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:

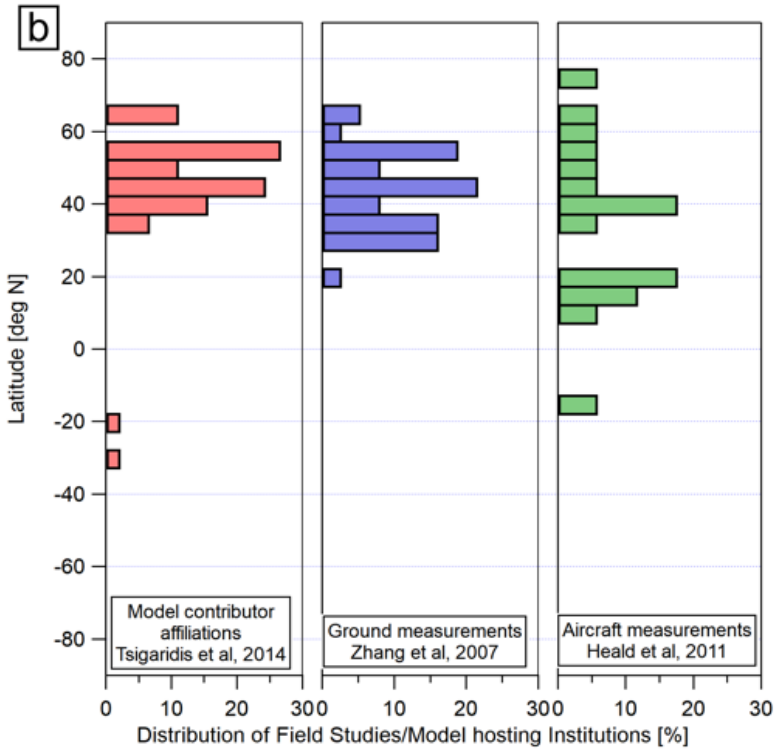
"*OA concentrations have a strong seasonal and zonal variability, with the highest levels measured in the lower troposphere in the summer and over the regions influenced by the biomass burning from Africa (up to 10 ug sm^{-3}).*"

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.

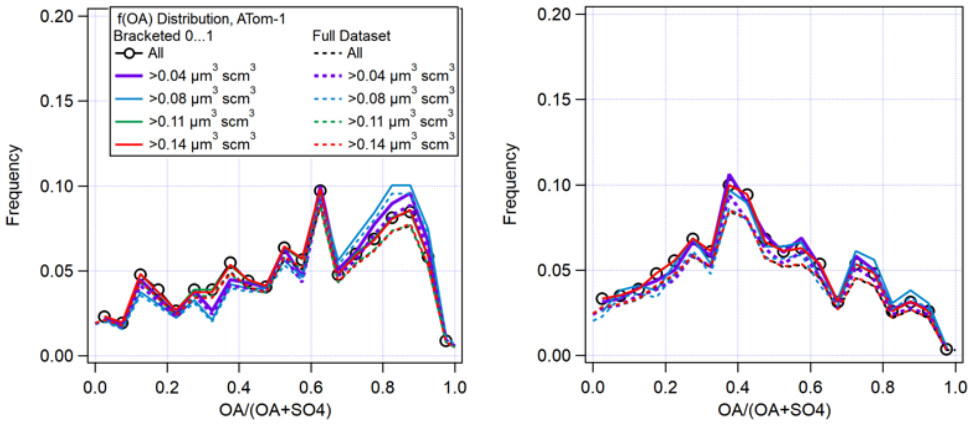
R1.7) Line 426-429: The averaging procedure you used is not clear. If the values are < 3 sigma detection limit, shouldn't you replace the value by zero (so as not to bias the average high)?

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.

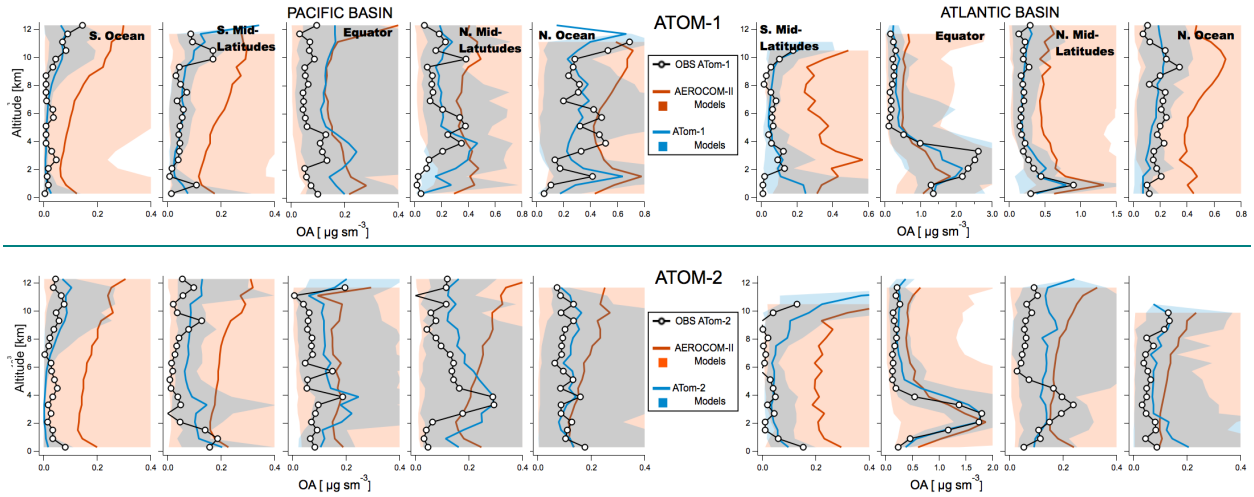
A1.8) We have revised this text to make the point clearer:

"*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)."

