Response to Anonymous Referee #2 for "An evaluation of global organic aerosol schemes using airborne observations"

We thank the referee for their comments and suggestions. Below are our responses (in red) to each of the **comments (in black),** including the proposed changes to our revised manuscript. Given extensive additions to the SI as a result of review, we do not quote text changes below but provide a track changes version of the main text and the SI. For details about the OA schemes that are unchanged from their original implementation, we have chosen to refer the reader to the requisite study rather than suggest these developments were part of our work. However, in order to limit any confusion, we have now included these details in the SI.

Use of ground measurements. It is not clear why the authors restrict their model evaluation to airborne measurements when there are also ground measurements available in the same areas (US, Europe) for the simulated periods. This is especially problematic because they conclude that the simple modeling scheme is especially attractive for health studies. They also suggest that based on the airborne measurements both schemes appear to overpredict OA in the lower troposphere. This comparison (at least with the IMPROVE and EUCAARI measurements) should be added to the revised paper.

We agree that a comparison with ground observations would provide a more constrained understanding about exposure. However, this is not the focus of this study and we chose to limit our analysis to aircraft observations for the following reasons:

1. Ground networks are only available in a few locations (US and Europe)
2. These ground networks use different instrumentation; our goal was to use consistent AMS measurements.
3. We could have used globally distributed AMS observations, such as in Jimenez et al. (2009), however comparisons of global models with surface sites are more susceptible to representation errors and sub-grid meteorology that are both challenging to address. We specifically designed the study to focus on the regional constraints offered by airborne measurements around the world that sample OA under a range of conditions.
4. Health impacts are not the focus of this work (i.e. we do not focus on surface concentrations or exposures), therefore observational exploration throughout the full troposphere seemed best suited for exploring the OA budget.

We have modified the main text by adding more details explaining our rationale.

Treatment of POA. It is not clear how the POA emissions and the POA atmospheric chemistry are treated in the two models. Some major issues:

(a) Why is the POA emission rate (based on Table 2) in the simple model 21.8 Tg/yr and in the complex model 55.4 Tg/yr? Based on the paper there is a 27 percent increase of the emissions while these numbers suggest a 150 percent increase.

In the simple scheme, 50% of the POA is emitted as EPOA (Emitted Primary Organic Aerosol) and 50% is emitted as OPOA (Oxygenated Primary Organic Aerosol) to approximate the near-field aging of EPOA. Total OC emissions are 31.2 TgC. Thus, given the OC:OM ratios of 1.4 and 2.1 assumed for EPOA and OPOA respectively, total POA emissions in the simple scheme are 21.8 Tg EPOA and 32.8 Tg OPOA for a total annual POA emission of 54.6 Tg. In the complex scheme, all POA is emitted as gas-phase EPOG after scaling the same inventory used in the simple scheme by 27% to account for the extra gas-phase material. Total primary emissions in the Complex Scheme are thus solely from EPOG gas phase emissions and amount to 55.4 Tg. We agree that these details can be difficult to follow and have therefore added more information to the main text (and corrected the simple emissions from 54.3 Tg yr^{-1} to 54.6 Tg yr^{-1}). We have also included more information in the SI.

(b) What is the assumed saturation concentration of the POA in the complex model? What is its enthalpy of evaporation?

As described in Pye et al. (2010), EPOG saturation concentrations are 1646 and 20 $\mu g\ m^{-3}$ at 300 K. As EPOG ages to OPOG, the volatility of the reaction products decreases by a factor of 100 (Pye et al., 2010; Grieshop et al., 2009). An enthalpy of vaporization of 50 kJ/mol is assumed in this study. We have added these details in the SI with the appropriate references.

Anthropogenic SOA. How is this simulated in the complex model? Why is so much less than in the simple model? The production of SOA from IVOCs in the complex model is not described well. There is little information about how the IVOC emissions (shown in Table 1) have been estimated and the corresponding yields used. The sentence regarding the use of naphthalene as a proxy (lines 169-171) does not clarify this issue.

The anthropogenic SOA in the complex scheme is as described in Pye et al. (2010). To address the reviewer's questions, we have added more detailed information on the complex OA scheme (and the formation of SOA from IVOCs and aromatics) to the SI and have also expanded the discussion in the main text to provide greater clarity and address the discrepancy.

POA lifetime: Line 306. This is quite difficult to understand and it appears to be counterintuitive. The information about the POA treatment is limited. For example, the saturation concentrations of the two products used (lines 152-153) are not given. Also the conversion of the EPOG to the OPOG is not explained and the corresponding parameters for the reactions and volatilities are missing. Finally, there is no information about the removal of the corresponding organic vapors (e.g., assumed Henry's law coefficients). The authors should explain physically how a 27 percent increase in emissions (with some of them in the gas phase) leads to a more than 50 percent increase of the burden of particulate POA.

