Response to Reviewers' Comments

Hyun-Kyu Lee, James A. Anstey, Hye-Yeong Chun, Shingo Watanabe, Francois Lott, Zhaoyang Chai, Yixiong Lu, Qi Tang, Jinbo Xie, Dong-Chan Hong, Seok-Woo Son, Federico Serva, Pu Lin, Martin B. Andrews, Neal Butchart, Aleena M. Jaison, Jeff R. Knight, Scott Osprey, Hiroaki Naoe, Kohei Yoshida, Yoshio Kawatani, and Jadwiga H. Richter

2 November 2026

Dear Editor and Reviewers,

Thank you very much for your constructive comments/suggestions on the manuscript entitled, "Analysis of systematic biases in equatorial resolved and parameterized wave forcing of the quasi-biennial oscillation in a multi-model ensemble" (egusphere-2026-2856). We have carefully considered the reviewers' comments and revised the manuscript accordingly. The revisions include (i) renaming "NoNudge" as "IntQBO," (ii) adding ERA5 results and ensemble variability of the QBOi models to the instability analysis, (iii) including the momentum budget imbalance in Figs. 7–9, and (iv) adding a discussion of the reliability and limitations of resolved wave forcing in ERA5. We also identified and corrected a data archiving error in BCC-CSM2-MR and updated the corresponding results. Below, we indicate the original comment of the respective reviewer in blue and our answer is denoted in black.

We hope the current manuscript may now be acceptable for *Weather and Climate Dynamics*.

Sincerely,

Hyun-Kyu Lee

Reviewer #1's Comments

General comments:

This article analyzes the QBO forcing in simulations from the QBOi project. The analysis includes the "NoNudge" simulations, ie. the original simulations, as well as "ObsQBO" simulations, where a nudging scheme for the zonal wind in the equatorial stratosphere is used to effectively assimilate the QBO from ERA5. A set of diagnostics is applied to understand better how the wave mean-flow interaction works and how this can be split in contributions by different wave types. The article is mostly descriptive and is useful because it provides a general view on the problems in simulating the QBO in this class of models model. It extends the existing collection of publications on the QBOi simulations.

The article is generally well written and the structure leads from the broader features to the finer details. Figures should be produced in a better resolution or in a scalable vector format, because some figures, e.g. Fig. 5, contain a large number of panels, so that zooming in is necessary.

The minor comments given below should be considered to improve the details of the article.

We sincerely appreciate the constructive suggestions. We have carefully considered the comments and revised the manuscript accordingly. Regarding figure resolution, the original figures were produced at resolutions exceeding 300 DPI, but their apparent quality was reduced during PDF conversion. The original figure quality is preserved in the revised submission.

Minor comments:

1. L34 and L112: "… NoNudge and ObsQBO …"

These names could be chosen better. As the focus is on the QBO, and one set of experiments simulates the QBO internally, as currently used for NoNudge, it seem straightforward to use the acronym "InternalQBO" or shorter "IntQBO". The other experiments have a nudged QBO, which should resemble the observed QBO but still isn't the real QBO. Anyway a natural acronym would be "NudgedQBO" or shorter "NdgQBO". So "IntQBO" and "NdgQBO" would be a nice pair. Please consider to change the acronyms.

Thank you for the suggestion. Following the reviewer's suggestion, we have changed "NoNudge" to "IntQBO" throughout the revised manuscript. However, "ObsQBO" is the experiment name defined in the protocol paper (Anstey et al., 2026), and changing it to "NdgQBO" could compromise consistency with other QBOi phase-2 papers. We have therefore retained "ObsQBO."

2. L37: maybe better: "… stratospheric zonal-mean zonal winds in the simulations toward that in ERA5 …"

It is corrected as suggested. [Line 39]

3. L44: If it is the case, it would be worth to mention that the vertical momentum flux across a level near the tropopause is statistically the same, because the sources of waves in the NoNudge and ObsQBO are the same, and only the filtering and thus the momentum transfer to the QBO jets are different. If however the sources are systematically different between NoNudge and ObsQBO, than please mention this instead. (And of course, if the filtering at lower levels is inefficient, then the momentum will be transferred at higher levels, where conditions are favorable, as you know.)

