

Response to Reviewer #2 (anonymous)

This paper evaluates how well twelve climate models from the QBOi phase 2 project reproduce tropospheric teleconnections associated with the QBO. Three experiments are analyzed: one with a freely evolving (internally generated) QBO, one in which the QBO is nudged toward observations, and one in which the QBO is suppressed. The results show that prescribing a realistic QBO improves some teleconnections, particularly the Holton-Tan effect. However, most teleconnections remain too weak compared to observations. The models generally underestimate key relationships, including the QBO's influence on the frequency and timing of SSWs, and they fail to reproduce the observed connection between the QBO and the NAO. The impact of the QBO on the North Pacific subtropical jet is also poorly represented. Overall, the findings suggest that the primary limitation is not simply the simulation of the QBO itself, but the models' ability to accurately represent the dynamical pathways through which QBO signals are transmitted from the stratosphere to the troposphere.

I find the paper to be well written and of clear interest to the readership of WCD. The analysis is relevant and thorough, and most of the results are insightful. I did not identify any major flaws in the study. I therefore recommend that the paper be accepted after the authors address the comments outlined below.

The authors would like to thank Reviewer 2 for kindly taking time to carefully review our manuscript. Responses to specific points are given below.

Main comments

I find the presented analysis somewhat unsatisfying in its current form. The central question—namely, why models struggle to reproduce QBO teleconnections—is not addressed. In particular, the paper does not fully exploit the potential of a multi-model framework by examining differences between individual models and assessing whether these differences are related to how each model represents the QBO itself or to other model characteristics (e.g., resolution).

Instead, the analysis focuses predominantly on multi-model means, which provides no insight into the underlying reasons for the relatively weak teleconnection signals. The manuscript frequently attributes discrepancies to “model biases,” but this term is vague. It would be very helpful if the authors could identify which specific biases (e.g., in QBO amplitude, vertical structure, wave forcing, or background circulation) are most relevant, and whether these can be linked to differences in teleconnection strength across models. I recognize that such an analysis is challenging and potentially beyond the scope of the present study, and I appreciate that this limitation is acknowledged as an important direction for future work.

The reviewer is correct we did just focus on the multi-model mean as now noted in the final paragraph of the Introduction. Inter-model differences were considered in our analysis though in general the findings were inconclusive. Therefore, we thought it best not to include that analysis in the current paper. Nonetheless, the paper in our opinion makes an important advance as for the first time we are able to identify the role QBO biases have in the poor representation of QBO teleconnections in models.

Lines: 198 - 200

Specific comments I find some of the explanations given in the introduction somewhat unclear or unphysical. For example:

81: The entire sentence starting with “This zero-wind line, ...”

Response: We have removed this sentence.

Lines: 85 - 87

94: “This results in ...”

Response: The sentence has been rewritten for clarity.

Lines: 98 - 101

96: “This modulates the concentration of ...”

Response: The sentence has been rewritten for clarity.

Lines: 101 - 103

99: “Regardless of the ...”

Response: We have removed this sentence.

Lines: 105 – 106.

105: “If the PNJ ...”

Response: We have removed this sentence.

Lines 111 – 112.

118: “For example ...”

Response: This sentence has been rewritten for clarity.

Lines 125 – 127.

83: “Under QBOE ...”: Please explain at what level the zero-wind line is measured and provide a reference for this statement.

Response: Explanation and reference added.

Lines: 87 – 90, 982 - 984

147: “Of relevance to this study ...”. I would make a new paragraph here and keep it together with the next paragraph. Also, check the language/grammar of this paragraph.

Response: Suggested change made.

Line: 162

175: A single-sentence paragraph is insufficient to motivate the study. This section should be expanded to better frame the scientific objectives and relevance of the work. For example, this could be done by better connecting it to the previous paragraphs.

Response: Single sentence paragraph removed, and the relevant content integrated into the next paragraph.

Lines: 192 – 194

187: Explain “AMIP-like” and/or provide a reference.

Response: The experiment description has been completely revised and now includes suitable reference for ‘AMIP’.

