Responses to Reviewers

We would like to thank the reviewers for their thoughtful views and valuable comments. Below is our response to each of the comment. The point-to-point responses are below with the reviewers' comments in BLACK, our responses in BLUE, and change in the manuscript in GREEN.

Response to RC 1

General comments:

1. The list of possible CMIP6 pairings of chemistry and no-chemistry models is incomplete: EC-Earth3 / EC-Earth3-AerChem could be added. My anticipation is that it would be worth adding this pair to the analysis.

Response: Thanks for the suggestion.
As described in van Nojie et al. (2021) (https://gmd.copernicus.org/articles/14/5637/2021/), the ozone field in EC-Earth3-AerChem is constrained using the CMIP6 forcing dataset from Checa-Garcia et al. (2018) (https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GL076770). Specifically, the mixing ratios of ozone in the stratosphere are nudged towards zonal mean fields calculated from the three-dimensional input data sets provided by CMIP6 (Checa-Garcia et al., 2018). Therefore, by comparing the differences in the 4xCO2 response between this pair of models, we cannot infer the effects of the stratospheric ozone response. Thus, we decided not to add this pair to our analysis.

2. Furthermore, the authors state that there are differences other than the treatment of chemistry between these pairs. That is true for half the pairs but not the other. Perhaps something more profound can be said about how these other differences (resolution of middle / upper atmosphere, height of the model top, and tuning of the non-orographic gravity wave drag scheme, that characterize the CESM2 and GFDL pairs) affect model behaviour. To my understanding there are no substantial differences in anything other than chemistry between the HadGEM3/UKESM1, SOCOL4/MPIESM, and GISS pairs.

Response: Thanks for pointing it out.
Indeed, for the other three pairs (CESM2/CESM2-WACCM, CESM2-FV2/CESM2-WACCM-FV2, GFDL-CM4/GFDL-ESM4), they have other differences other than chemistry such as model top height. However, they share similar patterns in terms of the comparison between chem and no-chem with those pairs without major differences. This indicates that different chemistry scheme contributes the most to the chem/no-chem difference and other model differences play a minor role.

3. It is clear to me that most of the large role of climate-ozone interactions is due to the fact that in no-chemistry models the prescribed ozone field is not changing with the changing state of the atmosphere in the experiments considered here, unlike e.g. in "historical" simulations where ozone is amongst the external-forcing fields varying with time. Maybe this can be discussed, and whether the results of this study could motivate changes to the experiment definitions of 4xCO2 and 1pctCO2, where for no-chemistry models ozone could be made to change consistently with the evolving CO2 forcing, much like in "historical" simulations in future iterations of CMIP.

Response: Thanks for the suggestion.
Indeed, in CMIP 7, for abrupt-4xCO2 and 1pctCO2 experiments, it would worth trying to change the ozone forcing to the ozone field simulated in chem models for the corresponding experiment.

We added the following sentence to the conclusion section (P27 lines 428-429):
"Therefore, it might worth using the ozone field simulated in chem models as the forcing for no-chem models in future model intercomparison projects such as CMIP7."

Minor comments:

1. Table 2: As noted, the EC-Earth3 /EC-Earth3-AerChem pair can be added here.

Response: We decided not to add this pair due to the reason listed in the response to general comment #1.

2. Figure 1: Similar patterns of change were found by Morgenstern et al., ACP, 2018 (their figure 10), using CCMI1 models. They also documented similar inter-model differences to those seen here. However the mechanism discussed in the text (NOx production changes under climate change) may not have been represented in the older CCMI models, hence the pronounced increases in tropical-tropospheric ozone were not simulated. Perhaps this is worth a mention.

Response: Thanks for the suggestion.
Compared with Morgenstern et al., ACP, 2018, we think the tropospheric ozone increase doesn't seem more prominent in our analysis, but we added the follow sentence to the manuscript to discuss the potential role of NOx (P9 lines 173-174):

"A similar pattern was simulated in some of the CCMI1 models (Morgenstern et al., 2018), even though not all those models fully represent NOx production changes under climate change."