Thank you for the comment. We find that the vertical component of the Eliassen–Palm flux (EPF), averaged between 100 and 130 hPa near the tropopause, is approximately 0.7 mPa regardless of nudging. Therefore, the differences in EPF divergence (EPFD) in the QBO regime between the two experiments can be attributed primarily to differences in wind filtering rather than tropospheric wave sources. We have explicitly stated in the revised abstract that differences in tropospheric wave sources between the two experiments are negligible. [Lines 46–47]

4. L65: "… In phase 1 of the Atmospheric Processes And their Role in Climate (APARC) QBO initiative (QBOi), which provided the first comprehensive multi-model intercomparison of QBO resolving GCMs, including an analysis of QBO wave driving (Holt et al., 2022). …"

The article by Holt et al. is from 2020, not 2022 (also on L79). And their work was "… in phase 1 of the Stratosphere–troposphere Processes And their Role in Climate (SPARC) …". SPARC became APARC only later. Please clarify this, for example in a footnote.

L69: Also Bushell et al. is from 2020, not 2022 (also L86). Please check the publication years in all references!

Thank you for the comment. We have clarified in the revised manuscript that APARC was formerly known as SPARC. [Line 69]

Regarding the publication years, we are aware that Holt et al. (2022) and Bushell et al. (2022) were first published online in 2020. However, both articles appeared in Volume 148, Issue 744 of QJRMS in 2022. We have therefore retained 2022 as their official publication year.

5. L85: "… the drag …"

Does "the drag" mean parameterized drag, then please write so. Very high resolution models may produce explicitly the gravity waves producing such drag, but Holt et al. probably analyzed CMIP-type model simulations, which do not simulate gravity waves explicitly, and instead depend on a parameterization for their drag.

It is corrected as suggested. [Line 88]

6. L88: "…, or all of these."

should be extended to make it clear that this concerns the resolved as well as the parameterized wave spectrum, for example: "…, or all of these, in the resolved and/or the parameterized waves."

It is corrected as suggested. [Lines 91–92]

7. L105: "… In the Exp1-ObsQBO experiment, the zonal-mean zonal winds in the models are nudged toward the zonal-mean zonal winds of ERA5 in the equatorial stratosphere. Equatorial waves in the troposphere are not constrained by the nudging. However, because full-field nudging is applied to three models …"

The nudging of the QBO needs some latitude and pressure (or height?) structure for the nudging strength (or nudging time scale), and some times also a target QBO structure in latitude and pressure. As the nudging details are important for the result, it is worth to add here some description. Please also state that only the zonal wind is nudged, if this is true. And if this is not true, please add information on other fields which are nudged. (I hope nobody nudged temperature!)

The details of the nudging procedure are provided in the QBOi phase-2 protocol paper (Anstey et al., 2026), in which Figure 1 shows the spatial variation with latitude and pressure of nudging timescale and it is described that only the wind is nudged (temperature is not nudged). We have explicitly noted this in the revised manuscript. [Lines 110–111, 118–119]

8. L114: "Table 1: Summary of the QBOi phase-2 models used in this study.…"

Please add a brief explanation of the difference between "MIROC6.1, p1" and "MIROC6.2, p2".

Thank you for the good suggestion. The two MIROC configurations employ different convection schemes, and this is clarified in the revised manuscript. [Lines 129–130]

9. L115: "… H denotes that the model provided high-frequency data (6-hourly) …"

Please specify also the output data frequency of all other models.

Following the reviewer's suggestion, the output data frequency for all models has been specified in the revised manuscript. [Lines 126–127]

10. L133: "… by subtracting the climatology from the monthly averaged zonal wind at 30 hPa …"
Which kind of climatology is subtracted, the climatology of the annual mean or the annual cycle? What is the range of the normalized zonal wind, and which original zonal mean wind strengths correspond to the threshold values of +0.5 and -0.5? Please clarify this.
In this study, we subtracted the climatological annual cycle, and we have clarified this in the revised manuscript. [Line 146]
For ERA5, the normalized zonal wind ranges from −2.0 to 1.5. The original zonal winds corresponding to the normalized thresholds of +0.5 and −0.5 vary across the 12 calendar months, ranging from −1.6 to 4.2 m s^{-1} and from −19.2 to −10.5 m s^{-1}, respectively. The corresponding values vary among models and experiments.

