Manuscript # WCD-2022-14 Response to Editor and Reviewers

García-Franco, JL., Gray LJ., Osprey S., R. Chadwick and Z. Martin

Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford. Parks Road Oxford, United Kingdom OX1 3PU email: jorge.garcia-franco@physics.ox.ac.uk

---

1. Reviewer 1

This work by Garcia-Franco et al. looks at the relationships between the QBO and tropical climate in observations and centennial pre-industrial CMIP6 simulations with one coupled climate model. The connections are difficult to diagnose from observations so long simulations are useful. The paper is overall interesting and covers many topics, but the authors should check the consistency of the symbols and names used (see comments by line).

Many thanks for your time reviewing our manuscript and for your constructive criticism. We have addressed your comments and we hope you find the manuscript to be improved and suitable for publication. In this document we provide a point by point reply to the comments of both reviewers, where your comments are shown in black, and our response in blue.

1.1. Major comments

1.1.1. It can be confusing to read different acronyms for the same quantities. The units reported in the plots should be verified.

The revised manuscript now consistently uses acronyms defined in the method section for the easterly (E) and westerly (W) phases of the QBO, as well as El Niño (EN), La Niña (LN) and Neutral (NN) ENSO conditions. The units in all plots now consistently use [], as suggested below by this reviewer.

1.1.2. Given the central role of model simulations, more information on its skill at simulating QBO and ENSO should be provided. For example, how realistic is the QBO amplitude at 70 hPa for this specific model?

The first version of the manuscript discussed the performance of the model in the tropical troposphere based on available CMIP6-wide assessments in section 2.2. The revised manuscript adds a second paragraph (at the end of sectxtended ENSO seasonion 2.2) that provides specific information about the model representation of ENSO and the QBO. Weaknesses and strengths of the model are mentioned in this section. For example, the revised manuscript also discusses the realism of the QBO amplitude with emphasis on bias at the 70 hPa level Bushell et al. (2020). This bias is key for our study and is discussed further in the final section of the manuscript.

1.1.3. Apart from composite differences, some climatologies should be discussed.

The relevance of the climatological representation of the tropical features is stressed more in section 3.4 as well as in the discussion section in the revised manuscript.

1.1.4. In the introduction reference to Geller et al., 2016 on gravity wave changes would fit.

We have added the reference mentioned in several instances of the revised manuscript.

1.1.5. Model-dependence of the results should be stressed, since different configurations of a single model are analysed and QBO/SST biases may play a big role. The causality analysis on how the QBO influences ENSO is not very convincing as it stands.

The discussion section now emphasizes that these results could be model and bias-dependent. We agree that this manuscript does not provide a causality analysis on the QBO-ENSO relationship. However, given the non-stationarity of the QBO-ENSO relationship in the observations and the lack of any model study reporting that this relationship exists in a climate model, our findings are relevant and strongly emphasize the need for targeted model experiments that will causally disentangle these relationships.

1.1.6. I guess the authors should also say something about the frequency of LN/EN events during neutral QBO (QBO-N).

The revised manuscript now makes a note of the frequency of LN/EN events during neutral QBO periods, which are 0.258 and 0.263 month month^{-1}, respectively, so we report them to be roughly evenly distributed.

1.1.7. The section about monsoons should be revised and maybe shortened, since QBO surface impacts may be very dependent of any QBO bias. For example, Giorgetta et al. 1999 (cited) nudged to QBO, so it was realistic in their case.

The section on monsoons and the ITCZ has been revised to emphasize the importance of model biases, both in the troposphere and the stratosphere. The section on monsoons has been shortened to two paragraphs. A paragraph in the discussion emphasizes how these impacts could be model and bias-dependant.

1.2. Specific comments

L52, maybe 'on the convective process'?

Reworded as suggested.

L55, define 'CMIP', rephrasing L62

Done

L63, are GWs tied somehow to sources?

The model can sustain GWs of sufficiently large scale to be resolved by the resolution of the model, these resolved waves are tied to convective sources and the mean flow (Walters et al., 2019), however, waves of scales that are too small for the model grid are parametrised using a spectral sub-grid parametrisation scheme (Scaife et al., 2002) which in the current version of the model is parametrised so that the wave momentum flux is a function of total precipitation at each grid-point (Bushell et al., 2015).

L76, both monthly means?

Yes, we have specified this in the text.

L82, is there a reason for not using the standard 0.25x0.25?

We have chosen to use a coarser resolution than the default 0.25x0.25 to balance the computational cost of the vertical integrals using daily means for the mass overturning circulation and maintaining a reasonable resolution (0.75x0.75) for the rest of diagnostics.

L83, it is a bit strange to put the (generic) link only for ERA5; I would move to data availability with direct links for all datasets (for ERA5 `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressurelevels-monthly-means?tab=overview`) and proper citation `https://confluence.ecmwf.int/pages/viewpage.action?pageId=197704114`

We have moved the link to the data availability section, specifying the links (through the DOI) to monthly and hourly download landing pages.

L90seq, define 'N' and 'ORCA' for the components resolution.

We have defined 'N', whereas ORCA is a standard and generic name for the tripolar mesh grid used in the NEMO model.

L91, UKESM or UKESM1?

We have chosen to use UKESM1 as the model name and reference the pre-industrial control experiment with the acronym UKESM N96-pi throughout the study.

L94, So 3 simulations in total? Would be good to state that you have two models with lower resolution and one with better resolution, which is the main interest.

We have made this point clearer in the text.

L96, I would move to 'data availability' or similar

Done.

L98, Here or later I would add something about some relevant model properties (e.g. both models have more spectral power in 2-3 years compared to observations). Also ocean resolution seems to be important for mean biases, and the realism of the ITCZ should be mentioned as well.

In the last paragraph of section 2.2 we describe the strengths and weaknesses of the Met Office models, with particular emphasis on the medium-resolution configuration of HadGEM3 GC3.1 The revised manuscript includes details on the biases of the QBO (e.g. more power at longer periods than observations) and on the realism of the ITCZ placing these biases in the context of current state-of-the-art CMIP6 models. The oceanic resolution of the model is stated in the first paragraphs of section 2.2; the relatively high oceanic resolution in GC3 N216-pi may be part of the reason why the GC3 N216-pi is top ranked in most model assessments in the tropics. However, we cannot comment further as we are not aware of a study that systemically analyses how oceanic resolution impacts model's performance in the tropics in CMIP6.

L105, What about UKESM1? Not sure why only HadGEM is mentioned.

There are several reasons for this: (1) the results of the manuscript are primarily presented for the medium resolution configuration of HadGEM3. (2) The assessment studies cited in this paragraph only seldom used the UKESM model so we cannot place UKESM into the wider context of tropical biases as we did for HadGEM3. (3) UKESM uses the same dynamical core as HadGEM3 which means that in many contexts of tropical climate, the climatological biases are very similar to HadGEM3 in pre-industrial control experiments (García-Franco et al., 2020).

L111, years or months?

Months. Corrected

L117, Which levels? Above you just mention 70 hPa.

We have specified that we use the range of 10-70 hPa.

L119, '1' and '2' are subscripts

Done.

L124, The product you use (GPCP?) for this index is providing convective and stratiform precipitation separately? Or is it a total precipitation? If not, remove convective (here and also in all instances following). Can you explain why using a precip-based IOD index rather than the standard SST-based one? Please add a reference if it was used before.

