Response to Reviewer 1

Recommendation: Minor revision

This is a useful summary of the experimental design for Phase 1 of the APARC QUOCA project, including a detailed set of tables that specify the model output requested. It is suitable for publication once the authors address the specific comments below.

➔ Thank you very much for your insightful and constructive review. Our responses to your specific comments are indicated below.

Specific Comments (line number):

(66) "a dampening of": This should be "damping", as in "reducing", or "dissipating"—not as in "moistening" (dampening)

➔ We thank the reviewer for catching this error. We have changed all incorrect references to"dampening" to "damping". Please see the revised manuscript.

(71) "former claims": "previous claims" would be better.

➔ We have changed "former claims" to "previous claims", as recommended by the reviewer. Please see the revised manuscript.

(118) Figure 2, caption: "dampened" à damped

➔ We thank the reviewer for spotting this error. This has been fixed in the revised manuscript.

(*141) "dampened" à damped*

➔ We thank the reviewer for spotting this error. This has been fixed in the revised manuscript.

(146) "dampening" à damping

➔ We thank the reviewer for spotting this error. This has been fixed in the revised manuscript.

(150) "whether other models exhibit similar vertical structure": The version of WACCM used by Calvo et al. (JAS, 2015) does not appear to have the same response in the deep branch.

➔ True – we thank the reviewer for her/his careful analysis. We have modified the sentence by adding "preliminary comparisons with other models suggest that the ozone feedback, particularly on the deep branch, may be model-dependent (Hufnagl et al. (2023), Calvo et al. (2025))". Please see the revised manuscript.

(179) "anomalous triggers (volcanoes, ENSO, wildfires)": However, among these, only ENSO is intrinsic to ocean-coupled simulations. Volcanoes and wildfires are boundary conditions that can be specified (or left out) as desired.

➔ We have retained the reference to volcanoes and wildfires, but we have changed the wording to more faithfully convey the difference between internal variability versus anomalous triggers. The rephrased clause reads as "internal variability intrinsic to atmosphere-ocean coupling (i.e., ENSO) and anomalous triggers (e.g., volcanoes, wildfires) in ozone." Please see the revised manuscript.

(184) "input4MIPs": This needs a link and/or reference.

➔ Thank you. We have added a link. Please see the revised manuscript.

(189) "at least ozone needs to run interactively": Note that the QBO-ozone perturbation above ~20 hPa will be driven mainly by changes in NOX, which are due to advection. So, interactive NOX and its precursors (N2O, tropospheric generation of NOX via lightning) should also be part of any INT experiment.

➔ We thank the reviewer for making this excellent point – indeed, several of the authors on this manuscript are interested in understanding precisely the NOx contribution to the QBO-ozone feedback at these levels. As shown in co-author Ming's recent study, NOx variations are essential, not only in reproducing the amplitude of the ozone peak above ~20 hPa, but it also has nonlocal effects on the ozone and temperature QBO patterns at lower altitudes (through modifying the column ozone aloft) (Ming et al. (2025)). At the same time, we intentionally do not ask that models use schemes with interactive NOx and its precursors as we want to include as many models as possible (including those with interactive QBOs, but linearized ozone mechanisms). One of the working groups in support of the QUOCA effort will be focused on assessing how well linearized ozone schemes capture various aspects of the ozone-QBO feedback (in particular, its vertical structure). Incorporating models with simplified ozone mechanisms (that do not have explicit NOx coupling) is therefore intentional. We now make these points explicit in the manuscript.

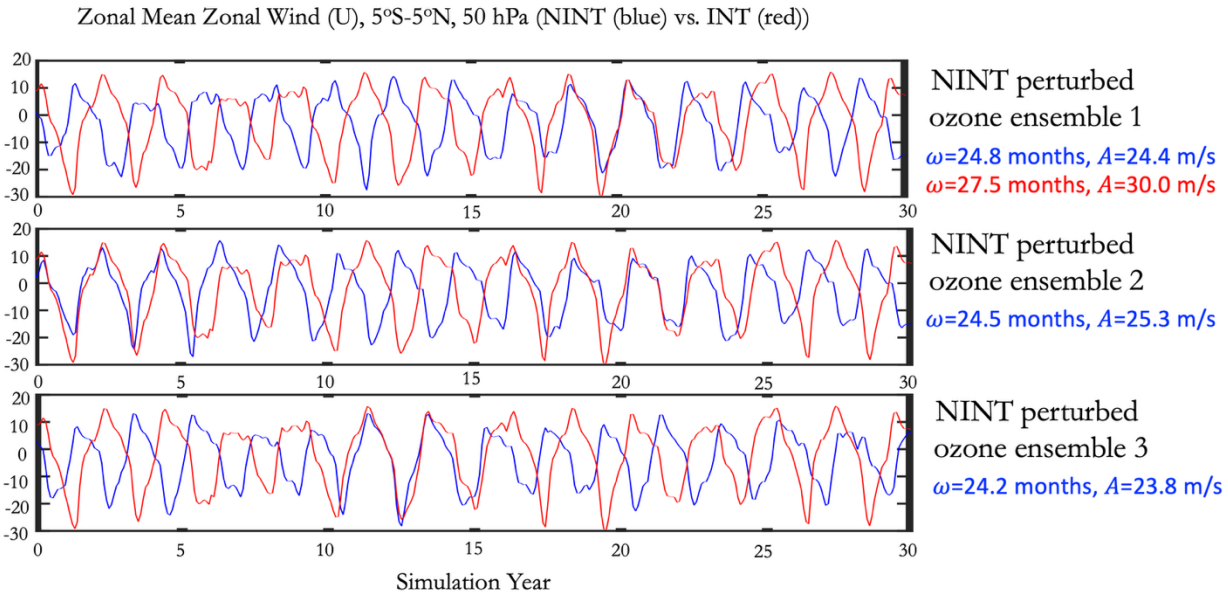
Ming, Alison, Peter Hitchcock, Clara Orbe, and Kimberlee Dubé. "Phase and amplitude relationships between ozone, temperature, and circulation in the quasi-biennial oscillation." *Journal of Geophysical Research: Atmospheres* 130, no. 4 (2025): e2024JD042469.