11. L135: "Second, the descending westerly (DW) and descending easterly (DE) are defined …"
Do mean QBO phases? Like: "… and descending easterly (DE) phases are defined …"
It is corrected as suggested. [Line 149]

12. L144: "… of the five-month smoothed zonal wind …"
You mean "… of the five-month smoothed zonal mean zonal wind …"?
It is corrected as suggested. [Line 157]

13. L170: "A half level is identified at the altitude where …"
Does this mean: " Then the highest half level below the upper level is identified where …"
Thank you for pointing this out. We have clarified this sentence in the revised manuscript. [Lines 183–185]

14. L216: "… In the NoNudge experiments, none of the models reproduce the westward winds of ERA5 between 10 and 50 hPa (Fig. 1a). …"
This should be reformulated. All models produce westward winds between 10 and 50 hPa! The issue is that none of the model profiles of westward wind reaches or exceeds the westward wind profile of ERA5.
Thank you for the suggestion. It is corrected in the revised manuscript. [Lines 234–235]

15. L217: "… At 10 hPa, the multi-model mean (MMM) …"

From here on the multi model mean is used to represent the "typical" model behavior. A multi model median would be another option, which is probably more "typical" if a few models are outliers. Please insert in this section or in the previous section a discussion on the statistics to be used to represent the typical model behavior, which currently is the mean, though a median would be another good choice.

Some of the following discussion about the influence of outliers on the MMM could be reduced if the median was used.

Thank you for the suggestion. Although the median is less sensitive to outliers, we have retained the multi-model mean (MMM) because our aim is to evaluate systematic biases while accounting for the contributions of all models. We have added a discussion of this choice to the revised manuscript. [Lines 120–122]

16. L252: "Metrics of the vertical and latitudinal extent of the QBO are shown in Table S1. …"

As these quantities are discussed here, it would be better to include this table in the main article.

Following the reviewer's suggestion, Table S1 has been incorporated into the main text as Table 3 in the revised manuscript.

17. L280: "… all models except BCC-CSM2-MR exhibit QBO periods of 28 months for the FFT method."

Table 3 reports and FFT period of 25.4 months in ObsQBO of BCC-CSM2-MR. What is the reason for this "failure"? Fig. S1 shows the correct number of cycles (~17) and thus it is surprising that the FFT amplitude, which is also averaged over levels, and which is described as the more robust method, results in a substantially shorter period.

Thank you for the insightful comment. During the revision process, we identified a data archiving error in ensemble member 3 of the BCC-CSM2-MR ObsQBO experiment (Fig. R1), which distorted the ensemble mean. We apologize for this error. After excluding this erroneous ensemble member, the model exhibits a QBO period of 28 months, consistent with the other models. We have excluded the third ensemble member of BCC-CSM2-MR and updated all corresponding results throughout the revised manuscript. The impact of this correction on the MMM is negligible.

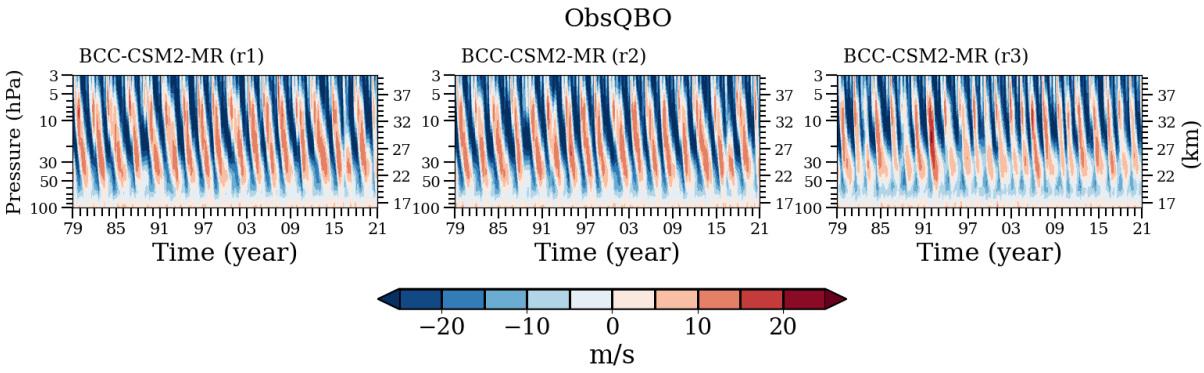


Figure R1: Zonal-mean zonal winds averaged between 5°N and 5°S in BCC-CSM2-MR.

