The authors would like to thank the reviewer for your time and effort in reviewing our manuscript entitled "QBOi El Niño Southern Oscillation experiments: Teleconnections of the QBO". Above all, the authors are deeply grateful for the many insights gained by reading the papers recommended by the reviewers. We will incorporate your valuable comments and suggestions in our revised manuscript to address your concerns. Our reviewer responses and revision plan are shown in blue text whereas reviewer's comments are shown in black. Individual responses are as follows.

Black: Reviewers comments

Blue: Authors response to the reviewer

To Reviewer 2:

RC2: 'Comment on egusphere-2025-1148', Anonymous Referee #2, 01 May 2025 reply

review of "QBOi El Nino Southern Oscillation experiments: Teleconnections of the QBO" by Naoe et al

This study aims to examine how QBO teleconnections are modulated by ENSO using a multi-model ensemble of QBOi models. The specific simulations examined are simulations in which SSTs are either climatological, El Nino, or La Nina, which allows for examining potential nonlinearities between QBO teleconnections and ENSO teleconnections. The use of ~10 models allows for assessment of model sensitivity and robustness. The authors examine four different QBO teleconnections - polar vortex response, subtropical jet, tropical precip, and Walker Cell. They conclude that the QBOi models generally fail to simulate the first three of these teleconnections, and hence it is difficult to conclude anything as to the possibility of ENSO and QBO teleconnections interacting. They find a robust effect of the QBO on the Walker Cell, however, the specifics of the QBO phase and season with maximum impact differ across the models.

While the paper should eventually be publishable in WCD, major revisions are needed first.

major comments:

R2-1. For the first three teleconnections where the models generally fail in the multi-model mean, there are still several models which are relatively more successful in capturing the observed response. There is no discussion of why there is spread across models for two of these teleconnections (vortex response and subtropical jet), while there is a very limited discussion of the third (namely Figure 10). The paper should include a detailed discussion for all three teleconnections as to possible causes of the intermodel spread in how well the models are doing. This could be similar to Figure 10, but instead of T100hPa, the authors could consider horizontal or vertical resolution, the mean state of the vortex or subtropical jet position, meridional width of the simulated QBO, strength of the QBO in each model in the lowermost stratosphere, strength of the QBO in the mid-stratosphere, etc. All of these factors could plausibly be linked to why some models are better than others, and by exploring all of them the paper might be able to provide some insights to model developers as to what needs to be improved.

We appreciate your helpful suggestions. In the revised manuscript, we will include more discussion and analysis of why almost all the models do not reproduce teleconnections examined.

a) Discussion for QBO teleconnections to polar vortex

ENSO modulation of the QBO in the QBOi El Nino Southern Oscillation experiments is investigated by a core paper of Kawatani et al. (in revision; https://egusphere.copernicus.org/preprints/2024/egusphere-2024-3270/)

One of general characteristics is that in the lower stratosphere, the westerly phase duration is generally longer in the La Nina simulations compared to the El Nino simulations. The downward propagation of QBO westerly and easterly phases to the lower stratosphere is more rapid during El Nino, which is a common characteristic among the models. However, QBOs in some models are irregular, from a simple, time-height cross-section of the monthly and zonal mean zonal winds over the equator in the El Nino and La Nina simulations for each model, as shown in Figure 2 of Kawatani et al. It is found that the QBO in the ECHAM El Nino experiment is irregular, with occasionally occurring in downward phases of easterlies and westerlies. The QBOs in GISS and LMDz La Nina experiments are more irregular, and westerly phases sometimes fail to propagate into the lower stratosphere.

Next, the QBO teleconnection problems to the stratospheric polar vortex teleconnection were investigated in more detail by previous studies (Bushell et al., 2022; Anstey et al., 2022). Figure 3 of Anstey et al. (2022) showed January correlation between vortex strength and equatorial wind at different altitudes for all models that performed CTL and for reanalyses. The strength of the QBO teleconnection to the NH winter stratospheric polar vortex was shown to correlate with the amplitude of the QBO at 50 hPa, which is the altitude that shows the strongest correlation with the vortex in observations. Most models show a statistically significant correlation at some altitudes, but the altitudes of peak correlation differ among models.

Figure 4 (b) of Anstey et al. (2022) showed January QBO-vortex correlation using 50 hPa QBO, versus QBO amplitude at 50 hPa. Models with weaker 50 hPa QBO amplitude show weaker correlation in January between the 50 hPa QBO wind and the polar vortex, consistent with the hypothesis that unrealistically weak low-level QBO amplitude can weaken the teleconnection.

