Reply to anonymous referee #1

General Comments:

Review of Zhang et al (2021) GMD

The article presents a comprehensive comparison of three chemistry scheme – GOCART, RACM_GOCART and RACM_SOA_VBS – against each other and observational data from the ATom-1 flight dataset. Compared to GOCART which uses simple sulfur chemistry and aerosols and prescribed oxidants, RACM_GOCART represents an increase in the sophistication of the chemistry scheme by using the RACM chemical mechanism including interactive oxidants. The RACM_SOA_VBS scheme appears to use this more advanced chemistry and an updated SOA scheme using a VBS approach for SOA.

The authors first perform a global comparison of the GOCART and RACM_GOCART scheme for sulfate, PM2.5, H2O2, NO3 and OH to investigate the influence of the updated chemical mechanism. The authors highlight a large difference in sulfate aerosol between the schemes, attributing this to differences between the prescribed oxidants in GOCART and the interactive oxidants in RACM_GOCART. The global distribution of SOA simulated by the RACM_SOA_VBS scheme is also discussed.

Then all three schemes are compared to observations from ATom-1 flights with separate sections devoted to the flight data over the Atlantic Ocean and continental United States by comparing hourly model data to high frequency flight data.

The flight over the Atlantic encounters separate regions with high biomass burning emissions, from central Africa, and dust emissions from the Sahara. Chemical (CO, O3, OH, CO and H2O2), aerosol (EC, dust, sea-salt) and meteorological variables are compared. The RACM_SOA_VBS exhibits improved EC performance compared to the other two schemes, attributed to a more realistic wet deposition approach. All schemes are generally able to reproduce the latitudinal and vertical profiles of key chemical tracers although they also exhibit high bias in CO at low attitudes around the biomass burning region.

From the flight data over the United States, particular attention is paid to California and Kansas. RACM_SOA_VBS again is much better than the other schemes at high altitudes. Low altitude EC and sulfate are high biased in all schemes, attributed to discrepancies between real and simulated emissions, while the two schemes which use interactive oxidants, RACM_GOCART and RACM_SOA_VBS, are worse than GOCART with its prescribed oxidants for high altitude sulfate.

Finally, correlation analysis is performed for sea-salt aerosol in the GOCART scheme and key species produced from biomass burning. Model sea-salt is generally high biased and while the EC/CO ratios in the model and observations compare favorably, model O3 production appears to be underpredicted. This is attributed to the uncertainty in the associated VOC emissions.

This study makes good use of observational data and presents the model-observational comparisons in a clear way for the most part. I also commend the authors' efforts to compare a wide of range of chemical, aerosol and mereological data and highlight the link between oxidants and aerosols. However, I do not believe this paper should be published until revisions have been undertaken to the text and further investigation performed. Some of these revisions involve checking that figures are labelled correctly, sections are numbered correctly, units are provided and references to figures in the text are correct. I provide a full list of my comments below.

Reply: We really appreciate the reviewer's very helpful comments and suggestions. The paper has been revised throughout based on all of the specific comments listed by the reviewers, including the text, figures, section number, unit and references, etc.

Specific Comments

Section 2
2.2.1 - Given the importance of aerosol in this study, could you clarify what you mean by "bulk" aerosols for BC and OC? Does this mean that they are simulated to have as a single mode?
Reply: Yes. The bulk aerosol means they are speciated mass and speciated particle size with external mixing, no microphysics in the aerosol simulation.

2.2.2
Typographical error P6 Line 5 in citation
The phrase "rate constants and product yields from the most laboratory measurements (Stockwell et al., 1997)" is unclear - do you mean that most of the rate constants and product yields are derived from Stockwell et al (1997)?
Reply: Revised. See P6, L11.

To give the reader an idea of the complexity of RACM, it would be helpful for you to state:
- the number of chemical species

Reply: Revised and corresponding table and descriptions have been added. GOCART: 19, RACM_GOCART: 68, RACM_SOA_VBS: 103. See P5 L33 and the table in P25.

- the number of reactions

Reply: Revised and corresponding table and descriptions have been added. GOCART:4, RACM_GOCART:214, RACM_SOA_VBS: 233. See P6 L14 and the table in P25.

Can you also provide a description of the SOA scheme so that the update in the RACM_SOA_VBS scheme is clear? Do biogenic species such as isoprene make SOA in the atmosphere?
Reply: Yes, isoprene is making SOA. The scheme is the same as one implemented in WRF-Chem, which is described in detail by Ahmadov et al., 2012. We also provided a brief description of the SOA scheme in the revised manuscript. See P6 Section 2.2.3.