18. L293: Please combine Fig. 2 and Fig. S2 as two panels in a single Figure.

Thank you for the suggestion. We have combined Figs. 2 and S2 into a single figure in the revised manuscript.

19. L315: already Horinouchi et al. (2003) observed this large variation between model (https://doi.org/10.1175/1520-0469(2003)060%3C2765:TCCAUW%3E2.0.CO;2)

We have cited Horinouchi et al. (2003) in the revised manuscript. [Line 338]

20. L323: "The differences between the ObsQBO and NoNudge experiments are very small in both the symmetric and antisymmetric spectra. … Consequently, nudging applied in the equatorial stratosphere has a limited impact on the tropospheric source, allowing the stratospheric wave forcing to be examined with minimal influence from changes in tropospheric sources."

This is an important though possibly unwanted finding of this analysis. This should be part of the abstract too, because it effectively makes the interpretation of the later results easier. (NoNudge to ObsQBO differences in the stratosphere are not related to differences in the tropospheric convectively coupled wave sources.)

Thank you for the suggestion. We have revised the abstract to explicitly state that the impact of stratospheric nudging on tropospheric wave sources is negligible. [Lines 44–45]

21. L371: “Table 4 shows the modulation of instability in five models …”
For this discussion it would be helpful to have the same diagnostics also form ERA5, which provides the nudging target for the ObsQBO simulations, but which does not generate its QBO by means of a nudging procedure. First, this would provide a sense of what should be a realistic value. Is it closer to ca. 4 units as for example in the MIROC6.1 simulations, or less than 1 unit as in E3SM? Further this would provide information on the role of the QBO jet structure vs. the assimilation method. Therefore please add ERA5 to Table 4 and discuss the model results also with respect to this source.
Thank you for the suggestion. We have added ERA5 to Table 5 and included the corresponding discussion in the revised manuscript. [Table 5; Lines 399–404]

22. L380: ″... the increased frequency of $q\varphi$ ...″ → ″... the increased frequency of negative $q\varphi$ ...″ ?

It is corrected as suggested. [Line 410]

23. L381: “… Consequently, a realistic WQBO structure could be an important factor to accurately reproduce the MRG waves in the equatorial stratosphere. …”
Is there any simulation in which the nudging degrades the QBO jets? Or is this about the type of nudging: zonal mean nudging vs. full field nudging?
Our discussion refers to the structure of the QBO jets. In this study, nudging weakened the westerly QBO jet in 7 out of 10 models (Table 2). Furthermore, nudging improved the representation of the frequency of instability across models, regardless of whether zonal-mean or full-field nudging was applied (Table 5).

24. L399: “The momentum budget of the QBOi models is examined through the individual TEM components (Eq. 2). …”
In Eq.2, there exists also the imbalance term Ximbalance that should include effects by diffusion and in the case of nudging additionally the nudging tendency Fnudging. Please include Ximbalance in your discussion and in the related figures. It would clarify the importance of Ximbalance in comparison to the internal forcing by EPFD. (Alternatively, if available from the nudged simulations, show Fnudging.) The Ximbalance or Fnudging term could be discussed after that on Xparam.
Thank you for the good suggestion. We have included $\bar{X}_{imbalance}$ in Figs. 7–9 and added the relevant discussion to the revised manuscript. Because $F_{nudging}$ is not available from several models, it could not be included in this study. [Figures 7–9; Lines 461–464, 469–471, 475]

25. L409: "… which corresponds to the slowdown of the downward propagation of the QBO induced by tropical upwelling. …"
There should be a hint to the role of the secondary meridional circulation of the QBO, which modifies the advection, and especially in the case of the westerly phase can reduce the advection effect of the general tropical upwelling to nearly zero. I think that is also what we can see in Fig. 8.
We agree with the reviewer's opinion. We have briefly discussed the role of the QBO-induced secondary circulation in the revised manuscript. [Lines 445–447]