As with the previous points, we refer the reader to the original paper on this scheme (Pye et al., 2010) for a complete description. In light of the reviewer's comments, we have expanded the discussion in the main text and have included the information below in the SI:

EPOG saturation concentrations are 1646 and 20 μg m^{-3} at 300 K. As EPOG reacts with OH and ages to OPOG, the volatility of the reaction products decreases by a factor of 100. The reaction rate for POG to OPOG (K_{OH}) is 2 x 10^{-11} cm^3 $molec^{-1}$ s^{-1}

The 27% increase in emissions are all in the form of gas-phase POG (with a OM:OC of 1.4) that is hydrophobic and is advected higher in the troposphere before forming OA where it is less likely to be deposited. It is then aged to OPOG (with a OM:OC of 2.1). This sensitivity to temperature (and thus altitude) results in the formation of organic aerosol aloft, in a manner that the simple scheme cannot simulate due to its non-volatile representation, and leads to higher concentrations in the upper-troposphere where the aerosol experiences longer lifetimes.

The fraction of gas-phase OA precursors wet deposited is dictated by the liquid to gas ratio for a grid-box at any given timestep. For a soluble gas 'i', this ratio is calculated based on the following relationship:

$$\frac{C_{i,L}}{C_{i,G}} = K_i^* * L * R * T$$

where K_i^* is the effective Henry's law constant that is calculated using the van't Hoff equation (Jacob et al., 2000), L is the cloud liquid water content, R is the ideal gas constant and T is the local temperature. Each organic gas-phase species has an associated Henry's law solubility constant (in M/atm), volatility constant (in K) and pH correction factor which is defined in the GEOS-Chem species database.

EPOG Henry's Law Solubility Constant = 9.5 M atm^{-1}

OPOG Henry's Law Solubility Constant = 1 x 10^5 M atm^{-1}

Lifetimes of OA components: A wide range of global average lifetimes (from 3 to 11.5 days) is predicted. This is rather surprising for particulate matter components that should have similar size distributions. A discussion of how removal is parameterized and an explanation of these unexpected differences among different components are needed.

We agree with the comment and were similarly surprised by the wide range in lifetimes. Closer inspection suggests that this is due to a few different factors. We have expanded our discussion in the main text and the SI to cover these points.

1. Spatial Distribution in Aerosol Types – Species emitted over marine / tropical regions experience a higher risk of being deposited via wet deposition than aerosol over more temperate regions.
2. Hydrophilic nature of the aerosol – EPOA is treated as hydrophobic, rainout is turned off and the rate constant for conversion of cloud condensate to precipitation is halved, leading to a longer lifetime against wet-deposition. Oxygenated OA (OPOA and SOA) are treated as hydrophilic and thus have shorter lifetimes.
3. Vertical Distribution – The complex scheme treats OA as semi-volatile. As a result, both POA and SOA species are often in the gas phase in the warmer parts of the troposphere and are less likely to be deposited. These gaseous vapors are advected to the upper

> troposphere where they then form OA aloft. These aerosol species then have a longer lifetime against deposition since they are less sensitive to meteorology in the lower troposphere.

We refer the readers to Wesely (1989), Zhang et al. (2001), Jacob et al. (2000), Liu et al. (2001) and Amos et al. (2012) for details on dry and wet deposition in GEOS-Chem and have included more details in the SI.

Uncertainty of measurements. The authors site a 34-38 percent uncertainty of the measurements that they use. However, this is not taken in the evaluation of the two models. Could the model errors and biases be due to the corresponding measurement errors? Could most of the R2 difference from unity just be due to these errors? These issues should be addressed and taken into account in the corresponding conclusions. For example, if one of the model was perfect but the measurements had a 35 percent uncertainty what would be the R2 between measurements and predictions?

This is an interesting point. 33% of the modeled data-points fall within the observed uncertainty. We have added a statement to the conclusion that addresses this. If the model was perfect but you add noise (normally distributed with a mean scale factor of 1 and a SD of 0.14) to the observations, the R^2 is 0.97. Thus, measurement uncertainty is unlikely to be a major driving factor in the model-obs R^2.

Averaging times. The evaluation of any model does depend on the timescale investigated. This timescale appears to be 10 min. If this is the case, it should be mentioned in all graphs, tables but also in the text. An analysis at different timescales (say 1 hr and campaign average) could provide some additional insights. This would allow the authors to include all their measurements (avoid excluding the highest values).

We have added more explicit references to the 10 min timescale in the main text and referenced this in the figure captions. In addition, campaign average statistics are available in Table 2. We chose not to average over a more extended period given that the spatial variance in aircraft observations provides useful information that we prefer not to degrade.