(192) "monthly annual cycle of three-dimensional ozone fields": This would be clearer as "annual cycle of three-dimensional, monthly-mean ozone fields".

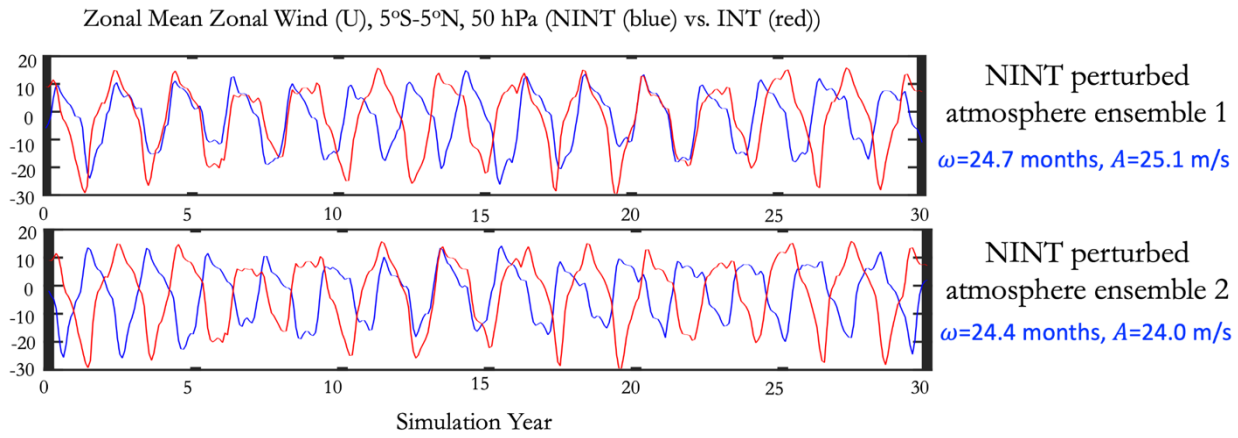
➔ We agree with the reviewer and have changed the wording as suggested. Please see the revised manuscript.

(193) "The 90-year-long PD-INT experiment … three successive 30-year … climatologies": You may want to offer some motivation for using a separate 30-year climatology for each NINT experiment instead of driving all of them with a single 90-year climatology. I would not expect 30-year slices of the 90-year INT run to be significantly different. That is, the statistical properties of 30 vs 90-year climatologies derived from the INT run ought to be statistically identical.

➔ The reviewer makes a good point that one might not intuitively expect that the climatological ozone distributions generated from three distinct 30-year ozone periods would introduce as much intra-ensemble variability, compared to more standard practices for ensemble generation. However, preliminary experiments that we performed using the GISS E2-2 model (shown below) comparing this ensemble approach with a more conventional approach show that the resulting spread in the QBO is similar in magnitude. More precisely, the figure below shows that three NINT ensemble members (blue), generated using the proposed approach (i.e., each constrained with a distinct ozone climatology derived from non-overlapping 30-year windows from the PD INT integration) all show a reduced QBO period (by ~2.7-3.3 months) and reduced amplitude (by ~4.7-6.2 m/s).



When we compare the intra-ensemble spread in this response to the responses in NINT simulations constrained with *same* ozone fields, but with slight differences in their initial atmospheric conditions (specifically, random perturbations in their tropospheric temperatures that are at most order ~1°C) we also find a reduction in QBO period with a similar spread (between ~2.8-3.1 months) and QBO amplitude (~4.9-6 m/s). This is shown below.



We conclude, therefore, that the ensemble spread generated through use of these different ozone fields is similar to that generated from more standard approaches. We have a brief description to the manuscript explaining our motivation for using this approach. Please see the revised manuscript.

(209) "to ensure that the chemistry is sufficiently spun up": I do not understand this. The entire 90-year PD-INT run should be properly spun up because you are already asking for a 10-year spin-up before starting the 90-year PD-INT run. It seems to me that the real purpose of picking different times within the 90-day PD-INT run to initialize the FT-INT runs must be to ensure that each member of the PD-NINT ensemble is an independent realization. Am I misunderstanding this?