26. L589: "… the maxima and minima of the zonal wind between 200 and 85 hPa, averaged over 5°N and 5°S …"
Does this mean maxima and minima of u5N-5S over all levels in the range and all longitudes, or maxima and minima of the zonal mean of u: u5N-5S over all levels?
Thank you for pointing this out. The phrase "the maxima and minima of the zonal wind" refers to the maxima and minima of the zonal-mean zonal wind. We have clarified this in the revised manuscript. [Lines 635–636, 648]

27. L612: "… When the onset of the easterly wind occurs, the TEM components of QBOi models in ObsQBO generally capture the characteristics of ERA5, although individual models diverge in their equatorial wave forcing (Figs. 16m–16p). …"
This is not entirely the case. A critical difference exists in the delayed forcing by XPW + XGW. (Fig. 16k). In ERA5 and MERRA2, this forcing peaks at month -1 and 0, but in the model simulations this forcing peaks later between months 0 to +2.
Thank you for the comment. The relevant discussion has been included in the revised manuscript. [Lines 660–662]

28. L637: "The correlations between WF30-50 in the ObsQBO and A50 in the NoNudge …"
Probably should be: "The correlations between WF30-50 and A50 in the ObsQBO …"
Although we respect the reviewer's comment, we confirmed that the original text is correct. In the ObsQBO experiments, the winds are nudged toward ERA5, so the QBO amplitude is externally constrained rather than determined by the simulated wave forcing. We therefore did not examine the relationship between wave forcing and QBO amplitude within the ObsQBO experiments.

29. Data Availability

For the QBOi phase 2 model outputs it is mentioned that these are hosted by the NERC Centre for Environmental Data Analysis (CEDA), but no information is given on the accessibility. Are all these data accessible? How can the data be accessed and downloaded? This should be explained.

The QBOi phase-2 model data are available upon request from the QBOi coordinators. We have explicitly stated this in the Data Availability section of the revised manuscript. [Lines 768–769]

Reviewer #2's Comments

Overall comments:

In many places in the paper, differences from ERA5 are interpreted as model errors. Words like "excessive," "insufficient" and "underestimated" relative to the reanalyses implicitly treat the reanalysis values as correct. That's reasonable for the QBO winds, which observations constrain, but less so for wave forcing, which they don't. Would be good to really emphasize that resolved waves in ERA5 are known to be too weak compared with satellite and radiosonde estimates.

Thank you for the comment. We acknowledge the uncertainties in the representation of resolved waves in ERA5. Specifically, the amplitudes of resolved gravity waves in ERA5 are smaller than those in satellite observations (Lear et al., 2024), and ERA5 shows limited skill in capturing high-frequency variability observed by superpressure balloons (Bramberger et al., 2023). However, previous studies have shown that planetary-scale waves are realistically represented in ERA5 when compared with observations (Fan et al., 2022; Ern et al., 2023). Therefore, we conclude that ERA5 remains a valuable reference dataset for evaluating planetary-scale equatorial waves in the models. To address this issue, we have added Section 4.3 to the revised manuscript, in which we discuss the reliability of planetary-scale waves and the uncertainties associated with resolved gravity waves. [Lines 722–738]

Specific comments:

1. Section 2.6: To increase reproducibility it would be good to make these things explicit:

1) Was the 90-day window stepped by 90 days with no overlap, was there an overlap, or was it stepped daily as a sliding window? You could add a small description like "applied every 30 days, centered on each calendar month, resulting in X spectral windows" or whatever you did.

Thank you for the comment. The 90-day window was shifted at monthly intervals and centered on each calendar month, resulting in approximately 60 days of overlap between adjacent windows. We have added this description to the revised manuscript. [Lines 211–212]

2) Since you state you multiplied by a factor of 1.5 to maintain original variance, this suggests you used a quarter sine taper instead of sine squared, for example. But this should be explicitly stated by adding to the end of "The structure of the window function involves a sine component for the first 30 days, unity throughout the middle 30 days, and a cosine component for the final 30 days." Something like:

$w(t) = \sin(\pi t/60)$ for $0 \le t < 30$; 1 for $30 \le t < 60$; $\cos(\pi(t-60)/60)$ for $60 \le t \le 90$ (t in days)

Thank you for the suggestion. It is rewritten in the revised manuscript as suggested. [Lines 213–214]

3) "We then conduct a two-dimensional Fourier transform on these variables to obtain the zonal (k) and frequency (ω) spectra at each latitude." Add: and pressure level. Then also state the vertical grid and resolution of the model data used for the wave analysis, and whether any interpolation was done before computing fluxes and divergences.

