Authors' response to Editor's comments on "QBOi El Nino Southern Oscillation experiments: Overview of experiment design and ENSO modulation of the QBO" by Y. Kawatani et al.

We thank the Editor for carefully itemizing the concerns raised and offering valuable suggestions for improving the revised manuscript. Our responses to each of the Editor's comments are included below. For clarity, we have included the Editor's comments in *blue italics* and our responses in regular font. Note that our responses here refer to numbered "general discussion points" (I-V) that are included in our responses to the Reviewers' comments.

Dear Prof. Kawatani and co-authors:
Both Anonymous referees have posted their comments on your manuscript (WCD 2024-3270). As per WCD policy, you are now to post a response on how you will address the referee's comments – after which I will make a decision on the manuscript.

Both reviewers have made excellent comments on the manuscript and call for major revisions which I agree with, particularly concerning the issues of "what is new? (Reviewer #2)", the unexplored impact of unrealistically large forcing on the response (Reviewer #1), the lack of a discussion of agreement between the simulated QBOs and those observed (particularly the period) (Reviewer #1), and the use of very old reanalysis data throughout the manuscript (both Reviewers). To provide guidance in revising the manuscript so that it is acceptable for publication in WCD, below I itemize the issues that I expect will addressed in a revised manuscript. I will also post these on the WCD page for the manuscript. `

We appreciate your efforts in thoroughly itemizing the issues.

1. Reviewer #1 asks for an examination and justification of the impact of applying an ENSO forcing with an unrealizable amplitude. This should be addressed in the revised manuscript. One way to do this quantitatively is to run some of the models for ~20 years with realistic forcing and show how this impacts the wave forcing, say by producing Figs. 11 and 12 and comparing the amplitude of the wave forcing in the 3x El Nino with that with a 1x El Nino (and ditto for La Nina forcing).

We address this concern in our response to the Reviewers, specifically in our general discussion point (II). Your reference to “3 x El Nino” and “ENSO forcing with an unrealizable amplitude” is not completely clear to us. We constructed our El Niño forcing SST anomalies by first compositing over all El Niño events which produced a result appropriate for a typical moderate El Niño. Then, to amplify the effect, we multiplied that composite SST field by a factor 1.8. This choice avoids an “unrealizable” amplitude as our imposed peak NINO3 anomalies are then similar to those in the strongest observed El Niño events. The same considerations led us to multiply our composited La Niña SST by a factor of 1.4.

We will add to our revised manuscript a series of calculations that analyze the El Niño vs La Niño contrast in QBO amplitude and period as simulated in earlier AMIP runs by each of the models included in our study. Actually, this will amount to a repetition of Taguchi’s (2010) analysis of the observed record. This directly addresses the issue of analyzing model runs with realistic SST anomalies. The comparison of the AMIP results with results from our runs with larger (on average) “annually repeating” SST will be interesting. Also, in alignment with the Editor’s comment #3 below, we will include in this analysis the seasonal effects as revealed in Taguchi’s analysis.

2. Reviewer #1 calls for a necessary discussion of a deficiency in half the model models to simulate a QBO with a period that is consistent with that observed (see Reviewer #1 comments on Lines 266-76), and I agree. The reviewer’s argument shows that four of the models have QBO periods that are unrealistic (EC-EARTH: 20, MIROC-AGCM-LL: 26, MIROC-ESM: 21). I note that GISS is on the edge of being disqualified by this measure, and the unrealistic QBO in the La Nina simulations is a reason to add this model to the disqualified list. Hence, three of the six models in which wave driving is examined in Figs. 13 and 14 have QBOs with unrealistic periods. These issues should be mentioned in the abstract, in section 3, and in the discussion/summary in section 5.

In our revision we will include a discussion of these issues in line with your suggestions here.

3. Reviewer #2 asks whether the changes in the QBO period are sensitive to the method used to define a QBO cycle and whether ENSO impacts all phases of the QBO simulated by the models. I expect the method used to provide a robust period. However, it is not clear from the analysis presented whether El Nino accelerates all phases of the QBO, as is seen in the observations – or whether it impacts certain phases (such as the downward *propagation of the westerly shear zone). In most of the models, the slope of the constant phase lines in the vertical-time plots with El Nino forcing experiment is indistinguishable from the slope in the La Nina forcing experiment (e.g., in Fig 2 for CESM1, EMAAC, MIROC-AGCM, MIROC-ESM and MRIESM2). Only in EC-Earth3.3 is the slope of the phase lines steeper with El Nino forcing than with La Nina forcing, as is also the case in the observations (Taguchi 2010, cf Fig. 9a with 9b). Repeating the straightforward analysis of Taguchi on the model results in this study would add considerable information on this issue.*

We appreciate your helpful suggestions. In the revised manuscript, we will incorporate new analysis results based on the methodology described in Taguchi (2010). Please see our response to the Reviewers general discussion point (III) for further details. As noted above, we plan to apply this analysis to our "annually repeating" runs and to earlier AMIP runs conducted with the same models.