We use the convective precipitation from ERA5, not from GPCP, to better diagnose the effect of the QBO on deep convection. This index was used because the dipole signal is diagnosed for precipitation and not for SST in the model data.

L125, I'd use same style for EN3.4, with []

Done.

L133seq, This symmetry seems strange (given the ENSO asymmetry and QBO stalling) can you provide numbers?

Good point, indeed neither ENSO or QBO numbers are symmetrical. The point of these sentences was to highlight how the length of these simulations renders large composite sizes and provided rough numbers. The revised manuscript states the specific numbers for each phase to clarify this point.

L135, This is 'observed' for ERA5? Can you provide the values for HadSST? It is useful to compare model/observation statistics.

SSTs and ENSO indices are obtained from HadSST, so these 'observed' quantities are obtained using SSTs from HadSST and QBO winds from ERA5.

L140, Maybe start with 'When estimating correlations, they are...'

Done.

Fig 1, 'mm day-1' in brackets, or move 'pr' to title

Done.

L159, Please comment on the ITCZ realism.

This is a good point, ITCZ biases in the ITCZ limit the realism of the diagnosed impacts in the model, the reader is referred the reader to section 3.4 which describes and discusses the biases and their influence on the diagnosed response.

L162, Add reference

Done.

L206, I guess would be useful to have a table in the method section with the different numbers for ENSO and QBO. Why 120, does it have a special meaning?

We have removed the 120 count reference.

L209, But the wet anomaly in the Pacific and dry in the Atlantic are more marked with ENSO included. This is also seen in Fig5.

Good point. The revised version now makes this distinction, i.e., there are some regions that exhibit a robust QBO response that is independent from ENSO, whereas the response in the equatorial Pacific and Atlantic is a function of both ENSO and the QBO.

Fig5, If regression coefficients are re-scaled (caption), then a prime is missing in a&d titles. See Supplement as well.

Done.

L214, (1) $->$ (Fig. 1)

Done.

L216, it was EN3.4 before The revised manuscript now uses EN3.4 as the acronym for the EN3.4 index throughout.

L221, why no significance in FigS3?

The significance hatching has been added to this plot.

L225, mention Gray et al., 1992

Done.

Fig6, I'd use E and W for QBO in (b). Moreover I would define once all the acronyms (EN, LN, E, W) in the methods and be consistent throughout (no 'ea', EN3.4, etc.). Suggest NE or NN for Neutral ENSO. Moreover, would it be easier to read the plot ordering the boxplots as LN/NE/LN ? Why not showing E and W phases separately for the amplitude?

The revised figure uses E and W as the rest of the manuscript (no 'ea'). The rest of the manuscript now uses the acronyms suggested by the reviewer consistently.

Unfortunately, we didn't understand the suggestion "would it be easier to read the plot ordering the boxplots as LN/NE/LN", perhaps the reviewer meant to say the NN composite should be in the middle of LN and EN boxplots. We have created such a figure and found little improvement to the readability of the figure.

Finally, the amplitude was not separated in the figure because there was no ENSO modulation of the amplitude regardless of QBO phase, so we chose to use the version of the figure that is more concise.

L238, Have you stated which level are descent rates for? From the methods I got that the amplitude is integrated in the 10-70 layer, but descent rate is by level.

In the methods section, a brief description is given as to the calculation of the descent rates. The calculation is done following Schenzinger et al. (2017) by finding for each time step the level of zero wind line and computing the height difference of this zero-wind level for consecutive months, i.e., this calculation is not for a single level.

L246, See Geller et al 2016 about GW variations.

Thanks for this suggestion. Our evidence suggests that the GW variations or at the very least the ENSO influence on the QBO through the GW variations is less pronounced in this model which is interesting.

L252, So the frequency would be for example (# months EN) / (# months W)? Maybe mention that IOD will be considered later?

Yes, exactly, we have clarified this in the Table caption and we mention that the IOD will be considered later.

L260, ENSO3.4 -¿ EN3.4 (or maybe ENSO) Done.

Fig 7, [] missing around mm day-1 (check other plots as well). I guess IOD-prc is same as IOD?

Done, the IOD-prc reference has been removed, now we refer to the IOD or the IOD index (based on convective precipitation) consistently.

L266, write months in full. Can you elaborate on how the difference model/obs depends on the ENSO evolution in the model (e.g. Lengaigne et al., 2006)? Also worth nothing how the model index amplitudes are 2-3 times smaller than obs.

The revised manuscript writes months in full.

We agree that the amplitude difference between model/obs indices should be noted in the manuscript (and we have rewritten this section accordingly). However, by boostrapping the long simulation into periods of similar length to the observations, one can find amplitudes in the differences of these indices of similar strength to observations, i.e., certain 36-yr parts of the model simulation agree with the magnitude of the observed differences.

Fig8, as before, why 'convective'? Why now using a higher confidence level?

ERA5 and the simulations output the diagnostic of convective rainfall which is the diagnosed precipitation resulting from the convection scheme-only. The revised manuscript clarifies that we use convective precipitation to investigate more directly a possible link between deep convection and the QBO, for which we use ERA5 and model data. The confidence level written in this figure caption was a typo, as the same confidence level is used throughout the study.

L275, but could this be model-dependent?

Yes, different horizontal resolutions lead to different tropospheric biases in the monsoons which could change the specific QBO-ENSO impacts for each region. The key message from this figure in our view is that the combined influences appear to be relevant in these models, which means that in some cases part of the non-linearity of ENSO impacts is associated (to a modest extent) with the QBO.

L280, Please avoid the mix of abbreviations and months in full The revised manuscript now uses months in full.

L286, Maybe the Indian Ocean sector, rather than IOD?

Replaced as suggested.

L293, why '.'? Rephrased.

L295, atmospheric circulations. However, the model biases should be noted.

Done. Model biases are now more explicitly mentioned in the discussion section as stratospheric and tropospheric biases need to be carefully considered in all of our results.

L300, How are these longitudes selected?

These longitudes were selected based on the results from the previous sections, i.e., Figures 4 and 5.

Fig9, Only convective, stratiform rainfall removed? Is panel (b) indicating a double ITCZ bias? Can you comment in the text?

Yes, in both cases only precipitation resulting from the convection scheme in the model (convective precipitation) is shown. Panel b) shows a double-ITCZ bias but a true "double" ITCZ is only found for a limited period of time in boreal winter (García-Franco et al., 2020). The revised version now makes this comment.

L317, remove 'rate'

Done.

Fig10, define acronyms MSD, NAM. For more direct comparison you could mask values over oceans? Do you know why the regions show very net boundaries in some cases? Compare with Lee and Wang, 2012 their Fig4

Done. The net boundaries are found by design in the model to look more closely at land monsoon regions and compare with the observational dataset. As this reviewer points out, one alternative would be to mask values over ocean.

Fig11, I am confused by vector sizes. They are 3 or 0.3 10-2 Pa s-1, but their lengths do not differ by a factor 10. Please clarify. Also the plots are quite busy, can you try improving them?

Thank you for pointing out this confusion. This confusion arises because the vector sizes in the climatology plots do not correspond with the vector sizes key in the QBO anomaly plots, this means that the two vector sizes are not related to each other and their lengths are not meant to differ by a factor of 10. We have clarified this confusion in the figure caption. Additionally, we have improved the plot by selecting a smaller vertical extent to more clearly show the differences, and we have also removed the hatching for the significance and instead only shade/color the significant differences and finally, the new plot also uses a slightly modified color scale to make the plots less busy.