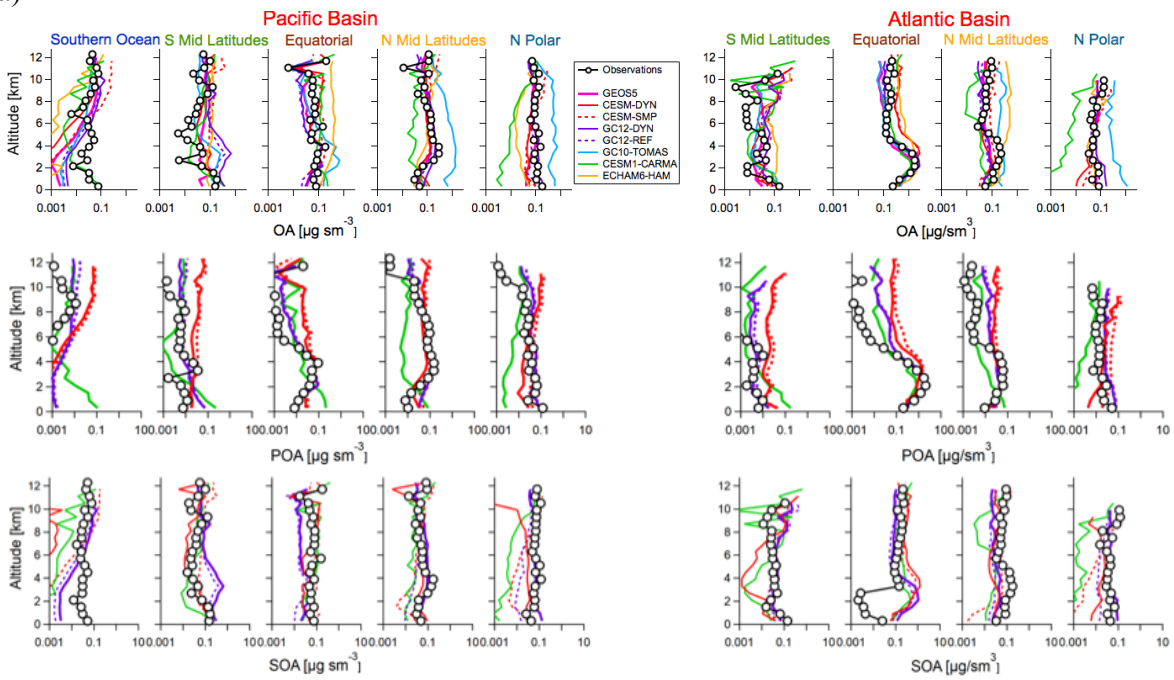


***Caption Fig. S5**: Similar to figure 4 but on a linear scale.*

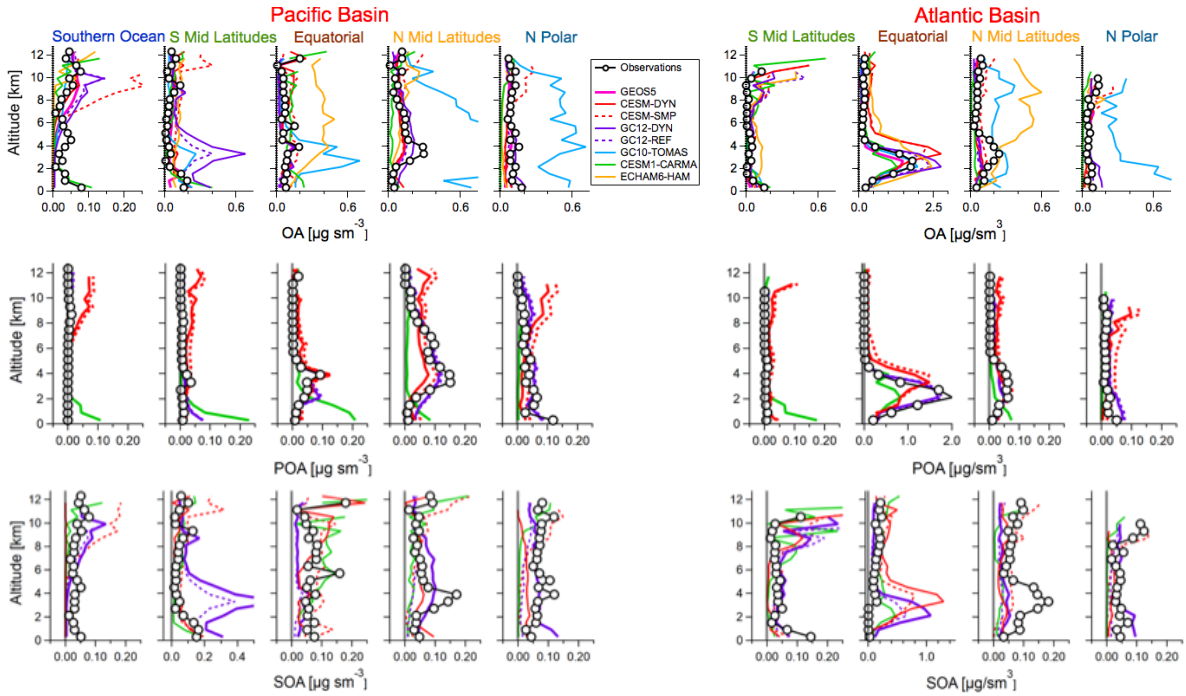
R1.10) Line 587-588: I would reference Fig S6 here, since it is on a linear scale. And you should change S7 to linear scale.

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*)."