In order to answer Reviewer #2 and Editor questions, we examine whether model performance of (a) QBO amplitude and/or (b) climatological polar night jet strength is related to the ability of model to capture the QBO-induced polar vortex responses (Fig. R2-1), assuming that the HTR relationship (polar vortex) route of the QBO teleconnection can be influenced by these two factors. The QBO amplitude is defined as the root mean square of the deseasonalized zonal wind time series at 50 hPa, multiplied by $\sqrt{2}$, following Dunkerton and Delisi (1985). Most models show poor performance of QBO amplitude at 50 hPa, in agreement with previous studies (Bushell et al., 2022; Anstey et al., 2022). It seems that larger QBO amplitudes at 50 hPa for models have larger polar vortex responses compared to the other models (but sometime wrong sign). Fig. R2-1(b) shows that climatological polar vortices in NH winter can be reproduced with their strength, with comparable strength to ERA5.

These results are consistent with the hypothesis that unrealistically weak low-level QBO amplitude can weaken the teleconnection.

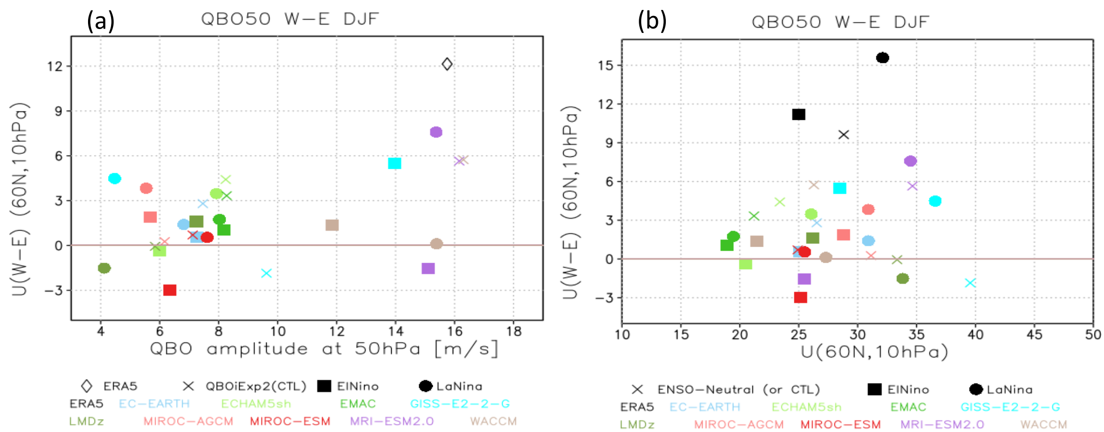

Fig. R2-1. (a) Relationship between QBO amplitude at 50 hPa and composite difference of zonal-mean zonal wind (QBO50 W-E) at 60N and 10 hPa for CTL, El and LN experiments plus ERA5 (1959-2021) in units of m/s. The QBO definition index at 50 hPa is used. (b) Relationship between NH wintertime climatological zonal wind at 60N and 10 hPa and composite difference of zonal-mean zonal wind (QBO50 W-E) at 60N and 10 hPa for CTL, EN, and LN experiments including the ENSO neutral, El Nino, and La Nina winters for ERA5 in units of m/s.

b) Discussion for QBO teleconnection to subtropical jet

Given that the subtropical jet route of the QBO teleconnection can be influenced by (a) the QBO amplitude and/or (b) the climatological position of the subtropical jet (Garfinkel and Hartmann, 2011), we examine whether model performance in simulating these two factors is related to the ability of model to capture the QBO-induced shift of the Asian-Pacific jet (APJ) (Fig. R2-2). Here, the QBO amplitude is defined as the root mean square of the deseasonalized zonal wind time series at 70 hPa, multiplied by $\sqrt{2}$, following Dunkerton and Delisi (1985) and Bushell et al. (2022). The climatological position of the APJ is defined as the latitude of the maximum zonal-mean wind averaged over the APJ sector (150°E–150°W).

Consistent with previous studies (Bushell et al., 2022; Anstey et al., 2022), most QBOi models underestimate the QBO amplitude. Only two models show a comparable QBO amplitude to the reanalysis. However, model biases in QBO amplitude do not affect those in the QBO-APJ connection (Fig. R2-2a). Models with larger QBO amplitudes do not necessarily exhibit stronger jet responses, nor do models with smaller amplitudes consistently show weaker responses. Similar result is also found in the APJ position (Fig. R2-2b).

