Response to Reviewer #1 (Prof. Garfinkel)

This study presents first results from the QBOi phase 2 experiments in which the QBO is nudged to observations while the rest of the atmosphere is allowed to freely evolve, with the hopes of understanding how teleconnections improve with a "perfect" QBO. While there is a modest improvement in teleconnections both to the polar night jet and to the subtropical Pacific, the signals are still weaker than those observed. The remaining difference could be either to model deficiencies in the teleconnections processes themselves, or to the fact that part of the observed signal is a fortuitous alignment of internal variability.

This study is clearly on its way to being publishable in WCD, however moderate revisions are needed as detailed below. My main critique is that I'm not fully convinced that the problem is with the teleconnection process itself, while the authors appear to conclude otherwise; see details below. I'm going to sign my review, as the papers I'm referencing in my review probably give my identity away in any event.

Chaim Garfinkel

The authors thank Prof. Garfinkel for kindly reviewing the manuscript. Specific points raised are addressed below.

general comments:

1. The paper doesn't directly show that the nudging leads to a reasonable QBO in the lowermost stratosphere and instead cites Anstey et al in prep (line 225). I realize how this can happen with community efforts; however, I think it is important to demonstrate that the lowermost stratospheric QBO (both amplitude and meridional width) are reasonable in the nudged runs, in order to confirm that the nudging is acting as intended and that we should expect teleconnections to be better. (For the purposes of this paper on teleconnections, only the lowermost stratosphere is what is relevant.). This concern is motivated by previous work which suggests that on the rare occasion that free-running models simulate a reasonable QBO in the lowermost stratosphere, the PNJ modulation is actually realistic in the extended winter mean (Rao et al 2020) even if the seasonality within winter is still biased.

Response: The protocol paper describing the experiments and how effective the nudging is, is now available in the public domain (https://egusphere.copernicus.org/preprints/2026/egusphere-2026-1165/egusphere-2026-1165.pdf). Results in that paper address the issue raised here. In addition, a further paper analysing the QBO per se in the phase 2 experiments has been submitted to WCD . (https://egusphere.copernicus.org/preprints/2026/egusphere-2026-2856/)

One way to bypass this concern (and also the lack of any figure documenting the QBO itself in this paper) is to focus on regressions of the PNJ against QBO winds, instead of a compositing perspective. A regression approach will "correct" for any deficiency in the QBO and instead focus directly on whether teleconnection processes are too weak. It would be interesting to compare regression coefficients of the QBO with the PNJ in Exp1 vs. in Exp1-obsQBO, and if they differ this would imply that the nudging is somehow interfering with teleconnection processes in models (e.g., the nudging could be messing with the waves, despite the fact it is

zonal mean only, by perhaps inducing artificial wave reflection, though I would think the 5-day timescale is weak enough to mostly avoid this).

Response: The protocol paper (https://egusphere.copernicus.org/preprints/2026/egusphere-2026-1165/egusphere-2026-1165.pdf) demonstrates that the nudging has no direct effect in the models on the climatology and interannual standard deviation in the extra-tropics. In response to both your suggestion and that of Reviewer 4, we calculated regressions between the PNJ and QBO winds when sub-sampling the multi-model ensembles with a QBOE-W sample size equivalent to observations. A comparison of Exp1 and Exp1-ObsQBO indicates that the strength of the model teleconnections are largely unchanged, and that the enhanced PNJ response in Exp1-ObsQBO is primarily the result of the increased amplitude of the nudged equatorial QBO. This is now reflected in the main text.

Lines: 47 – 49, 378 – 384, 772 - 774

On a related noted note, some QBOi-phase2 models use full-field nudging and not zonal-mean nudging, and it might be helpful to see if those models have a systematically different QBO->PNJ effect than the ones using zonal-mean nudging only. Some of the downward impacts from the polar vortex to the near-surface in the SNAPSI runs are stronger in the full-field nudging runs (e.g., Australia T2m in Feng et al 2025) for reasons that aren't yet fully clear. To be specific, it would be helpful to add additional panels to figure 2 where you separately show the multimodel mean for full-nudged vs. zonal-mean nudged.

Response: Differences between the full field and zonal mean nudging were considered but obtaining useful conclusions was limited by the small sample size of the full field results . Likewise, it is difficult to generalise from the specific case studies of SNAPSI to the multi-year simulations used in QBOi phase 2.

2 (but to some degree an extension of 1): I'm a bit confused as to how to interpret figure 2. Doesn't this figure show that the correlation of the QBO winds with U1060 is nearly perfect in the multi-model mean? There are still issues with January specifically, but if I mentally perform a wintertime average the correlations seem very reasonable. See comment #1 about using a regression approach rather than a compositing approach - based on Figure 2 I would guess that a regression approach would likely show that the models are doing a very good job in the winter mean, and that other than a specific issue in January, there is no evidence for a systematic inability to capture QBO->PNJ teleconnection. Rather, the bias is in whether the models are simulating the correct amount of variance, either in the PNJ or in the QBO. (A regression version of Figure 2 in addition to the correlation version currently used would also help clarify this point.) If the authors agree with this interpretation, then the overarching conclusion reached in this paper (i.e., models struggle with the PNJ teleconnection) may need to be toned down.