2.2.3
It is not entirely clear how the RACM_GOCART and RACM_SOA_VBS schemes differ in their chemistry. It appears that the RACM_SOA_VBS schemes uses the RACM scheme plus the SOA_VBS but the sentence (P6, line 16-18): "The RACM_SOA_VBS scheme includes photolysis reactions for multiple species, full nitrogen and VOC (anthropogenic and biogenic) chemistry, inorganic and organic aerosols." differs from the description of RACM_GOCART in Section 2.2.2, leading to confusion.

Reply: The major difference between the RACM_GOCART and RACM_SOA_VBS is in the aerosol scheme part, RACM_GOCART is using the bulk aerosol scheme of GOCART, the RACM_SOA_VBS is using the modal aerosol scheme MADE (Modal Aerosol Dynamics Model for Europe) with the VBS (Volatility Basis Set) approach based on SOA scheme. The GOCART scheme has much fewer aerosol species compared to the MADE_SOA_VBS scheme. We have clarified it in the revised manuscript. See P6 Section 2.2.3.

If RACM_SOA_VBS does use the same chemistry, these sentences should be either made to be identical
or the sentence in Section 2.2.3 removed. As the key update in RACM_SOA_VBS is the SOA VBS scheme, a short description of the VBS approach should be added. This should include how SOA precursors are generated - i.e. prescribed or produced by oxidation of biogenic/anthropogenic species and if so, which ones. For example, in Section 2.3.4 [sic], organic condensable vapours (OCVs) are referenced in relation to the VBS but they make little sense if additional detail is not provided Section 2.2.3.
Reply: We have rewritten Section 2.2.3 to make it more clearly, including a short description of the SOA_VBS approach.

A table showing the similarities and differences between the 3 schemes should be added to aid readers. An
additional column which highlights the added value of each update (e.g. relative to GOCART RACM_GOCART shows impact of improved chemistry) would be highly beneficial.
Reply: The 3 chemical schemes are relatively independent, that are not the upgraded from one to the other. However, they are the combinations of different aerosol/gas phase chemistry schemes. We have added a table to show the comparisons of these 3 chemical schemes in the revised manuscript. See the table in P25.

2.3.4 - this should be renumbered to 2.2.4
Reply: Revised. See P6 L35.

P6, Line 24 - missing "Model"
Reply: Revised. See P7, L5.

Section 3
General comment: Please change all time units from "..Z" to the time in UTC (e.g. 00Z to 00:00 UTC)
where UTC has been defined on its first usage as Coordinated Universal Time.
Reply: Revised. See Section 3.

Figure 2 - typographical error with units (assume micrograms/m3) although note that it is correct in
caption list.
Reply: Revised.

Information should be provided in the caption and at the first mention of Figure 2 in the text that the simulated model data (presumably?) corresponds to the surface concentration.
Reply: Added. See Section 3.

The caption in Figure 2 is also confusing. You state a 120-hour forecast – does this mean you have taken
an average over 5 days? Yet you also state a single time point – I do not understand how these relate.
Reply: 120 hour forecast is not an average over 5 days, however, an instantaneous forecast result at 120th hours. The forecast is from 00:00 UTC July 29th, 2016, and it is the instantaneous predicted results at 00:00 UTC August 3rd, 2016. We have clarified and modified the descriptions in the revised manuscript throughout in Section 3.

The description of how the chemical forecasts is performed is confusing. When you say the chemical species are "cycled every 24 hours from their output", do you mean that the final chemical state of the previous day is used to initialise the chemical state of the model for the next day?
Reply: Yes. We have clarified it in the revised manuscript: "The retrospective daily forecast is using the cycling way for the chemical fields since we don't have the data assimilation system for the chemical model. It means that the meteorological fields are initialized by the GFS meteorological fields every 24 hours, while the chemical fields from the last output (forecast at 24:00 UTC) are used as the initial conditions of the current forecast (00:00 UTC), except the stratospheric O_3 above tropopause which are from satellite derived fields available within GFS." See P11 Section 4.

P8, Line 39 onwards - The statement that concentrations in excess of 100 micrograms[/m3] are simulated in "wide areas of south Africa, areas of south Asia and eastern China" should be reconsidered. Concentration > 100 micrograms/m3 are only simulated in small part of these regions.
Reply: Revised. See P9 L15.