These results suggest that neither the QBO amplitude nor the APJ position explains the inter-model spread in the QBO-APJ connection. Other factors, such as transient and stationary eddies, may determine the QBO-APJ connection in the model. This possibility needs to be explored in a future study.

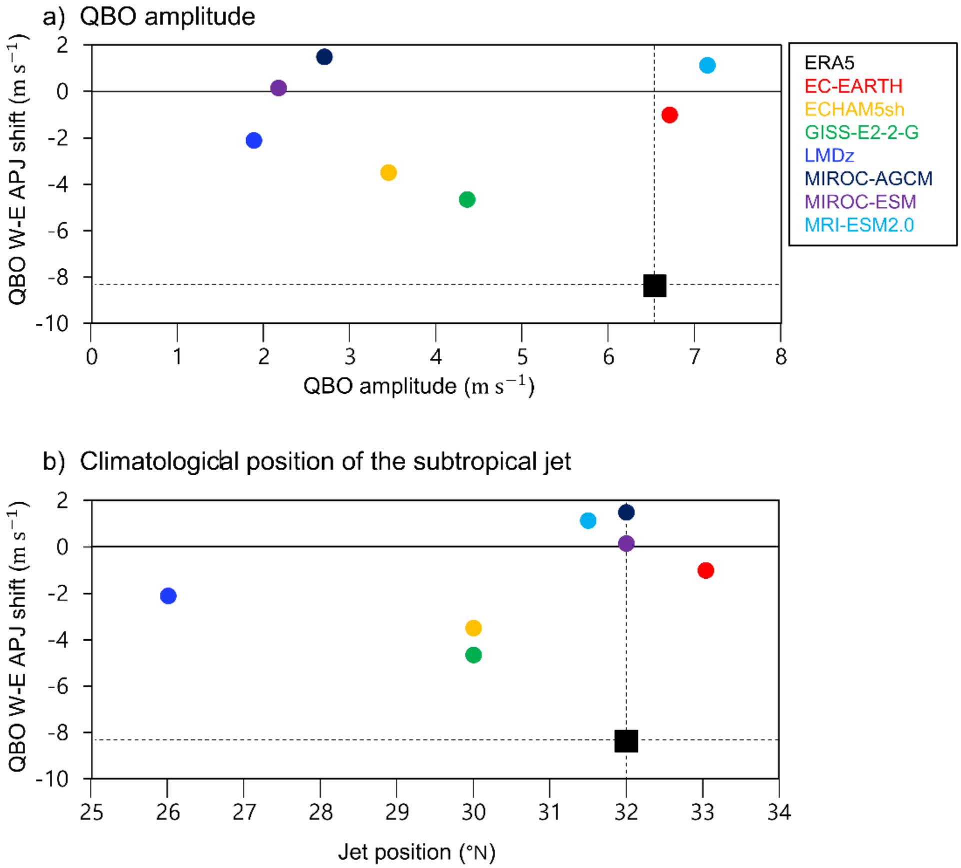

Fig. R2-2. Relationship between the QBO-induced APJ shift index and (a) QBO amplitude, and (b) subtropical jet latitude during ENSO-neutral (CTL) years.

Refs:

Garfinkel, C. I., and Hartmann, D. L.: The influence of the quasi-biennial oscillation on the troposphere in winter in a hierarchy of models. Part I: Simplified dry GCMs, J. Atmos. Sci., 68, 1273–1289, 2011.

Dunkerton, T.J. and Delisi, D.P. (1985) Climatology of the equatorial lower stratosphere. Journal of the Atmospheric Sciences, 42, 376-396.

Bushell, A.C., Anstey, J. A., Butchart, N., Kawatani, Y., Osprey, S.M., Richter, J. H., 20 others: Evaluation of the Quasi-Biennial Oscillation in global climate models for the SPARC QBO-initiative, Q. J. Roy. Meteor. Soc., 148, 1459–1489, 2022.

