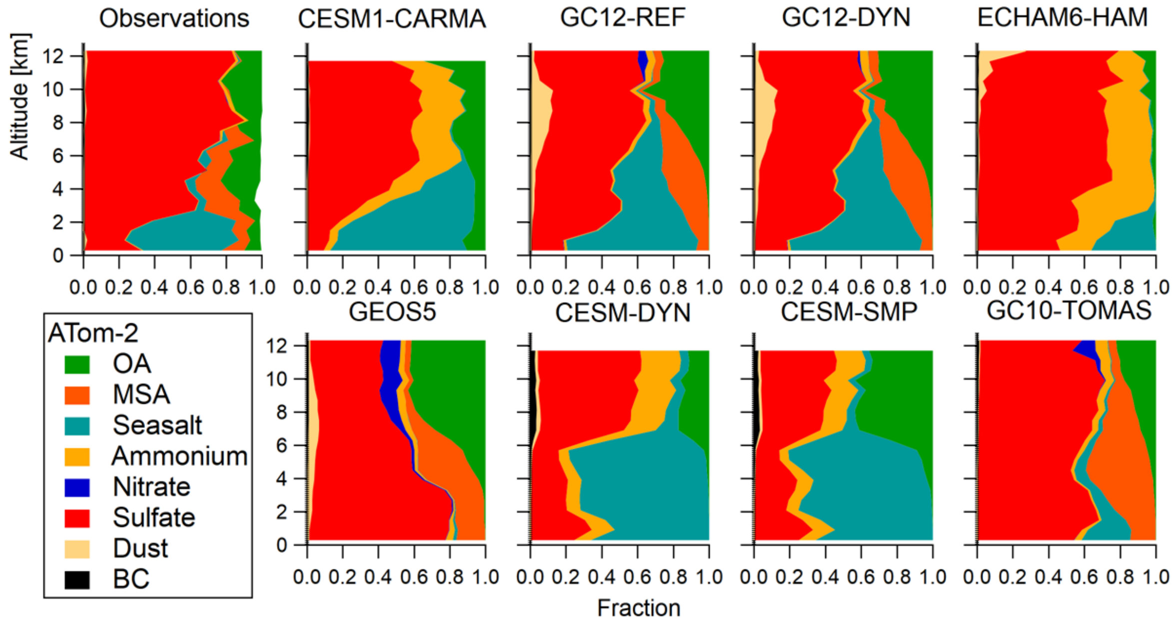
Reddington, C. L., Yoshioka, M., Balasubramanian, R., Ridley, D., Toh, Y. Y., Arnold, S. R. and Spracklen, D. V.: Contribution of vegetation and peat fires to particulate air pollution in Southeast Asia, Environ. Res. Lett., 9(9), doi:10.1088/1748-9326/9/9/094006, 2014.

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?