Line 207, 205 – 243, 1194 – 1195.

196: The current naming convention (e.g., “Exp1”, “Exp1-ObsQBO”, “Exp1-NoQBO”) is somewhat confusing as the repeated use of “Exp1” does not add meaningful distinction in the context of this paper. While there may be internal or project-specific reasons for this nomenclature, it reduces readability. It would improve clarity if the authors adopted simpler and more intuitive labels throughout the manuscript, such as “FreeQBO”, “ObsQBO”, and “NoQBO”.

Response: The experiment description has been completely revised.

We prefer to maintain the experiment short names for consistency with the companion papers. Exp1 is a generic term used through all QBOi phases to describe the base (control) experiment similar to the way CMIP uses the term “AMIP.” For continuity we would prefer to keep it. Also, it is not a requirement for a model to spontaneously generate a QBO in Exp1 hence replacing it with “FreeQBO” would be a misnomer.

231: Define boreal mid-winter. Also, in the caption of Fig. 2a you say early-mid boreal winter.

Response: “boreal mid winter” changed to “December and January”. Caption also revised as appropriate.

Lines: 264, 283 – 286.

235: What is meant with “climatological zonal-mean zonal wind”. A climatology over all months and years, or over each month individually?

Response: This has been clarified. It is the climatological monthly zonal-mean zonal wind.

Line: 270.

Figure 2: The caption should explain what is shown in the figure. In particular, the contour lines in the right panel. Also, how is early-mid boreal winter defined?

Response: The caption has been rewritten.

Lines: 283 - 287

Figure 3: The individual panels are much too small and hard to see. Also, what is the sample size of QBOE and QBOW?

Response: The sample sizes have been added to each panel in figures 3 and 4 and the figures has been expanded to fill the page width.

Lines: 301 – 307, 323

269: What are the criteria for the composite analysis? Presumably, each model has a somewhat different QBO climatology. So, do you use thresholds specific to each model, or do you calculate the composite QBOE-W response across all models regardless of the intermodel differences?

Response: The text has been revised to clarify the details for the WCD reader. Note in the nudged simulations the inter-model differences in the QBO are very small.

Lines: 311 – 312.

327: “These results might ...”: I do not agree with this interpretation. None of the model sub-samples reaches the observed PNJ response. So, why then should the relatively large observed response be due to sampling uncertainty?

Response: The reviewer makes a valid point. In light of this, and also reviewer 1’s suggestion we have modified the text.

Lines: 373 – 374.

329: “Anstey et al. ...”: Why do you expect that each experiment has a different variability? Could you explain this better?

Response: It is possible that the nudging directly impacts on the amount of interannual variability of the PNJ and in turn the modulation of that variability by the QBO (i.e. the HTE) is likely to be impacted. However, Anstey et al. shows there is very little change in the amount of interannual variability. The sentences have been revised in an attempt to improve clarity.

Lines: 376 – 378.

338: “This suggests that ...”. I do not entirely agree. The models do not faithfully reproduce the HT-effect. Also, some of the correlation could be due to having identical boundary conditions.

Response: All three experiments have the identical boundary conditions and forcings, so the comparison between ObsQBO and NoQBO is reasonable. The reviewer is correct that models do not faithfully reproduce the HT-effect, but the correlation is a measure of co-variability and is not affected by magnitude. We have softened the sentence to suggest that the observed interannual variance “may be accounted for” for rather than “is accounted for” and made a few associated changes.

Lines: 392 – 396.

345: I would also explain the PNJ correlations between the different experiments to shed light on the role of the boundary forcing in driving PNJ variability.

Response: The paragraph has been revised (please see above). An additional comment is included when the table is discussed highlighting that time-varying boundary conditions may play a role in driving PNJ variability.

Lines 428 – 429.

349: What do you mean with “This”? The “fact” from the previous sentence? Also, can you explain in more detail why this could be linked to the signal-to-noise paradox?

