Review of the article *QBOi El Nino Southern Oscillation experiments Part I: Overview of experiment design and ENSO modulation of the QBO* by Yoshio Kawatani et al.

General comments

This work is a continuation of the publication series produced by the QBOi project, based on an experimental protocol and several models known to simulate the QBO. This work is focused on the El Niño/La Niña effects on the QBO as simulated in 9 models. Most of the article describes the common and different features found in the different model simulations, and its structure follows the work of Kawatani et al. (2019), hereafter K2019, where they investigated the El Niño/La Niña effects on the QBO in MIROC models.

The experimental design chosen here is however deviating from that of K2019. Here they decided to uses amplified mean El Niño/La Niña SST anomalies. This makes any direct comparison to K2019 and to observations difficult. Whether or not the QBO response should be linear to the amplitude of the SST anomaly pattern is not discussed. Probably it would have been better to use the same anomalies as in K2019. (Or an entirely different design based on SST fields of selected El Niño and La Niña years.)

They find that El Niño/La Niña effects on the QBO period are qualitatively similar with respect to the period, with El Niño leading to a shorter period despite of the increased tropical upwelling in the tropical lower stratosphere, from which it is clear that El Niño must also produces a stronger wave meanflow interaction. No common response is found for the QBO amplitude.

An interesting part is the discussion and analysis of the reasons for the described results: The more equatorial precipitation and the weaker Walker circulation found during El Niño conditions. These features are found in all models, and they probably are independent of the skill of a model to simulate a QBO. The discussion of the wave meanflow processes is however more complicated, because of the rather different ways that resolved and parameterized waves generate the QBOs in the different models. And therefore not so much can be learned from this part, except that there exists still a considerable difference in the way that models generate QBOs. Further, as acknowledged by the authors, more detailed model diagnostics would be needed to learn more about the underlying reasons for the found behaviours. But this additional diagnostics was not part of the protocol, or the modelling groups could not produce these diagnostics.

Overall I think that the publication is worthwhile, as it creates a baseline for further work on this topic. Some minor corrections are needed before publication.

Detailed comments and questions

Abstract

L40 Stratosphere-troposphere Processes And their Role in Climate (SPARC) …
As we know QBOi has been started as a SPARC project. But SPARC has changed its name to APARC and QBOi is now listed as an APARC project. Maybe it is worth to add a remark or a footnote on this aspect.

L45 … models -models … should probably be … models. Models …

1 Introduction

L64 … that QBO facilitates … → … that the QBO facilitates …

L140 … Conducting a common ENSO-QBO experiment across a range of QBO-resolved climate models could help elucidate the role of non-orographic GWP in driving the oscillation. …
The work of Richter et al. (2020) on the climate warming effects on the QBO unfortunately showed that the differences between GWPs are considerable and probably responsible for the rather different QBO responses to the warming. As it seems it was not possible to decide which GWPs were "wrong" or "right". Now a similar exercise is presented aiming at El Niño/La Niña variations in SST as the external forcing instead of a warmer SST and increased atmospheric CO_2. Why should we expect a scientifically more robust result if Richter et al. (2020) have shown that differences in parameterizing non-orographic gravity drag can lead to very different results? Simply because El Niño/La Niña cycles exist in the historical period for which the models have been tuned?

2. Model Description and Experimental Design

L179 … These factors bring the peak composite anomaly SSTs closer to the anomalies observed during the most intense El Niño and La Niña events. …
Using amplification factors is problematic. This makes a comparison to observations or to the work by K2019 difficult. It seems necessary to add some remarks about the linearity between the SST pattern amplitude and the response of the QBO. Can this be assumed?
Alternatively you could have chosen specific years with strong El Niño and La Niña SST anomalies. Then there would be no need to amplify the SST anomaly, and there would be less of a risk to construct an SST anomaly pattern that mixes the different types of El Niños, which are discussed in literature.

L199 … For clarity and conciseness, we will refer to these models as CESM1, EC-EARTH, ECHAM, EMAC, GISS, LMDz, MIROC-AGCM, MIROC-ESM and MRICESM1, respectively. …
The abbreviated model names are introduced here, but not used consequently. Tables, Figures, and also soe sentences use the full model names. Please decide whether short names shall be used or not. But if you decide to use short names, then please use these in all places: tables, figures, and text.

L203 … Launch levels for parameterized gravity waves varied across models, ranging from 450 to 700 hPa or 1000 to 100 hPa. …
To which model(s) do the two pressure ranges relate? Please clarify.

Table 1. lunched level → launch level

Table 2. What does the entry for GISS-E2-2G and Residual stream function (✓5-1115) mean?

L243 … from the ERA-Interim (ERA-I; Dee et al. 2011) reanalyses …
Why is ERA-I used for this comparison, when ERA-5 is now available? Newer reanalyses are generally improved compared to earlier ones.

