Response to Reviewer #3 (Prof. Lesley Gray)

This is a very useful paper documenting the impact of correcting biases in model representation of the lower stratospheric QBO on teleconnections to the extra-tropics and troposphere. The main conclusion of the paper is that several aspects of the teleconnections are improved by correcting the QBO bias but there remain many significant aspects that are still not reproduced by the models. The paper is therefore useful in highlighting areas that require further investigation and potential model improvements.

The authors thank Professor Gray for these positive remarks.

I believe the paper is suitable for publication but it requires fairly extensive (but hopefully not onerous) revision.

Major comments.

1. The title suggests only extra-tropical teleconnections are examined but there are some interesting results of teleconnections with the sub-tropical jet and tropical rainfall, so I suggest a slight revision of the title to reflect this.

Response: The subtropical jet is generally considered to be at ~30°N-35°N which is in the extra-tropics. There is also a companion paper analysing tropical teleconnections. Therefore, to avoid confusion we prefer to keep the title quite simple and general, leaving the abstract to provide additional detail.

2. With regard to figure 2, I think it would be useful for the reader to also see the equivalent plot for the multi-model Exp1 results. The text states that more models achieve the H-T effect in Exp1-ObsQBO but it would be good to show this in fig 2.

Response: Fig 2 shows the ensemble mean and is only viable because all ensemble members for Exp1-ObsQBO have near identical nudged QBOs. In Exp1 the QBO is part of the internal variability and would be removed or significantly reduced in the ensemble mean.

3. Figures 2 and 3 clearly show a reversal of the H-T effect in late winter (Feb/March) in the ERA5 data which is not reproduced by the multi-model mean in figure 4. This needs mentioning / discussing both here and later (line 354) when you discuss table 2 (and probably also when you discuss the failure of the models to reproduce the timing of SSWs), including references to earlier studies that have examined this reversal. Based on figure 3, it also seems surprising to use the DJF comparison (lines 325-327) to compare ERA5 with the multi-model mean because the ERA5 data have strong negative PNJ values in Dec/Jan counterbalanced by very weak (possibly positive) values in February, whereas the MME clearly shows a weak negative value in all three months. This should perhaps be noted in the text. A better comparison would be to compare DJ and FM averages, which I suspect would more clearly highlight the ERA5/MME differences.

Response: The reversal in sign response in ERA5 for March is now noted. We have included a reference to the paper by Lu et al. (2020) in Section 3, a paper which was already referenced elsewhere in the manuscript. We agree with the reviewer that the differences between ERA5 and MME are likely due to seasonality (e.g. timing of SSWs) therefore it is important to check that after removing the differences in seasonality with a seasonal mean (e.g. DJF) if ERA5 and the MME agree or disagree. The two-month average doesn't eliminate differences in seasonality hence it is not clear what it offers in addition to the monthly means.

Lines: 290, 297 – 298, 315, 332 – 333, 415 – 418, 769.

4. The analysis finds that the QBO influence on the NAO is weak and mostly insignificant in the models. Did you explore whether the results were sensitive to the definition of the NAO? It would be interesting to see maps of surface pressure correlation (as e.g. in Gray et al. 2018) to see if there is any correlation at all in the Atlantic. It may be that a signal is evident in the subtropics but absent at higher latitudes because of the higher background variability there.

Response: Given the relatively weak extra-tropical stratospheric response to the QBO in the MME we didn't consider it appropriate to devote too much space to the NAO response. As noted in our response to Reviewer 1, the lack of MME NAO signal is most likely not because of the chosen definition of the NAO but more the fact that the models show different seasonality of the QBO-PNJ coupling and its propagation to the surface. We looked at the two-station NAO index (Azores, Iceland) and found that the results were consistent. The two-station equivalent of Figure 11 has been included as Figure A4, and the text has been modified accordingly. A detailed analysis of the NAO response could stretch to a whole paper, including correlation maps, which is very much encouraged, but is beyond the scope of this manuscript.

Lines: 600 – 601, 620 – 621, 867 – 869.

5. All figure captions need to be re-visited and revised where appropriate. Figure captions should provide information on what is plotted. Currently, the majority include interpretive statements that are more appropriate in the main text.

Response: The interpretive statements were included to allow the reader to gain a quick understanding from the figures-alone without having to refer to the main text. These have been removed as requested.

Figure captions: 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13.

Minor comments

Line 51: 'influence the latitudinal profile …' – in what way? In good agreement with obs?

Response: Revised to "improve its influence on……"

Line: 53

Line 64: do you mean 1 hPa here? Surely this level is dominated by the SAO?

Response: Changed to "around the stratopause," consistent with the Anstey et al (2022) QBO review.

Line: 68.