P9, Line 1 - concentration should have units of 100 micrograms/m3
Reply: Revised. See P9 L14

P9, Line 5 - it is said that sulfate is not shown yet Fig 2 (d-f) are described as plots of sulfate. This needs to be clarified. If the intension is to show the sulfate, then it does appear that there is a good spatial correlation between the regions of lower PM2.5 in RACM_GOCART and regions of lower sulfate.
Reply: Revised as "see Fig.2 right column".

"We find the reductions of sulfate are about 10 µg/m3 on the order of 40-80% over the eastern US and are up to 40% over eastern China in RACM_GOCART (Fig. 2b)."While this statement appears acceptable for China, the part about the eastern US is confusing. Does this is mean that a reduction of 10 µg/m3 corresponds to a 40 80% reduction in sulfate or the 10 µg/m3 reduction occurs over 40-80% of the eastern US?

Reply: It means a reduction of 10 μg/m3 corresponds to a 40-80% reduction in sulfate over eastern US.

No information about units is given for Figure 3 in the text or caption. Units of ppb is given the caption list but I doubt this is correct for some of the species, particularly for OH. It should also be stated if these are surface concentrations.
Reply: They are all surface fields. The figure captions and corresponding text have been revised throughout.

Given the acknowledged importance of OH in production of sulfate aerosol, the same analysis as presented in Figure 3 for 12:00 UTC should be presented, i.e. when OH concentrations over South America, Africa, Europe and a large part of Asia are non-negligible.
Reply: We thank the review's suggestions. The 12 UTC figures have been added, which shows the reduction over Africa, India and Asia. See P10, L4.

P9 Line 25-26. "The general spatial distributions are comparable to the satellite observation." This statement is vague and not supported by observational data in the study. Basic comparison to a readily available satellite product or observational data for some of the key urban areas (e.g TOAR dataset) should be made if this statement is to be retained. I am also surprised by the low (<30 ppb) O3 over China and India.
Reply: Well comment taken. We have removed this sentence. Fig.4 is a 5-days (120-hours) forecast of O_3 at 00 UTC and 12 UTC, it is early morning and dark evening over East China. The relative low surface O_3 is likely due to the O_3 titration during the early morning and nighttime periods. See P10 L16.

Figure 4's caption needs to be amended to include the units and clarify the data corresponds to surface O3 (I acknowledge that the units are given the caption list). The caption also has the same issue with confusion surrounding the use of the "120-hour forecast" and single time point – please clarify. (Same for Figure 5)
Reply: Revised.

While the general SOA distribution looks reasonable in Figure 5, it is hard to understand the impact of the VBS update in the RACM_VBS_SOA scheme without seeing a comparison to SOA from the RACM_GOCART scheme. Either comparison to that or to observation should be performed to provide context for these results. Can you provide an explanation as to why SOA over the Amazon, the largest global source of isoprene and monoterpenes, is lower than SOA over the Mediterranean sea?
Reply: As we answered in above section, the RACM_SOA_VBS is not an update of RACM_GOCART scheme and they are two different schemes, especially in the aerosol part. For the aerosol scheme, the RACM_GOCART is using the GOCART aerosol scheme, so there is no SOA in the GOCART scheme. While both RACM_GOCART and RACM_SOA_VBS are using the gas-phase chemistry scheme of RACM. The globally emission of biogenic emission is still based on the old MEGAN version. So we would expect there are still some biases for the SOA simulation over some regions, which will be improved by updating the biogenic emissions.

Section 4
Section 4.1
P10 line 22 and throughout, EC is not defined - may mean elemental carbon or organic carbon and its
relationship to BC should be made clear
Reply: Defined. See P11 L18.

P10, Line 23 - which model?
Reply: Added. See P11 L17.

You should differentiate between RACM_GOCART and RACM_SOA_VBS to say explicitly that the RACM_GOCART performs well at low altitudes but it high biased at high altitudes while the RACM_SOA_VBS approach performs better.
Reply: We thank the reviewer's suggestions. Corresponding statements have been added. See P11 L23.

I do not agree that the latitude-height profiles are captured "very well". For the RACM_GOCART and RACM_SOA_VBS schemes there is roughly a 100 ppb high bias in the lowest 2 km in the tropics and this should be made much clearer. I acknowledge you state there is a low bias above ~ 6 km which is a fair statement but the sentence you mention this in is confusing to read. I would suggest you split it in up into two sentences which deal with the high altitude and low altitude sections separately.
Reply: Revised. See P11 L35.