Thank you for pointing this out. It is corrected as suggested in the revised manuscript. [Lines 214–215] In this study, no additional interpolation was performed before calculating the wave fluxes and their divergences. The vertical grids and resolutions of the QBOi models are described in Fig. 2 and Table 3 of Anstey et al. (2026), and also their Table C5 which clarifies that 6-hourly fields for equatorial wave analysis were requested on a set of pressure levels that is representative of the actual model resolution. This reference to Anstey et al. (2026) regarding the model resolutions was already included in the original manuscript and can be found in the revised manuscript [Lines 118–119].

4) Are these performed on daily data? 6-hourly? 3-hourly for reanalyses?

The equatorial wave separation was performed using 6-hourly data for the QBOi models and 3-hourly data for the reanalyses. We have explicitly stated in the revised manuscript. [Lines 497–498]

2. Line 290: There's no mention of why the models that do not reproduce a realistic QBO even when nudged might not do so. BCC-CSM2-MR, CAS-ESM and E3SMv2 still miss ERA5 amplitudes when nudged. With a 5-day timescale, the leftover error likely reflects each model's forcing bias, which is useful information the paper could use. BCC and E3SM are also in the wave analysis, which weakens the premise that their background winds are "fixed."

Thank you for the comment. We have added a brief discussion of the remaining QBO amplitude biases in BCC-CSM2-MR, CAS-ESM, and E3SMv2 to the revised manuscript. [Lines 311–314]

Although residual differences in the background winds may affect the diagnosed wave forcing, nudging substantially reduces the biases found in the IntQBO experiments and thus provides improved background wind conditions for the wave forcing analysis.

3. Table 4: Two suggestions:

1) Add a row for ERA5 to show what the number "should" be. And in this case it might make sense to coarsen ERA5 first, depending on how you calculate the frequency.

Thank you for the suggestion. Following the reviewer's suggestion, we have added the ERA5 results to Table 5 (previously Table 4) and included the corresponding discussion in the revised manuscript. [Table 5; Lines 399–404] Throughout this study, the QBOi model and reanalysis data were analyzed at their original horizontal resolutions, without additional spatial interpolation. We therefore did not coarsen ERA5 for this particular diagnostic, in order to maintain a consistent analysis procedure across the study.

2) Show the ensemble spread for the models with 3 members.

We have added the ensemble spread for the QBOi models. [Table 5]

4.Line 370 and Table 4: To increase the reproducibility, it would be good to make these things explicit:

1) Is this done on 6-hourly, daily, or monthly mean data? This matters a lot since negative gradients come from the curvature term, and time averaging smooths curvature.

The calculation was performed using 6-hourly high-frequency data. It is stated in the revised manuscript. [Line 396]

Since instantaneous data were used, the results were not affected by the smoothing effect of temporal averaging.

2) What does the percentage count? Is it the fraction of all grid point and time samples in the defined region, the fraction of time steps with any negative value in the region, or something else? These would give very different numbers.

The percentage represents the spatial average of the monthly frequency of negative values at each grid point within the defined region. It is stated in the revised manuscript. [Lines 425–426]

3) Are ensemble members pooled or computed separately and then averaged? And does LMDz6's NoNudge value, from only one run, rest on a smaller sample than its ObsQBO value?

The values were calculated separately for each ensemble member and then averaged across ensemble members. The LMDz6 IntQBO experiment has one ensemble member, whereas the ObsQBO experiment has three (Table 1). We have explicitly stated in the revised manuscript. [Lines 396–397]

4) How are N^2 and the vertical derivative computed?

The vertical derivatives, including those used to estimate of N^2, were computed using second-order centered finite differences on irregular vertical grids, following Fornberg (1998).

5. Section 3.3:

1) Again from the TEM section, are the TEM diagnostics performed on daily data? 6-hourly? 3-hourly? Figs 7-9 include all 11 models and you stated that only 5 have 6-hourly output. Were there submitted TEM diagnostics for each model or did you compute them? If so, on model levels or interpolated pressure surfaces?