L330, Mention the QBO biases which may be important

The revised version of the manuscript now discusses the weak amplitude bias of the QBO in the lower stratosphere which is found in all the CMIP6/QBOi models with varying degrees of magnitude. The HadGEM3 model is amongst the best in this respect, yet the bias is noteworthy still. The mention of the QBO biases is done in several parts of the manuscript including in this section referenced by the reviewer.

L335, If you integrate to the top, then the integration bounds are swapped and $0- > p_{top}$ (or p_{surf})? Gravitational acceleration (g) rather than constant (G)? How do you compute the divergent component of zonal wind?

Thanks for this comment as the text was wrong. As the reviewer mentions, the integration bounds would be swapped. The revised version now states in the text that the methdology is to integrate *from the top level to the surface*. We use the acceleration term (g) for consistency with previous studies (see e.g. Yu and Zwiers, 2010). The divergent component of the wind is computed using the python library windspharm (Dawson, 2016) ; the revised manuscript now makes clarifies this part.

L351, To me some QBO/ENSO superposition can also be seen from the plots.

Agreed, as in most plots a superposition of ENSO impacts can be seen, which is why for all the plots we show a Neutral ENSO version of the figure. In this case, the revised manuscript now notes this remark by the reviewer, i.e., that some superposition can be seen.

L406, or role of QBO bias...

The revised manuscript discusses the weak amplitude bias of the QBO in the lower stratosphere which is found in all the CMIP6/QBOi models with varying degrees of magnitude. The HadGEM3 model is amongst the best in this respect, yet the bias is noteworthy still. The mention of the QBO biases is done in several parts of the manuscript including in this section referenced by the reviewer.

L416, have you ever mentioned TRMM in the text?

We have removed this reference to TRMM.

L420, 'observations' -¿ 'variables'

Done.

L422, revise. you speak about days, I understand the input data is monthly mean, so it this weighting already built in? Does the MOHC model have 360 day calendar?

This formulation was written for the most general case, which is the case of the computation of the composites of observational data. However, as the reviewer points out, in the case of the MOHC the calendar is 360-day which means there is no need to consider the "day" part of this formulation.

L435, I guess the 'i' subscript is redundant with one predictor? Same in Fig S3

Yes, in the simplest cases (Figs. S3 a, c) there is only one predictor. The equation has been rewritten and the plot title updated.

L440, State that summation is $j = 1...N$, as X_0 appears already

Done.

L446, Is there a stray A3?

Typo, the A3 reference has been removed.

L513, why uppercase? Corrected to lower case.

2. Reviewer 2

The paper seeks to understand the link between the tropical stratospheric QBO and variability elsewhere in the tropical atmosphere (and on ocean SST). The tools used include observational/reanalysis output since 1979, and several preindustrial control runs from various versions of the Met Office model. This is an interesting subject and the foundation of a paper that could eventually be publishable in WCD is clearly present. Many of the arguments are not convincing in their current form however, and while I don't think these critiques are insurmountable, addressing them will require some major rethinking and rewriting.

We thank this reviewer for their time reviewing our manuscript. We found your concerns and questions to be very useful to revise and improve our submission. Below we present a detailed point-by-point response to your concerns and questions. Your comments are shown in black and our responses in blue.

2.1. Major comments

2.1.1. My main criticism is that the authors argue there is a connection between the QBO and ENSO, but the evidence provided is not strong enough (and in this reviewer's opinion the authors' claims are actually incorrect). First, the observational period covered by this paper only begins in 1979, however high-quality radiosondes have tracked the QBO since 1953, and reliable information on the ENSO state is available even earlier. Studies that have used the entirety of the observational record have reached an opposite conclusion of that reached by the authors. Specifically, in the period before 1979, there were more easterly QBO events simultaneous with El Nino. This has been noted by at least three papers (Garfinkel and Hartmann 2007, Hu et al 2012, Domeisen et al 2019), none of which were cited in this paper. The net effect is that the observed connection between ENSO and the QBO is non-stationary, and (cherry-) picking a limited subset of the full observational record can lead to misleading (and erroneous) conclusions. Over the entirety of the observational record (at least until 2018, the last year considered by Domeisen et al 2019), the correlation was essentially zero.

Thank you for this comment. To address this comment we have analyzed the period before 1979 using the zonal winds at 70 hPa from the Freie Universität Berlin (FUB) radiosonde dataset, available at `https://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/index.html#access`. This dataset covers the 1953-2020 period using radiosondes launched at Canton Island, Gan/Maledive Islands and Singapore. Furthermore, a reconstructed index at 90 hPa, described in Brönnimann et al. (2007), was used to investigate an even longer period (1930-2020), this dataset is available at `https://climexp.knmi.nl/data/iqbo_90.dat`. Using these two new indices, the observational part of Figure 3 in the original manuscript was reproduced for different periods, shown in Figure 1 for detrended HadSST data.

Figure 1 confirms the suggestion of this reviewer that in the 1953-2020 period there is little-to-no difference (QBO W-E) in the equatorial Pacific due to the cancellation of a negative signal in 1953-1979 with a positive signal in 1979-2020. However, the longer period (1920-2021) shows large agreement with the shortest 1979-2020 period, except in the South Atlantic. A warmer central-eastern Pacific is observed

