

Dear Prof. Naoe and co-authors:

Thank you for your third revision on the manuscript (WCD 2025-1148). I appreciate that you and your co-authors have corrected the grammar and presented a manuscript that is now clear of ambiguities due to language problems. In doing so, I was able to find a few minor (but important) issues that should be easy to address quickly. I am recommending minor revisions to allow you to make these modifications (or to argue why certain modifications are not appropriate).

Please feel free to contact me if you have questions.

Regards, David

Note: all line numbers refer to the revised text, not the Author tracked changes. Text appearing in the revised manuscript is in *italics*; my suggested changes are in **red**.

- As stated on lines 303-304, *“no model reproduces the observed strength of the Holton-Tan relationship between the phase of the QBO and the strength of the polar vortex across all three experiments”*. That is true, but I don’t think it goes far enough. Following that sentence, I suggest you insert the following sentence: **“From Fig. 2, only three models reproduce the observed relationship in the CTL and EN experiments, and only one model reproduces the observed relationship in the LN experiment.”**
- Lines 305-306: the correlation between U60 and Ueq50 in LN observations is positive (and statistically different from zero), while in the GISS model LN experiment is zero, so I don’t think it is fair to say this is a good fit compared to observations. So I would remove GISS from the list on line 306 and change to read **“In LN three models....”**
- Change lines 38-39 in the Abstract to reflect the above results. Change *“... respectively). In LN, four out of nine models reproduce the observed Holton–Tan relationship within half of the observed amplitude.”* to read **“The strength of the Holton-Tan relationship between the phase of the QBO and the strength of the polar vortex seen in observations is reproduced in fewer than three models under ENSO neutral conditions and by one model under EN conditions. In LN, three out of nine models reproduce the observed Holton–Tan relationship, but with less than half of the observed amplitude.”**

- Line 241, change to read “... of the statistical test  $\alpha$  is adjusted ...”
- Fig. 1, titles on panels c, f and j should probably read CTL49EN39LN39, CTL69EN81LN81, and CTL87EN99LN99, respectively.
- The sentence on lines 354-356 (*In summary, **across** all three experiments, models generally show.... observed response.*) is still ambiguous. If this statement is meant to say “In summary, for **each** experiment, models generally show ...”, then it isn’t supported by Figs. 2 (or 3): for the EN experiment, Fig. 2, five of nine models produce the opposite response to that in observations. I would change this sentence to read “In summary, **across the CNT and LN experiments**, models generally show.... observed response.)”
- Line 412, change to read “... also show later final warming dates in LN ...” to read “... also show **median final warming dates that are later in LN ...**”
- Line 414, change “fail to show earlier final warming dates” to read “**show later final warming dates**”
- Line 445, change “... on the APJ. Only ...” to read “... on the **Asian-Pacific subtropical jet (APJ)**. Only ...” so the reader doesn’t have to search 10 pages back to recall what APJ means.
- Line 548, change to read “... precipitation response **to the phase of the QBO across models or experiments.**”
- Line 642, change to read “... expect models **that have** larger TTL ...”
- Line 650, change to read “.. underestimate **QBO TTL** temperature anomalies ...” because zonal wind, in itself, isn’t the zero-order control on convection.
- Lines 735-736, change “...exhibit a tropospheric signal characterized by upper-level (100 hPa) westerly and lower-level (700 hPa) easterly anomalies during various seasons from May to November. This pattern suggests ...” to read “...exhibit a **tropospheric signal characterized by upper-level (100 hPa) westerly anomalies during various seasons from June to November, while about half of the models and**

the observations show lower level (700 hPa) easterly anomalies. This pattern suggests ...”

- Delete the two sentences on lines 783-785 (“*The observed .... the observations.*”) because it is irrelevant to what follows in this paragraph.
- Lines 808-809, change to read “... show weaker responses. Hence, neither the ...”
- Line 811, change to read “The impact of the QBO on the troposphere is examined, focusing ....”
- Line 812, change to read “... to the QBO phase in the ...”
- Line 820, change to read “... is strongest in observations over the...”
- Line 829-830. I am not sure what is being argued here. Yes, the SST forcing (common to all the models) constrains the circulation in the equatorial (lon-height) plane, and so each model has a very similar mean state that is being acted upon by the QBO. Hence, one might expect the QBO W minus QBO E response to look the same across the models for *the same experiment*. But the SST for the LN experiment is different from the SST for the EN experiment, and so the mean state that QBO acts upon in the LN experiment is different from that in the EN experiment. And yet, the QBO W minus QBO E response is very similar in the LN and EL experiments (compare the colored contours in Fig. 11 to those in Fig. 12). This implies that the impact of the QBO phase is *not* terribly sensitive to changes in the SST. Also, on Line 831-2, it is stated that precipitation may be less constrained by the experimental setup – but we know that tropical Pacific precipitation is heavily constrained by the SST, which is common to all models for each experiment. Altogether, I find this paragraph confusing or even troublesome, and so I would drop it.
- Line 858, change “... vortex coupling arise from consistently weak QBO amplitudes at lower levels in the equatorial stratosphere biases...” to read “vortex coupling arises from consistently weak QBO amplitudes at lower levels in the equatorial stratosphere, biases...”