Authors' response to reviewer comments on "Tropospheric ozone in CMIP6 Simulations" by Paul T. Griffiths, Lee T. Murray et al.

This paper evaluates the simulations of tropospheric ozone from the "preindustrial" (1850), through the present day, to 2100 undertaken as part of the CMIP6 chemistry climate model simulations. This should inform the next IPCC report, and is part of an ongoing multi-decadal project to provide this information for the IPCC reports. The timings for the submission of the paper is mainly driven by the IPCC timescales.

There is utility in publishing this paper. Having a new assessment of both the performance of the current generation of chemistry-climate models and their variability is useful. I would suggest publication after some changes.

There are however some disappointments inherent in this paper which are symptomatic of the CMIP process for tropospheric composition. The comments below are more directed to the wider CMIP community than the authors specifically. 1) The tropospheric chemistry modelling community appears to be disengaging from this process. Looking at the ACCENT (2006), ACCMIP (2013) and the present study there is a linear decay in the numbers of models taking part in this tropospheric ozone budget aspect. Interpolation would suggest that there will be no models engaging in the process by around 2023. It would be useful for the CMIP community to consider why this is the case, and think about how the outside community is valuing its activities. 2) Papers very similar to this have been being published for the last decades. The authors refer to Young et al. (2013) and Stevenson et al. (2006) as the precursors to this, and there are previous activities which are very similar going back to chapters in the 2001 IPCC report and earlier. It is not obvious that the models' ability to simulate ozone is getting any better over this timescale. One of the conclusions from this paper is the present day mean O3 burden (ACCENT to CMIP6) has only changed by 3% from 15 years of research. 3) It is also of concern that the tools used to analyse these models has not changed in these almost twenty years and the explanations for model differences have similarly not evolved from being a combination of chemistry, emissions, deposition and transport. Perhaps the authors would want to consider whether there needs to be advances in diagnostic techniques before the next model comparison exercise in the conclusions / discussions?

These points are not issues associated with this paper specifically and the authors don't need to reply to these questions but it may be useful for the wider community to think about this.

Conclusions

The community has been around the cycle of IPCC reported model comparison exercises for tropospheric ozone multiple times now over the last two decades. Figure 8 shows that for the preindustrial to the present day the model prediction (well the multi-model mean) hasn't essentially changed since the ACCMIP evaluation. The explanation for the spread between these models also hasn't really changed. It is some indistinct combination of different emissions, chemistry, deposition and transport in the model. It might be useful to the community for the authors to provide a potential vision of how things should change going forwards. Will the CMIP7 version of this paper look exactly the same as this? If not how should we make advances in the future?

Authors' response:

We thank the reviewer for the stimulating comments and share his concerns. We would like to respond to the introductory and concluding remarks together.

It may well be that the tropospheric modelling community is less engaged than previously. This may be a result of the increased scope and complexity of the CMIP6 data request, or may reflect the exclusion of chemical transport models from CMIP6 in favour of coupled general circulation/chemistry-climate models. The requirement of the CMIP6 DECK to include idealised CO2 experiments results in an exclusion of models that don't include a coupled carbon cycle.

The centennial scale experiments required for CMIP6 are undoubtedly expensive to perform, and this, coupled with the large number of sub-projects, may have meant that modelling centres have had to be conservative in their participation in order to conserve and prioritize available resources. The timeline of CMIP6 was perhaps also an issue: the expectation that CMIP6 would run over a number of years has meant that the timeline for data availability could perhaps have been clearer. Certainly additional experiments, for example AerChemMIP experiments, that would have contributed to this assessment paper were not available at the time of writing.

We share the reviewer's concern about the progress in the field, and there is little change in the estimates of the present-day ozone burden from CMIP5 to CMIP6. But CMIP6 has made a large step forward in the availability of diagnostic data for calculation of ozone budgets which allow us to understand model diversity through use of consistent production (o3prod variable), destruction (o3loss) and physical removal via dry deposition (dryo3) variables made available through the Earth System Grid Foundation data archive. There is a consistent definition of the tropopause available, which improves model estimates of ozone burden, and is objectively more accurate than the pre-industrial chemopause used for CMIP5. Our ability to diagnose the reasons for inter-model differences is constrained by the lack of required diagnostic output, and should be a focus of future MIPs. Full chemical diagnostic output is not necessarily needed for the whole period, but it is clear from this study that it would be useful to have it for selected time periods, such as 1850, the recent historical period and at the end of the experiments in 2100, where we are particularly interested in the inter-model range of estimates. We have noted this in our revised summary section.

