We thank Owen Cooper, Chris Smith and an anonymous reviewer for their useful comments that have helped improve the manuscript. To address these comments we have made substantial additions to the text, particularly to add discussion, figures and tables covering the radiative impacts of the ODS changes. The effective radiative forcing from the stratospheric ozone recovery is significantly larger than expected from previous studies and from calculations based on the traditional stratospheric-temperature adjusted RF. To further understand the physical processes behind the meteorological adjustments, we have added further discussion, figures and tables covering changes in humidity and surface albedo. We hope that these additions to the paper will make the importance of the scientific outcomes clearer to readers.

Owen Cooper

General comments:

The paper begins with an excellent review of ozone radiative forcing, which will be a very helpful reference for the scientific community (especially Figure 1).

Thank you.

Abstract.

Here the main findings on ozone RF are reported as: "We find robust changes in ozone due to future changes in ozone precursors and ODSs. These lead to a positive radiative forcing of 0.27±0.09 Wm-2 ERF, 0.24 ± 0.021 W m-2 offline SARF, 0.29 ± 0.10 Wm-2 online IRF." But to be clear, these numbers are the changes in RF between 2015 and 2050, correct? Should these results be reported as "delta ERF", for example? Would ERF for 2050 be estimated as 0.47 +/- 0.23 W m-2 (from IPCC AR6) plus 0.27±0.09 W m-2 (this study), to equal 0.74 +/- 0.31 W m-2 ?

We have now clarified that the forcings are 2015 to 2050

In this study we do not report these as changes since 1750, but use 2015 as the baseline. Care would be needed in combining our results with those from IPCC AR6, so we do not recommend this in this paper.

Line 74

It would helpful to provide a little more background on the SSP3-7.0 scenario, which is generally described as being driven by "regional rivalry". While NOx and CH4 emissions go up in this scenario through 2050 (see figure 6.18 of IPCC AR6 WG-I), there are clear regional differences, especially for NOx (see figure 6.19 of IPCC AR6 WG-I). For example, NOx emissions decrease quite strongly by 2050 in North America, Europe, Russia, Central Asia, and Pacific OECD.

We have added more detail here:

"through increases in methane, NOx and other ozone precursors - although note NOx emissions decrease in OECD countries (Szopa et al., 2021)."

The information in Figure 2 is a very handy reference, but I think it would work better as a table so you can list important items for each estimate such as: TOA vs tropopause, ERF vs IRF, starting and ending year. There seems to be an error in the mid-point of ozone RF for IPCC AR 2, which is plotted as 0.5 W m-2. However, Page 20 of IPCC AR2 states: "Changes in tropospheric ozone have potentially important consequences for radiative forcing. The calculated global average radiative forcing due to the increased concentration since pre-industrial times is 0.4 (±0.2) Wm-2."

Thank you for your suggestion, we have added the information in figure 2 as a table in the supplement.

The value for SAR has been corrected – thank you for spotting this.

In terms of comparing the model output to observed ozone changes (1995 through 2021) a new paper by the HEGIFTOM Working Group will soon be available for open review on EGUsphere: Van Malderen, R., Z. Zang, K.-L. Chang, R. Bjorklund, O. R. Cooper, J. Liu, C. Vigouroux, E. Maillard Barras, I. Petropavlovskikh, T. Leblanc, V. Thouret, P. Wolff, P. Effertz, A. Gaudel, H.G.J. Smit, A. M. Thompson, R. M. Stauffer, D. E. Kollonige, D. Tarasick (2024), Global Ground-based Tropospheric Ozone Measurements: Regional tropospheric ozone column trends from the HEGIFTOM homogenized ground-based profile ozone datasets, submitted to ACP (TOAR-II Community Special Issue) The analysis provides long-term ozone trends for many regions around the world, based on merged datasets (ozonesondes, lidar, IAGOS, FTIR). In my opinion, this is the best observation-based summary of global tropospheric ozone trends. The results are similar to the findings of IPCC AR6 (Gulev et al., 2021), except there is now evidence for a decrease of tropospheric column ozone in the Arctic, and there is a clear drop in ozone in 2020 and 2021 that coincides with the COVID-19 economic downturn.

Thank you for drawing our attention to this dataset which will be useful for the community. As we do not compare against observed changes in our study we will not cite this paper here.

Line 333 Please provide more information on the time-slice definition. Is this a 10-year time-slice centered on 2015 (i.e. averaged over 2010-2019)?

We have now clarified that this is a continuous year 2015

Chris Smith

This paper provides results from a multi-model experiment of ozone radiative forcing (using three metrics: instantaneous, stratospheric adjusted and effective) for 2050 relative to 2015 under the SSP3-7.0 scenario. This addresses an important gap in the literature that was not available during the IPCC Sixth Assessment on the likely future evolution of ozone radiative forcing, and will be a valuable resource to the community that will be referenced for years to come.