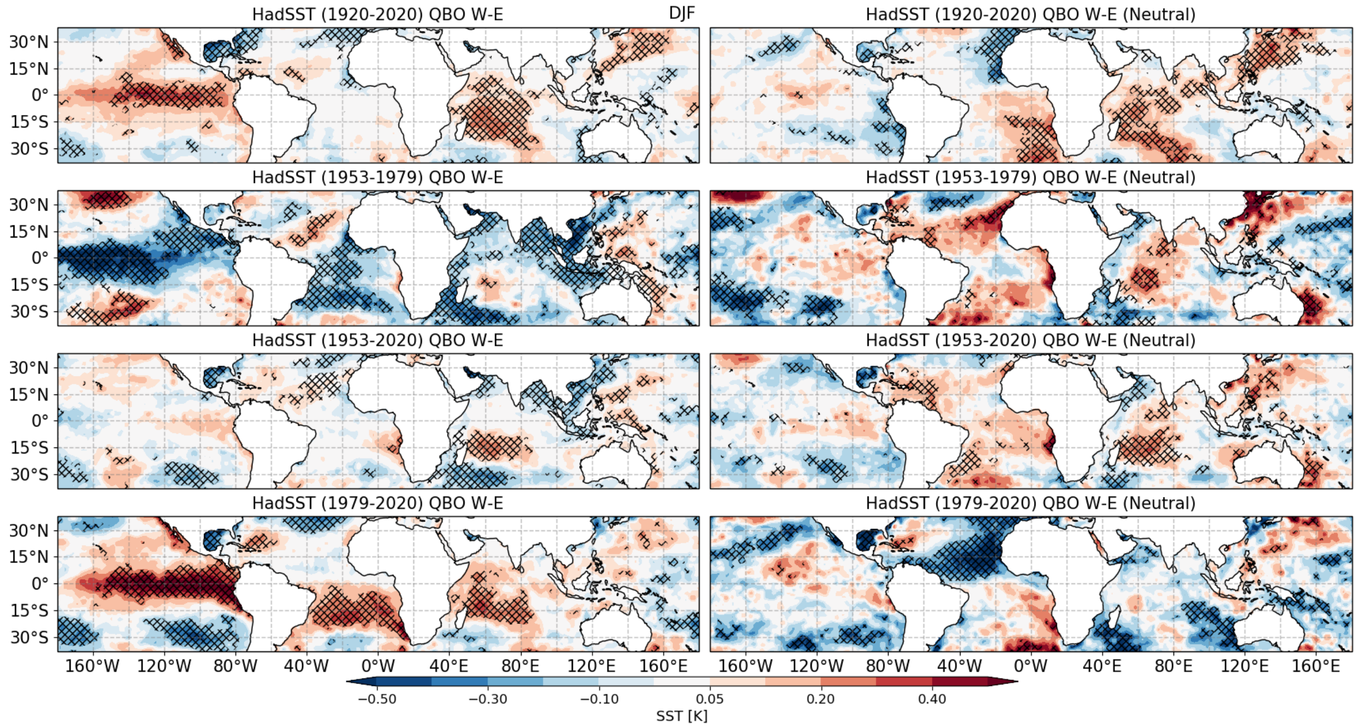

Figure 1: DJF QBO W-E mean SST difference (QBO W-E) in HadSST using different periods. The right column shows only Neutral ENSO months.

in both periods for QBO W-E differences. A warmer western Indian Ocean in QBOW compared to QBOE is observed in almost all panels.

The changes to the Pacific response are further investigated by plotting the time-series of the QBO W-E differences in the EN3.4 SST index (Figure 2) for model and observational data over different sliding windows. This comparison shows that, in the model, the EN3.4 and IOD responses exhibit decadal variability with the EN3.4 response turning negative in some short periods but being overall positive. In the observations, the results confirm the suggestion by the reviewer that in the 1960-1990 period the sign of the QBO-ENSO relationship was negative but became positive in recent decades. Interestingly, the longer record in the QBO reconstruction by Brönnimann et al. (2007) at 90 hPa matches well with the Singapore and ERA5 datasets. The reconstruction data suggests that earlier in the record, 1930-1960, the relationship was also positive.

The net impact of this new evidence over the manuscript is that the revised version addresses much more directly the non-stationarity of the ENSO-QBO relationship in both model and observational data. The SST composites (Figure 3 in the original manuscript) are now shown to emphasise the non-stationary nature of the relationship over the seasonal variability of the response.

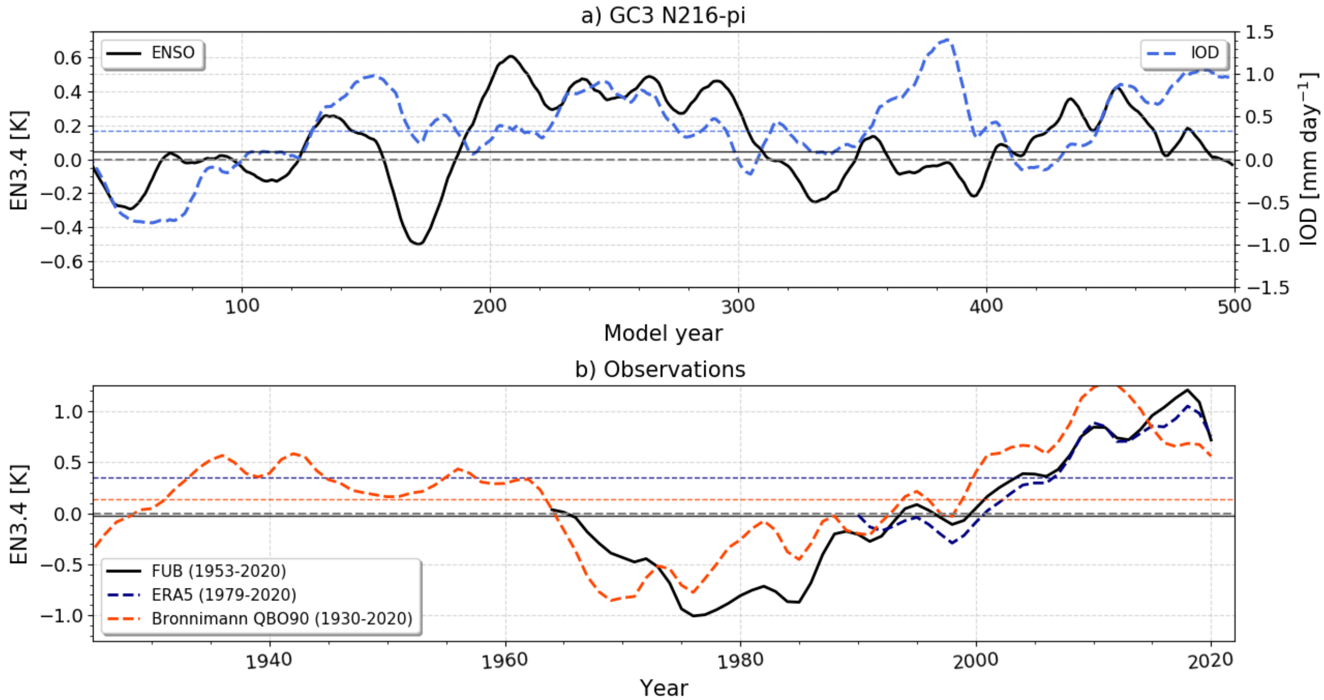

Figure 2: Time series of QBO W-E differences in the EN3.4 [K] in (a) GC3 N216-pi and (b) observations using the HadSST dataset and multiple indices for the QBO. In (a) the time-series is constructed by computing the W-E differences in sliding windows of 30 year periods, note that the IOD time-series is also plotted. In (b) the time-series is constructed by computing the 12-yr sliding average, so the first point of each time-series is plotted 12 years after the first observation.

2.2. *Second, the modeling evidence presented by the authors for a relationship between ENSO and the QBO is also misleading and perhaps wrong. The authors consider several different simulations from one model, however Rao et al 2020 (not cited) recently considered the connection between ENSO and the QBO in 17 different CMIP5/6 models. Rao et al found that some models simulated a connection of the same sign as that found in this paper. However other models simulated an opposite effect. Notably, the two MetOffice CMIP6 models considered by Rao et al 2020 had opposite responses (their Figure 11n and 11r). The multi-model mean effect was essentially null in Rao et al 2020. Thus, it is conceivable that the MetOffice models examined in this study do indeed simulate a connection between ENSO and the QBO, however this relationship does not appear to be generic, and future work is needed to unravel the causes of model disagreements.*

We understand this reviewer's concerns because the evidence presented by our manuscript contradicts the results by Rao et al. (2020) that used several models. However, we have investigated these differences and found that this is the result of (1) the use of a different index and (2) the use of a single ensemble member by Rao et al. (2020). Several indices have been tested in the literature but for tropical teleconnections most studies conclude or use the 70 hPa index because this level best diagnoses changes to the tropopause region (Gray et al., 2018, Serva et al., 2022).