Line 72: worth adding also that the QBO influences tropical precip, also less well understood, given that you mention it in the abstract.

Response: The influence of the QBO on tropical precipitation is now mentioned in the appropriate place further on in the introduction in response to a comment by reviewer 1.

Lines: 134 – 141.

Line 84: typo - 'it lies AT approximately 6 degrees N'. Also, how can the QBO be westerly if the zero wind line is in the N hemisphere? Do you mean 6 degrees S?

Response: The phrase has been removed from the revised text.

Line: 85 – 89.

Line 107: 'the frequency of occurrence for SSWs …. For QBOE winters …..' Which months do you use to define 'winter'? DJF? Which months you use matters, given that SSWs tend to occur later in QBOW years.

Response: This is just reporting results from Baldwin et al. (2021) and Gray et al. (2024) and the WCD reader is referred to those papers for the full details.

Line 115: I suggest replacing 'the slowly varying QBO might improve NAO predictability' by 'improved model representation of the slowly varying QBO and its teleconnections might improve NAO predictability'.

Response: Changed as suggested.

Lines: 121 – 122.

Line 147: I misunderstood this text at first, and didn't realise that the 'two sets of experiments ….' referred to the Anstey et al. (2022b) study and not to this current one. It would maybe help if you started a new paragraph at 'Of relevance….' and merged it with the following paragraph which is still discussing the Anstey paper.

Response: Revised as suggested.

Lines: 161 – 162.

Line 155: when you say ‘ensemble-mean’ here, do you mean model-mean too?

Response: No, we mean ensemble-mean.

Line 175: this paragraph was clearly written before the addition of analysis of the QBO impact on tropical rainfall and the subtropical jet. You also need an extra paragraph somewhere in the introduction to introduce the tropical rainfall / STJ issues and mention relevant previous studies.

Response: In response to reviewer 1 we have included additional text relating to QBO-related tropical precipitation anomalies earlier in the introduction. The paragraph you mention has now been removed, the essence of which has been incorporated in the following paragraph starting ‘Section 2’.

Lines: 134 – 141, 196 – 202.

Line 211: ‘suggested adopting ….’ – this is too vague. Did all models follow this suggestion?

Response: This whole paragraph has been removed in light of Anstey et al. (2026) becoming available in the public domain. Additional explanation has been moved to the start of section 2.

Lines: 205+, 237 – 243.

Line 230: I think you need to say (in the text) how the PNJ is defined.

Response: The PNJ definition has been included.

Lines: 440 – 444.

Line 249: the figure caption is meant to explain what is plotted in the figure. It should not be a statement of interpretation.

Response: Caption revised as suggested.

Line: 283.

Line 252: I think you need another couple of sentences at the end of section 2 that describes how you derived the multi-model mean and also the statistical significance estimates. Also, perhaps how you defined the NAO and why you chose the method you used (other studies have used EOFs and/or pressure differences between Iceland / Azores); presumably you used your NAO method because it was easiest to extract from the model data?

Response: How the multi-model mean and statistical significance are derived is given in Section 3. The NAO definition has been moved from the caption in Figure 10 to Section 5 in response to the previous request. The broader area definition of NAO can help to capture

model NAO better in general, given that the centres of action in different models will differ. We also looked at the QBOE-W results using a two-station NAO index and now include those results in Figure A4 – they are consistent with the original results. Using an EOF based definition would be better but that is difficult to do as a multi-model mean.

Lines: 292 – 294, 583 – 584, 600 – 601, 620 – 621, 867 – 869.

Line 262: see previous comment about caption 2. The first sentence of this caption is an interpretation rather than description of what is plotted. Also, the text describing how the significance is derived should be in the main text of section 2.

Response: The caption of Figure 3 has been revised. The derivation of statistical significance has been moved to where it first occurs in the text.

Lines: 292 – 294, 302 – 307.

Line 271: the text 'of -2ms-1 December-March' does not make sense.

Response: Changed to "of -2ms-1 over the four months December to March"

Lines: 313 – 314.

Line 272: I think you should also comment that the MME does not capture the H-T reversal at end of the winter (Feb/March).

Response: Comments have been added which note that the models (both free running and with a nudged QBO) failure to capture the reversal in the sign of response in March.

Lines: 315 – 316.

Line 290: did any of the separate models capture the late-winter H-T reversal?

Response: None of the individual models for Exp1 managed to capture the reversal in the sign of PNJ response in March. For Exp1-ObsQBO, only one model (BCC-CSM-MR) managed to capture the transition from negative to positive PNJ response from February to March.

Line 302: typo – observe

Response: corrected

Line: 348.

Line 314: see previous comments on figure captions. The first sentence of the caption is an interpretive statement, not a description of the figure.

Response: All figure captions have been revisited and revised in response to the reviewer's earlier suggestion.