Thank you for your comment.

The title and abstract don't make it clear that only a single scenario is considered: we trust that the future will follow SSP3-7.0. Either call out SSP3-7.0 specifically or frame in more general terms: "Climate forcing due to future ozone changes in a high emissions scenario..." for example. Why was this scenario selected – presumably because it had the biggest signal? In the abstract, please state the 2050 relative to 2015 timeframe.

Since the focus of the paper is the metrics and methods rather than the specifics of the scenario we prefer not to lengthen the title. The key point of the paper is not the final radiative forcing number, but a comparison of the different methods to calculate radiative forcing.

We now explain in section 1 why this scenario is chosen. "This scenario is chosen as it has the largest increase in tropospheric ozone (Keeble et al., 2021; Turnock et al., 2020) ..."

We now clarify the timeframe: "... forcing from 2015 to 2050 of ..."

Line 56: I think that tropospheric ERF is AR6 for 1750-2019 was +0.41 W m-2 and not +0.45 W m-2. Not that we mentioned it in the AR6 text, but I did do the trop/strat split. Does the Skeie et al. number include climate effects on ozone forcing? More generally, I am supposing that the experimental design here, using 2015 SSTs, does not include the climate effects on ozone either. SSP3-7.0 is a degree or more warmer in 2050 than in 2015 in many models so the effects are probably not insignificant.

AR6: " The contributions to total SARF in CMIP6 (Skeie et al., 2020) are 0.39 ± 0.07 and 0.02 ± 0.07 W m–2 for troposphere and stratosphere respectively ... The dataset is extended over the entire historical period following Skeie et al. (2020), with a SARF for 1750–1850 of 0.03 W m–2 and for 2010–2018 of 0.03 W m–2" – so this gives 0.45 W m–2 for the troposphere.

The Skeie et al. number does include climate effects. The experimental design here does not. This could be important for the comparison with CMIP6 ozone in section 3.2. We have added the caveat: "...although the CMIP6 ensemble include changes in climate which are excluded in the TOAR-RF simulations.

The section starting 1.1 on radiative forcing should possibly be promoted to a level-1 section (section 2). Possibly also section 1.2, but could sit under the radiative forcing header. I also thought that section 1.1 was a bit textbook and may not be required for this paper, but the coordinator of TOAR-II likes it, so it's probably a matter of taste and my familiarity with the topic.

Thank you for the suggestion. We have renumbered as suggested.

Line 270: "present day" in GISS means 2015? In all cases where "present day" is mentioned, please be specific on the year(s) (e.g. line 337).

"present day" has been changed to "2015" throughout

Line 337: where do these climatologies come from? Are they model-specific or centrally provided?

DMS concentrations climatologies are model-specific.

Lines 348-349: “the models’ respective radiation and cloud schemes…”: will changes in aerosols (that are not fixed) affect clouds, which will affect the ERF?

The aerosols only affect the chemistry schemes. Connections between aerosols and radiation and aerosols and cloud microphysics have been turned off. This has been clarified:

“… continue to see year-2015 atmospheric concentrations of greenhouse gases and cloud condensation nuclei”

Line 473: 298.3 ± 8.3 DU – any observations to compare this to? Figure 6 has data from NIWA, perhaps this could be compared?

This comparison has been added to section 3.2.

“the TOAR-RF ensemble mean is systematically biased high relative to observations (283.5±1.1 DU) by 11 DU”

Lines 545-548: I’d also say the historical trend of CMIP6 models compared to the obs is quite good, even if biased high.

This comparison has been added to the text:

“The historical trend in CMIP6 ozone agrees well with the observations”

Figure 8: the 150 ppb ozone tropopause forcings agree quite well between models. Is this expected and/or worth a comment?

A comment to this effect has been added:

“There is close agreement between models in the tropospheric forcing, but more model variability when including stratospheric changes.”

Very minor, editorial things

Line 51: reference after full stop

Fixed

Line 79: write out equation on a new line

Fixed

Line 116: comma after full stop

Fixed

Line 310: Walters citation as author (year).

Fixed

Line 545: Bodeker Scientific doesn’t show up in the references.