Figure 3 illustrates both of these methodological impacts on the results by comparing the 100 hPa

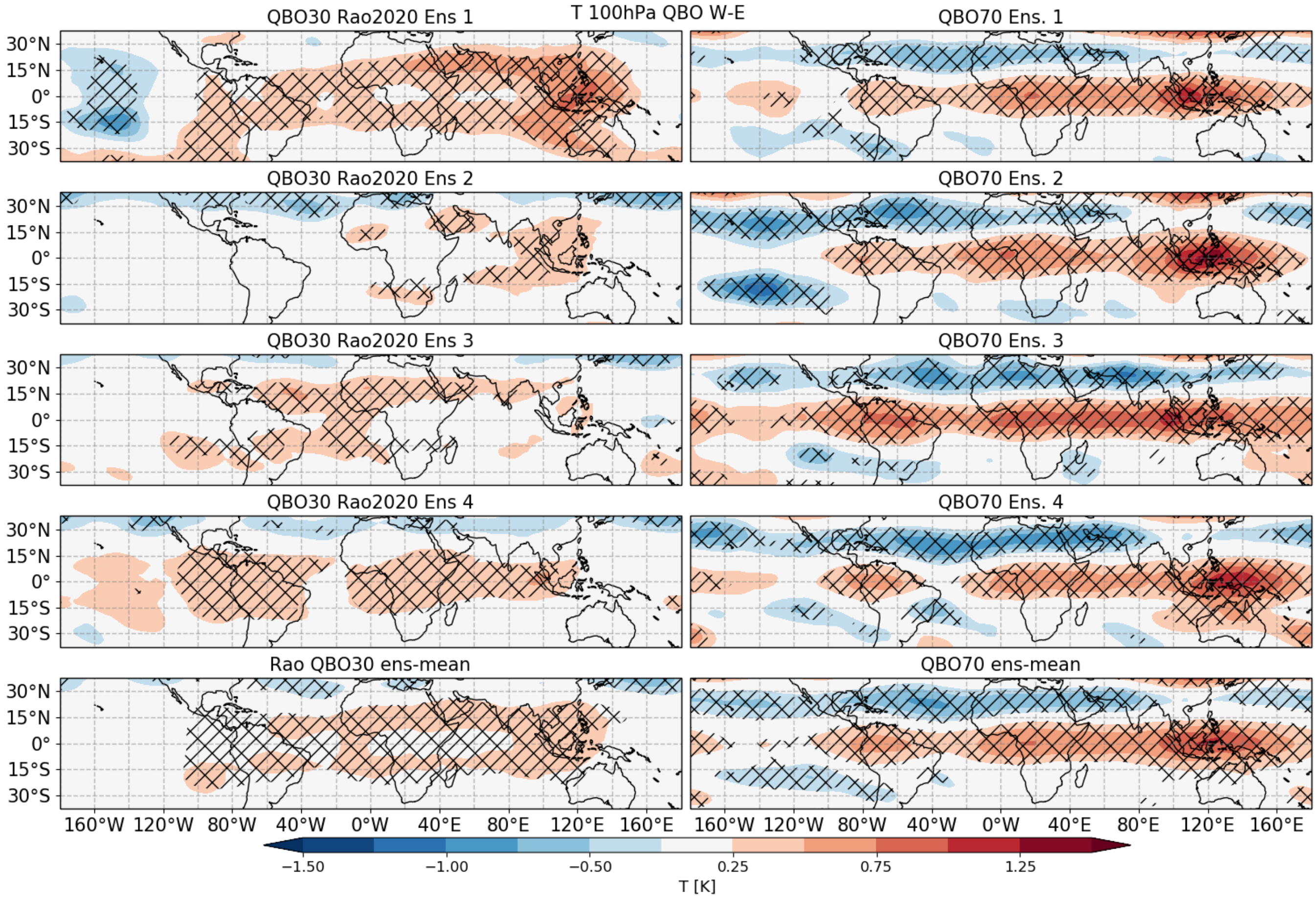

Figure 3: DJF mean air temperature (at 100 hPa) differences (QBO W-E) for GC3 N96-historical simulations using (left) a 30 hPa index defined as in Rao et al. (2020), and (right) a 70 hPa index as in the main manuscript for each individual ensemble member and (bottom) ensemble-mean. Hatching denotes statistically significant differences to the 95% confidence level.

temperature (near tropopause temperature) differences associated with QBO using each index and each ensemble member. Note that we use the index exactly as defined in Rao et al. (2020), i.e., averaging continuous DJF seasons and defining westerly and easterly phases when the 30hPa index is above or below 7.5 m s^{-1}. The diagnosed QBO impact over the tropopause temperatures is very weak, if any, using the 30 hPa index and is much better captured by the 70 hPa index. Secondly, there are notable differences between ensembles, likely because of the effect of historical forcing in these simulations.

Given that the literature suggests that the QBO influences the tropical troposphere through the effect of the QBO in the tropopause region, the 30 hPa is unlikely to capture any such tropical effect and will lead to misleading results. The 70 hPa index better captures the QBO influence in the tropopause region and is therefore better suited for the purpose of the submitted paper.

The reviewer also notes that, in Rao et al. (2020), UKESM and GC3 N96 historical simulations show opposite responses. This is also mainly due to the use of a single ensemble member and the choice of index. Figure 4 shows that the first ensemble member of UKESM-historical, used in Rao et al. (2020), shows a