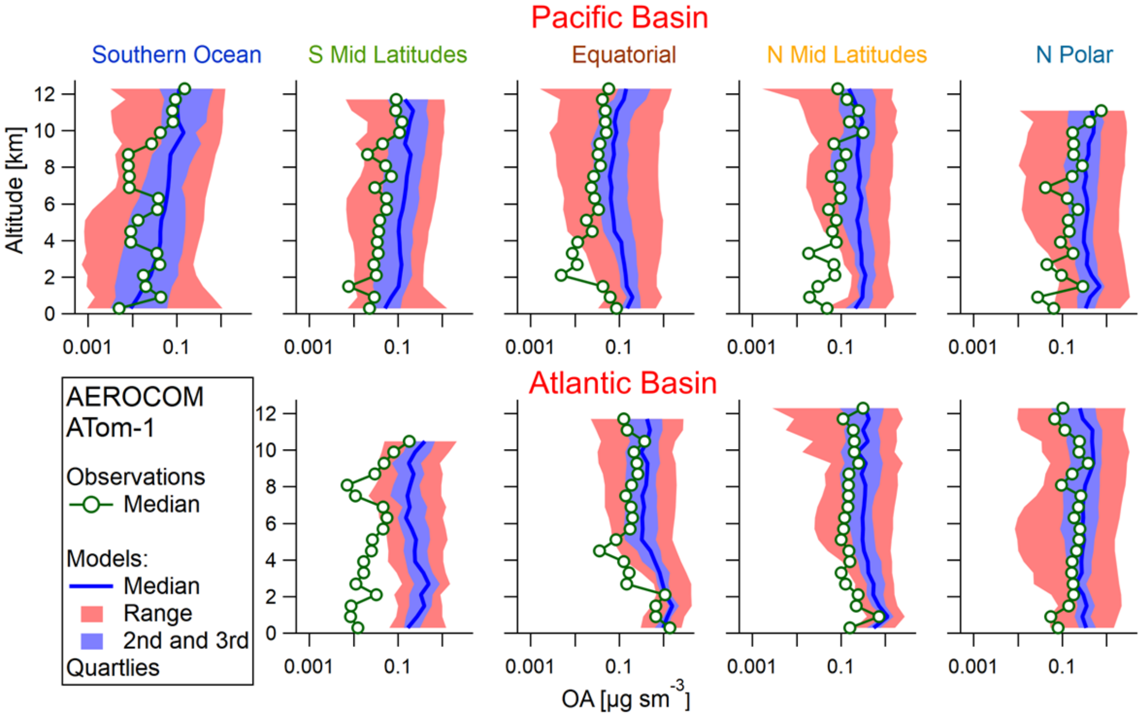
A2.25) For the model-measurement comparisons of the AeroCom-II and ATom model ensembles, we have compared both the box plots for various regions using medians (Figure 3) and the vertical profiles using the means of each ensemble (Figure 4). The results of those analyses are consistent, and show a factor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

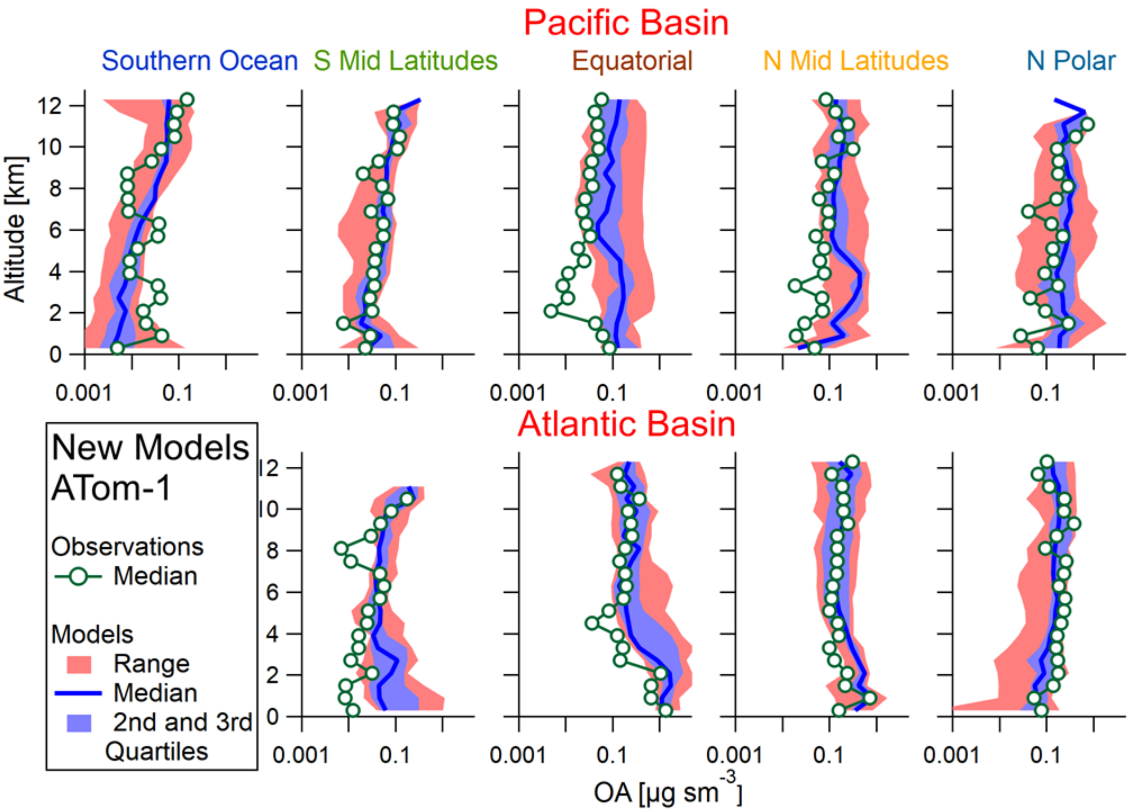
This is now explained in the revised manuscript:

"*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).*"

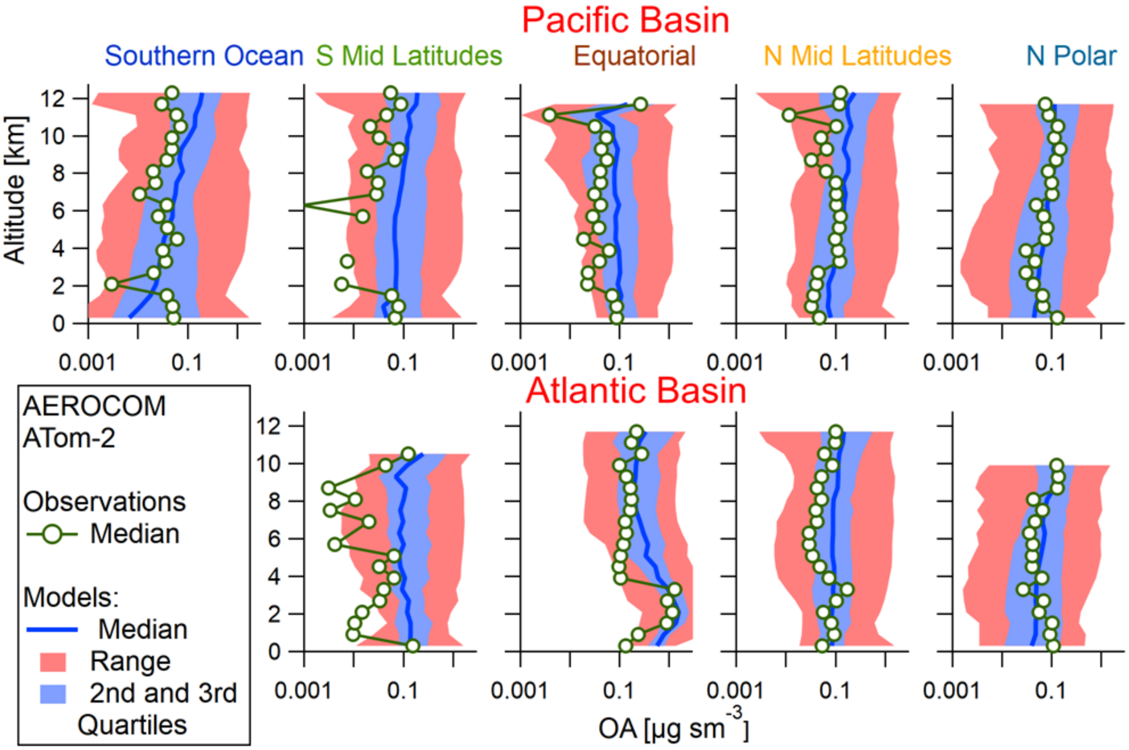
For the evaluation of the individual models the statistics are shown in Table 2. As suggested by the reviewer we have in addition compared distribution plots of OA mass concentrations for the observations and various models.