Anstey, J. A., Simpson, I. R., Richter, J. H., Naoe, H., Taguchi, M., Serva, F., 23 others: Teleconnections of the quasi-biennial oscillation in a multi-model ensemble of QBO-resolving models, Q. J. Roy. Meteor. Soc., 148, 1568–1592. https://doi.org/10.1002/qj.4048, 2022.

c) Discussion for QBO teleconnection to tropical precipitation

Several potential biases likely influence the tropical route of QBO teleconnections. Most proposed mechanisms linking the QBO to the tropical surface rely on interactions between the lowermost stratosphere and the uppermost troposphere. A key bias common to many models, including those used in this study, is a weak QBO amplitude in the lower stratosphere, which limits the effectiveness of stratosphere–troposphere coupling processes (Oueslati et al., 2013; Richter et al., 2020; García-Franco et al., 2023). Additionally, models exhibit persistent tropospheric biases, including the double Intertropical Convergence Zone (ITCZ) and unrealistic rainfall intensity distributions. These biases typically stem from model parameterizations, notably in convection and cloud microphysics schemes (Hagos et al., 2021). The combination of these stratospheric and tropospheric biases likely weakens the QBO signal reaching the tropical troposphere and contributes to inter-model differences in both the timing and spatial manifestation of the teleconnection. This helps explain why some models produce stronger signals during certain seasons or in particular regions compared to others.

Hagos, S. M., Leung, L. R., Garuba, O. A., Demott, C., Harrop, B., Lu, J., & Ahn, M. S. (2021). The relationship between precipitation and precipitable water in CMIP6 simulations and implications for tropical climatology and change. *Journal of Climate*, *34*(5), 1587-1600.

Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020). Progress in simulating the quasi-biennial oscillation in CMIP models. *Journal of Geophysical Research: Atmospheres*, *125*(8), e2019JD032362.

Oueslati, B. and Bellon, G.: Convective entrainment and large-scale organization of tropical precipitation: Sensitivity of the CNRM-CM5 hierarchy of models, J. Climate, 26, 2931–2946, 2013.

On the topic of Figure 10, what is the correlation and slope of the best-fit line? Is the relationship statistically significant?

On the topic of Figure 10, the correlation of the best-fit line for all the data is −0.48 with a p-value of 0.02 according to a t-test of the Pearson correlation coefficient, indicating that the relationship is significant. However, these metrics are sensitive to the experiment, since a larger (in magnitude) correlation coefficient is found for El Niño conditions ($r = -0.82$) and a lower coefficient for La Niña experiments ($r = -0.2$).

R2-2. For the QBO signal in reanalysis, do you try to regress out a lingering signal of ENSO before plotting wqbo minus eqbo? Line 476-479 seems to indicate you don't do this, and it isn't clear whether this is done for the other teleconnections either. If this is not done, then comparing the observed signal to the model signal isn't a fair comparison as there will still be a residual signal from SSTs.

First of all, we do not consider ENSO to be a confounding factor in this study because our simulation explicitly isolate ENSO conditions. We know the idealized QBOi simulations cannot be directly compared to e.g. ERA5. Thus, this point will be acknowledged in the revision.

We do not see clear advantages of regressing out an ENSO signal, compared to compositing on ENSO phase. In lines 476-479, for example, our composites of the observed precipitations (here GPCP, not a reanalysis) are made for El Nino and La Nina events. In this way, our comparison is 'apples-to-apples'. Regressing out an ENSO signal would make our analysis incomparable to our experiments.

R2-3. For the fourth teleconnection examined, the Walker Cell one, the authors adopt a completely different methodology than for the first three. Why for this section only do you play with the season and pressure level, but for earlier sections you don't? For the first three the models did a poor job, and now for this teleconnection they appear to be doing ok. Is this success for the Walker Cell just because you are giving the models lots of opportunities to succeed? Why not use this methodology for earlier sections too? Either way, the fact that a single paper is using very different methodological approaches for different sections is confusing, and leads to the (in my opinion misleading) impression that the models are much better at the QBO-> Walker Cell connection than the others.

Thanks for your comments and suggestions. To provide more context and clarity in the Walker circulation section, we also apply the same methodology used in the earlier sections, using a fixed QBO definition (in terms of season and vertical level) across all models. Our initial analysis reveals a coherent response in the Walker circulation compared to other sections. Figure R2-3 shows the initial results for LN experiment and other figures will be included in the revised supplementary material for consistency. Given this encouraging result, we explore whether the signal can be further strengthened by tailoring the QBO definition to each model, as a way to account for differences in how models represent the QBO vertical structure and timing.

One possibility of these more coherent responses is that the zonal circulation in these SST forced simulations is similar enough amongst models, due to the SST forcing, that the response is relatively more similar, whereas other aspects of the response, such as tropical precipitation, the polar vortex and the subtropical jet may be less constrained by the experimental setup. It is also plausible that the mechanisms that drive the Walker cell response are better represented in these models, given the relatively large static stability anomaly shown in the results, one could reasonably suspect that this mechanism could be large enough in the models to produce a consistent response.