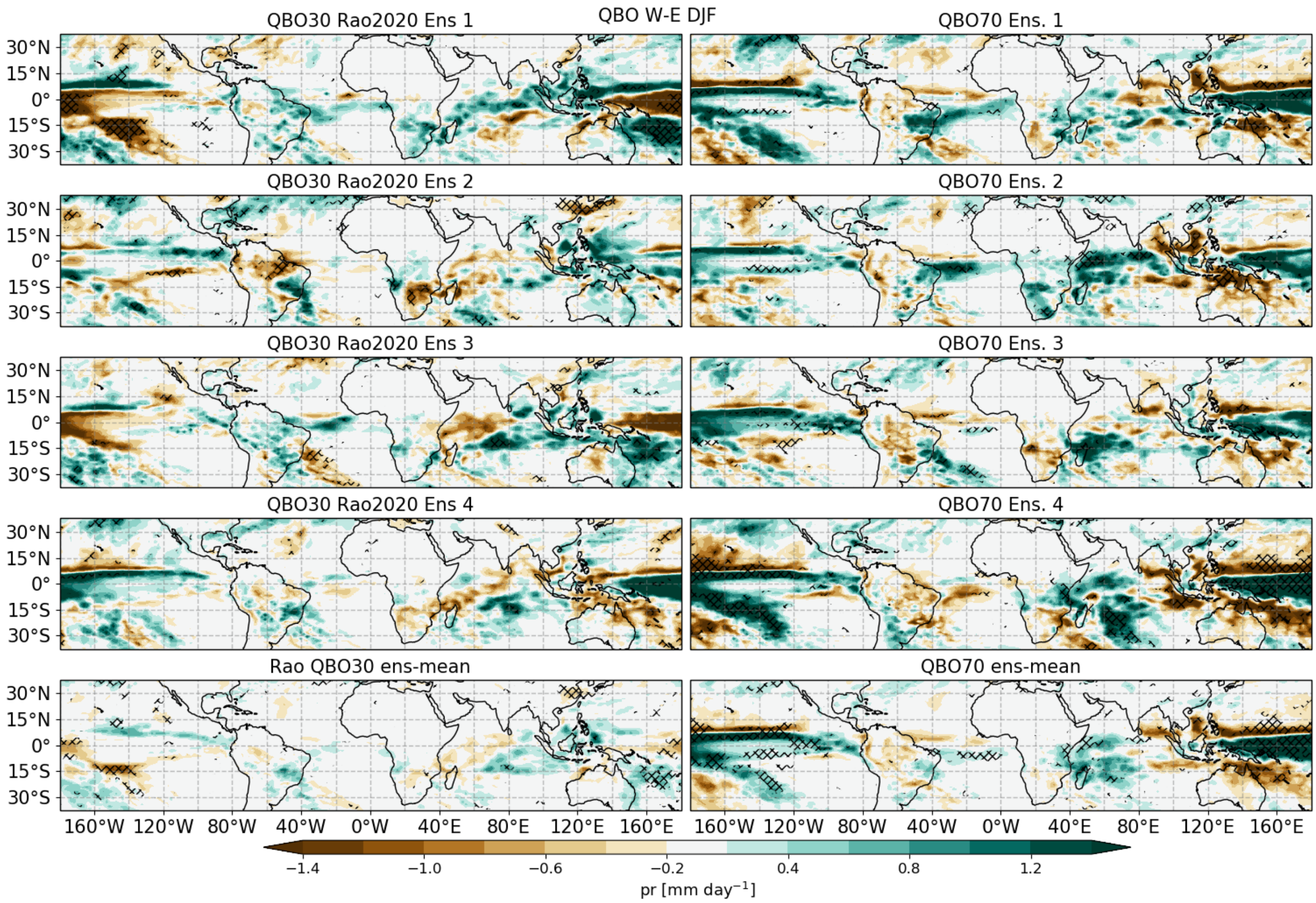

Figure 4: As in Figure 3 but for precipitation in UKESM historical simulations.

different response to the other ensemble members, according to the 30 hPa index. However, the 70 hPa index shows a consistent response across the ensemble members which agrees very well with the response diagnosed in the main manuscript. Some differences amongst members are also observed with the 70 hPa index which also suggests that multiple ensemble members are needed to diagnose this response in runs with external forcing. Note, for example, that Serva et al. (2022) also uses constant forcing simulations in their study. These results suggest that our approach of using a 70 hPa index and a long pre-industrial control simulation with no external forcing is better suited to understand tropical teleconnections. The implication from this evidence is that the results by Rao et al. (2020) are different to this manuscript because of the choice of index and the use of a single ensemble-member.

2.3. *The net effect of these criticisms is that I don't think it is particularly informative or meaningful to study the tropical atmospheric response to the QBO unless and until the ENSO signal has been regressed out. The authors indeed do perform such a regression, and they also additional examine ENSO neutral years only, which is great! But the analysis earlier and also later in the paper is suspect to this reviewer. Stated another way, the authors themselves note that the observed response to the QBO depends sensitively on whether neutral ENSO only is examined, so why even show the observed response before removing the ENSO influence?*

We appreciate and understand this reviewer's concern and suggestions given the available evidence to the reviewer at the time of their review. However, given the evidence provided in this response, as well as the evidence presented in Figures 10 and 11 of Serva et al. (2022), in addition to the evidence in the original manuscript, we contend that reporting and describing the QBO-related precipitation response prior to removing the ENSO influence is key to highlight a QBO-ENSO relationship in the model. This response is found in most QBOi models (Serva et al., 2022) and warrants further investigation because if the same precipitation pattern response in the equatorial Pacific is observed in several models and under different configurations with different forcings then this evidence points to a QBO-ENSO link that should not be ignored.

2.4. *My suggestion is to focus on the results where ENSO is "removed" much more or exclusively (as they do indeed contribute to the scientific discourse), and significantly shorten the rest of the paper. At the very least, the discussion of the figures 1, 2, 3, 7 needs to be rewritten.*

The results reported in the revised manuscript (Figs. 9-12) are shown for ENSO Neutral years as well as for all years, and in all cases the response does not change. We have chosen not to focus solely on the QBO response independent of ENSO but rather show the mean QBO W-E response (Fig. 1) which agrees with the findings of Serva et al. (2022) and which strongly suggests a link between the QBO and ENSO. Instead of ignoring these links suggested by both our Figure 1 and the multi-model finding of Serva et al. (2022), we have chosen to investigate and describe the nature of this relationship in more detail as the original manuscript intended but taking into consideration the non-stationary nature of the simulated and observed ENSO-QBO relationship (Figure 2). After characterising that this relationship is statistically significant and not straightforwardly explained by an upward control of the QBO by ENSO (sections 3.2 and 3.3 in the manuscript), we describe the impacts to monsoons, Walker circulation and ITCZ with and without the influence of ENSO.

2.5. Finally, Rao et al 2020 also consider the response of OLR and precip to the QBO with the ENSO signal regressed out, and find a wide range of responses across the models. Particularly perplexing to this reviewer is that Figure 7n/8n and 7r/8r of Rao et al consider the OLR and precip response in two different versions of the Met Office models, and find if anything opposite results. The present paper focuses mainly on the higher resolution runs which were not analyzed by Rao et al, however the authors should include in the supplemental material additional figures for the other model versions for most of the figures in the paper.

We have shown in a response above that these differences between our study and the results in Rao et al. (2020) can be accounted for by the choice of index and the single member approach of that study. In addition, we have added additional figures from the lower resolution runs matching Figures 2, 3, 7, 8, 10 to the Supplementary material, as suggested by this reviewer.

2.6. Specific comments

My general comments mentioned four very relevant papers that appear to have not been cited. Please add them as appropriate throughout the manuscript.

We have added the references suggested by the reviewer and the revised manuscripts cites them in several instances.

The authors attempt to remove an ENSO influence throughout by forming eQBO and wQBO composites during ENSO neutral years only. Note that this doesn't guarantee that the mean ENSO index during the wQBO and eQBO composite are actually identical. Can the authors compute the mean of the Nino3.4 index for these composites, in order to confirm that any ENSO influence is removed?

The length of the simulations is such that the EN3.4 index is 0.0 in the Neutral composites. Specifically for QBOW NN the mean index is 0.004 [K] and for QBOE NN -0.005 [K]. For the observed period of 1979-2019 the mean EN3.4 was -0.06 for QBOW NN and -0.01 [K].

An ENSO index can be removed also by linear regression, e.g., linearly regressing out variability associated with the Nino3.4 index, as done in figure 5. The authors seem to prefer to examine neutENSO conditions instead. There are pros and cons for both methods, and it would be worth noting in the text if results are different for either method of attempting to remove the ENSO influence.

In our view, the two approaches should render the same answer except for two factors. One is non-linear impacts associated with ENSO diversity, which may mean that removing the influence of ENSO by linear regression may not be the best approach. Second, Figure 8 in the main manuscript suggests that the QBO may be a modulator, a modest modulator to be precise, of ENSO impacts in a non-linear fashion. This means that the QBO influence may be different depending on the ENSO phase and this influence may only be observed by plots similar to Figure 8 and not by linear regression methodologies.

Line 55 add Rao et al

We have added Rao et al as a reference.

Line 69 the the

Corrected.

Figure 1, 4, and 8: This figure looks fairly different from figure 8a, 8n, and 8r of Rao et al. Particularly perplexing is that 8n and 8r of Rao et al, which focus on two different versions of MetOffice models, do not agree with each nor with any of the panels here as best as I can tell. There are certainly many methodological differences between the studies (whether/how ENSO is removed, the season analyzed, historical vs. PI control), but if the results are so sensitive to these choices then the overall effect may not be particularly robust.