3. Figure 2: I find this figure hard to parse. A suggestion might be to calculate dO3/dT as a function of latitude and pressure for the various models and display that. Where these two

quantities do not highly correlate, this could be made NaN. Might that be a more intuitive way of displaying this information?

Response: Thanks for pointing it out.
dO3/dT as a function of latitude and pressure shows a similar pattern as depicted in the current version of Figure 1. However, we prefer to keep the current way of displaying the actual data since, if displayed along with Figure 1, one can infer the sensitivity and the actual change in the variable of interest. Moreover, it enables direct comparison between models by comparing the slopes. It also clearly shows the differences between different latitude bands and also different layers of the stratosphere, which helps readers understand the different dominant drivers of ozone responses. Therefore, we prefer to keep the current figure. We also updated the colormap to make it easier to read.

Here is the updated Figure 2:

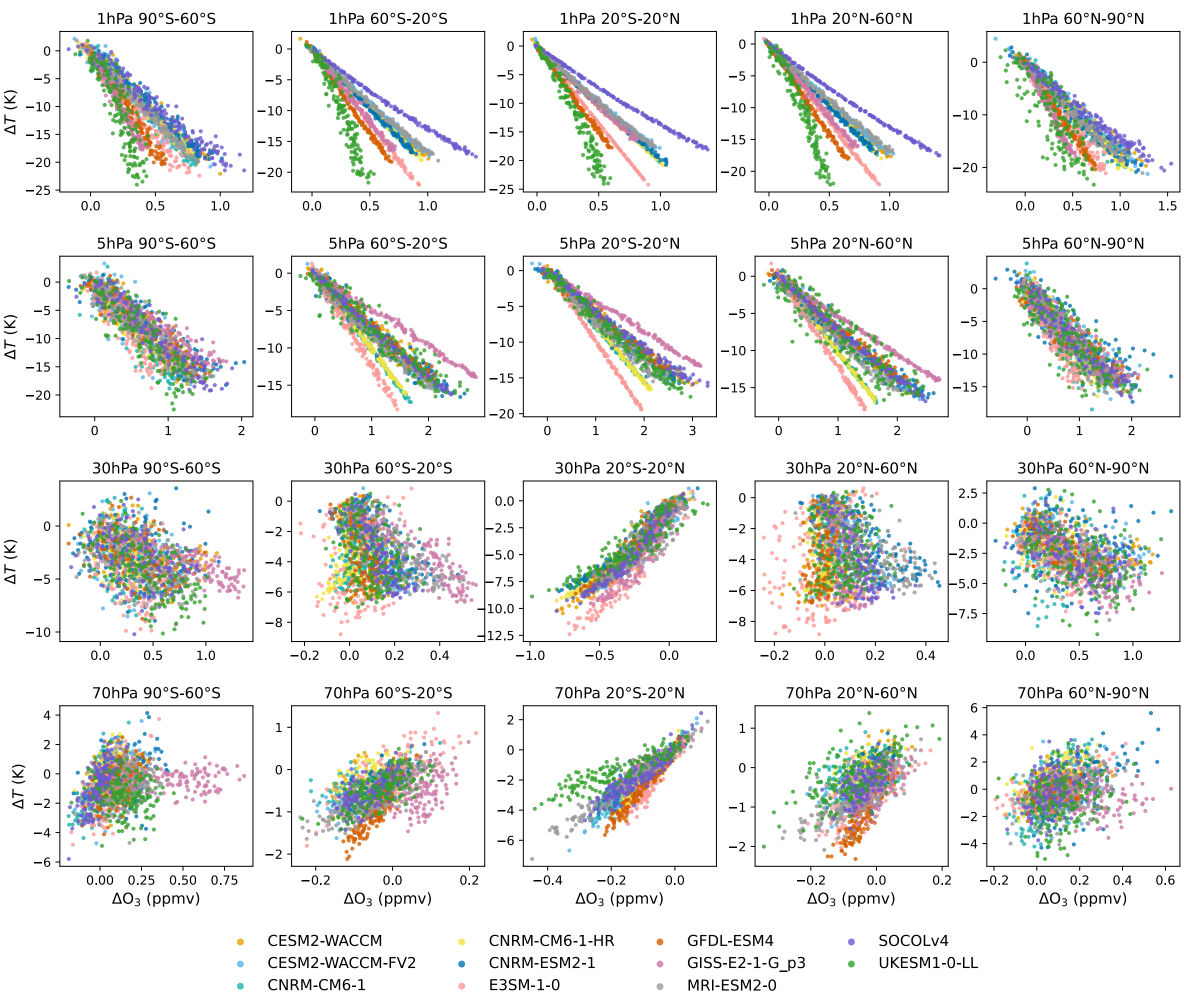

Figure 2. 150-year-long annual-mean ozone response to temperature change in stratosphere at different pressure levels and latitude bands based on the 1pctCO2 experiment.

4. Figure 3: Indeed the relatively weak dependence of ozone on temperature is because of the low abundance of halogens in a PI world. There is no way the dots can be visually attributed to a particular model (not in my print-out, at least). Perhaps again a different way of displaying this can be considered?

Response: Thanks for the suggestion.
We updated the colormap to make it easier to differentiate between models. We further denote the value of R2 for each of the model in the legend to help readers understand the plot. The fitted lines show the correlation between ozone and zonal wind response, R2 indicates how strong this correlation is supported by the data, which are the main information we want to convey through this plot.

Here is the updated Figure 3:

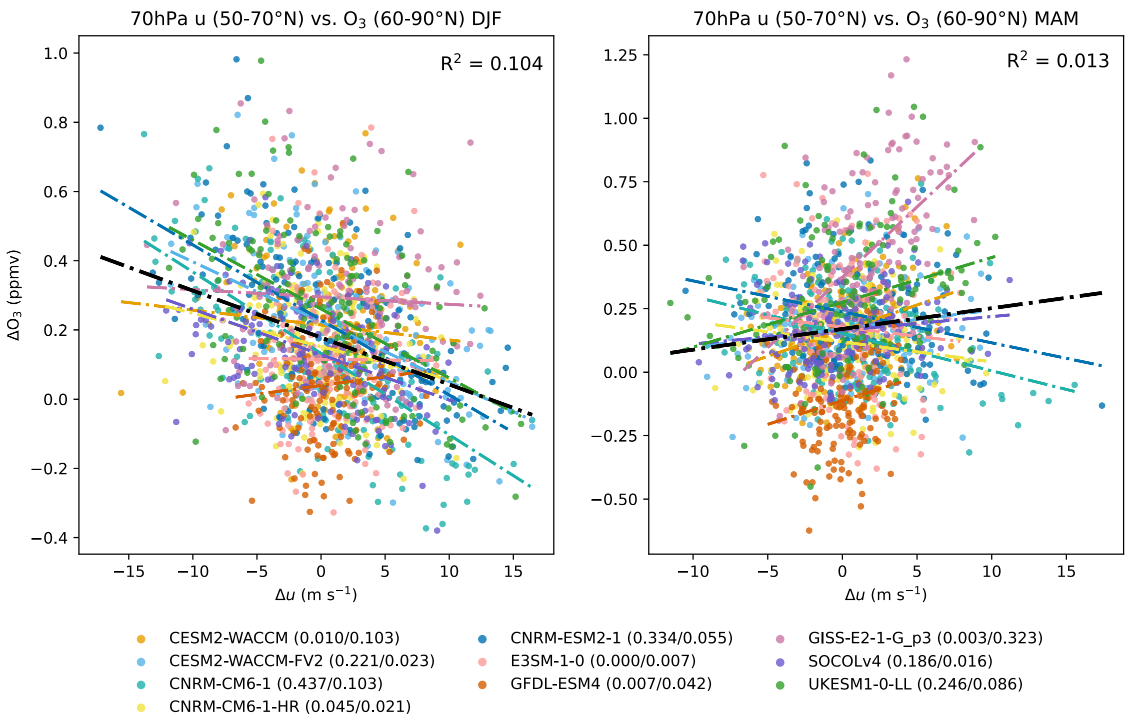

Figure 3. 150-year-long seasonal-mean ozone response in 60-90N to zonal wind (u) change in 50-70N at 70hPa in MAM and DJF for 1pctCO2. Fitting lines retrieved from linear regression are plotted as dash-dotted lines with the corresponding color for each model. The thick black line is fitted using data from all models with the corresponding R2 denoted in the upper right corner of the plot. R2 values for each model are denoted in the legend for DJF and MAM respectively.

5. Figure 5: This figure is also similar to Morgenstern et al., ACP, 2018, their figure 11, showing essentially the same: Unambiguous increases in TCO in the northern extratropics, model-

dependent signs of the tropical TCO trends due to cancellations, and a large spread of the ozone change over Antarctica.

Response: Yes, Figure 5 in our manuscript is similar to Figure 11 in Morgenstern et al., ACP, 2018. We added the citation to this paper in the following sentence in the manuscript (P15 line 258-260):

"These results are consistent with the analysis of the data from four CMIP5 models (Chiodo et al., 2018), including also the response in the NH being larger than that in the SH due to a stronger BDC (Butchart et al., 2014). They are also largely consistent with a previous study using CCMI-1 data on the sensitivity of ozone to GHGs (Morgenstern et al., 2018)."

Response to RC 2

General comments:

The authors pointed out that Arctic ozone increase when the Arctic stratospheric vortex weakens. The study uses multiple CMIP6 models to analyze the relationship between ozone and the polar vortex; however, the models differ significantly in their simulations of polar vortex strength, suggesting that there may be uncertainty in key processes within the models. The authors suggest