Response: We are not sure what is meant by the correlation been "nearly perfect" in the multi-model mean? Sure, the correlation is of similar size to that for ERA5 but that is not unexpected given that the multi-model averaging removes most of the internal variability. Nonetheless the signal is much weaker as can be deduce from the results shown in Figure 6. The protocol paper (now available at

https://egusphere.copernicus.org/preprints/2026/egusphere-2026-1165/egusphere-2026-1165.pdf) confirms that the models do simulate the correct amount of variance in the PNJ. The nudge ensemble members have near identical QBOs (again see the protocol paper) which of course have the same QBO variance as ERA5.

3. I'm also confused as to how to interpret the results of Table 2 in the context of figure 4 and 5. Figure 4 and 5 indicate that the QBO in Exp1 leads to a significant vortex response (albeit weaker than that of Exp1-ObsQBO). On the other hand, Table 2 indicates this experiment nonetheless has no skill. Is it surprising that skill is now essentially zero even as the HT effect is still there, just weaker? To be specific, I would be interested in seeing an analogous panel on Figure 2 but for the Exp1 runs: is the correlation weaker than for Exp1-ObsQBO? If it isn't weaker, then this would indicate that the primary problem is that the QBO itself is too weak in Exp1.

Response: Figure 4 considers the difference in the composite response of QBOE and QBOW. The correlations in Table 2 assess the ability of the experiments to capture interannual variability of the ensemble-mean timeseries (which includes times when the QBO is neutral). Figure 5 simply assesses whether a subset of QBOE and QBOW randomly chosen from the across the experiment period are capable of matching the observed PNJ response. Hence there is no conflict between the results in Figures 4 and 5, and Table 2. The text has been revised to highlight that the approaches differ and that a significant composite PNJ response does not necessarily lead to significant predictability.

Lines: 386 – 388, 431 – 433

4. Section 5: it would be nice to see a map of slp or near surface U (or similar). It could be the signal doesn't project strongly onto the NAO pattern you predefine, but there still could be a surface signal. Also, I couldn't find the description of the NAO definition used in this paper. Are you allowing for each model to have a slightly different NAO loading pattern? Does the QBO downward impact project onto a slightly modified version of the NAO loading pattern?

Response: The connection of the zonal mean wind towards the middle troposphere is shown in Figure 3 (ERA5) and Figure 4 (QBOi experiments). The NAO definition was contained within the caption of Figure 10 but has now been moved to the start of section 5. 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 intend to perform a detailed analysis of the NAO response, or lack thereof, in the future. This would stretch to a whole paper and is beyond the scope of the current manuscript.

Lines: 583 – 584

5. The discussion near line 575-580 of the causes of the subtropical jet response seems very speculative to me - please explain in more detail why these tropical precip anomalies should matter. Garfinkel et al 2011 (already cited) instead argue that the subtropical Pacific response is related to downward arching of zonal wind anomalies which then interacts with baroclinic eddies, and actually the Introduction (line 124-130) and Discussion section (line 677-680) of this

paper seems to focus on this downward arching mechanism and not the tropical precip mechanism. Note that Garfinkel et al 2011 (the dry model one) explicitly rejects the possibility that the subtropical jet modulation arises from a tropical convection route, but of course in other models or in observations this pathway cannot be ruled out just based on the dry model results. If you want to focus on this tropical convection mechanism, I suggest including it in the Introduction and Discussion.

To be specific, can you make figure 12 (or 13) for a range of pressure levels from 70hPa to 300hPa? If the Garfinkel et al mechanism is relevant, there should be a continuous connection from the QBO region itself to the subtropical Pacific in the upper troposphere.

Response: We thank the reviewer for highlighting the omission of any mention in the Introduction of the tropical convection route. Text has been added to the introduction highlighting this as a possibility and the reason why tropical precipitation anomalies might matter. Like most of the teleconnections in this paper we consider the co-variability with the QBO but, for the sake of brevity, do not deep-dive into the detailed analysis. Follow-up studies that explore the mechanisms in depth are planned.

Lines: 134 – 141, 801 - 802

6. I have a stylistic suggestion, motivated by the fact that the discussion of what may cause the remaining difference in teleconnection strength/position between the nudged runs and observations seems somewhat disjointed. Namely, some parts of the text focus on issues in teleconnection processes (PNJ response) while others focus on the possibility of internal variability or other processes aliasing the observed response (subtropical jet). My guess is that this reflects the intuition of whichever co-author wrote that particular section, though Figure 5 makes the case that internal variability isn't enough for the SPV response and there is a genuine bias in dynamical processes. Either way, it would seem that both possibilities are relevant to all of the teleconnections (though maybe the authors think otherwise! I would welcome being corrected.) If that is indeed the case, the paper could probably be better organized by laying down both possibilities already in the introduction section, then shortening/cutting as much as possible the discussion of these two possibilities from the Results section, and instead revisiting it in detail in the Discussion section in a more systematic way. [I think the paper would be clearer if this suggestion was followed, but it's entirely up to the authors.]