This paper reinforces the point that having the correct diagnostic information is a crucial part of this work. We have added a section to the summary to indicate additional diagnostics, and a potential application of other diagnostic approaches for the calculation of the tropospheric oxidant that may be useful (Bates and Jacob, 2019; Edwards and Evans, 2017). These different approaches may be valuable and should be assessed carefully before CMIP7.

Experimental design is just as important, particularly for understanding the causes of ozone burden change. CMIP6 features a number of useful attribution experiments, mostly in AerChemMIP, to understand the role of e.g., aerosol precursors, VOC and NOx emissions changes (Thornhil et al., 2020). As the reviewer notes, it would be good to complement these with experiments designed to understand the other important aspects of ozone modelling, namely transport and chemistry. On the first point, we should like to see idealised tracer experiments performed, and the inclusion of specific idealised tracers may give a great deal of insight into the model differences driven by dynamics. CCMI employed idealised tracers (Orbe et al., 2018) which have been used to quantify

model differences in circulation, particularly inter-hemispheric transport times and the effects of convection. E90 tracers (Prather et al., 2011) can be useful for understanding the effects of stratospheric circulation and, while included in CCMI, are not available in CMIP6. For the second point, understanding the different model sensitivities is important. We note that the sensitivity of different models to emissions has been quantified through the use of multi-model perturbed parameter ensembles in understanding ozone burden (Wild et al., 2020), and that HTAP simulations use a variation in emissions to ascertain sensitivity (Fiore et al., 2009). Tagged source experiments may also be useful (Butler et al., 2018). These strategies can be used effectively to quantify the causes of diversity in model response to different perturbations - either in model inputs or in model parameters. Targeted studies of key processes, such as Hardacre et al. for dry deposition (Hardacre et al., 2015), or as in the ATom project for photochemical ozone models (Prather et al., 2017), would be valuable. We note that it may also be possible to use the forthcoming AerChemMIP experiments to look at model response to certain emissions changes in a multi-model sense (e.g., Allen et al., submitted).

We can't speculate too much about the CMIP7 paper, but we expect that the goal of CMIP7 will be to calculate the ozone radiative forcing, which will require continued improvement to the estimate of the pre-industrial ozone burden. Future assessments will need to quantify and understand the roles of biogenic VOC emissions, NOx sources, biomass burning emissions and to provide assessment of changes. We note that the increasing use of Earth-system models in CMIP6 means that models are increasingly becoming more complex with greater representation of process-level feedbacks. An example is dry deposition of ozone which is increasingly treated as an interactive deposition process that couples to land cover and vegetation. Biogenic VOC emissions are another, which are increasingly treated in an online sense. There will be a need to better quantify the source of inter-model variation in such processes, and hence careful choice of diagnostic output is required.

We would urge the community to engage further on these points, particularly in exploiting the CCMI experiments that are scheduled between CMIP deadlines. It may be advantageous to exploit the CCMI runs further to focus on the process-based understanding of composition and to quantify inter-model differences in these experiments and so allow CMIP to be retained for the lengthy centennial coupled experiments.

Well designed, well thought out experiments with targeted and novel diagnostics has the potential to gain more traction with the community. This aspect has been recently discussed extensively by Archibald et al in the TOAR budget paper discussing the role of CMIP DECK style experiments geared towards identifying the roles of changes in chemistry, emissions, deposition in the models leading to differences in modeled tropospheric ozone burden/budget.

Specific comments.

> Relationship to other CMIP6 papers There are a number of papers submitted to a number of journals based on these CMIP6 simulations. It would be useful to provide some explanation in this paper as to where it is expected to sit in relation to the other papers.