Line 338: this is a nice result.

Response: Thank you for the compliment.

Line 348: do any of the models display inter-annual variability as large as the observed?

Response: Broadly, all the models have the correct interannual variability as shown in Anstey et al. (2026) which is now in the public domain. A lot of this variability is removed with the ensemble averaging even when the simulations all have the same nudged QBO. In reality the cycle-to-cycle variability of the QBO is part of the natural internal variability and would be lost when averaging over realizations.

Line 354: please see major comments (point 3) on the late-winter reversal. I don't understand the comment that the late-winter response in observations is being masked by other influences – are you suggesting that the late-winter reversal is not real?

Response: No, the observed late winter reversal is indeed real. The text has been revised to remove the suggestion.

Line: 417 – 418.

Line 361: see previous comments on figure captions.

Response: All figure captions have been revisited and revised according to the reviewer's earlier suggestion.

Line 366: these early winter correlations in Exp1 and Exp1-noQBO are interesting / puzzling. Do they suggest that SST/SI influences may be important in early winter?

Response: Yes. The SST/SI and other forcings are the only other external source of predictability in these experiments. An extra sentence has been added.

Lines: 428 – 429.

Line 376: I suggest you reverse the caption heading, since you discuss frequency of SSWs first.

Response: Section heading modified.

Line: 439.

Line 436: is your conclusion then that the HTE is not simply due to the occurrence (or otherwise) of SSWs? Do you think it more likely the accumulation of relatively smaller wave impacts?

Response: It is generally accepted that the HTE is not solely due to the occurrence of SSWs. More likely this result indicates that there are issues with the morphology of SSWs in the model and/or deficiencies in mechanism producing a HTE in the models.

The paragraph containing the text from lines 436-440 (original manuscript) has been removed, and Figure 8 and its discussion have been significantly reworked. As discussed in the new text, filtering the models according to those that produced a Holton-Tan effect does result in a marginally significant, positive QBOE-W SSW frequency difference for both Exp1 and Exp1-ObsQBO.

Lines: 468+, 535 – 539.

Lines 468-475: this is consistent with figures 3/4 that show the models do not capture the late-winter H-T reversal.

Response: See our response to major comment #3 above.

Line 482: I think this section contains a useful set of analyses, even though the results are weak and mostly insignificant. I think it's important to report null results. You could also point out the differences between figures 3 and 4 in the extent to which the responses penetrate into the troposphere at high latitudes. The model responses do not penetrate as deeply as the ERA5 response. In fact, the responses in January and February appear to penetrate further into the troposphere in Exp1 than in Exp1ObsQBO, which is interesting.

Response: This paragraph has been revised to make the important link to figures 3 and 4.

Lines: 586 – 587, 590 – 591.

Line 485: describe how you define the NAO (it does not belong in the caption to figure 10).

Response: The definition of the NAO has been moved from the figure caption to the main text.

Lines: 583 – 584.

Line 528: please see major comment 4 regarding an additional surface level pressure analysis, which would also be helpful to this section.

Response: Please see response to major comment 4.

Line 536: please tell the reader here that you have switched to showing W-minus-E rather than E-minus-W, and explain why.

Response: We have changed to QBOE-W, modifying figures 12 and 13, and the text as appropriate.

Lines: 646 – 655, 669 – 671, Figure 12, 690 – 694, 703 – 704, Figure 13, 718 – 720, 800.

Line 541: presumably when you say neither are significant, you mean significantly different from zero?

Response: Yes. The text has been updated to confirm that 'neither are significantly different from zero.'

Line: 650 – 651.

Lines 556-563: nice results here.

Response: Thank you. We appreciate the compliment.

Line 584: maybe say that it's an E-W dipole?

Response: Thank you for highlighting the need for a more precise description. The text has been modified to reflect that is an east-west dipole.

Lines: 703 – 704.

Line 601: is this a weakening of the Walker circulation under QBOW?

Response: We thank the reviewer for identifying a misleading nuance in the text. The responses are associated with QBOE-W and not a specific QBO phase. The text has been modified to reflect this.

Lines: 717 – 720.

Line 635: you state 'extra-tropical' here, but you have also looked at the STJ and tropical rainfall.

Response: Please see our response to major comment 1.

Line 646: I think a sentence to say that the models do not reproduce the late-winter H-T reversal would be appropriate here.

Response: Sentence added.

Line: 769.

Line 693: ditto comments about the model failing to capture the late-winter H-T reversal, which is likely linked to their failure to capture the SSW frequency and timing response.

Response: Given that this is speculation we prefer not to put too much emphasis on this in the final paragraph. It has been mentioned a number of times previously.

Line 701: I suggest removing the word ‘quite’ here.

Response: Removed as suggested.

Line: 829.