The vertical profiles of the two mechanisms are similar and, as they have the same chemistry scheme, this bias could be a result of the chemistry, vertical transport and/or emissions. Your suggestion regarding biomass burning injection height appears sensible. Can you provide any additional information show where the BB emissions are injected? Given that tropical CO is high biased from the surface, could it be that the BB emissions are added at the surface?
Reply: The biomass burning emission is originally at the surface layer after reading it from the input data. However, the model includes a plume-rise module, which is able to redistribute the biomass burning emission based on the meteorological variations instantaneously (every time step) and widely used in other models. Unfortunately, the model did not archive the injected emission distribution in the retrospective or real-time run. We do agree the reviewer's comments, that bias may related to horizontal and vertical transport (fire emission injection height) since the places with large bias is not the fire emission source region, however the downwind areas. The discussion and descriptions have been modified. See P12 L1-3.

It is interesting that O3 is well modelled at the surface despite the high CO bias, and it could be the case that the mechanisms are getting the right answer for the wrong reasons (i.e. O3 is good because CO is too high). It would be instructive to add the observed and modelled NO2 concentrations, if available, as this would provide further information about the injection height of BB emissions and ozone production efficiency of the mechanisms.
Reply: We thank the reviewer's very good suggestions. CO is only one of the precursors for O3 productions, VOC is also very important for O_3 production. While the global biogenic emission in the current model is very old version and may not be good enough, we are trying to include newer

biogenic emission in the future, and hopefully it will help to identify this issue and improve the model performance. Unfortunately, we do not have the observed NO_2 concentrations to compare with the model results. But we will keep on paying attention to it in the future if any NO_2 observation is available. See P12 L8-11.

P12, Line 4 - do you mean the biomass burning, in which case it would be Figure 9(a) or dust and so Figure 9(b).
Reply: Yes. Revised. P12 L23 and the Figure caption of Fig.9.

Looking at Figure 9, the injection height of the BB emissions could be an issue and, as mentioned previously, I think it would be useful to see the NO2 concentrations because if these are quite similar between mechanisms and observations, it would solidify the claim that the mechanisms reproduce O3 well. In addition, if there are any observations of other biomass burning tracers such as acetonitrile (I assume this is not a tracer in the model), this would also be useful to plot since it would provide some information about where the biomass burning emissions are actually reaching up to.
Reply: We thank the reviewer's very good comments. We do have the comparison for NOx which is about 80% NO_2 on the average (See Figure A below in the biomass burning event as Fig.9a). The NO_x profiles from models are quite comparable to the observation except some overestimate at the upper levels. Also, there is not acetonitrile in both the model and observation. But that is a very good suggestion to investigate the similar issue in the future if these biomass burning tracers are available in the model and observation.

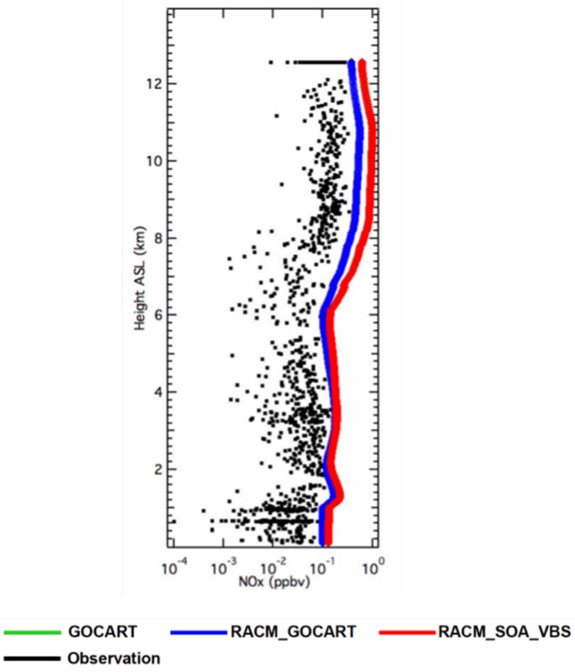

Figure A. ATom-1 observations and model vertical profiles of NOx in the biomass burning events. The biomass burning plume is from August 15, 2016, profile #16 near 20°S.

Section 4.2
Section 4.2 P12 Line 31 please change anthrophonic to anthropogenic
Reply: Yes. Revised. P13 L29.