We have shown that these differences are due to our better suited choice of index and type of simulation. In addition, Figure 4 has reproduced the results of Rao et al. (2020) using the 30 hPa index and the results of the main manuscript (which uses a pre-industrial control experiment) in the corresponding historical experiments. This result increases the confidence in our result, as it is found in both pre-industrial and historical experiments, while also accounting for the disagreement between UKESM and GC3 N96 in the study of Rao et al. (2020). While a 30 hPa level index may be the right index to diagnose extra-tropical impacts, the weight of the literature (Huesmann and Hitchman, 2001, Collimore et al., 2003, Gray et al., 2018, Hitchman et al., 2021, Serva et al., 2022), as well as Figure 3, suggest that influences in the tropical tropopause region associated with the QBO are best characterised by indices in the 50-70 hPa levels. Therefore, we disagree with this reviewer that our results are due to a large number of methodological sensitivities and these are simply due to the choice of index. For that reason, we have added a sentence in section 2.3 to emphasize the importance of the choice of index and indicate our reasoning to choose 70 hPa over 30 hPa index.

Line 184-185, 250-270 see my general comments about the ENSO-QBO relationship. These sentences are not representative of the entirety of the published literature or other runs of the model used in this paper.

Prior to the publication of Serva et al. (2022), we would have agreed with this reviewer that the entirety of the literature did not suggest a robust link between ENSO and the QBO. However, their Figure 11 strongly suggests a link because the majority of the QBOi models analysed in that study show the same response as found in this study, i.e., stronger (and southward displacement of) precipitation in the East Pacific ITCZ during DJF.

Figure 3: please use as much as possible of the 1953-2022 period for the observational composites. I expect the resulting figure to be rather different to what is shown here, which will necessitate a rewrite of the accompanying text.

Figure 1 in this document shows that the response in the 1953-2020 period is different to the 1979-2020, as suggested by the reviewer. However, a longer record (1920-2020) agrees best with the most recent period. Similarly, in several periods the model also shows a negative relationship just as in 1953-1979 in the observed record. For this reason, the revised manuscript addresses the non-stationary ENSO-QBO relationship and focuses less on the seasonal variability of the QBO footprint on SSTs, so the new Figure 3 shows the different SST differences in the different periods as well as the model simulation results.

Table 1: please add the other model versions to this table

We have added the data for UKESM-pi and GC3 N96-pi to the table.

Section 3.4: Garfinkel and Hartmann 2011 (already cited) discuss changes in convective precipitation and OLR over monsoon regions the ITCZ in response to the QBO with fixed SSTs. Note that Garfinkel and Hartmann 2011 also performed some targeted experiments in which the QBO profile nudged towards was modified (line 409).

We have included this reference in our discussion of monsoon regions and the ITCZ. However, note that because both the monsoon and ITCZ dynamics are strongly driven by SSTs, the coupled simulations used in this study are not directly comparable to the fixed SST results. The suggestion to note this paper in our discussion of targeted experiments has also been added in the revised manuscript.

Section 3.4: Hu et al 2012 discuss Walker circulation changes in response to the QBO. Please include in your discussion.

We have included this reference in our discussion, as well as in the introduction of the main manuscript.

Figure 11 caption discusses panels g and h, which don't appear to exist.

Corrected.

3. Reviewer 3

The paper is in general well written and the subject is interesting. However, many aspects of the methodology is not explained in enough details. I also find that the authors sometimes over-interpret the differences they find in the composites. I find that the paper needs some major improvements before it can be accepted for publication.

We thank the reviewer for their time reviewing our manuscript, particularly for their suggestion of improving our description of the methodology. Below we present a detailed response to your main concerns and we describe how we addressed your specific concerns. Your comments are shown in black and our responses in blue.

3.1. Specific comments

l110. I guess this description also is valid for the models and not just ERA5.

Yes, this is the case. We have made this clear in the revised of the manuscript.

l130: I don't understand this weighting. Why is this important and how important is it? In particular the weighting with the number of days in each month cannot be important. The annual-mean composites seem to long-term means.

Thank you for this comment. The original manuscript detailed the weighting to obtain seasonal means because in contrast to several studies that use a seasonal-mean index and average entire continuous seasons (e.g. Rao et al., 2020), our composites were constructed from monthly-mean values which may be unevenly sampling months per QBO or ENSO phase. For example, a MAM composite for QBOW may have more March samples than May samples compared to QBOE due to the QBO seasonal phase-locking. However, the description of this weighting was clearly confusing, and the manuscript may have overstated how important this weighting is so we have decided to remove the appendix section and simply state that the composites were calculated from monthly means by weighting the number of individual months in a season for a composite to ensure all months in a season contribute equally to the seasonal-mean.

l135: I don't think I understand these counts. For example, in observations you have 62 QBOW El nino months in a 40 years period while you have 376 months in 500 years for the model. But 62/40 is very different from 376/500. Does the model behave differently from observations or have I misunderstood something?

Thank you for this good observation. This is mostly because the frequency of QBO phases is lower in the model due to the weak amplitude bias of the QBO in most models, including the UK Met Office model (Bushell et al., 2020). This means our counts of QBOW and QBOE are less frequent for the model than for observations lowering the ratio of QBOW El Niño months to the total amount of months. In the revised manuscript we have noted this lower ratio and provided this explanation.

l140: More details should be given here. Is it individual years or months that are resampled? The time-scale of both the QBO and the ENSO are much longer and this should be reflected in the resampling procedure. If this is not done, the significance will be overestimated.

Thank you for pointing out this issue with the description of the significance testing. The revised manuscript clarifies how the significance testing is done for composite differences in (a) observations, (b) the simulations and additionally, (c) for the regression analysis. For (a), the bootstrapping method was used by creating composites from randomly selected months in way that keeps the structure of the target composite, i.e., the random composite has to match the size and structure of an observed composite. For example, for a MAM QBO W-E significance test, random March points are drawn from the entire observed record to match the sample size of March in QBO W and E, separately. The same is done for April and May. Then, the seasonal average is computed for each random sample. The difference between the mean of both random samples is computed (1 iteration) and the process is repeated (10,000 times) to generate a distribution. Instead of the bootstrapping method, two sided t-tests were also used for the observational composite differences and no appreciable differences were found.

l165: The signal in the model seems weaker and more confined than in observations.

Agreed, we now have noted this in the text.

Figure 3: I would say that the signal is in general weaker in the model than in observations and that there are considerable differences between model and observations.

The revised manuscript has a new Figure 3, but the discussion in text now considers this comment by this reviewer.

l224: It is not correct that multiple linear regression assumes that the independent variables are orthogonal. But they cannot be linearly dependent.

We have have corrected this statement as suggested.

Figure 6: The authors should briefly mention what the box plot shows: median, std. dev. etc.

Done.

l240-250: How is the statistical significance of the differences in Fig. 6 estimated? The spread seems very large.

We have specified that the statistical significance in all cases is estimated using Welch t-tests for the model simulations and boostrapping with replacement for the observations.

l249: Christiansen et al. 2016 (doi:10.1002/2016GL070751) suggests that strong warm ENSO events change the phase of the QBO. Is there evidence for this in the model?

We have added a sentence in this section to note that in contrast to the model used in Christiansen et al. (2016), the results in the Met Office model do not indicate a phase shift of the QBO after strong ENSO events.

Table 1 Why are the errors smaller for the observations than for the model? Thank you for pointing this out. The errors were wrongly calculated for the observations. These errors were being calculated exactly as for the models, i.e., by bootstrapping with replacement the population repeatedly into 36 yr samples which means that for observations this would render very small errors that are not representative of the observed statistical uncertainty. In the revised manuscript, the errors are calculated using a different approach to better characterise the observed spread which resulted in larger errors than in the original manuscript and comparable to the errors in the models.

l250-255: I don't understand this paragraph. Why do you look at the pdfs? The K-S test tests if the pdfs are different, and not necessarily if the averages are different.

