

Reviewer 1:

Novelty of the study:

Thus, the novelty of this work requires further clarification. The authors should explicitly state what is new or improved compared to previous studies. Earlier research has already shown that QBO teleconnections tend to be too weak in models. The present study largely revisits this conclusion using the recent availability of large ensembles from models with spontaneously generated QBOs.

Moreover, the discussion and summary rely heavily on findings from previous studies, with limited new insights. Overall, the manuscript concludes that the use of large ensembles provides a clearer picture of how well models simulate QBO teleconnections; however, this contribution needs to be more clearly articulated in terms of its originality and added value.

The key point of this paper is that the conclusion from many previous studies - QBO teleconnections are systematically too weak - is an oversimplification. This is stated clearly in the abstract and Discussion. More generally, the abstract highlights many aspects of this work which are clearly different from any previous paper:

- 1. The magnitudes of these teleconnections are statistically indistinguishable from those observed in two of the models; this is a notable improvement from previous work that analyzed small ensembles.**
- 2. These large-scale teleconnections lead to surface temperature and precipitation anomalies over the mid-latitude continents, including an impact on western North America surface temperature which appears to have not been noted before.**
- 3. Remarkably, one of the models simulates a connection between the QBO and the Madden Julian Oscillation that mimics observations, although it remains too weak. This is nonetheless notably better than many previous studies using climate models**
- 4. Finally, the LESFMIP simulations allow an exploration of external forcings impacting the magnitude of teleconnections. Among these experiments, greenhouse gas forcing is seen to significantly influence the subtropical wind horseshoe of the QBO.**

All of these issues are discussed in the Discussion section at length.

We have also added to the introduction:

“Here, we revisit these relationships using a factor of 20 to 100 more model output, while also discovering new surface impacts of the QBO that appear to not have been noted before Here, we revisit these relationships using a factor of 20 to

100 more model output and isolating how external forcings may modulate their strength.”

We have added to the Discussion section:

“These discrepancies stand out more clearly in our study than in any previous work due to the huge amount of data examined.”

“The downward arching wind response in the subtropical Pacific sector also strengthens after 1970 in the hist-GHG experiment, likely because the QBO's lower stratospheric MMC strengthens in response to a reduction of lower stratospheric static stability forced by increased GHGs.”

“Overall, the use of large ensembles has allowed us to develop a clearer picture of how well models simulate QBO teleconnections while revealing surface impacts over North America that appear to not have been noticed before”

Technical errors:

The current study requires technical corrections. The figures appear to have been prepared in a rather casual manner, particularly with respect to font sizes, consistent notation, titles, color bar levels, and axis labels. Also, there is random use of abbreviations in the text without introduce them. The authors should ensure consistency in notation and presentation throughout the manuscript.

Some of these issues have been highlighted in the specific comments; however, it has not been possible to identify all of them due to the large number of inconsistencies.

We have corrected the issues raised by the reviewer as described below, and also made similar edits to other figures.

Specific comments:

Second paragraph 2 (33-43). Also, emphasize the routes associated with these three distinct mechanisms.

For all three mechanisms, we briefly discuss the route by which the remote influence occurs. Since this paper is not about diagnosing mechanisms, we do

not think it is appropriate to provide a detailed description, especially since the paper is already long.

L77 Use the QBO E and QBO W instead of eQBO and wQBO .

We use eQBO and wQBO throughout the paper.

L118 remove “also”

We prefer to keep “also” in this sentence

L119 “NH”, Abbreviations should be defined at first use by providing the full term. Replace “(all but IPSL6)” with “ but except IPSL6”

corrected

L125 “the mean meridional circulation of the QBO” with “the QBO mean meridional circulation (MMC)”, and further use of this abbreviation.

Changed as suggested

L130 “on the other hand”. Move it to the beginning of the sentence.

Changed as suggested

L148 -149 In ERA5...., rewrite the sentence with more clarity.

Changed to “In ERA5, on the other hand, the response peaks in December and January, but is no longer significant in February.”