The model outputs used in this study were provided on pressure levels. For the QBOi models, we used monthly mean TEM diagnostics submitted by each model. To maintain consistency across models, we used these monthly mean outputs even for models that provided high-frequency data. Offline calculations using high-frequency data yielded similar results (not shown). We have explicitly stated in the revised manuscript. [Lines 431–434]

2) Figs 7-9: The panels show zonal wind, tendency, advection, EPFD, and parameterized GWD. There could be a panel for the nudging or imbalance term. The nudging tendency appears in Equation 1, but appears to be absorbed into the imbalance term in Equation 2 and not shown. Since you state that the improved QBO in ObsQBO is driven by the nudging forcing, its magnitude and structure by phase for each model would directly quantify the forcing deficit you want to measure. It would strengthen the paper to show it I think, and to state whether it was archived by the models or computed as a residual.

Thank you for the good suggestion. Following the reviewer's suggestion, we have added $\bar{X}_{imbalance}$ to Figs. 7–9 and included the corresponding discussion in the revised manuscript. [Figs. 7–9; Lines 461–464, 469–471, 475] $\bar{X}_{imbalance}$ is calculated as a residual of the TEM momentum budget.

6. Ln 700: There's no code availability statement, and it's unclear whether the QBOi data at CEDA is public.

The QBOi Phase 2 model data can be accessed upon request by contacting the QBOi coordinators. This has explicitly stated in the Data Availability section in the revised manuscript. We have also added a Code Availability statement. [Lines 768–771]

7. Figs. S3 and S4: Seems there are errors in the captions, should be referring to Fig. 6?

Thank you for the comment. It is corrected in the revised manuscript.

8. Fig. S7: BCC seems to dominate the five-model means. Its Kelvin flux at 70 hPa is nearly twice ERA5's. The "about 60% larger than ERA5" Kelvin forcing near 10 hPa could be checked without BCC.

The MMM of Kelvin wave forcing at 10 hPa, excluding BCC-CSM2-MR, is 0.14 m s day^{-1}, approximately 55% larger than the ERA5 value of 0.09 m s day^{-1}. Thus, the larger Kelvin wave forcing in the MMM relative to ERA5 is not solely attributable to BCC-CSM2-MR. We have explicitly stated in the revised manuscript. [Lines 538–539]

9. The correlations in Figs. S9 and S10 rest on 4–5 models and are dominated by single models (CAS-ESM in S10). I'm not sure you should use them as quantitative support.

Thank you for the comment. The correlation coefficients obtained from models employing the source-dependent gravity wave parameterization do not change substantially when CAS-ESM is excluded (Figs. S8 and R2). Nevertheless, we acknowledge the limitation imposed by the small sample size and have revised the manuscript to interpret these relationships more cautiously. [Lines 693–694, 698–699]