> Is there a separate paper discussion stratospheric O3? Ozone radiative forcing? OH? CH4 etc. There is some nods to some of these papers but it isn't clear how this is likely to fit in with the other papers. Could the authors provide information about the other papers currently going through review which touch on this topic (stratospheric ozone, OH etc).

Done. We have added references to available papers published or in discussion.

> Models 4 models are described in section 2. But I only see 3 models in figures etc. Section 2.1 says that the 'ozone evaluation' uses 4 models but the 'budgets' use 3. But it is unclear therefore why the GISS model doesn't appear in Figure 1,8 and 15. It is unclear which models are being included in which metrics. Could each metric please indicate whether it is calculated from the multi model mean of 3 or 4 models? As dis- cussed earlier this is a small number, especially given previous evaluations. Could the authors give a little bit of context? Are there fewer models engaging in the whole CMIP process, or is it just this tropospheric chemistry aspect aspect? Was the minimum criteria for inclusion solely providing a tropospheric ozone concentration or were there others?

This has been addressed in the revised manuscript. 5 models are now included and are included in each figure. The reviewer is correct that we used a subset, focusing on those models which include interactive tropospheric ozone and provided data.

> Model description It would be useful to have a table outlining the model configurations. Sections 2.0.1-2.0.4 give differing bits of information about the models and understanding what is the same and what is different between the models is difficult. There are only 3 or 4 models so it shouldn't be too difficult to pull the useful bits of information from the models on things like – resolution, anthropogenic emitted species, lighting emissions scheme, soil nox emissions, biogenic emissions schemes, treatment of aerosols, heterogenous chemistry in a standardized format etc.

Done. We now include a table in a supplementary section to this point.

> Model's representation of tropospheric processes It would be useful for the authors to comment on whether these models represent our best understanding of atmospheric chemistry and, if not, what could the implications of this be. These models are by their very nature fairly conservative in what processes they include and their complexity of representation. But they likely miss some significant processes such as tropospheric halogens, and a complete representation of organic chemistry, heterogenous chemistry etc. It would be useful to have some comments (probably in the discussion) of what this might mean for the conclusions drawn here.

Done. We will add a section on key uncertainties with regards to missing processes to the summary.

> Future and past emissions It would be useful to have a description of the ssp370 emissions – what are the assumptions about how the world gets to 2100? To those embedded in the IPCC process this might be obvious but to those who are not it is hard to know what this scenario is and what it assumes about the state of the world etc.

Done. We have added a section on the SSP370 pathway.

It would also be useful to have a sense of how the results from these simulations compare to the world predicted by the previous round of model assessments with the RCPs. IIt would be useful to mark the multi-model mean O3 burdens for 2100 found from the last round of CMIP model experiments on Figure 8 for example.

This is a difficult point. Firstly, the models used to generate the data for CMIP5 necessarily differ from those used here, which makes comparison difficult. We also use a different tropopause definition, as noted in the text. Secondly, the RCPs used for AR5 do not correspond well to the updated pathways used for AR6 (Figure 2 in Rao et al., 2017) , and neither RCP6 or RCP8.5 correspond well to SSP370 used here. Finally, we have analysed only a single pathway, SSP370, due to data availability at the time of writing, and the range of tropospheric ozone burdens would not necessarily be comparable to the inter-RCP range in CMIP5. We prefer to leave this to a follow-up paper focusing on the different CMIP6 pathways that will be available once more model data come online.

Similarily, how do the preindustrial anthropogenic emissions differ from those previously used in these assessments?

Done. These are given in the Hoesly paper -particularly Figure 2 - which gives a comparison with AR5 - to which we have added a reference, as well as to a recent paper that compares emissions estimates used in CMIP6 (Elguindi et al., 2020).

Acronyms There are quite a few acronyms used in the paper. These tend to alienate readers so it would be useful to see if some of them can be removed especially when they are only referred to a few times after being defined. Can the full wording also be used when the definition of the acronym is well away from its usage (on a different page etc). It took me a while to workout what BDC was on page 16.

We have removed as many acronyms as possible, and checked that they are defined on first use.

Specific comments Abstract. Line 10. It's not clear whether the large differences between the models (30%) is referring to the burden or the budget.

We have clarified this point.