Fixed

Line 576: superscript -2

Fixed

Anonymous Reviewer 2

General Comments

The manuscript compares across different metrics and methods to estimate ozone radiative forcing by first presenting a synthesis of prior work and then calculating the radiative forcing from 2015 to 2050 using the current generation of Earth system models. The manuscript documents inconsistencies in approaches in prior work as well as unique configurations in some models that complicate a straightforward inter-model comparison. This detailed documentation is invaluable to the modeling community and is a critical piece for interpreting some simulations and possibly the next round of multi-model studies. It does, however, lead to a lengthy manuscript in which major conclusions may be missed. At the same time, providing some additional context to the abstract and conclusions may help a reader understand the importance of the work. Two general suggestions:

1. Articulate more clearly the key messages in the abstract/conclusions with some short synthesis statements that provide slightly broader context. For example, the authors may wish to consider the following questions: How much confidence is there in estimates of ozone forcing and the stratosphere versus troposphere contributions as reported here and in prior work? Given the emphasis on this prior work in the introduction, I expected the authors to conclude by comparing their results to that earlier work. Do the conclusions drawn here have implications for the interpretation of historical radiative forcing estimates for ozone? How important is ozone relative to other greenhouse gases? How do the emission trends driving the ozone forcing in this scenario compare with other future scenarios?

We have revised the conclusions to include the context that this calculation is larger than that assumed in the IPCC AR6 and would make ozone the second most important contributor to radiative forcing over the 2015-2050 period in this scenario. We also draw attention to the calculations that the ERF metric indicates much larger climate effects than SARF for stratospheric ozone recovery.

2. Justify the use of SSP3-7.0. For example, is this the scenario that regional emissions have followed most closely in the last decade or simply the one that all the models ran? What are NOx, NMVOC, and CO emission trends as well as ODS and methane in this scenario? Consider showing a plot with zonal mean 2050-2015 emission changes (and ODS) to highlight regional differences.

We have now explained that the choice of SSP3-7.0 is a pragmatic choice designed to give the largest ozone signal. The key point of the paper is not the final radiative forcing number, but a comparison of the different methods and metrics to calculate radiative forcing. For this reason we do not discuss the trends in the different precursors or the regional differences.

Specific Comments

1. The abstract should include the time frame and scenario over which the forcings are calculated.

We have now clarified that the forcing is over the period 2015 to 2050 and for the SSP3-7.0 scenario.

2. Please clarify what appears to be conflicting findings: Checa-Garcia et al. (2018) report that tropospheric ozone cools the lower stratosphere (line 117) but Figure 9g shows warming in the lower stratosphere in response to increasing tropospheric ozone (fODS).

Thank you for pointing this out. We now discuss in section 4.7 and 5.3 that the modelled temperature changes can differ from the FDH calculations.

3. Pages 2-3: Clarify focus here is not only on tropospheric ozone.

We have clarified that changes in tropospheric and stratospheric ozone are quantified.

4. Line 42: Include example of a non-cloud adjustment the first time this term is mentioned.

We now give water vapour and albedo changes as examples.

5. Figure 1: Clarify if this is for tropospheric ozone only?

We have clarified that this is for ozone changes up to 0.1 hPa

6. Section 2.1: Different information is provided for different models (number of chemical species, reactions given for some but not others). Consider a summary Table with consistent information provided for all models. This table could be main text and detailed text describing unique model aspects could move to supplement to shorten main text.

Thank you for the suggestion. We have now included such a table and moved the more detailed model descriptions into the Supplementary Material.

7. Lines 341-342: Was stabilization achieved in all the runs?

We clarify that the spin ups were long enough for the models to stabilise.

8. Section 2.4: mention here that IRF is also calculated (line 797 points to IRF kernels but only SARF kernels are introduced)

We now mention that an IRF kernel is used too.

9. Figure 6: Is the tropopause used for the observations consistent with the definition in the models? Clarify that the TOAR value is from this study.

The comparison is for total column ozone.

We have added clarification that the TOAR value is from this study.

10. Sections 3.3/3.4/3.5 and 4.3/4.4: Add a final summary sentence or short paragraph to convey the key message(s) emerging from the findings reported in each section?

We have added summary sentences to these sections.

11. Lines 589-591: Given the different cloud fields, is this finding meaningful?

This has been clarified to explain that the meaningful finding is that the different cloud fields mean that the fluxes are noisy

12. Line 653. GEOS-Chem reads in the cloud fields, so rather than a zero response it's the same clouds being used as input.

Agreed, this has been removed from the text and the cell in the table has been changed to "N/A".

13. Lines 720-722. So tropospheric ozone precursors and ODS (&N2O?) contribute equally to ERF in 2050 in this scenario? Figure 12 might better illustrate this by using hatching as in Figure 8. Consider adding this point to the abstract.

Thank you for your suggestion on the bar chart. We have revised this to better illustrate the comparison.

We agree on the importance of the ODS contribution, and have added this point to the abstract.

Technical Corrections

1. On Figure 5, is the topography used to determine where to plot as a function of pressure the same in the different panels? Around 40S and 90N there are some differences.

Thank you for pointing this out. There was an error in how topography was used, which has now been corrected.

2. Bodeker Scientific, 2024 is not in the bibliography

Fixed