Your suggestion of model emissions being high biased seems reasonable. However, I would like to see a short run where BC and sulfate emissions are scaled by 0.5 to test whether this is indeed a major driver of the model high bias. If this leads to improved model performance, it would strengthen the argument that the simulated chemistry is doing a good job. It is interesting to note that the GOCART scheme, which uses prescribed H2O2, OH and NO3, performs better for high altitude sulfate. However, given your suggestions regarding the possible high bias of model emissions, it is not clear whether the GOCART prescribed oxidants approach is getting the right answer for the wrong reasons. This makes conducting a 0.5x scaling test even more important. The profiles of CO, HCHO and O3 from the 0.5x scaling should also be discussed since you suggest the emissions bias may affect them too.

Reply: Before the model finally go to real-time run, we have performed the emission scaling factor several times as regression test. And the model should pass this regression test before starting the retrospective and real-time forecast. The modeled concentration is almost linear to the emission scale factor. The higher of the emission would results into the larger tracer concentrations. This is also the popular way used to tune the model by giving a scaling factor to the emission. However, this should be normally based on some scientific published evidence, and it is dangerous to apply one scaling factor globally. Due to the uncertainties in different emission inventory, for the retrospective and real-time forecast, considering the expensive computation time, we normally do not run the sensitivity experiments by adding random scaling factor into the emission globally since it may cause big uncertainties in the emission globally and the results can't be used for retrospective and real-time forecast evaluation to compare with observation. Instead, we tried to use different anthropogenic and biomass burning emission inventory to quantify the impact of different emission on the forecast and quantify the emission uncertainties, which is more valuable for both the retrospective and real-time forecast. It can also help to provide more information to the emission data group to improve the emission inventory. This idea is similar to apply scaling factor into emission to quantify the impact. See Figure B, here we are applying two different anthropogenic emission inventories of HTAP and A Community Emissions Data System (CEDS, Hoesly et al., 2018) into the model of GOCART scheme. The CEDS anthropogenic emission is much stronger than HTAP over California for SO2 (see Figure C). Thus, you can see significant enhancement in sulfate concentration near the surface of California when using CEDS anthropogenic emission, which due to the major precursor of SO2 for sulfate productions. The other case is the OC by using different biomass burning emission of MODIS and Blended Global Biomass Burning Emissions Product (GBBEPx) as a preliminary testing runs for the same GOCART scheme, however coupled with FV3GFS model. Obviously, with much stronger GBBEPx emissions over California and US, would result into much larger OC concentrations near the surface of California and US.

For the background testing by applying a scaling factor to prescribed H_2O_2, OH and NO_3, we don't think is a workable way for the sensitivity experiments. Because the H_2O_2, OH and NO_3 are coming from the climatological fields of GMI model which are no linear correlation however, they are in a chemical balance as a triangle. Any scaling factor applying to them would break the chemical balance and cause uncertain issue in the chemical reaction, which is out of the normal chemical reactions. To better identify the uncertainties in the prescribed H_2O_2, OH and NO_3, what we do was trying to compare them with the ATom-1 observation. Please see our answers in next item.

We have included all of the descriptions and figures into the revised manuscript. See Section 4.2 in P13 L34 to P14 L11.