L 152, “larger uncertainty in observation”, please specify them briefly

We mean that the error bars on the observed response are very large in February. We have simplified this sentence to “Conversely, in March, the regression coefficients in HadGEM3 and in MIROC6 are significantly larger than those based on ERA5. In February, it is not possible to confidently identify a model bias when comparing with ERA5.”

L171 “ Northern Hemisphere” Please refer to the comment at line 119, where the abbreviation “NH” is used without first defining the full term. Insert a full stop after ‘Europe’ and begin a new sentence.”

Changed as suggested, though we prefer a comma after Europe. We have shortened the sentence to make it tighter.

L172 North Atlantic Oscillation, the abbreviation NAO has already been introduced.

Changed as suggested

L173 “previous work...” Include one or two most relevant citations instead of referring to the Introduction.

Changed as suggested

Corrections in figures:

Fig.1 Interchange the left and right sides of the y-axis, i.e., place the pressure scale on the left side and the height scale on the right side. Also, ensure a clear separation between the y-axis label and the unit text.”

We prefer leaving the y-axis as is. We have increased the separation between km and the y-axis label for figures 1-3 and 13

Set the colorbar scale at appropriate intervals to avoid crowding or spacing. In subsequent Fig. 3, the left panel shows only the extreme values on the colorbar, whereas the right panel includes intermediate values as well.

Changed as suggested for figures 1-3, and also figure 10 and 13

Colorbar scale unit is “m/s per 10 m/s” but for the supplement (Figure S5) “m/s”. Same inconsistency in other figures and their supplementary.

This colorbar scale unit is correct as is. Figure 1 is showing regression, while the supplement is showing a composite approach. No changes made.

Proper use of title. Text Uzaregress is confusing, do you mean “Uza regression”? Same for Tzaregress ?

Changed as suggested

In Fig.2 caption “x-es”?

Changed to “stippling”

Fig,4 Shift panel (a1)---.. at title level.

We prefer to keep it inside the panel

Revise 'Uza60N10hPa' to 'Uza 60N 10hPa'. The spacing used for longitude-latitude–pressure variables is inconsistent; please ensure a consistent format is maintained throughout the manuscript.

Changed as suggested, and also for figure 5

Main title should start capital letter i.e. “regression” to “Regression”.

Changed as suggested (also for figure 5). We have also made the font properties similar for the ERA5 row and for the LESFMIP rows.

Fig.6 The box size for the calculation can be extended over the entire significant region, rather than being restricted to a very small core region.

We prefer to focus on the core region in order to more closely match the region of significant anomalies in ERA5 (which is shifted to the west of that in the models). See Supplemental Figure S10. We have expanded the region 5degrees further east.

Fig. 10 Use the same colour code as in Figure 9

It is standard practice in IPCC reports and in journal articles in our field to use a different color scheme for precipitation than for other variables. No changes made.

S10 -S12 Display the x- and y-axis labels with adequate spacing for better readability (Figs. 6, 7, 9, 12 too)

We prefer to not include the longitude and latitude labels to keep the figure cleaner

The above suggestions are only indicative; please review all figures carefully to ensure consistency and improved presentation.

Reviewer 2:

Comments/suggestions by line number:

Fig. S1 caption: state the value of the blue & red line contours shown below 30 hPa.

We have added “The $\pm 2.5\text{m/s}$ isotach is included in blue and red below 30hPa”

136-137: If this refers to the positive (red) regression coefficients, is CNRM the exception (not IPSL)?

We are referring to the blue (negative) coefficients, as these track the QBO winds themselves and their downward horseshoe to the troposphere. This is now clarified.

Fig. 4 caption: from the panel d5 title it seems to be the same region as for the models (220-240E, 55-70N).

Thanks for catching this! corrected

157: insert " (any of December, January or February)" after "available data" to help clarify these are the available months for this test (if I've understood correctly).