Page 16. The OPE calculation is very interesting. This shows a much larger range than I would have assumed. The authors argue that that it is the differences in the background VOC mixing ratio in the model to explain this. It might not be that simple and they don't really provide any evidence to support this. Differences in the chemistry schemes may play a role here in a number of ways. Choices about the chemical rate constants, mechanistic choices and the speciation of VOCs used could all cause differences They do show that there are differences in the bVOC emissions (Figure 1) but there isn't any other evidence to support their argument about this being due to background VOC mixing ratios (they discuss NMHVOCs in one sentence and then in the next use VOC; is in the other is there a subtly in their argument about CH4 here that I'm not getting). There are also substantial differences in the mean NOx concentrations being calculated which would also influence the OPE. Without some additional evidence the explanations of the model performance appear to be somewhat of a throw away comment.

Done. We have updated this figure with LNOx data that have since come online. The modified figure shows that the intermodel range of OPE values is much smaller when done as a calculation based on

total NOx emissions, which indicates consistent chemistry/response, although the GISS model remains an outlier compared to the other models due to its much higher production term.

Tables. Table 1+2 These table are currently without units. It would be useful to include some additional information. The ozone burden would be useful as would the mean lifetime (Burden/(L+DD)). The table seems quite long. Reporting fewer times would not change the story.

Done. Burdens and lifetimes have been added.

Figures.

Figure 1. Could this be expanded to include CH4 concentrations or emissions and the anthropogenic emissions of VOCs, SO2 etc? This would help to put the rest of the paper into context.

Done

It would also be useful to know what the models are predicting for OH concentration. I realise this that might be being covered in more detail in another publication but it is hard to understand the impacts of O3 without understanding the influence of OH. If there are other papers covering other areas of the model simulations it would be useful to understand which papers will be covering which activities.

The data the reviewer requests are in Stevenson et al. for the historical period and we have added a reference to this published paper. References to other papers accepted at the time of writing have been added.

Figure 2. Extra dot before C.E.

Done

Figure 4. The markers are too small to see the colours. I'd suggest that they are just filled back squares or circles. This seems to be discussed before figure 3? Figure 5. Can you explain MMS, MMM in the caption text.

Done. Caption text shows the mean and other data.

Figure 8. Can the models be described in the caption box as well as in the text. It would make it easier to understand.

Done, as requested.

Figure 14. It would be more useful to know the deposition velocities in the model than the fluxes here. In trying to attribute change it is hard to know whether it's the differences in the O3 concentrations calculated by the model which are causing the differences or the changes in the land surfaces or assumptions about land surface which are causing these differences.

Done. We have removed this figure.

Figure 15. I'm not sure that the units are appropriate here? The units say Tg(O3)/yr but shouldn't the model resolution be taken into account here? The text says that the models are at their native grid so a model at 2x2.5 resolution compared to one at 1x1 would have 5 times as much ozone production in each gridbox which would make it look much redder even if the integrated ozone production was the same? Similarily, will this plot also tend to over emphasise the poles in the budget as it given them equal weight on the plot as the tropics?

Thanks for the comments. The integrated net ozone production from each model is now on a common grid.

Figure 16. Can this be converted into two plots? One of ozone production and one loss? It is a bit busy at the moment.

Done

Figure 17. Can the scale on the plots be changed to reduce the emphasis on the stratosphere and increase the emaphasis on the troposphere? 300 ppbv of O3 is pretty high? Page 16. There are a lot of acronymns here which don't I think make the document transparaent. BDC is defined much earlier in the document making understanding difficult.

This plot means to demonstrate ozone changes in both the lower stratosphere and the troposphere and we wish to retain this emphasis.

Data availability. Can the ESGF be spelt out in more details and a website given?

Done.

Bibliography

Bates, K. H. and Jacob, D. J.: An Expanded Definition of the Odd Oxygen Family for Tropospheric Ozone Budgets: Implications for Ozone Lifetime and Stratospheric Influence, Geophys. Res. Lett., 47(4), doi:10.1029/2019GL084486, 2020.

Butler, T., Lupascu, A., Coates, J. and Zhu, S.: TOAST 1.0: Tropospheric Ozone Attribution of Sources with Tagging for CESM 1.2.2, Geosci. Model Dev., 11(7), 2825–2840, doi:https://doi.org/10.5194/gmd-11-2825-2018, 2018.