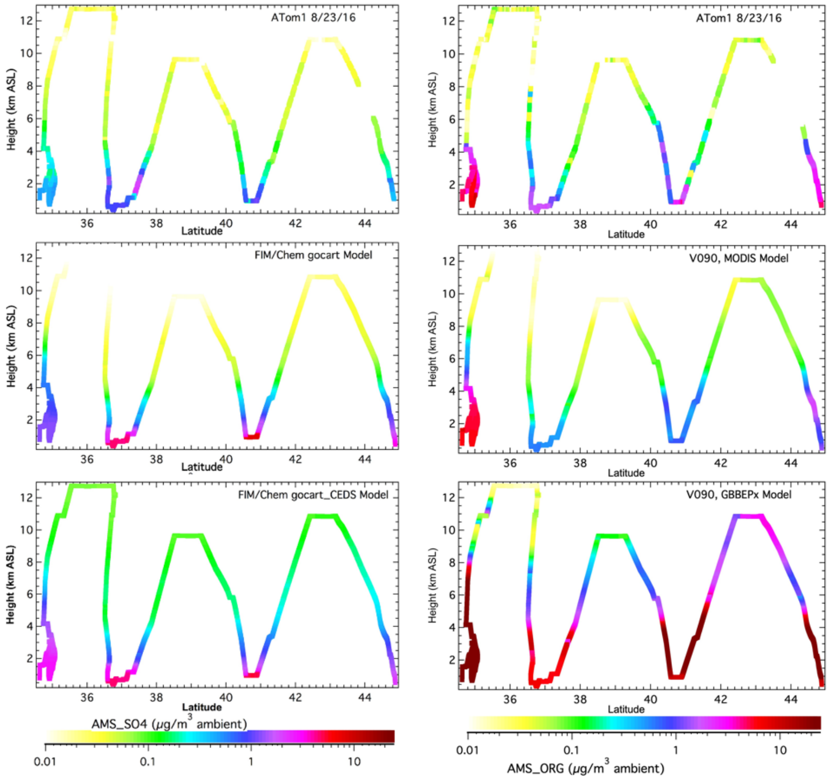

Figure B. Height-latitude profiles of sulfate and OC for ATom-1 and model forecast with GOCART scheme over United States on August 23rd, 2016 for ATom-1 (top), using HTAP (middle left) and CEDS anthropogenic emission (bottom left), and using MODIS (middle right) and GBBPEx (bottom right) biomass burning emission.

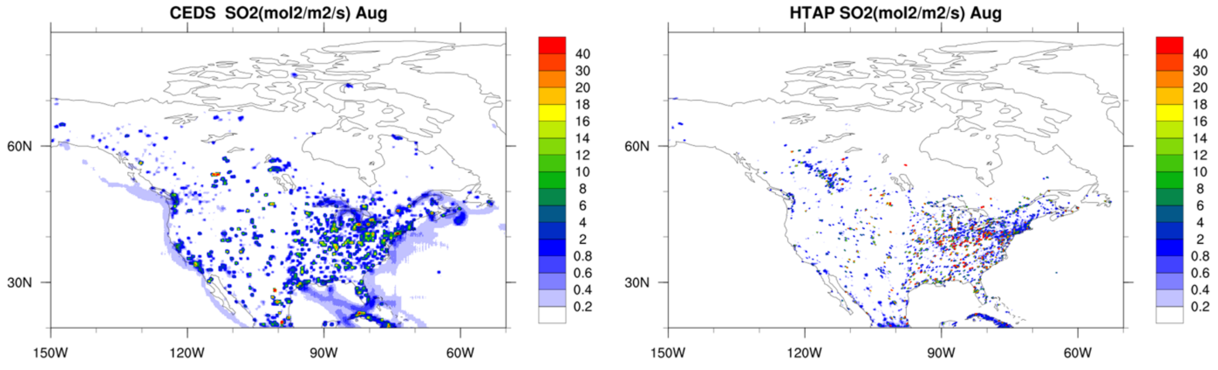

Figure C. Anthropogenic emissions of CEDS and HTAP for SO_2 on August.

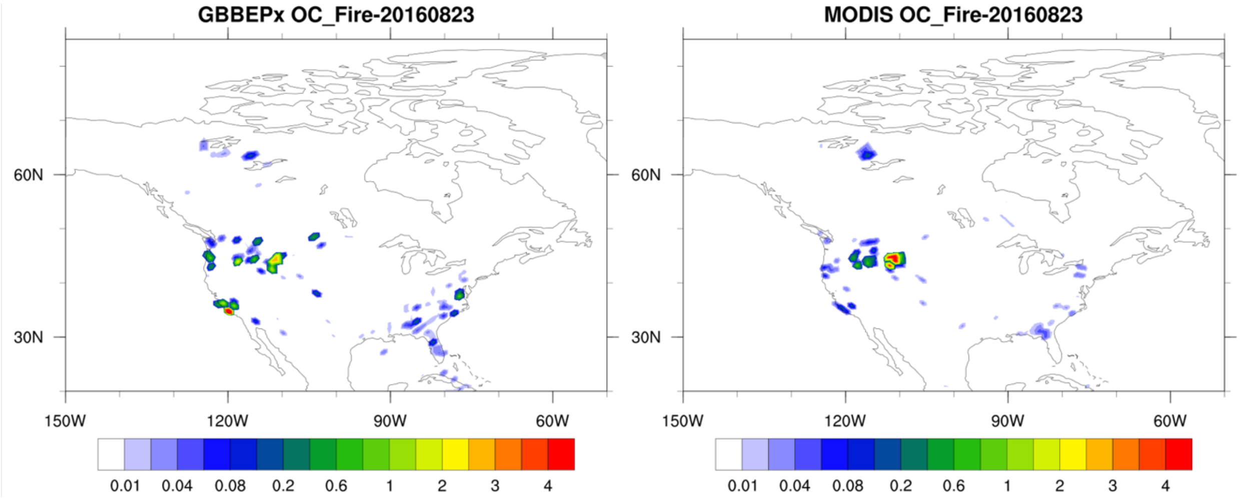

Figure D. Biomass burning emissions between GBBEPx and MODIS for OC (Unit: ng/m2/s).