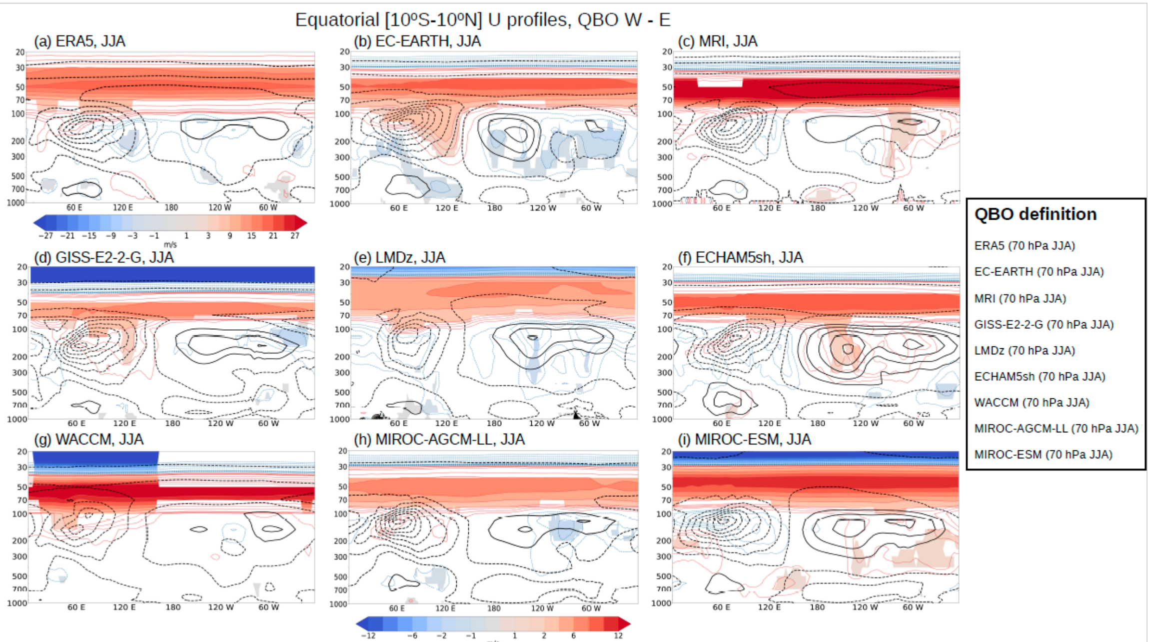

Figure R2-3. Climatology (black contours) and QBO Westerly (W) minus Easterly (E) differences (shading and colored contours) in equatorial zonal wind profiles, averaged over 10° S-10° N, from the LN experiment for the QBOi models. Black contours are drawn at 4 m s-1 intervals, and colored contour follow the same scale as the shading, as indicated in the color bar. The target season is JJA for all models, with the QBO phase defined at 70 hPa during JJA. Only statistically significant zonal wind differences at the 95% confidence level are shaded.

(Figures for CTL and EN experiments will be added in the supplementary material.)

R2-4. To be specific, previous work which allowed for different vertical levels to define the QBO can lead to very different conclusions as to whether models capture the HT effect of the polar vortex. See Rao et al 2020a. It could also be that the seasonality of the HT effect differs from one model to the next. It would be interesting to see if the QBOi models still struggle to represent the HT effect if the authors adopted Rao et al's methodology.

As described in the method section, we do model-observation comparison by applying the same QBO phase definitions to the models that are optimal for observed teleconnections, in order to determine if observed teleconnections are manifested in the model runs. Thus, we use 'standard' indices (e.g., 50-hPa equatorial wind for the QBO), without adjusting them on a model-by-model basis, for all analyses presented in this article.

In Rao et al. (2020a), on the other hand, their QBO was defined at 30 hPa instead of 50 hPa, because some models largely underestimate the QBO magnitude in the lower stratosphere and for some models the QBO is difficult to detect at 50 hPa, and because the westerly phase lasts nearly as long as the easterly phase at 30 hPa.

These mean that our method is rather focuses on the identification of model biases by optimizing the observed teleconnections while Rao et al.'s method is to focus on the detection of model QBO signal, i.e., by maximizing the models' signals. In Garcia-Franco et al. (2022), when looking at Rao et al. (2020), which used QBO definitions at 30 hPa, they demonstrated that this level was not the most suitable for the tropical route.