Response: Unfortunately, this paragraph wasn't very clear to the authors. We surmise that the reviewer is concerned with separating the impact of biases in the QBO from other model biases. As a result of this remark, we have revised the introduction to make it clearer that our study focuses on the role of the biases in the QBO. In turn, the discussion section already acknowledged the possible role of other model biases, though a detailed analysis of these is quite beyond the scope of the present study.

Lines: 177, 199, 755 – 756, 765

minor comments

line 49: "robust and insignificant" is an oxymoron. I wouldn't use this expression in the abstract, as there isn't space to explain the subtleties behind why this combination makes sense. (I think you intend for "robust" to refer to across model agreement, and "insignificant" whether anomalies meet 95% significance using standard statistical tests, right? You might want to state this explicitly in the methods section.)

Response: The sentence has been reworded to remove reference to 'robust' as you are right, the two words are challenging to reconcile. Have also done similar in section 4 the Conclusions section.

Lines: 50, 780

line 209: also shown in the ozone-only LESFMIP experiments (Garfinkel et al 2025)

Response: Reference to Garfinkel et al., 2025 added.

Lines: 211, 1055 - 1057

line 302: observe -> observed

Response: Corrected.

Line: 348

line 327: "is an overestimate" could be worded more carefully. Do you intend "might reflect aliasing from unrelated variability"? See general comment 6 above.

Response: We have considered your suggestion as well as a similar comment from reviewer 2 and have changed the text as recommended.

Lines 373 – 374.

line 350: on S2S timescales, there is no evidence for a S2N paradox in the stratospheric polar vortex (Garfinkel et al 2024)

Response: It is true that Garfinkel et al 2024 demonstrates no evidence of a S2N paradox in the stratospheric polar vortex out to around 7 weeks, however, the QBOi phase 2 experiments consist of 40-year model runs implying that the period of influence of the initial conditions is not a significant factor. It is clear from Figure 6 that the PNJ signal for the ensemble mean is much smaller than the model spread for these long experiments. The text has been modified to highlight that these experiments are of much longer timescale than sub-seasonal forecasts, that Eade et al's paper covers seasonal-to-decadal, and a reference to Garfinkel et al. 2024 has been added.

Lines: 403 – 407, 1051 - 1053

line 414: "QBOE vs QBOE" one of these should be QBOW

Response: The paragraph containing this text has been deleted in response to a change in Figure 8.

Line: 504 – 509

Figure 10 seems to have lower resolution than the other figures.

Response: Figure 10 has been improved with a higher resolution version with colours that match those in Figure 11.

Lines 610 – 615.

Figure 13: it would be interesting to include the eastern Indian Ocean in these figures. Overall the pattern matches very well the recent simulations of Schwartz et al 2026, including a tripole pattern in the tropics

Response: It would be interesting, but we prefer to maintain the focus of the discussions on the North Pacific.

Feng, Kexiang, Jian Rao, Chaim I. Garfinkel, Amy H. Butler, Weihua Jie, Tongwen Wu, Peter Hitchcock, Eun-Pa Lim, Andrew J. Dowdy, and William Seviour. "Can stratospheric nudging improve surface predictability? Insights from the 2019 Southern Hemisphere sudden stratospheric warming." npj Climate and Atmospheric Science 8, no. 1 (2025): 353.

Garfinkel, Chaim I., Jeff Knight, Masakazu Taguchi, Chen Schwartz, Judah Cohen, Wen Chen, Amy H. Butler, and Daniela IV Domeisen. "Development of the signal-to-noise paradox in subseasonal forecasting models: When? Where? Why?." Quarterly Journal of the Royal Meteorological Society 150, no. 764 (2024): 4417-4436.

Response: Reference added.

Garfinkel, C. I., Avisar, D., Osprey, S., & Smith, D. (2025). The response of the QBO to external forcings: implications for disruption events. Journal of Geophysical Research: Atmospheres, 130(22), e2025JD044438.

Response: Reference added.

Schwartz, C., Garfinkel, C. I., & Chen, W. (2026). Simulated tropical troposphere response to the QBO: Effect of vertical resolution, gravity waves parameterization, and boundary forcing. Geophysical Research Letters, 53, e2025GL120711. https://doi.org/10.1029/2025GL120711

Rao, J., C. I. Garfinkel, and I. P. White, 2020: Impact of the Quasi-Biennial Oscillation on the Northern Winter Stratospheric Polar Vortex in CMIP5/6 Models. J. Climate, 33, 4787–4813, https://doi.org/10.1175/JCLI-D-19-0663.1.