Review of Extratropical teleconnections in a multi-model ensemble nudged towards the observed QBO by Andrews et al.

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.

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.
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.
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.
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.
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.

Minor comments

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

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

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

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?

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.

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'.

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.

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

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.

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

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

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

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?

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.

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

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).

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

Line 302: typo – observe

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

Line 338: this is a nice result.

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

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?

Line 361: see previous comments on figure captions.

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?

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

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?

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

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 Exp1-ObsQBO, which is interesting.

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

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

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

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

Lines 556-563: nice results here.

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

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

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

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

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.

Line 701: I suggest removing the word 'quite' here.