that this relationship is stronger in winter but weaker in spring on the interannual timescale. But from the perspective of seasonal variation, a weakening of the barrier in early spring may lead to enhanced transport, so why is the response weaker in this period? Actually, the breakup of polar vortex associated with final warming during early spring is also closely related to the transport barrier effect. The authors shall investigate the connection of breakup time of polar vortex in early spring to ozone changes, instead of using March-April-May mean, which may mask this relationship. In a short, I think the sentence of 'the transport barrier role of the polar vortex is generally weaker in spring than in winter' is not appropriate. In addition, the Antarctic polar vortex is stronger and more stable than the Arctic polar vortex, why is there no discussion of how changes in the Antarctic polar vortex respond to ozone feedbacks?

Response: Thanks for the comments and the suggestions.
Indeed, averaging over a long time span might blur the relationship, and thus it is not appropriate to infer the response of Arctic stratospheric vortex from this analysis. In order to study the breakup time of polar vortex, one would need to align the data relative to the final warming date of each individual model, which is out of the scope of this paper Therefore, we revised our discussion of Figure 3 to emphasize this caveat as follows (P12 lines 225-228):

"The breakup of the polar vortex may lead to enhanced transport of ozone to polar region, but averaging over MAM may mask this relationship. Investigation of the breakup time of polar vortex and how it changes under climate change would need to be considered for each models, which is out of scope, but which merits further investigation."