Response: The line beginning with ‘The fact’ has been removed to aid clarity. Added the pointers which are indicative of the signal-to-noise paradox, and changed ‘likely linked’ to a more accurate ‘indicative of’.

Lines: 402 – 407.

359: This is not a good explanation. Why should “other influences” like ENSO only affect the observations and not the simulations?

Response: We agree, it more likely due to models not getting the mechanism correct and/or having the incorrect seasonality (e.g., timing and frequencies of SSWs). The sentence has been revised to reflect that the models aren’t able to capture the weakening and reversal of the PNJ response seen in observations.

Lines: 415 – 418.

Table 2: Can you explain/comment on the rather large and significant correlations between the observed PNJ and the PNJ in Exp1 and Exp1-NoQBO? This must be due to the BCs.

Response: A comment has been inserted in the main text regarding the possibility of the time-varying boundary forcings playing a role in driving PNJ variability in November and December.

Lines: 428 – 429.

Figure 7: How are the SSWs defined? Also, the figure indicates that the SSW frequency in ERA5 is ~1. I find a SSW frequency which is much lower, more like 0.6-0.7, depending on the time period. Lastly, the DJF U1060 strength in the ERA5 is ~30 m/s, and not ~23 m/s.

Response: Thank you for highlighting the need to specify the SSW definition in this manuscript. Text has been added to explain the definition and alludes to the increase in SSW events compared to the 60°N 10 hPa definition (e.g. Baldwin et al. 2021).

Thank you, also, for your careful analysis of Figure 7. The ERA5 strength in the figure has been corrected. The plot now shows results for Exp1-ObsQBO (which yields similar correlations as for Exp1 and Exp1-NoQBO). As for the SSW frequency for ERA5: as noted in the text we are using a different, somewhat more flexible SSW definition than the traditional one based on 10hPa 60°N zonal wind reversal, which naturally increases our SSW count, but also accounts for different model biases in the vortex.

Lines: 440 – 444, 487

409: “If we ...”. Please provide a short explanation why doing this might be useful.

Response: Since Figure 8 has been significantly modified so that it does not show the ENSO-filtered results (in response to a comment by Reviewer 4), this part of the text has also been changed, and focuses on the Holton-Tan ensemble results instead, including a short explanation.

Lines: 493 – 495, 504

Figure 8: The figure is not very clear. Perhaps, to reduce the amount of information, could you just show differences in SSW-frequency between QBOE and QBOW? Also, perhaps use different signatures for the different models. This would allow identifying whether some models reproduce the observations better than others. Also, why is El Nino not shown, and La Nina shows only SVM?M?

Response: In response to comments by yourself and Reviewer 4, Figure 8 has been significantly reworked to *only* show the QBOE-QBOW differences in SSW frequency. The individual experiments are now indicated with markers, the same as used in Figure 7. The ENSO-filtered results have also been removed leaving only the full multi-model and Holton-Tan ensembles.

We mention in the text that filtering by ENSO phase was performed but that it did not produce significant results due to the very large uncertainty introduced by sampling such a small subset of the 1979 – 2020 boreal winters.

Lines: 521 – 523, 499 – 503.

546: It is interesting that the observed and simulated latitudinal position of the STJ agree. Is this true when taking the zonal mean over the entire globe, or only over the North-Pacific? If this is for the entire globe, then perhaps calculate the observed and simulated STJ positions over the North-Pacific, and if there is disagreement, this may change your interpretation of Figure 12.

Response: The latitude of the climatological STJ in the manuscript is for the North Pacific. The text has been updated to clarify this point.

Lines: 656

Figure 12: Could you use the same experiment names as in the paper?

Response: We have replaced Figure 12 with a QBOE-W version with the correct experiment names.

Lines: 676

Figure 12 caption: “The QBOW and QBOE groups are each deseasonalized ...”. It is not exactly clear what is done and what is meant here. I assume the zonal wind, but, model-by-model, or for the multi-model mean? Or did you deseasonalize the u-wind used to define the QBOW and QBOE groups?