Model performance statistics. The discussion is based on R2 and normalized mean bias (NMB). I think that the errors (both absolute and normalized) should be included in the analysis. For example, they appear to be quite useful in the discussion related to Figure 12.

The choice of metrics can be somewhat subjective and an initial draft of our study did include multiple metrics. However, we ultimately chose to limit ourselves to NMB and R^2 in order to maintain consistency and readability. Absolute differences and ratios are included in the SI (Figure S4). We found that the conclusions of our manuscript do not depend on the choice of metrics and have added a section to the paper explaining our choice of metrics along with their descriptions.

Role of enthalpy of evaporation. Given that a lot of the measurements were collected at relatively low temperatures the predicted partitioning of the semi-volatile OA components will depend on the assumed enthalpy of evaporation. It is not clear which values are used in the present work. The sensitivity of the conclusions of the paper to these assumed values should be discussed.

We assume an enthalpy of vaporization of 50 kJ mol^{-1} and have included that information in the main text and SI.

Underestimation of OA. Is it still appropriate to say (abstract and introduction) that models significantly underestimate observed OA concentrations in the troposphere?

The most recent studies including both models here tend to overestimate OA levels. Some additional discussion of the recent OA modeling work is needed here.

Certain recent studies with the GEOS-Chem model (such as Marais et al. 2016) have been shown to match the magnitude of SEUS observations, but since the scheme was developed and optimized for this region, this cannot be viewed as an independent test. The simple scheme as implemented here has not been previously evaluated in the literature (the older OA scheme in GEOS-Chem used different parametrizations) but previous evaluations of OA with GEOS-Chem have demonstrated a tendency to underestimate OA at the global scale (Heald et al., 2011; Spracklen et al., 2011; Hodzic et al., 2016). Our statement is intended to convey a general historical context (particularly in the GEOS-Chem model) as opposed to serving as a blanket statement for all regions.

Sources of bias. The complex model tends to overestimate OA levels. It is not entirely clear which attributes of this model (at least compared to the simple one) are contributing to this tendency for overestimation. Is it the isoprene SOA that is included in different ways in the two models? Something else? Some additional discussion is needed.

The discrepancy between the simple and complex scheme is the result of a number of factors and can be attributed to different reasons depending on the location. As Figure 11 shows, the complex model overestimates in many regions, but underestimates in others, and the reasons for these biases depend on the regional factors which are discussed in the main text. In general, Figure 4 shows that the burden in the complex scheme is higher generally because of more isoprene SOA and primary OA.

Sesquiterpenes. It is not clear how the models treats SOA from sesquiterpenes. In Figure 1 this pathway is missing. In line 168 they are mentioned as a source of SOA. The yields used are surprisingly low at least for the simple model: they are equal to those of the monoterpenes, despite the much larger size of these molecules. The yields used in the complex model are not clear. The same applies to their importance as SOA precursors in these simulations.

Thank you for pointing out this omission. In the simple scheme sesquiterpenes are assumed to have a 10% SOA yield. In the complex scheme, they are treated using the VBS framework (described in Pye et al. 2010) and have an average yield of 42% at 298K with a C_{OA} of 10 ug m^{-3}. In both cases, the resulting SOA is lumped with SOA from other terpenes (TSOA). We have clarified the description in the main text and added the relevant details to the SI.

Organonitrates. The treatment of organonitrates requires some additional explanation given that they are implicitly included in the SOA formed in the parameterizations of Pye et al. (2010). Do they represent additional SOA mass or do they describe a subset of the SOA predicted by the parameterizations used? Could this be contributing to the overestimations? Finally, there has been considerable work quantifying the organonitrate PM levels in Europe (Kiendler-Scharr et al., GRL, 2016) during the simulated EURAARI period that is not used here.

The complex scheme uses an explicit mechanism to model the formation of organo-nitrates based on work by Fisher et al. (2016). The Pye et al. (2010) paper does not explicitly model SOA from this pathway, instead using lumped SOA yields within the VBS framework to model the SOA resulting from oxidation with NO_3. The simple scheme parametrization does not explicitly account for organo-nitrate formation. We have outlined the mechanisms used in this study in the Supplementary Information and also point the readers to the appropriate references for further detail.

Regarding new developments in model organo-nitrate treatment, the focus of this work is not specifically on organo-nitrates so we have not incorporated more recent developments in the field. We have added a reference to Kiendler-Scharr et al. (2016) in the discussion, suggesting that a more constrained treatment of organo-nitrates could improve model performance over Europe.

Model performance in Europe. The parameterizations used in both models have been based, to a large extent, on US field measurements. The relatively poor performance of both models in Europe during EUCAARI requires additional attention. The comparison with the ground sites would be quite helpful here. What is the source of the high predicted concentrations which are inconsistent with the measurements?