4. Though not explicitly discussed by the Anonymous Reviewers, the changes in the period of the QBO due to the phase of ENSO (El Nino vs. La Nina) in six of the nine models are small compared to that observed, despite the 3x forcing applied. This should be noted in the abstract, in section 3, and noted and discussed in the summary in section 5. For example, the observational analysis in Taguchi (2010) suggests that a QBO in a perpetual El Nino would have a period of 25 months – 7 months faster than during a perpetual La Nina (a 26% change). Only three of the models in this study feature this amount of change (even under 3 times the observed ENSO forcing), only one of which has a GWD parameterization that responds to changes in convective activity. [Interestingly, all three of these models are also the only models to have an average QBO period that is consistent with that observed (~28 months).]

We appreciate your insightful comments regarding the changes in the QBO period due to the ENSO phase. You correctly point out that the changes in the QBO period in six of the nine models are small compared to observations. We agree that this discrepancy should be addressed in the revised manuscript.

As you suggest, we will add a note to the abstract and section 3, and provide a more detailed discussion in the summary (section 5). We will specifically address the observational analysis in Taguchi (2010), which can be interpreted as indicating a 26% change in the QBO period between perpetual El Niño and La Niña conditions. We will also discuss the fact that only three models exhibit this magnitude of change, and that these models are also the only ones with an average QBO period consistent with observations.

5. Both reviewers call for ERA 1 reanalysis data to be replaced with ERA 5 data throughout the manuscript and I agree. Also, Reviewer #2 provides references to recent literature that is relevant to this study.

We will use ERA5 in the revised manuscript. Please see our general discussion point (V) in our response to the Reviewers for further details.

6. Reviewer #2 questions what is learned from this study, given that it is already well established that gravity waves that cannot be explicitly resolved in (even high resolution) AGCMs are important for the driving of the observed QBO and that the response of the QBO to forcing is sensitive to the gravity wave parameterization scheme. The reviewer laments that this phase of the project did not deliver on the promise of a quantitative analysis of the spectral properties of the wave driving in each of the models, which would have made the current study novel. Though I am sympathetic to the Reviewer's concerns, I do see value in the current study, but the revised manuscript should persuasively argue for the merits of the study, given the superficial nature of the analysis. [Certainly, the inability of 4 or 5 of 9 models to simulate a QBO with a realistic period is further evidence of the sensitivity of the QBO to the parameterization of gravity waves (see comment 3 above). See also point 7 below.]

We thank the Editor for acknowledging the value of this study. We will strengthen the manuscript by more clearly articulating the merits of our work, recognizing the limitations of our current analysis. As noted earlier, we also will quite significantly expand the scope of our study by including analysis of the QBO effects of ENSO as simulated in earlier AMIP runs. We also acknowledge the Editor's point that the inability of several models to simulate a realistic QBO period further emphasizes the QBO's sensitivity to gravity wave parameterization, and we will emphasize this connection in our revised discussion.

7. Reviewer #2 has made some good suggestions to improve the figure presentations. Moreover, adding observational results (from ERA5) to Figures 11, 12 and 14 would add important observational evidence for how ENSO actually does affect the wave driving, *and provide important information for evaluating the efficacy of the ENSO impact on wave driving in the models. These plots would be sufficient to assuage Reviewer #2's comment "What is new?".*

Thank you for the suggestions. We will follow the Editor's advice and include ERA5 results in our revised Figures to provide observation-based context for the impact of ENSO on wave driving. We believe this addition will help address Reviewer #2's concerns regarding the novelty of the study.

8. Figures 8, 9 and 10 are not necessary for the paper. That AGCMs reproduce the observed zonal average changes in circulation has been documented over and again, and the changes in these figures and not useful/used in understanding the impact of ENSO on the QBO (Figs. 11 and 12 are sufficient). Similarly, the text on lines 443-504 should be deleted (it detracts from paper).

Thank you for your suggestions. We acknowledge your point that Figures 8, 9, and 10, and the corresponding text (lines 443-504), may not be strictly necessary for this paper. However, we believe these figures, illustrating the fundamental zonal wind, temperature, wave forcing, and residual circulation, are valuable and relevant. These figures align with those presented in K2019 and are intended to support future QBOi-ENSO research.
In light of your feedback, we are prepared to move these figures and the related text to the Supplementary Materials to streamline the manuscript and maintain a clear focus.

Finally, a minor comment on statistical significance: On Line 240, we find "Emphasis will be placed on … statistically significant at the 95% confidence level." But elsewhere you mention 99% (e.g., Fig. 3 caption and on Line 287). Which is it? Line 241-242 goes on to say "Statistical significance is determined using a two-sided Student's t-test, sampling the maximum individual yearly mean data (e.g., 100 data points for models with 100-year integrations) for both the El Niño and La Niña runs". This is fine for differences in the climatological mean, but not for discerning whether the period of the QBO is different in El Nino vs. La Nina, which has degrees of freedom equal to the number of QBO cycles (minus 1) in each respective regime. Using these degrees of freedom for each model, I find that all of the differences in Fig. 2 are indeed statistically significant at 99%.

You are correct in identifying the discrepancy regarding statistical significance levels. The 99% confidence level is indeed appropriate for Figure 3 and the QBO period differences, while the 95% level applies to the climatological mean differences. We appreciate your careful attention to detail and will explain this more clearly in the revised manuscript.