Response: We have clarified what has been done in the manuscript by adding the following explanation to the caption of figure 12, ‘The zonal wind for each ensemble member (within a given model and experiment) is deseasonalized to produce anomalies. These anomalies are then composited by QBOE and QBOW phases before subtraction.’

Lines: 679 – 681.

Figure 12 caption: Does the vertical dashed line show the latitude of the climatological (October-March?) STJ over the Pacific (?) seen in ERA5? Maybe clarify this.

Response: Thank you for highlighting the need for clarity. The latitude of the climatological STJ is calculated from October-to-March across all years over the Pacific in ERA5. The caption for Figure 12 has been updated accordingly.

Lines: 688.

579: Exp1 essentially shows no wind response that would be consistent with the obs. So, “weaker tropical and subtropical wind response” is an overstatement.

Response: We understand the potential for confusion that arises from imprecise wording. The text has been tightened up to state that there are weaker tropical and subtropical responses for individual models leading to no wind response in the ensemble mean.

Lines: 693 – 697.

612: What is “tropical (Nino3.4) precipitation”?

Response: This forms part of the text associated with Figure 14 which has now been removed following the suggestion of yourself and reviewer 4.

Line: 732.

Figure 14 and its discussion: This figure is confusing and difficult to understand. In my opinion, the figure shows that there is no relationship as the correlations depend on single outlier models. How significant are these correlations? In Fig. 14c, it seems the positive correlation would become negative if model EMS4 would be excluded. I strongly recommend excluding Figure 14 and its discussion from the manuscript. The paper is already relatively long.

Response: We agree with the suggestion and have removed Figure 14 and associated text.

Lines: 724 – 746.

635: It is stated that the sensitivity of QBO teleconnections to model biases in the equatorial zonal-mean winds has been investigated. This appears to be somewhat overstated, as the paper does not clearly identify or quantify the relevant model biases. It would be helpful if the authors clarified which specific biases are being referred to and how they are diagnosed.

Response: Thank you for highlighting this omission. The authors’ involvement with this paper and also Anstey et al., (2026) have resulted in this oversight. The bias-correction alluded to in this paper is reducing biases in the amplitudes of the vertical QBO zonal wind structure.

One of the main outcomes of QBOi phase 1 was that models tend to show similar biases in the QBO zonal-mean zonal winds which includes an underestimated amplitude at the lower altitudes of the QBO around 50 hPa. In QBOi phase 2 the Exp1-ObsQBO experiment relaxes the model zonal winds in the tropical stratosphere toward reanalysis winds, thus imposing a realistic (unbiased) QBO (Anstey et al., 2026). I refer the reviewer to the contents of Anstey et al. (2026) including Figure 4.

Clarifications have been added to the Introduction, Section 2 . The line in the conclusion already mentioned that the model biases in the equatorial stratospheric zonal mean winds has been investigated. This line has been rewritten to more concisely reflect the investigations of this manuscript.

Lines: 175 – 177, 198 – 199, 256, 755 – 757.

640: “... exhibited a negative PNJ anomaly ...”. You need to be more specific. This is for QBOE-QBOW conditions, I assume.

Response: Thank you. More specific wording has been inserted.

Line: 761.

643: Again, you mention very unspecific model biases.

Response: Thank you. More specific wording has been inserted.

Lines: 764 – 765.

Technical comments

84: "... lies at approximately ..."

Removed in response to another comment.

Line: 89.

125: "... on either side ..."

Fixed

Line: 132.

234: "through to" ?

Fixed

Line: 267.

302: "observed"

Fixed

Line: 348.

378: "winter frequency"

Fixed

Line: 447.

400: during > between

Changed to 'for' as this relates to the QBOE-W frequency difference not the difference between Exp1-ObsQBO and Exp1. Figure 8 has been modified to show QBOE-W difference in SSW frequencies.

Line: 478.

477: Please check this sentence, in particular the “consistent with the model’s”. Which model do you refer to?

Response: we have corrected the sentence by removing “model’s” and replacing with “weak MME”.

Line: 577 – 578.

539: “..., on either side ...”

Fixed

Line 648.