We thank this reviewer for their positive and constructive comments and their helpful advice for improving the manuscript. Our responses are given in the blue font. Where necessary, we refer to the revised version of the manuscript.

Review of egusphere-2024-111, entitled "Investigating the role of stratospheric ozone as a driver of inter-model spread in CO_2 effective radiative forcing" by Byrom et al.

General comments

The submitted manuscript investigates stratospheric ozone as a cause of inter-model differences in stratospheric temperature, and hence its role as a contributing factor to inter-model spread in CO_2 radiative forcing. The work aims to explore whether the stratospheric temperature dependence of instantaneous radiative forcing (IRF) extends to effective radiative forcing (ERF). Using the Norwegian Earth System Model 2, the authors explore the impact of stratospheric ozone perturbation and further stratospheric temperature on the magnitude of 4xCO2 ERF and its components (IRF and rapid adjustments) via a series of well-designed fixed-SST simulations. The authors found that the systematic stratospheric ozone perturbations barely influence the spread of 4xCO2 ERF, although the effects of systematic stratospheric ozone perturbations clearly show in both the IRF and stratospheric adjusted radiative forcing (SARF) via their base-state dependence. Meanwhile, the authors also found considerable uncertainty in the IRF calculations with different methods, stemming from differences in both radiative transfer codes and the vertical coordinates (e.g., model top pressure and vertical resolutions). These results may help to better understand the large spread in the IRF and further ERF, hopefully reducing uncertainty in estimates of climate sensitivity. Overall, I found the manuscript well-organized, well-illustrated, and a good addition to the understanding of state-dependent radiative forcing. I would recommend Atmospheric Chemistry and Physics consider this article for publication after minor revisions.

Major concern:

The authors found that the systematic stratospheric ozone perturbations barely influence the spread of 4xCO2 ERF, although the effects of systematic stratospheric ozone perturbations clearly show in both the IRF and SARF via their base-state dependence. Based on Figure 3, it is clear that the cloud adjustment offsets the stratospheric temperature dependence of the IRF and SARF. I would suspect that most cloud adjustments come from cloud changes at higher altitudes. It would be good if the author could decompose cloud adjustment into contributions from high, mixed, and low clouds, probably following Soden and Vecchi (2011). It would be even better if the authors had ISCCP simulator results available (it is totally ok if not). If the cloud adjustment comes from high altitudes, as suspected, the author could probably get the stratospheric temperature-dependent ERF by simulations only perturbing ozone within the upper stratosphere in the future.

Soden, B. J., and G. A. Vecchi (2011), The vertical distribution of cloud feedback in coupled ocean-atmosphere models, Geophys. Res. Lett., 38, L12704, doi:10.1029/2011GL047632.

This is a good point, and we agree with the reviewer that more attention should be given to the apparent impact of the ozone increases/decreases on the cloud adjustment. Unfortunately, ISCCP simulator results are not available for these NorESM2 simulations, so we couldn’t diagnose and decompose the cloud adjustment via that method.

However, following this comment (and similar comments made by the other reviewers) we decided to change our method for calculating the cloud adjustment to the adjusted cloud radiative effect method of Soden et al. (2008; https://journals.ametsoc.org/view/journals/clim/21/14/2007jcli2110.1.xml, also see our response to Reviewer 1). As shown in the updated version of Figure 3 (see revised manuscript), the magnitude of the cloud adjustment in each experiment is now much more similar, demonstrating that this adjustment doesn't offset the impact of ozone increase/decreases on IRF and SARF as implied previously.

Minor comments:

Lines 15-16: Are the host-model radiative transfer calculations referring to online or offline double-call calculations with the same radiative transfer codes? If yes, the authors could probably encourage model centers to provide online double-call for their simulations.

