

Reviewer 2 comments on the revised manuscript (26.7.2026)

I thank the authors for making substantial changes to the manuscript, and also for pointing me to the two submitted papers that addressed many of the concerns I had.

Thank you for taking the time to review the substantial changes made during the review process. Much appreciated.

The main concern I still have is with regards to Figure 2.

This figure gives me the impression that the models are doing a reasonable job at capturing the HT effect when the HT is evaluated as a correlation in Dec/Jan (in agreement with recently published Garfinkel et al 2026). I realize that this figure was included in the paper for a different purpose as you note in your response, and that lines 325-329 indicate that the regression coefficients are too weak by a factor of 4. But I'm still confused as to how to interpret figure 2 in the context of the rest of the analysis.

From your response, it appears that in panel b you are first computing the ensemble mean PNJ and then correlating it with the QBO. If that is what you are indeed doing, then it would make sense why the correlations in b are so strong. But then the comparison with panel a is not straightforward, and it is misleading to show the two panels side-by-side. Can you create a version of panel b where you correlate the QBO with the PNJ in each model and ensemble member separately, and only then average across members and models? This would allow for a direct comparison to panel a.

We note your concern about the comparison of the two panels. You are right that the ERA5 panel is effectively a single-member whereas the Exp1-ObsQBO panel is an ensemble-mean. The purpose of these diagrams in the 'Experiments, Models and Methods' section is to provide an indication that the optimal atmosphere pressure to measure the QBO is similar for ERA5 and the Exp1-ObsQBO ensemble. The reason why it is better to show the correlation for the ensemble mean and not the mean of the individual correlations is that it helps separate the QBO signal from noise enabling us to identify the best choice of QBO index to use in the paper. Figure 2b is only referred to in a single paragraph (L221 – 223) which is very clear and with no reference to comparisons.

We have removed the ERA5 contour lines from Figure 2b and removed the suggestion of comparison from the caption.

L238-243.

minor comments:

In the Discussion section and also scattered in the results (e.g. line 320-325), there is consideration of how to interpret model vs. obs differences: do they reflect observational sampling variability, or are they reflective of genuine biases in model's teleconnection processes, or both?

Large ensembles of model runs can help distinguish these possibilities - see e.g. Garfinkel et al 2026. Garfinkel et al 2026 indicates that at least in some models there is a genuine model vs. obs discrepancy in the HT effect that likely cannot be explained simply by unrelated variability in obs.

Thank you for the pointer to the paper. We note that Garfinkel et al. (2026) have much larger ensemble sizes (10 – 50) per model and much longer runs (165+ years) which allows more detailed conclusions to be drawn for some models. In comparison, the QBOi phase 2 experiments only have three members for each model with run-lengths of 40 years.

line 659-660 these kinds of runs already exist for 4 models, though regrettably there doesn't appear to be overlap with those models contributing to QBOi. See Martin et al 2023.

Thank you for highlighting this paper which focuses on the relationship, or lack thereof, between the QBO and the MJO in experiments that nudge the stratosphere above 150 hPa. It would be interesting, as a separate exercise, to get hold of these datasets, although neither mean surface pressure nor geopotential height appears to be available via the zenodo data-store hence no analysis of the NAO is currently possible with the publicly available data.

Garfinkel, Chaim I., David Avisar, Scott M. Osprey, Doug Smith, Jian Rao, and Jonathon S. Wright. "Revisiting the surface impacts of the QBO in the Large Ensemble Single Forcing MIP simulations: are teleconnections still too weak?." *Weather and Climate Dynamics* 7, no. 3 (2026): 1133-1152.

Martin, Zane K., Isla R. Simpson, Pu Lin, Clara Orbe, Qi Tang, Julie M. Caron, Chih-Chieh Chen et al. "The lack of a QBO-MJO connection in climate models with a nudged stratosphere." *Journal of Geophysical Research: Atmospheres* 128, no. 17 (2023): e2023JD038722.

David Battisti comments on the revised manuscript (26.7.2026)

Dear Prof. Andrews,

Thanks for addressing the issues raised by all three reviewers (and myself). I think the paper is almost there, but all three reviewers wanted a chance to comment on the revised manuscript, and so I am awaiting their responses before making a final decision.

I also took a careful read of the revised ms and have some suggestions I would like for you to consider. These are noted as comments on the pdf of the revised ms (attached below).

Cheers, David

Thank you for taking the time to carefully read through the revised manuscript.

L44: Modify 'models have' to 'models are nudged to have'.

Revised.

L70: Add references to Holton/Lindzen papers.

Added references to L70.

Added to reference list: L949-950 and L999-1000

L213-4: What does this mean? 'and that mean zonal-mean zonal wind model biases and inter-annual variability are removed from the equatorial winds in Exp1-NoQBO'

To provide clarity, the text, 'and that mean zonal-mean zonal wind biases and inter-annual variability are removed from the equatorial winds in Exp1-NoQBO',

has been replaced with 'while in Exp1-NoQBO the nudging removes both the inter-annual variability and biases in the climatological mean from the simulated equatorial zonal-mean zonal winds'.

L236-237: Remove text, '); Trenberth 1997, Gray et al, 2018) sea-surface temperature anomalies with a 5-month rolling average'

Removed.

L268-269: Add text, 'and a much too weak response in Dec-Jan'.

Modifications made to the text.

L288: statistically significant response?

Yes, as per Table A1. Have added 'statistically', L286.

L503-504: Suggest bracketing the sentence: 'In an attempt to improve the significance of the observed teleconnection a longer period (1959 – 2021) than covered by the experiments was analysed.'

Done.

L538: Remove the word 'western' from 'the occurrence of western disturbances'.

Done.

L595: Add the words, 'absolute value of the' prior to 'ONI'.

Done.

L627: Modify 'is improved' to 'is somewhat improved'.

Done.

L661-663: Suggest to remove sentence starting, 'Additional experiments', as it is not clear why this would be worthwhile.

Removed. Thank you. We agree that this sentence does not add value.