Both schemes largely under-predict OA concentrations over Europe, potentially due to an underestimate in terpene SOA. Unfortunately, to our knowledge there are no publicly accessible airborne AMS observations over Europe to further investigate this bias. We have expanded our discussion on this topic.

Model behavior during the winter. This is significant weakness of the paper given that similar modeling exercises suggest that the models have significant problems in the winter. The authors can use ground measurements to close this gap. They could also change the title of their paper to specify that this evaluation is only for the summer and spring.

Given the paucity of airborne AMS observations in the winter (with to our knowledge only one publicly accessible campaign during this season – WINTER 2015) and the computational constraints of running more simulations, as stated in our manuscript we do not feel that we can adequately characterize the full seasonal cycle. Introducing comparisons with ground observations over limited regions would skew our conclusions about wintertime OA, and would be inconsistent with the design of the study (see our response above). We have expanded on this aspect in the conclusions and have included a statement in the abstract in the interest of transparency and clarity.

Evaluation for CO and other gases. Given the importance of CO for the parameterization used in the simple model, an evaluation of the ability of the performance to reproduce measured CO levels is needed. A similar evaluation for isoprene (in the appropriate regions) would be helpful.

An evaluation of the model ability to reproduce gas-phase species was not the focus of this study and was deemed to be out of the scope of this analysis, particularly given that these observations were not available for all campaigns. We have provided global emissions numbers for CO, NOx, and isoprene, so that these can generally be compared to other studies.

Characterization of the two models. I found the characterization of the two models as simplex and complex rather misleading. Their differences are mainly in the processes that they simulate, the emissions that they use, and the parameters that they use. They are by no means a simplified version of the other. While I understand the need for names, I also think that the paper should include a table with the main differences in the two simulations. These differences should be the emissions (Table 1 is confusing right now), the processes simulated in one and not the other, the differences in the simulation of the same process (e.g., assumed effective yields for a given OA level), and the different parameters used. If the authors can attribute the differences in predictions of the two models to simplicity or complexity, this should be done carefully and should be properly supported.

Thank you for the suggestion. In order to provide clarity and limit any confusion we have added a table with the relevant details to the SI and also included a section in the main text explaining the difference in emissions.

Some additional points

Lines 42-43. This sentence is confusing.

We have modified the statement to limit confusion

Lines 73-74. The formation of SOA in some of these models was by no means instantaneous. The VOC precursors were required to react first.

While some types of SOA were formed via the reaction of VOC precursors, the studies we are referring to use a fixed yield approach to the formation of many types of SOA. We have modified the wording in order to ensure clarity.

Lines 131-132. The names of the species in the code are not needed here.

We include them because the use of EPOA and OPOA are not standard in the GEOS-Chem community when describing non-volatile OA, and we want to ensure that future users are clear about our definitions. Thus, in the interest of clarity, we prefer to retain this.

Lines 141-143. The use of the term yield is rather confusing. Does this mean for example for fires that the assumed SOAP emission is 1.3 percent of the emitted CO? Something else? In Figure 1 it appears that CO is producing SOA making this even more confusing.

We have modified the terminology to ‘co-emission’ in order to limit any confusion.

Lines 142-143. Please explain the biogenic sources discussed here. I am assuming that this means isoprene and the monoterpenes. Referring to Figure 1 could help.

We have expanded on the statement to provide more detail.

Lines 147-150. Please explain what do these tracers represent and how their use allows the independent adjustment of model parameters.

The default model only includes one SOAP and one SOAS tracer, making it impossible to differentiate the different sources. Our modified scheme separately tracers the SOAP and SOAS from fire, anthropogenic and biogenic sources in order to establish the relative contribution from each pathway.

Table 2. I find this table confusing. For example, the four POA emission rates do not add up to the total. It also mixes actual emission inputs with estimates from steady state calculations. May be the information could be clearer if it was replaced by two tables.

We have added a paragraph to the text to clarify some of the confusion about atmospherically formed vs directly emitted OA

Line 202: The 54.3 Tg/yr of POA emissions in the simple model cannot be found Table 2.

POA emissions (54.6 Tg /yr) are composed of direct EPOA + OPOA emissions but do not include OPOA steady state formation (which is included in the table). We have edited the main text to limit any confusion and corrected an error to change 54.3 Tg yr^{-1} to 54.6 Tg yr^{-1}.

Figure S1 should probably be moved in the main paper. If needed it can replace of the figures with the vertical profiles.

We have moved Figure S1 into the main text.

The conclusion regarding the overestimation of the aqueous uptake of isoprene should be added to the abstract.

We have added a statement to the abstract.