Clarified to “months in **DJF from the available data”**

188-189: Uncertainty in the observed response is so large (confidence intervals in Fig. 4b5) that it may not be possible to say that the models' seasonality differs from that observed.

For HadGEM3 we can conclude there is a discrepancy. The huge amount of HadGEM3 output available allows us to conclude that this discrepancy is robust, and not just sampling variability, as the error bars don't overlap. We have added this to the text.

205: typo: "Scandinavia"

corrected

222: Weaker westerlies? Based on the negative values in the Pacific boxes in Fig. 6 or Fig. S11. I guess some plots show stronger westerlies on the poleward side of this feature, but it looks to me like the negative blob is more consistent across all plots. (Line 231 also mentions weaker westerlies.)

Thanks for catching this. We indeed meant weaker westerlies. Corrected.

269-270: Not sure what features are being highlighted as robust in South Asia or East Asia. For HadGEM and MIROC at least there seems enough agreement across experiments that these patterns represent a real response, but the two models don't look very similar to me in South and East Asia.

We have modified the projection for this figure to focus more tightly on the tropics and subtropics. There is enhanced precipitation in South and East Asia in the models, and this is the point we are trying to make. However the details of where the precipitation increases is model dependent.

284: Fig S10 shows ERA5. Fig S12 uses 1970-onward but for tas. Was a figure mistakenly left out of the Supplemental?

We intended to refer to Figure S11 and not S10. This has been corrected. Thanks for catching this.

286: warm anomaly --> cold anomaly

Corrected. Thanks for catching this.

287: Meant to refer to Fig. S12 here? (in addition to Fig 9)

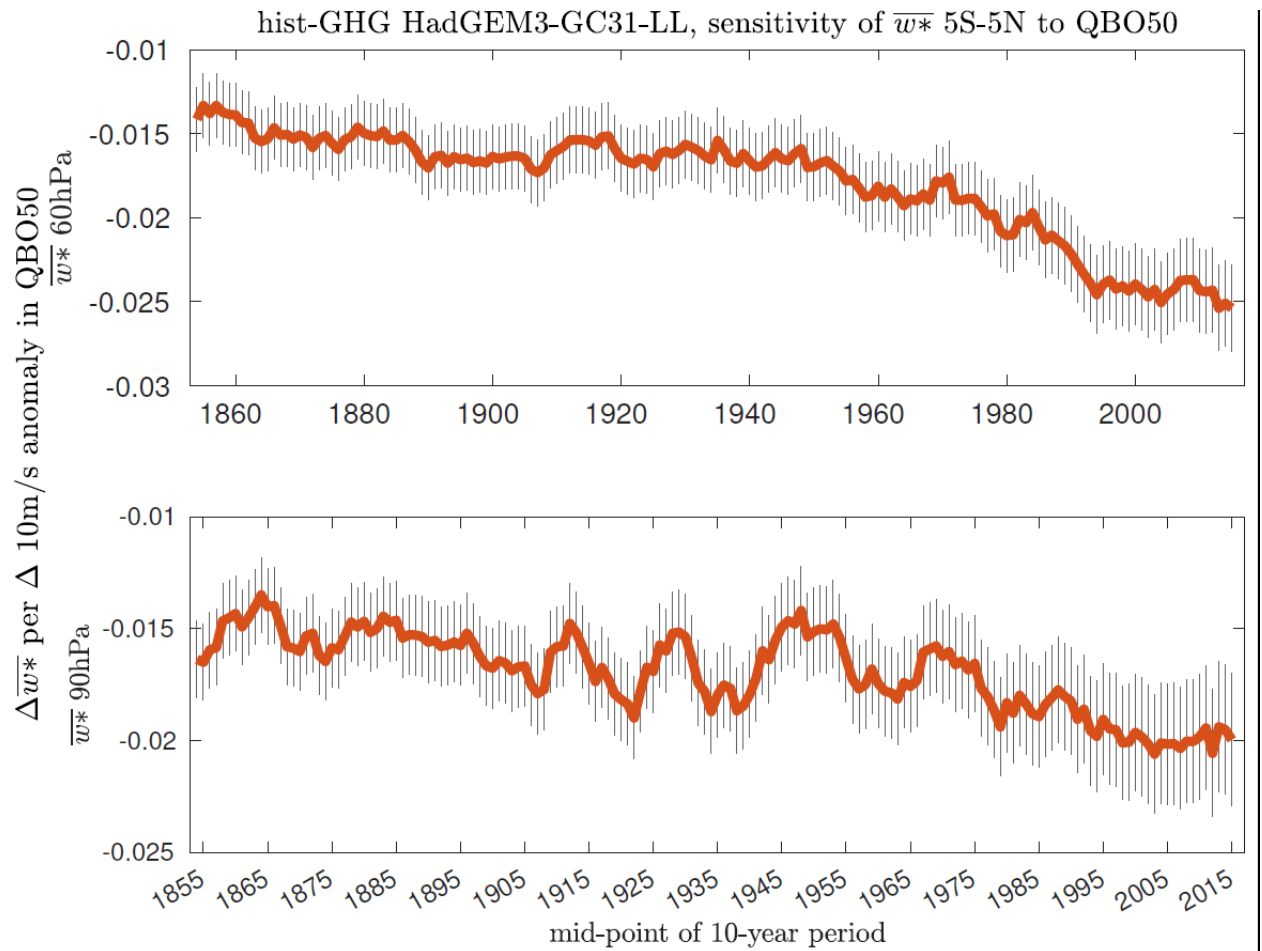
“and S12” has been added.