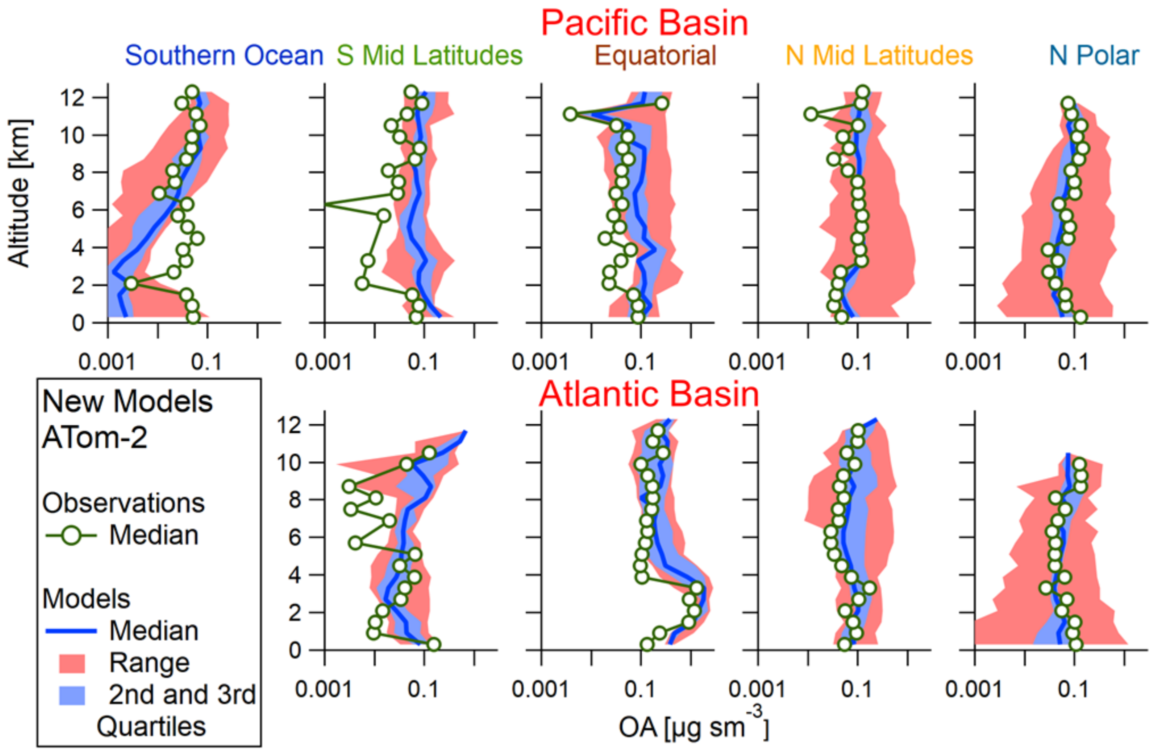
a)


Pacific Basin
Southern Ocean
S Mid Latitudes
Equatorial
N Mid Latitudes
N Polar
Altitude [km]
0
2
4
6
8
10
12
0.001
0.1
Atlantic Basin
New Models
ATom-1
Observations
Median
Models
Range
Median
2nd and 3rd Quartiles
OA [μg sm-3]

b)


Pacific Basin
Southern Ocean
S Mid Latitudes
Equatorial
N Mid Latitudes
N Polar
Altitude [km]
0
2
4
6
8
10
12
0.001
0.1
Atlantic Basin
AEROCOM
ATom-2
Observations
Median
Models:
Median
Range
2nd and 3rd Quartiles
OA [μg sm-3]

c)