We perform offline host-model radiative transfer IRF calculations and demonstrate the need for others to also use the host-model radiative transfer code to calculate IRF, but this could be achieved either online or offline. We have now added more text to the conclusion (lines 303-305) to make it clear that these calculations could be performed offline or online and have included a sentence to encourage modelling centres to provide IRF results:

"*Here we demonstrate that accurate calculation of IRF requires the use of host-model radiative transfer calculations, which can be computed either offline or online using double-call simulations. As noted elsewhere (e.g., Chung and Soden 2015b), we encourage modelling centres to make this diagnostic available with their simulations*."

Lines 16 and 81-82: I wondered why the authors chose to use the 50% increase and decrease in stratospheric O_3 concentration instead of doubling and halving stratospheric O_3 concentration.

This choice to go with a 50% increase and decrease was a relatively arbitrary choice, however we went with these perturbations in order to generate a somewhat similar effect on temperature either way (i.e., with an increase and with a decrease). A doubling and halving of O_3 concentration would also be interesting to look at.

Lines 17-19, 240-241 & 265-268: Is the effect of the spectral overlap of CO_2 and O_3 comparable to the effect of stratospheric temperature dependence? It would probably be interesting to have a simple test with LBL codes in the future.

We previously performed LBL tests (using the Oslo LBL code) to compare the relative importance of these effects. We found that a 50% reduction in O_3 leads to a +0.07 W m^{-2} increase in 4xCO_2 IRF (from 4.2 W m^{-2}), whilst a reduction in the temperature of -2K across the whole stratosphere leads to a 0.16 Wm^{-2} increase. Based on these results the effect of stratospheric temperature dependence is stronger than the effect of spectral overlap. Furthermore, as shown in Figure 2, decreasing stratospheric ozone by 50% results in widespread cooling of the stratosphere with ΔT values largely more negative than -3 K and peaking at -9 K. We have now added these results as a footnote on page 12 of the revised manuscript:

"Tests performed by the GENLN2 line-by-line (Myhre et al., 2006) show that a decrease in temperature of 2 K across the whole stratosphere leads to a 0.16 W m^{-2} increase in 4xCO_2 IRF, whilst a 50% reduction in stratospheric O_3 leads to a 0.07 W m^{-2} increase in 4xCO_2 IRF."

Lines 92-95: Can the authors provide more details for the PORT offline calculations? Are the calculations described here just offline double-call calculations? So, just two offline calculations using 1x and 4x CO2 concentrations with identical base-state from control simulation. Can the authors help to explain what "the simulations are run for 16 months with the last 12 months used to …" means?

Yes, this is correct – the PORT IRF calculation consists of two offline calculations using pre-industrial CO_2 concentrations and $4xCO_2$ concentrations that are each run with an identical atmospheric base state from the control simulation that is taken from the first year of output from the NorESM2 simulation.

The part that says, "the simulations are run for 16 months with the last 12 months used to …" was based on the recommendation in Conley et al. (2013): "*Running the model for 4 months prior to the 12-month period of the average is recommended, so that the stratospheric temperatures are in steady state before the time for a full-year (1 January–31 December) computation begins.*"

We originally also used PORT to perform stratospheric temperature adjusted RF (fixed dynamical heating) calculations and followed the above advice to run the model for 16 months (from 1st January year 1 – 30th April year 2) and use the last 12 months of the output to calculate the stratospheric temperature adjusted calculation (i.e. from 1st May – 30th April). We also derived IRF using 16 month long simulations and took the last 12 months of the 16 month output to calculate IRF, however we realise that this is not necessary given that an instantaneous forcing should ideally be calculated immediately following the perturbation. We have now derived all IRF results using the first 12 months of the simulation (i.e., 1st January – 31st December) and have amended the methodology. Note however, that this did not impact the IRF result for the 'Standard' experiment and only marginally changed the net IRF for the 'O3x1.5' and 'O3x0.5' experiments (by 0.01 W m^{-2}).