It would also be helpful to show how prescribed oxidants used in GOCART compare to observations and the oxidants simulated by the RACM_GOCART and RACM_SOA_VBS since this would provide further information as to the drivers of the sulfate high/low bias. For example, if the prescribed oxidants are lower than those in RACM_GOCART and RACM_SOA_VBS oxidants, it could explain the lower sulfate production and concentrations.

Reply: We do have the comparisons of OH and H_2O_2 in GOCART, RACM_GOCART and RACM_SOA_VBS with ATom-1 observation (see Figure E). We can not compare NO_3 because we don’t have the observation data. We can see that the prescribed OH is close to the ATom-1 observation, which may be the major factor contributing to get better sulfate production in GOCART. Also, from Fig. 2 and Fig. 3 in the manuscript, OH looks like the major factor impact on the sulfate concentration.

Considering the sulfur chemical reaction mechanism and the aerosol scheme in RACM_SOA_VBS is completely different to that in GOCART and RACM_GOCART, the comparison of oxidants may not be the only reason causing the differences, which still need more investigations. Though the sulfur chemical reaction mechanism is consistent in GOCART and RACM_GOCART, but the fields of OH, H_2O_2 and NO_3 are not from the prescribed oxidants used in GOCART. Different to the GOCART, the SO_2 and sulfate in RACM_GOCART are also impacted by the gas phase chemistry of RACM scheme, which is also difficult to quantity the impact from the OH, H_2O_2 and NO_3 and need further study.

We have included all of these descriptions and figures into the revised manuscript. See P14 L20-23.