We also added a similar plot as Figure 3 for Antarctic vortex and corresponding discussion in the appendix as following:

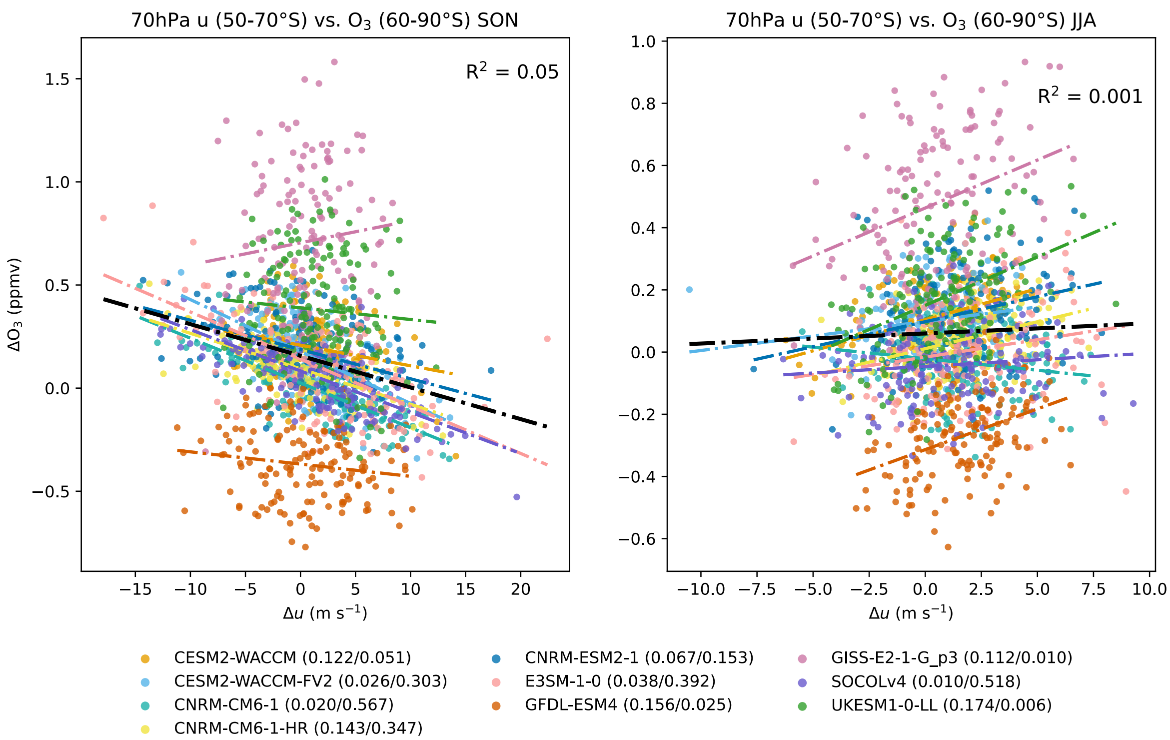

Figure B2. 150-year-long seasonal-mean ozone response in 60-90S to zonal wind (u) change in 50-70S at 70hPa in SON and JJA for 1pctCO2. Fitting lines retrieved from linear regression are plotted as dash-dotted lines with the corresponding color for each model. The thick black line is fitted using data from all models with the corresponding R2 denoted in the upper right corner of the plot. R2 values for each model are denoted in the legend for SON and JJA respectively.

Minor comments:

1. Line2: '…, and lead to' -> '…, leading to'

Response: Revised accordingly.

2. Line6: 'This work employs the latest data from Coupled Model Intercomparison Project Phase 6 (CMIP6), …' The comma after "CMIP6" is unnecessary.

Response: Revised accordingly.

3. Line10: 'We then explore the feedback exerted by ozone on climate'. This expression can be simplified as 'We then explore the ozone-climate feedback'

Response: Revised accordingly.

4. Lines11-12 ‘We find that the stratospheric temperature response is substantial, with a global negative radiative forcing by up to −0.2 W m−2.’ The radiative forcing responses of the different models have large variations, and in the text analysis shows that the largest radiative forcing is −0.19 W m−2 and is derived from the UKESM1-0-LL that does not perform well in any of the other feedback processes (including ozone response to 4×CO_2, ozone response to temperature change and SSW frequency change due to 4×CO_2), and I think that a clear range of global mean net radiative forcing should be included in the abstract.