Line 99: It would probably be better to use "temperature, water vapor, or surface albedo" instead of "stratospheric temperature, surface albedo or clouds", since there is no cloud kernel in the radiative kernels of Soden et al. (2008).

This has now been changed to read as "temperature, surface albedo or water vapour" (line 113).

Lines 141-144 & 157-159: It is great to see the conclusion that the magnitude of IRF varies notably between both experiments, demonstrating a dependence on the diagnostic method of choice, although this may not be very new. Even with the identical radiative kernel method, the IRF obtained from 17- and 19-levels are noticeably different. In particular, stratosphere adjustment has a strong dependence on the model top and probably vertical resolution. This is because the radiative flux perturbation due to the same temperature perturbation at higher altitudes (e.g., 1 hPa) within the upper stratosphere is larger than that of lower altitudes (e.g., 10 hPa) within the upper stratosphere. Meanwhile, the temperature cooling at higher altitudes within the upper stratosphere is usually also stronger than at lower altitudes within the upper stratosphere. Therefore, stratospheric adjustment (IRF) from 17-level kernels is smaller (larger) than that of 19-level kernels.

This is a good point to highlight that the magnitude of the IRF consequently varies between the use of the 17 and 19 level kernels.

Lines 144-145, 171-175 & 187-191: Since the accuracy of IRF calculation from the radiative kernel method depends on the differences of radiative transfer codes and the vertical coordinate resolution (ignoring the base-state dependence), here it would be great to isolate the contribution from the difference of radiative transfer codes by interpolating (and extrapolating if necessary for surface and probably for CAM5 kernels) the three radiative kernels onto the output resolution of NorESM2-MM and redoing kernel decomposition calculations. I believe there are native model grid versions available for the three radiative kernels. With the obtained difference, it may be easy to determine whether the error is acceptable. For NorESM2-MM with a low model top, the authors could even simplify the isolation process by using 17-level kernels for the kernel decomposition calculations. Actually, the sentence in lines 173-175 suggests the difference for kernels from radiative transfer codes is small.

We agree that this test would be useful to isolate how much different radiative transfer parameterisations impact the kernel results. However, we decided not to add these extra calculations to the manuscript to keep this section of the paper relatively concise with respect to our focus on the ozone experiments. As suggested, the reader can infer from the text that the difference between the magnitude of the stratospheric temperature adjustment is small.

Line 146: I was just wondering if the authors have any idea why there is a close agreement between cloud adjustment from the two different methods.

As noted above, we have changed our method for calculating the cloud adjustment. However, this still results in a very close agreement with the Smith et al. (2020) value but a more in-depth assessment of the two different methods would be needed to say more about why they agree so closely.

Lines 146-148: If I understood Smith et al. (2020a) correctly, the cloud adjustment in Smith et al. (2020a) was obtained by using APRP for SW and PRP for LW, and there is no liquid water path adjustment for the CO_2 cloud adjustment calculation. It would be great to double-check it.

This is correct, apologies for the misunderstanding. The text has now been updated (see lines 171-173):

"In Smith et al. (2020a) shortwave and longwave A_c *are estimated separately using the APRP approach and offline monthly-mean partial radiative perturbation calculations, respectively."*

Lines 205-206 & 221-226: These sentences suggest a ~8K temperature difference at 10 hPa, and correspondingly, there is a 0.8 W m^{-2} spread in the IRF. The resulting IRF sensitivity to temperature matches very well with the around -0.1 W m^{-2} K^{-1} (slopes) shown in Fig 1C and Fig S2 in He et al. (2023). As the online IRF difference (4 W m^{-2}) reported by He et al. (2023) includes both contributions from radiative transfer code difference and base-state difference, it may be better to compare the 0.8 W m^{-2} spread here with the ~2 W m^{-2} spread in offline IRF calculation with identical radiative transfer code (e.g., Fig 1B in He et al. (2023)).