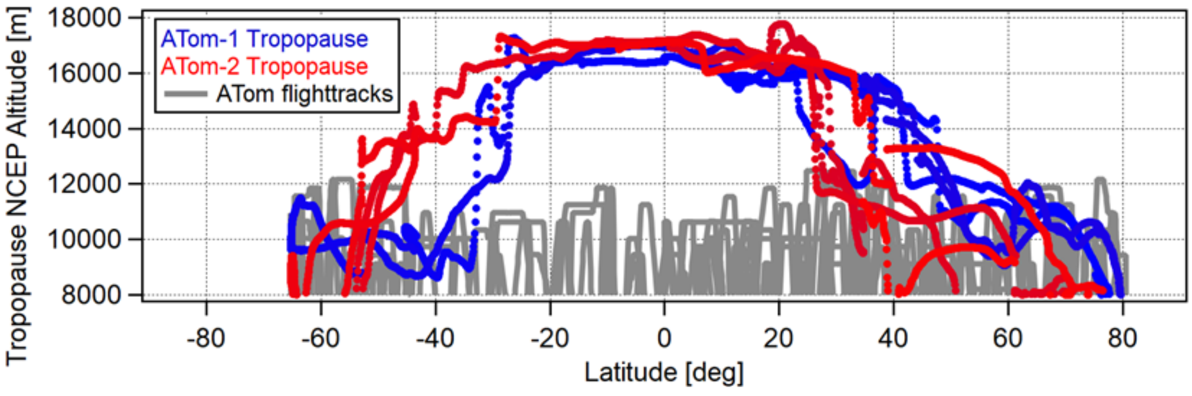


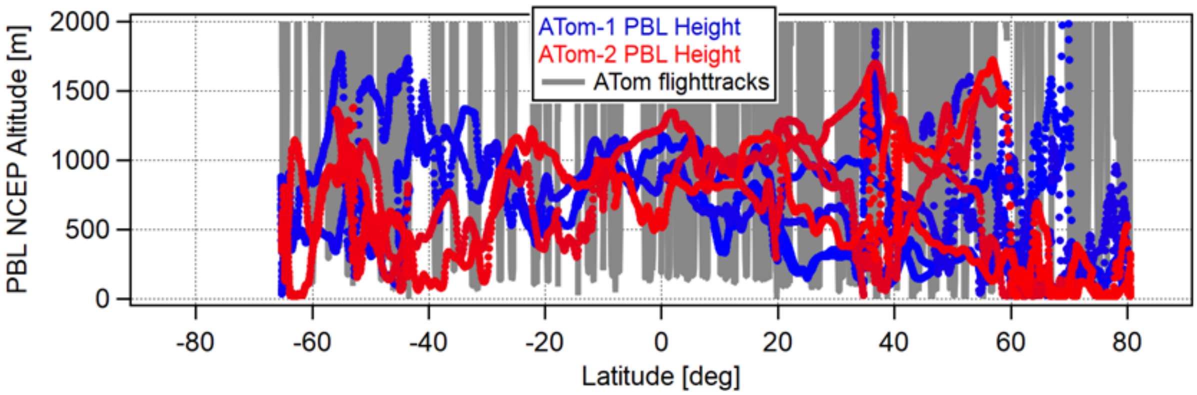
d)

***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.*"

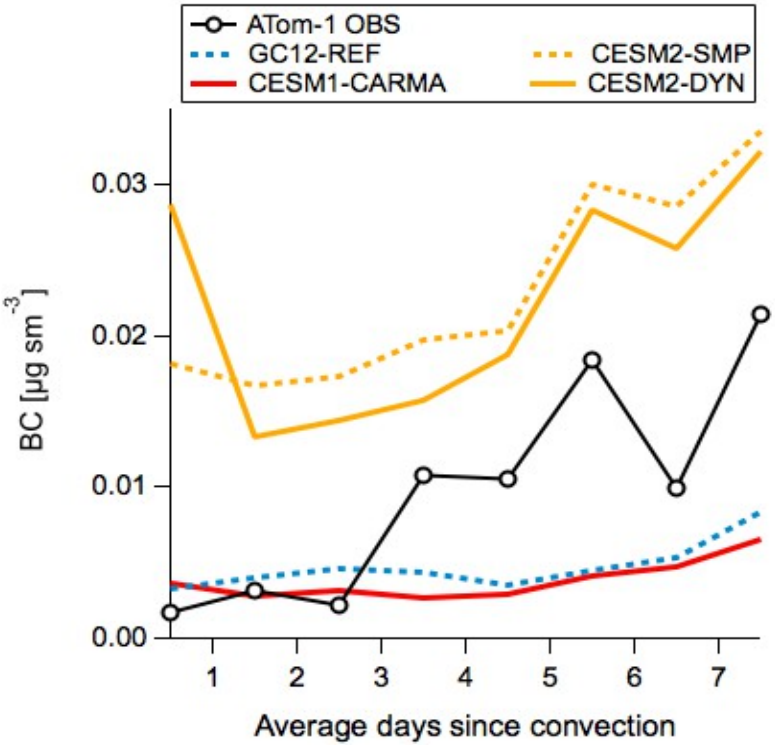
R2.33) Section 4.4: The POA to OA ratio is derived, not directly measured (e.g. line 685).

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.*