Response: Thanks for the suggestion.
We replaced "with a global negative radiative forcing by up to −0.2 W m−2" with "with a global negative radiative forcing ranging from -0.03 W m-2 to -0.19 W m-2".

5. Line152: ‘against’ -> ‘with’

Response: Revised accordingly.

6. Line158: ‘year 135 to 145’ -> ‘years 135 to 145’

Response: Revised accordingly.

7. Line172: It should be 200-240 nm in this reference.

Response: Revised accordingly.

8. Line 209: Figure 3 only reflects the correlation between ozone and zonal wind. How did you know that the polar vortex is weakening from Figure 3? Is it through the average zonal wind of each model?

Response: Thanks for bringing it up!
We mainly look at the correlation between the ozone and zonal wind response indicated by the negative trend instead of the absolute change, which is shown later in Figure 11. From this correlation, we propose that when the polar vortex is weakened (delta_u <0), the weakened barrier leads to more mixing of polar air with ozone-rich air, thus an increase in ozone abundance for most models.

We have revised the corresponding sentences as following to make it clearer (P12 lines 218-220): “Figure 3 shows that in winter, for most models, the weakening of the NH polar vortex reflected by the weakened zonal winds in 50-70N correlates with an increase of ozone in the Arctic (small but significant negative slope).”

9. Line223: in most locations -> in most regions

Response: Revised accordingly.

10. Line243: Decoupling -> Decomposing

Response: Revised accordingly.

11. Line259: 'during the last 80 years' perhaps it could be changed to 'over the subsequent 80 years'

Response: Revised accordingly.

12. Line354: Do you mean stratospheric ozone depletion or stratospheric ozone recovery? There are different behaviors of the polar vortex and jet stream under these two scenarios. Please clarify it.

Response: Since we are comparing chem/no-chem, the change would stem from the changes in ozone under 4xCO2, which is similar (but not necessarily the same as) future stratospheric ozone recovery.

13. Line410: Expanded AMOC as "Atlantic Meridional Overturning Circulation" when first introduced in the sentence, then used the abbreviations consistently.

Response: Revised accordingly.

14. Although the authors mention statistical significance tests (e.g., t-tests), there is limited information on the exact methods used. It would be useful to provide more details about the statistical.

Response: Thanks for bringing it up.
We add the following sentences in the Results section to explain how we did the t-test (P8 lines 166-168):
"We assume the timeseries of ozone concentration under piControl and 4xCO2 are independent samples with the same variance, then we compute the t statistic to see if the two samples have same mean value. This also applies to other variables we analyze hereafter."

15. The manuscript provides a detailed assessment of the long-term (150-year) ozone response to increased CO2. Meantime, the authors mention that ozone changes in the early stages of CO2 increase are characterized by rapid adjustment. Does this fast-adjusting response exhibit

nonlinearities or threshold points? Could this threshold point depends on whether the chem or no-chem model? Is there some consistency of threshold in the chem/no-chem models?

Response: Thanks for the question.
We do not find evidence of any nonlinearities or threshold points from the evolution of ozone in the 1pctCO2 experiment. And therefore, we don't expect any nonlinearities in the chem/no-chem models. There is also no clear evidence of non-monotonic behavior in the stratospheric circulation under increasing CO2 forcing, at least for GISS (Menzel et al., 2023 https://doi.org/10.1175/JCLI-D-22-0851.1).
However, the only aspect that introduces some non-linearity might be the AMOC, which collapses at different times across models and configurations, and in one case (GISS), the difference in the behavior may be related to ozone feedbacks (Orbe et al. 2024 https://doi.org/10.1175/JCLI-D-23-0119.1), but this has not been shown yet for other models.