This is a good point, thanks for highlighting this more suitable comparison. We have now amended the text (lines 261-269) to more usefully compare our results against those from He et al. (2023) Figure 1b and Figure 1c:

"The IRF across all three experiments ranges from 4.9 – 5.7 W m^{-2} *resulting in a spread of 0.8 W* m^{-2}*. This is smaller than the spread of 2 W* m^{-2} *(ranging from around 5 – 7 W* m^{-2}*) reported by He et al. (2023) for offline 'double-call' experiments of 4x*CO_2 *IRF calculated with a single radiative transfer code and base-states from the Atmospheric Model Intercomparison Project (AMIP) for 12 CMIP5/6 models (hence their spread is due only to differences in base-state). However, we find closer agreement between our IRF sensitivity to 10 hPa base-state temperature and He et al. (2023), who found a near -0.1 W* m^{-2} K^{-1} *relation between the spread in offline double-call experiments and air temperature at this level (Figure 1c in He et al., 2023). From Figure 2 it can be inferred that this matches very well with the ~7 K difference in base-state 10 hPa temperature between 'Strat*O_3*x1.5' and 'Strat*O_3*x0.5' and the corresponding 0.8 W* m^{-2} *spread in IRF."*

Meanwhile, I hope the authors can discuss cloud adjustment more. Apparently, we can see the stratospheric temperature dependence in both IRF and SARF. The offsetting effects of cloud adjustment are the reason why the stratospheric temperature dependence does not extend to the ERF. It feels like the cloud adjustment probably mainly occurs for high clouds, which could be closely related to the response of tropopause to ozone perturbation. It would be good if the author could decompose cloud adjustment into contributions from high, mixed, and low clouds, probably following Soden and Vecchi (2011). It would be even better if the authors had ISCCP simulator results archived. If the guess is correct, the authors could probably avoid the high cloud response (or tropopause response) by limiting the ozone perturbation within the upper stratosphere. In that case, the authors could probably get the stratospheric temperature-dependent ERF.

See earlier response (also below).

Lines 254-255: It is because of the offsetting effects of cloud adjustment.

As shown in the updated version of Figure 3, the magnitude of the cloud adjustment in each experiment is now much more comparable and so this adjustment does not offset the IRF as previously implied. The cloud adjustment in the 'Strat O_3x0.5' case now decreases by 0.14 W m^{-2} (from 1.68 W m^{-2} in the 'Standard' case), and whilst this does appear to offset the increased IRF (and the very slightly increased stratospheric adjustment), the upper bound of standard error falls within the lower bound of the 'Standard' and 'O_3x1.5' case, making it a lot more uncertain to state the offsetting role of clouds.

Lines 255-256: It could be expected, considering the similarities between CESM2 and NorESM2.

We agree with this comment and a similar point that was raised by Reviewer 1. We have now deleted this comparison from the text and supplementary, seeing as the comparison was somewhat limited and further calculations would now need to be performed (for the CESM2 cloud adjustment).

Lines 263-265: I wondered why the authors expect a large spread in the magnitude of stratospheric temperature adjustment. It looks like Fig S6 in He et al. (2023) shows almost no difference in the stratospheric adjustments obtained from piClim-4xCO2/piClim-control and amip-4xCO2/amip simulations.

This is a useful result from He et al. (2023), and one we had not seen prior to running these experiments. We now refer to this result in the text (lines 257-260):

"This corroborates experiments from He et al., (2023; Figure S6) that compare the size of stratospheric temperature adjustment after a quadrupling of CO_2 from two different base-states; the multi-model ensemble-mean difference in their adjustment is just -0.03 W m^{-2} (although the model range is considerably larger at around 0.5 W m^{-2})."

Due to the strong influence of ozone on stratospheric temperature we anticipated to see more of an effect of our idealised experiments on the ERF via both the impact on IRF and through the impact stratospheric adjustment, given that stratospheric cooling is the dominant adjustment for CO_2.