301: This argument seems to assume the vertical shear remains the same, but if the QBO weakens then the shear weakens. Is that decrease large enough to affect this conclusion? e.g., in eq. (2) could N and the T anomaly both decrease, leaving w^* about the same? Are there any runs of the same models with w^* available that could be used to check? (I don't imagine a large ensemble would be needed for this.)

Our proposed mechanism involves the weakening of background N , not the N associated with the QBO. The reduction in the background decadal mean N is due to changes in GHGs directly, and is a well understood thermodynamic/radiative process (Vallis 2015) in a warmed climate. This has been clarified.

The consequence of this is the w^* per m/s of QBO winds should increase. w^* is available for just one model - HadGEM3 -and for this model we now explicitly provide evidence for this effect. Below is the regression coefficient of equatorial w^* at 60hPa and at 90hPa per 10m/s change in QBO winds at 50hPa for rolling 10-year chunks of data in the hist-GHG experiment from HadGEM3. Note that the regression coefficient is always negative (i.e., eQBO leads to upwelling), but that the relationship intensifies in a warmed climate with higher GHGs as discussed in

the main text. The 95% uncertainty bounds on the regression coefficient as given by a Student-t test is indicated with a vertical line. Units are mm/sec per 10m/s. We have included this figure in the Supplementary Material.



312: "Prescribed ozone changes invigorate the QBO in both models": the QBO amplitude in wind and/or temperature increases? Is a reference to Butchart et al. 2023 appropriate? (<https://doi.org/10.1029/2023GL104401>)

Garfinkel et al 2025 showed the invigoration in QBO winds, and this is now clarified

We added a citation to Garfinkel et al and to Butchart et al 2023 here.

332: "leads to" --> "is consistent with"

Changed as suggested

375: Line 280 said that this increase is not statistically significant, so if this finding is mentioned in the conclusions I think at least that caveat should be mentioned.

“though not significantly” has been added

380-384: It seems to me this hinges on what exactly is meant by the term "teleconnection". If it refers to a response magnitude, e.g. size of a polar-vortex composite difference, then the models do underestimate these responses. If it refers to the mechanism(s) by which a QBO influence leads to a response, the regression results argue that the models do represent the mechanisms but underestimated QBO amplitude results in a weak response. If so, perhaps that point could be made more explicit here.

We agree this could be worded more precisely. We have changed “biases in teleconnections themselves” to “biases in the processes that connect QBO winds with remote regions”