Yes, the K-S test does not say anything about the mean values,, but because we constructed PDFs from bootstrapping, we are able to also estimate if the PDFs are different. In this case, this means that both K-S and standard t-tests were used and model QBO W-E differences were statistically different according to both tests. The revised manuscript now notes that both tests render the same result. The rationale of investigating the PDFs is that the KS is more suitable as there is no assumption about the shape of the distribution and does not only test differences in the mean but the overall shape of the distributions.

l260: Does this significance refer to the * in Fig. 7? There are very few *.

Yes, this significance refers to the few * in the Figure. We have found evidence that these signals are non-stationary (see Fig. 2 in this document) which may explain the relative scarce differences that pass the significance testing. In the revised manuscript, we emphasize that both QBO-IOD and QBO-ENSO relationships are strongly seasonally dependent (in the model).

l268: They are very often opposite in sign. Can you say that the numerical values are different within the error-bars?

They are significantly different in the MAM and SON seasons for the EN3.4 and IOD indices, respectively, which agrees with the results of the previous sections. However, as the reviewer rightly points out, the differences are not significant in other seasons and for that reason, the revised manuscript emphasizes that both ENSO and IOD relationships with ENSO are strongly seasonally dependent and in the discussion we also note that model features such as the seasonality of ENSO may be key to determine the specific seasons where the QBO influence is observed for each model.

Figure 8: The hatching is hard to see.

In the MAM season, the differences and the hatching are clearer, so the revised Figure more clearly shows the contrast of EN vs NN vs LN differences. Note that for the other two models, in DJF the differences are also clearer.

l275: The plots in Fig. 8 seem very similar to me. Are you sure the difference of differences are statistically significant?

Given that in GC3 N216-pi, the season that is observed to be more sensitive to the QBO phase is boreal spring, in addition to the fact that the strongest connections between ENSO and the QBO were also found in boreal spring (Figs. 2, 6), in the revised manuscript we use this season (MAM) to more clearly show the non-linearity discussed in the original manuscript. Note, however, that in the other models the corresponding plot is given in the supplementary material for DJF because in these models this is the season of strongest QBO-ENSO connections. The main point from this figure and analysis is that caution is warranted when removing the ENSO signal, as in some models in some seasons, part of the non-linear impacts of ENSO across the tropics may be explained by the QBO phase.

l370: What is the difference between the sentence 'When only ..' and the following sentence?

We have rewritten this sentence hoping the revised version makes our point more clear.

References

Brönnimann, S., Annis, J. L., Vogler, C. and Jones, P. D. (2007), 'Reconstructing the quasi-biennial oscillation back to the early 1900s', *Geophysical Research Letters* **34**(22).

Bushell, A. C., Anstey, J. A., Butchart, N., Kawatani, Y., Osprey, S. M., Richter, J. H., Serva, F., Braesicke, P., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Garcia, R. R., Gray, L. J., Hamilton, K., Kerzenmacher, T., Kim, Y.-H., Lott, F., McLandress, C., Naoe, H., Scinocca, J., Smith, A. K., Stockdale, T. N., Versick, S., Watanabe, S., Yoshida, K. and Yukimoto, S. (2020), 'Evaluation of the Quasi-Biennial Oscillation in global climate models for the SPARC QBO-initiative', *Quarterly Journal of the Royal Meteorological Society* pp. 1–31.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3765*

Bushell, A. C., Butchart, N., Derbyshire, S. H., Jackson, D. R., Shutts, G. J., Vosper, S. B. and Webster, S. (2015), 'Parameterized gravity wave momentum fluxes from sources related to convection and large-scale precipitation processes in a global atmosphere model', *Journal of the Atmospheric Sciences* **72**(11), 4349–4371.

Christiansen, B., Yang, S. and Madsen, M. S. (2016), 'Do strong warm ENSO events control the phase of the stratospheric QBO?', *Geophysical Research Letters* **43**(19), 10–489.

Collimore, C. C., Martin, D. W., Hitchman, M. H., Huesmann, A. and Waliser, D. E. (2003), 'On the relationship between the QBO and tropical deep convection', *Journal of climate* **16**(15), 2552–2568.

Dawson, A. (2016), 'Windspharm: A high-level library for global wind field computations using spherical harmonics', *Journal of Open Research Software* **4**(1).

García-Franco, J. L., Gray, L. J. and Osprey, S. (2020), 'The American monsoon system in HadGEM3 and UKESM1', *Weather and Climate Dynamics* **1**(2), 349–371.

Gray, L. J., Anstey, J. A., Kawatani, Y., Lu, H., Osprey, S. and Schenzinger, V. (2018), 'Surface impacts of the Quasi Biennial Oscillation', *Atmospheric Chemistry and Physics* **18**(11), 8227–8247.

Hitchman, M. H., Yoden, S., Haynes, P. H., Kumar, V. and Tegtmeier, S. (2021), 'An observational history of the direct influence of the stratospheric quasi-biennial oscillation on the tropical and subtropical upper troposphere and lower stratosphere', *Journal of the Meteorological Society of Japan. Ser. II* .

Huesmann, A. S. and Hitchman, M. H. (2001), ‘The stratospheric quasi-biennial oscillation in the ncep reanalyses: Climatological structures’, *Journal of Geophysical Research: Atmospheres* **106**(D11), 11859–11874.

Rao, J., Garfinkel, C. I. and White, I. P. (2020), ‘How does the quasi-biennial oscillation affect the boreal winter tropospheric circulation in cmip5/6 models?’, *Journal of Climate* **33**(20), 8975–8996.

Scaife, A., Butchart, N., Warner, C. and Swinbank, R. (2002), ‘Impact of a spectral gravity wave parameterization on the stratosphere in the met office unified model’, *Journal of the atmospheric sciences* **59**(9), 1473–1489.

Schenzinger, V., Osprey, S., Gray, L. and Butchart, N. (2017), ‘Defining metrics of the Quasi-Biennial oscillation in global climate models’, *Geoscientific Model Development* **10**(6).

Serva, F., Anstey, J. A., Bushell, A. C., Butchart, N., Cagnazzo, C., Gray, L., Kawatani, Y., Osprey, S. M., Richter, J. H. and Simpson, I. R. (2022), ‘The impact of the QBO on the region of the tropical tropopause in QBOi models: present-day simulations’, *Quaterly Journal of the Royal Meteorological Society* .

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., Bushell, A., Copsey, D., Earnshaw, P., Edwards, J., Gross, M., Hardiman, S., Harris, C., Heming, J., Klingaman, N., Levine, R., Manners, J., Martin, G., Milton, S., Mittermaier, M., Morcrette, C., Riddick, T., Roberts, M., Sanchez, C., Selwood, P., Stirling, A., Smith, C., Suri, D., Tennant, W., Vidale, P. L., Wilkinson, J., Willett, M., Woolnough, S. and Xavier, P. (2019), ‘The Met Office Unified Model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations’, *Geoscientific Model Development* **12**(5), 1909–1963.

Yu, B. and Zwiers, F. W. (2010), ‘Changes in equatorial atmospheric zonal circulations in recent decades’, *Geophysical Research Letters* **37**(5).