L247 … Importantly, the composite ERA-I and CMAP data were not scaled …
This is a kind of a flaw in the experimental design. If the response to the SST anomaly patterns is non-linear to the amplitude, then the applied scaling is hindering a direct comparison to observations or analyses. If, however the signal is linear, then the signals derived from ERA-I should be scaled like the SST patterns used for the simulations.

3. ENSO Modulation of the QBO and Climatological Mean Field Differences

L266 – L276 These lines discuss deficiencies in the structure of the simulated QBO, as occurring in El Niño or La Niña simulations of ECHAM, GISS, and LMDz. In my opinion it is necessary to point out another deficiency, which is an unrealistic period, although a regular pattern of downward propagating westerly and easterly jets is simulated. Taking the displayed 20 years (Fig. 2) of the El Niño and La Niña simulations together, we have 40 years for which on average (40years / 28 months) we would expect about 17 cycles. A count of the cycles shown in Figure 2 can now serve as an additional measure for the quality of the simulations. If we allowed a range of 15 to 19, then the following models (here excluding ECHAM, GISS and LMDz) would fail: EC-EARTH: 20, MIROC-AGCM-LL: 26, MIROC-ESM: 21. Please extend your discussion of problematic simulations in this direction, so that the reader knows from the beginning which model simulations need to be viewed a bit more critically.

L311 … Next, we consider ENSO modulation of QBO amplitude, which is known less robust … → … Next, we consider the ENSO modulation of the QBO amplitude, which is known to be less robust …

L323 … GISS, LMDz, and CESM1, all of which have variable GWP sources. …
I think it should be added that MIROC-AGCM-LL has variable gravity waves too, though these are explicitly simulated, within the given resolution, instead of parameterized. Thus variability of gravity waves not necessarily leads to a strong amplitude difference between El Niño and La Niña. And one needs to wonder if a strong change is indicating that the variability of gravity waves is important aspect for a GWP, or whether this effect is rather a result of other aspects of parameterizing gravity wave. Please add some thoughts on this problem.

L383 … although models tend to simulate the precipitation peak to the east of the observed one over the central Pacific in the El Niño run. …
It should also be mentioned that the precipitation peak in the model simulations is higher than in observations, which indicates that the local forcing by latent heat release in the simulations is higher than that explained by the observed precipitation. Quite likely this is related to the amplified El Niño/La Niña SST strength.

L422 … significantly deep westerly difference … → … significantly deeper westerly difference …

L432 … for (left top) ERA-I …
ERA-I is “center top”

4 Contrasting wave forcing and residual mean meridional circulations in El Niño and La Niña from QBOi models

L458 … The $\overline{X}$ term represents any other unresolved forcing. …
Do you mean here parameterized momentum diffusion and effects from numerical diffusion and damping operators?

L470 … in El Niño and La Niña simulations. …
La Niña simulations are nor shown, but differences of the El Niño and La Niña simulations.

L477 … La Niña c annual … → … La Niña annual …

L498 … which both use variable sources in their GWP, …
Do you mention this because you think that this is the reason for the differences? Often other differences in the formulation of the non-orographic gravity wave drag parameterizations can cause substantial differences already.

L509 … averaged over 20°S–20°N …
Maybe it is worth to explain why a band of 20°S – 20°N is chosen, while earlier diagnostics/figures used narrower bands. (I guess this is made in order to remove residuals of the secondary meridional circulation of the QBO.)

L510 … ranging from approximately 0.2 mm s^{-1} in MIROC-AGCM to approximately 0.4 mm s^{-1} in LMDz. …
This strong difference in the tropical upwelling implies also a strong difference in the strength of wave meanflow interaction that is necessary to simulate a QBO with a realistic period. This aspect is not discussed here, and maybe this El Niño/La Niña related article is the wrong place. Still it directly shows that the wave meanflow interaction must work at different strengths.

L522 … However, the specific altitudes at which $\overline{w}^*$ changes would most strongly influence the overall QBO period remain unclear. …
Sentence unclear.

L579 … While output data of parameterized gravity wave fluxes in LMDz were not available at the time of this analysis, this model, which also uses variable parameterized wave sources related to precipitation activity, showed similar structures affected by precipitation distributions (Dr. Lott, personal communication). …
Francois Lott is a co-author of this study. Please include the LMDz results in Figure 12.

L620 … parameterized wave forcing below is stronger …
What does “below” refer to? Maybe the sentence needs to be rephrased.

L639 … As discussed for Fig. 14, … → … As discussed for Fig. 13, …

5. Summary and concluding remarks

L685 … remained consistent …
“consistent” seems to be the wrong term, because this could have different meanings. (If El Niño/La Niña influences the ozone distribution, then the same ozone field cannot be consistent with El Niño and La Niña conditions at the same time.) “unchanged” would express more clearly that these fields simply have not been changed.