➔ Wow—excellent catch! This was a complete error on our part, and we thank the reviewer for pointing out our mistake! We should have written "Each ensemble member is initialized from a different point in the PD-INT experiment in order to sample different initial states in the atmosphere." We are grateful to the reviewer's scrutiny of this section and refer her/him to the revised manuscript.

(230) "projects onto": "affects" would be clearer.

➔ Thanks – we have rephrased as suggested. Please see the revised manuscript.

(238) "controlled more directly": Why "more directly"? The OX lifetime at and above 10 hPa is less than a month and becomes even shorter (1 day) by 0.5 hPa, so I think it is safe to say that OX (hence, ozone) is photochemically controlled at these altitudes.

➔ OK – we have removed the reference to "more directly". Please see the revised manuscript.

(250) "the QBO period reduces": this should be "is reduced" or "decreases" (reduces is a transitive verb).

➔ Thanks – we have rephrased as suggested. Please see the revised manuscript.

(262) "contributed submissions": "contributed results" or "submitted results" would be better.

➔ OK – we revised this phrasing, changing it to "contributed results". Please see the revised manuscript.

(284) "tropopause air pressure and …": Will there be a common definition of "tropopause"? Will it be the same in the Tropics vs. extratropical latitudes?

➔ We have not asked that all modeling centers use the same definition of the tropopause and we realize that this will add another layer of structural uncertainty when interpreting the results, especially those related to stratosphere-troposphere exchange. However, we do ask in the appendix entitled "Prescribing Stratospheric versus Tropospheric Ozone" that modeling centers capable of diagnostically distinguishing between tropospheric and stratospheric ozone both a) prescribe the climatological ozone fields from the PD-INT experiment in the stratosphere only when performing the PD-NINT experiment; and b) provide information about which exact tropopause they have used and details of the methodology used to calculate it.

(284) "Table B2": It would be useful to indicate in the table header that the quantities in the table are zonal averages.

➔ Not all the quantities in the table are zonal averages (e.g., prec, cod, convec_cloud_area_frac, cloud_area_frac). We assume, however, that the reviewer (based on another comment) would like this noted, regardless. We have added a clause to the table caption qualifying that most of the variables are zonal averages. Please see the revised manuscript.

(292) "subset of latitudes (15°S to 15°N)": ±20° might be more appropriate since forcing is substantial across this range of latitude (see, e.g., Garcia and Richter (2019), their figure 7).

➔ While we understand the physical motivation for extending the latitude range as suggested, we want to maintain as much consistency with the QBOi output request as possible (as this may enable incorporation of results from those experiments as well in new analyses – mixing diagnostic definitions would obfuscate these types of efforts).

(298) "including" à included

➔ Thanks for spotting this error! We have fixed this typo in the revised manuscript.

(306) "Table B4": Again, it would be useful to indicate in the table header that most of these are zonally averaged quantities, except for O3STE, AOD and ST8025. By the way, the table does not state what ST8025 denotes; ditto for e90. This information should be included in the "long name" column, as is the case for other fields.

➔ OK – we have added the qualifier about zonal averages to the caption, as requested. We agree that more descriptions of the ST8025 and e90 were warranted. We have added these to the table, now referring to these tracers as stratospheric and surface loss tracers with 25 day^{-1} and 80 day^{-1} lifetimes, respectively. Please see the revised manuscript.

(312) "these will be used to calculate the TEM … to verify consistency … Table B2": I find this a bit confusing. I presume you do not intend to calculate TEM quantities from the eddy fluxes (u'v', v'T', etc.) in Table B2 (plev42 grid) but will instead compare TEM fields calculated from the outputs specified in Table B3 (plevTEM grid) with the pre-calculated TEM quantities in Table B2 (vtem, wtem, psitem, epfy, epfz). Is that the idea? Note also that it is unlikely that the EP flux divergence, which involves the vertical derivative of epfz, can be calculated accurately from epfz output in Table B2 because the levels in that output are not native model levels. So, if one wants monthly div(F) on the plev42 grid, it ought to be included as a pre-computed field in table B2. I may be misunderstanding your intent, but I think this requires some clarification

- ➔ Correct – we intend to calculate TEM quantities based on the eddy fluxes specified in Table B3 on the plevTEM grid. We are not exactly clear, however, why this is confusing to the reviewer because we already specify that the flux output from Table B3 (not B2) will be used. Nonetheless, we have added the qualifiers "6-hourly instantaneous" and "monthly" to more clearly distinguish between the B3 versus B2 outputs. Please see the revised manuscript.

(316) "further below" à farther below

- ➔ Thanks – this has been fixed. Please see the revised manuscript.

(321) "6-hourly (3-D) instantaneous output … for only 10 years": If this is going to be used to analyze QBO forcing by, say, compositing with respect to QBO phase, 10 years may not be enough for statistical reliability since that interval spans less than 5 QBO cycles.

- ➔ We agree that it may be challenging to extract a robust signature, but we are intentionally limiting this output request to ten years so that it remains tractable. No changes to the manuscript.