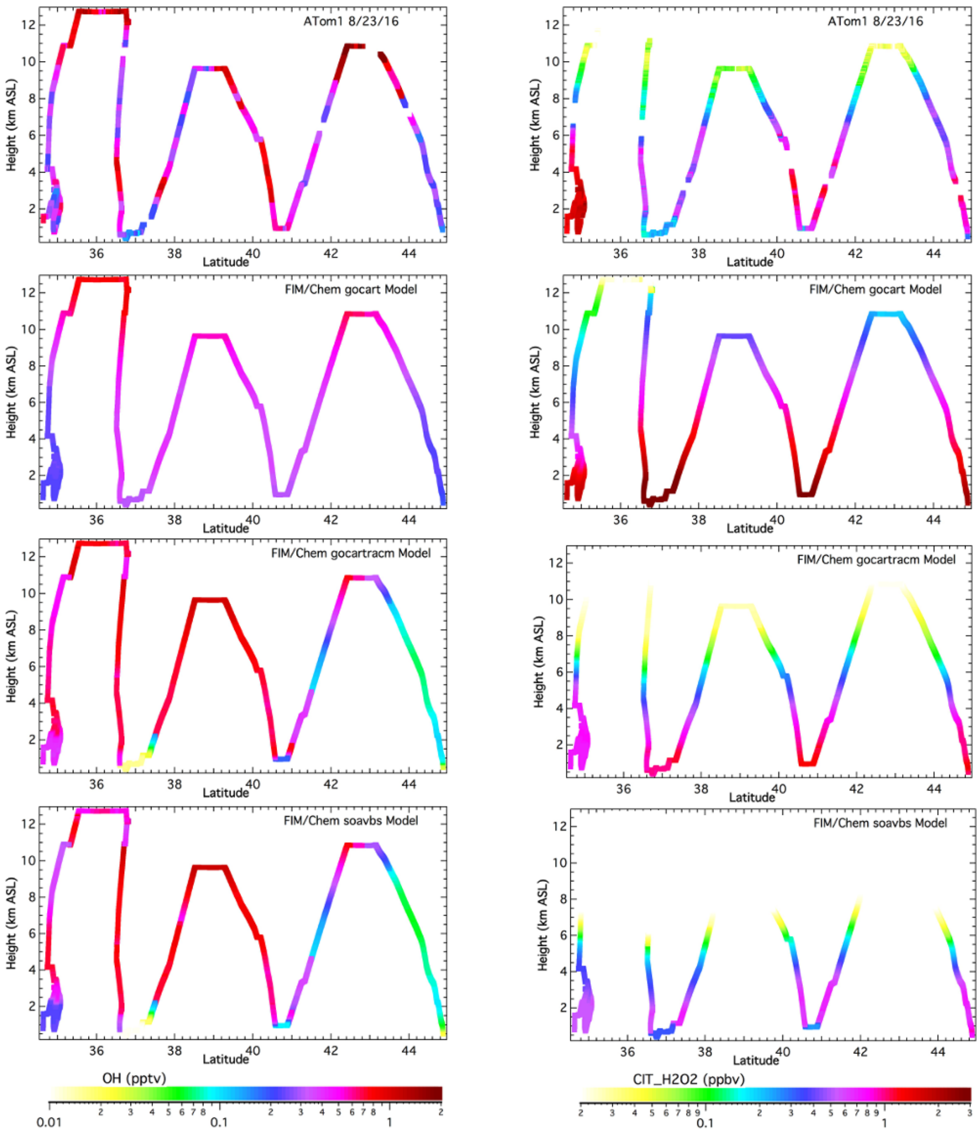

Figure E. Height-latitude profiles of OH and H2O2 over United States on August 23rd, 2016 for (a) ATom-1; (b) GOCART; (c) RACM_GOCART; and (d) RACM_SOA_VBS.

Section 6

Page 14 line 16 - The sentence starting "Compared with the .." should be restructured. The GOCART scheme is the simplest mechanism considered here and therefore should be the baseline for comparisons while here the RACM_GOCART is the base for comparison. Furthermore, the next sentence flips to have the RACM_GOCART scheme as the default scheme for comparison. Please be consistent and compare RACM_GOCART to GOCART in each.

Reply: Revised. P15 L32-33.

I would like to see greater structure to the conclusion, mirroring the structure of the main paper. You start with the global analysis which is fine, but this should be clearly signposted before each observational comparison - Atlantic, California and Kansas - are introduced in separate paragraphs with the final paragraph on the sea-salt and BB comparisons.

Reply: Revised througout. See Section 6.

Further to my previous comments on Section 4, please highlight the issue with tropical high biased CO at low altitude where you highlight the low bias above this point. You could also retract the statement that EC is "very good" with the GOCART and RACM_GOCART mechanisms, particularly as you go on to discuss how it is improved in the RACM_SOA_VBS mechanism. I think it would be fair to say that the EC is simulated well by GOCART and RACM_GOCART mechanisms up to 4 km but above this the mechanisms are high biased.
Reply: We thank the reviewer's very good suggestions. It has been revised. See P16 L11.

Further to the comment about BB emission injection height, a comment should be included regarding the NO2 profiles.
Reply: We thank the reviewer's suggestion. As we mentioned above, we don't have the NO_2 data in the observation for comparison. We do have the comparison for NO_x which is about 80% NO_2 on the average (See Figure A above in the biomass burning event as Fig.9a). The NO_x profiles from models are quite comparable to the observation except some overestimate at the upper levels. Also, NO_y is largely NO_x, especially in fresh fire plumes. To address this concern of the reviewer, we have added analysis and discussion based on Fig.17 in the revised manuscript, showing scatter plots of observed and modeled NO_y versus CO and O_3 versus NO_y for the tropical biomass burning signature seen between 22°S and 12°N below 6km altitude. We have also modified the text in Section 5 to include a brief discussion of what the reviewer has noted. See P15 L15-25.

Please change CH2O to HCHO for consistency.
Reply: Revised. See P16 L21.

Again, while the model anthropogenic emissions over California may drive the model bias for CO etc., the results of the 0.5x scaling tests should incorporated into the conclusion. Please also note the high bias over ~ 3km for sulfate with the RACM_GOCART and RAC_SOA_VBS schemes over the USA when you mention GOCART's good performance.
Reply: We thank the reviewer's comments. We have added the conclusion about using different anthropogenic emissions into the conclusions (similar effect as the 0.5xscaling tests). We also included the high bias over ~ 3km for sulfate with the RACM_GOCART and RAC_SOA_VBS schemes over the USA when you mention GOCART's good performance. See P16 L31.

Reference

Hoesly et al, Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). Geosci. Model Dev. 11, 369-408, 2018.