Edwards, P. M. and Evans, M. J.: A new diagnostic for tropospheric ozone production, Atmospheric Chem. Phys., 17(22), 13669–13680, doi:https://doi.org/10.5194/acp-17-13669-2017, 2017.

Elguindi, N., Granier, C., Stavrakou, T., Darras, S., Bauwens, M., Cao, H., Chen, C., Gon, H. A. C. D. van der, Dubovik, O., Fu, T. M., Henze, D. K., Jiang, Z., Keita, S., Kuenen, J. J. P., Kurokawa, J., Liousse, C., Miyazaki, K., Müller, J.-F., Qu, Z., Solmon, F. and Zheng, B.: Intercomparison of Magnitudes and Trends in Anthropogenic Surface Emissions From Bottom-Up Inventories, Top-Down Estimates, and Emission Scenarios, Earths Future, 8(8), e2020EF001520, doi:https://doi.org/10.1029/2020EF001520, 2020.

Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T.,

Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S. and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, J. Geophys. Res., 114(D4), doi:10.1029/2008JD010816, 2009.

Hardacre, C., Wild, O. and Emberson, L.: An evaluation of ozone dry deposition in global scale chemistry climate models, Atmospheric Chem. Phys., 15(11), 6419–6436, doi:https://doi.org/10.5194/acp-15-6419-2015, 2015.

Orbe, C., Yang, H., Waugh, D. W., Zeng, G., Morgenstern, O., Kinnison, D. E., Lamarque, J.-F., Tilmes, S., Plummer, D. A., Scinocca, J. F., Josse, B., Marecal, V., Jöckel, P., Oman, L. D., Strahan, S. E., Deushi, M., Tanaka, T. Y., Yoshida, K., Akiyoshi, H., Yamashita, Y., Stenke, A., Revell, L., Sukhodolov, T., Rozanov, E., Pitari, G., Visioni, D., Stone, K. A., Schofield, R. and Banerjee, A.: Large-scale tropospheric transport in the Chemistry–Climate Model Initiative (CCMI) simulations, Atmospheric Chem. Phys., 18(10), 7217–7235, doi:https://doi.org/10.5194/acp-18-7217-2018, 2018.

Prather, M. J., Zhu, X., Tang, Q., Hsu, J. and Neu, J. L.: An atmospheric chemist in search of the tropopause, J. Geophys. Res. Atmospheres, 116(D4), doi:https://doi.org/10.1029/2010JD014939, 2011.

Prather, M. J., Zhu, X., Flynn, C. M., Strode, S. A., Rodriguez, J. M., Steenrod, S. D., Liu, J., Lamarque, J.-F.,
Fiore, A. M., Horowitz, L. W., Mao, J., Murray, L. T., Shindell, D. T. and Wofsy, S. C.: Global atmospheric chemistry – which air matters, Atmospheric Chem. Phys., 17(14), 9081–9102, doi:10.5194/acp-17-9081-2017, 2017.

Rao, S., Klimont, Z., Smith, S. J., Van Dingenen, R., Dentener, F., Bouwman, L., Riahi, K., Amann, M., Bodirsky, B. L., van Vuuren, D. P., Aleluia Reis, L., Calvin, K., Drouet, L., Fricko, O., Fujimori, S., Gernaat, D., Havlik, P., Harmsen, M., Hasegawa, T., Heyes, C., Hilaire, J., Luderer, G., Masui, T., Stehfest, E., Strefler, J., van der Sluis, S. and Tavoni, M.: Future air pollution in the Shared Socio-economic Pathways, Glob. Environ. Change, 42, 346–358, doi:10.1016/j.gloenvcha.2016.05.012, 2017.

Wild, O., Voulgarakis, A., O'Connor, F., Lamarque, J.-F., Ryan, E. M. and Lee, L.: Global sensitivity analysis of chemistry–climate model budgets of tropospheric ozone and OH: exploring model diversity, Atmospheric Chem. Phys., 20(7), 4047–4058, doi:https://doi.org/10.5194/acp-20-4047-2020, 2020.