An investigation of seasonality of Holton-Tan effect using different pressure levels would be a certainly interesting topic. But we think that only after identifying existing model biases that are done in the present work, we can move on to the next step, such a study of seasonality drift of Holton-Tan relationship using the phase-angle technique. Moreover, present-day simulations might be more appropriate to perform this kind of investigation, to better compare models and observation-based datasets.

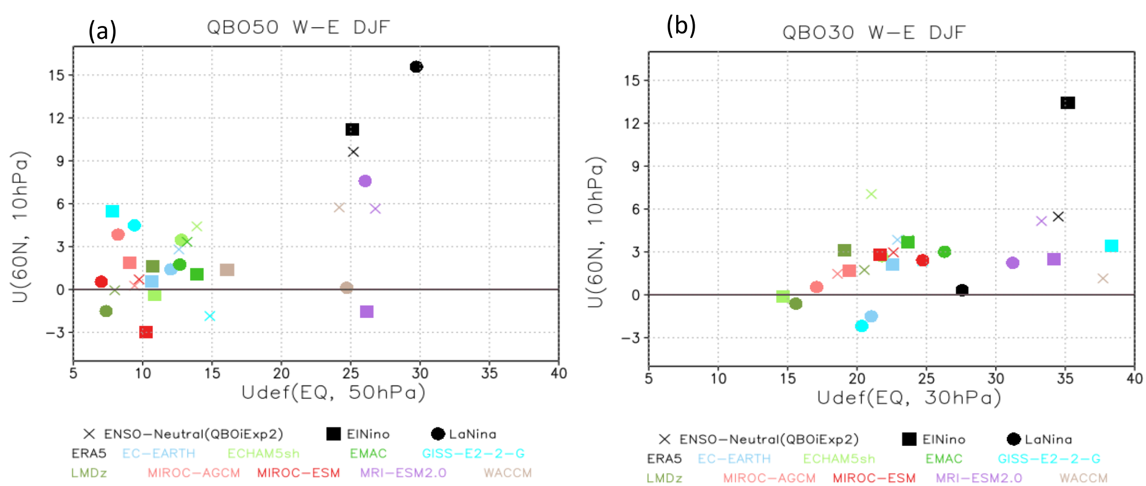

Fig. R2-4. Relationship of composite difference of zonal-mean zonal wind at 60N and 10 hPa with QBO definition index at 50 hPa (QBO50, left panel) and at 30 hPa (QBO30, right panel).

In order to answer Reviewer #2 and Editor questions, we check levels to define the QBO that are based on observational studies (i.e., at 50 hPa) and that are based from model specific (i.e., at 30 hPa), as shown in Fig. R2-4. Both panels (QBO50 and QBO30) show that most models underestimate QBOs and they are struggling to reproduce observed polar vortex responses to the QBO although some models largely underestimate the QBO magnitude in the lower stratosphere, and for some models the QBO is difficult to detect at 50 hPa, as already described in Rao et al. (2020). In observations, a QBO response is large in La Nina in the left panel (QBO50) while a QBO response is largest in El Nino in the right panel (QBO30).

As described in the Introduction, previous studies investigating the joint effects of QBO and ENSO on polar vortex variability in winter suggested that their interactions are nonlinear insofar as the Holton-Tan relationship is found to be significant in the La Nina phase but much weaker in the El Nino phase (Wei et al., 2007; Garfinkel and Hartmann, 2008; Calvo et al., 2009; Richter et al., 2011; Hansen et al., 2016). Thus, this means that our QBO-vortex responses in the observation classified in the ENSO phase using QBO50 is more consistent with previous studies.

minor comments:

1. Somewhere in the paragraph from lines 109 to 116, and also near line 124, Rao et al 2020b should be cited and discussed

Thank you for the helpful suggestion for this topic. We will revise the text to cite this reference as they explored and evaluated three dynamical pathways for impacts of the QBO on the troposphere.

2. Line 135: Trascasa-Castro et al 2019 and Weinberger et al 2019 should be cited and discussed

Thank you for the helpful suggestion. We will revise the text to cite these references about the relationship between ENSO and SSWs.

3. line 142-146: Ma et al should be cited and discussed

Thank you for the helpful suggestion. We will revise the text to cite this reference as QBO and ENSO have a nonlinear combined effect on North Atlantic surface pressure anomalies.

4. line 199: Pahlavan et al should be cited.

This reference will be included in the revised text.

technical edits aren't included in this round, but will be provided after the major comments are addressed.