16. The different colors in Fig.2 are hardly to see. Please redraw it.

Response: Thanks for pointing it out.
We have updated Figure 2 as following:

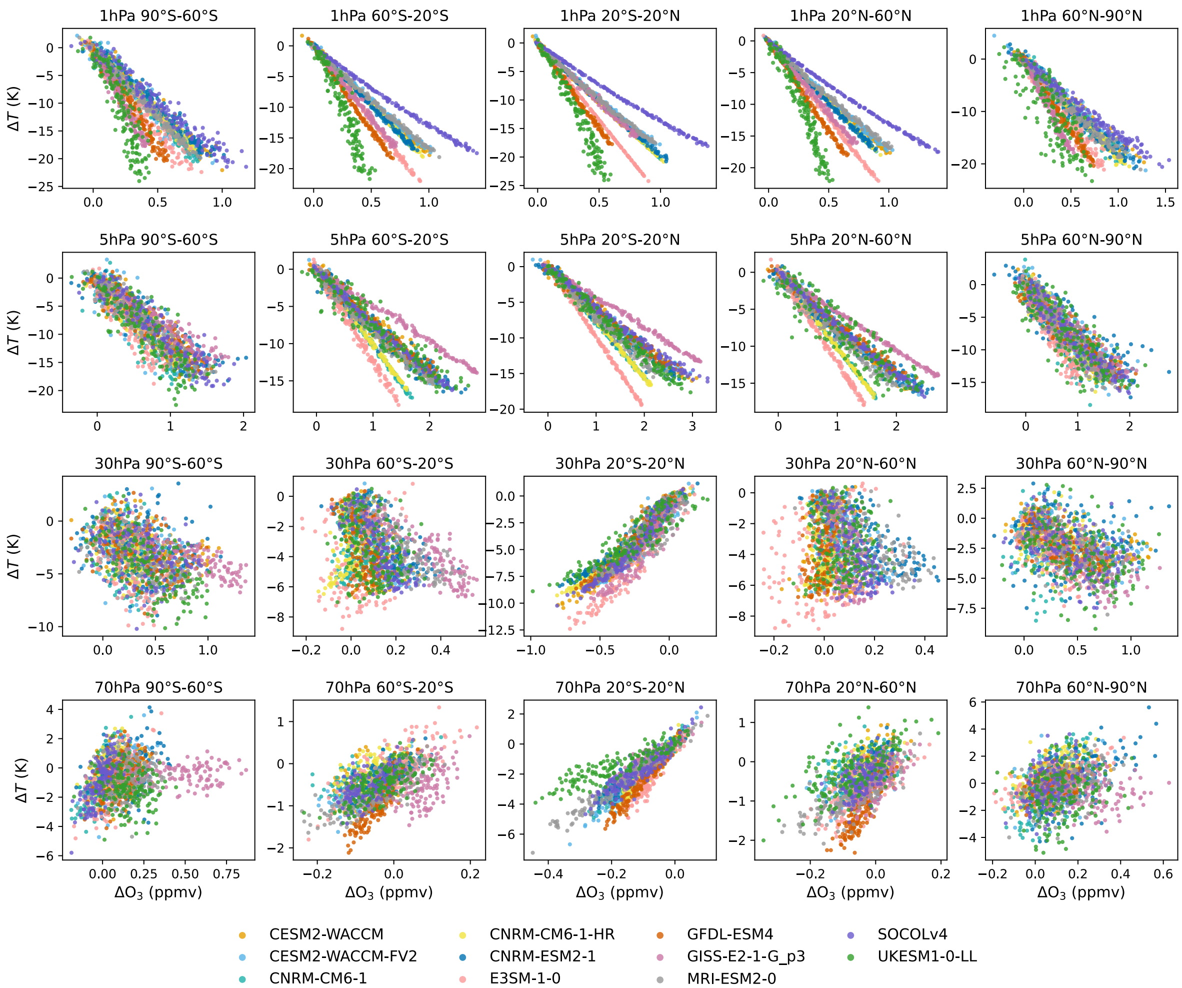

Figure 2. 150-year-long annual-mean ozone response to temperature change in stratosphere at different pressure levels and latitude bands based on the 1pctCO2 experiment.