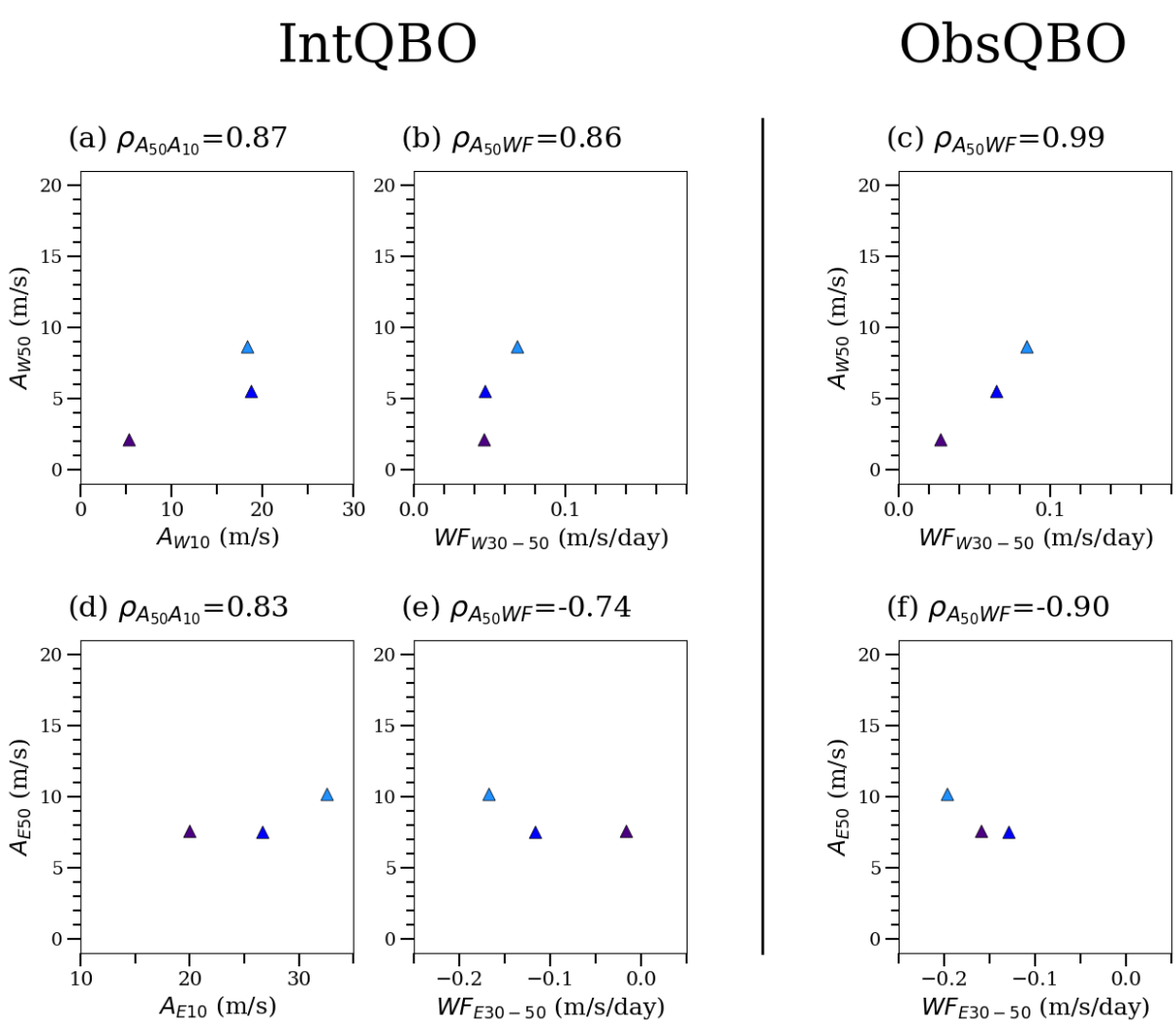


Figure R2. Same as Fig.17 but the models employing with the source-dependent gravity wave parameterization, excluding CAS-ESM.

References

Anstey, J. A., Butchart, N., Osprey, S., Kawatani, Y., Hamilton, K., Richter, J. H., Stockdale, T., Andrews, M. B., Chai, Z., Davini, P., Hong, D.-C., Huang, K., Jaison, A. M., Kerzenmacher, T., Knight, J. R., Lin, P., Lott, F., Lu, Y., Naoe, H., Serva, F., Simpson, I., Son, S.-W., Tang, Q., Watanabe, S., Xie, J., and Yoshida, K.: Experiment design, nudging protocol, and models participating in Phase 2 of the APARC Quasi-Biennial Oscillation initiative (QBOi), EGUsphere, 2026, 1–38, https://doi.org/10.5194/egusphere-2026-1165, 2026.

Bramberger, M., Goetz, D., Alexander, M. J., Kalnajs, L., Hertzog, A., and Podglajen, A.: Tropical wave observations from the Reel-down Atmospheric Temperature Sensor (RATS) in the lowermost stratosphere during Strateole-2, Geophys. Res. Lett., 50, e2023GL104711, https://doi.org/10.1029/2023GL104711, 2023.

Bushell, A. C., Anstey, J. A., Butchart, N., Kawatani, Y., Osprey, S. M., Richter, J. H., Serva, F., Braesicke, P., Cagnazzo, C., Chen, C. C., Chun, H.-Y., Garcia, R. R., Gray, L. J., Hamilton, K., Kerzenmacher, T., Kim, Y.-H., Lott, F., McLandress, C., Naoe, H., Scinocca, J., Smith, A. K., Stockdale, T. N., Versick, S., Watanabe, S., Yoshida, K., and Yukimoto, S.: Evaluation of the quasi-biennial oscillation in global climate models for the SPARC QBO-initiative, Q. J. Roy. Meteorol. Soc., 148, 1459–1489, https://doi.org/10.1002/qj.3765, 2022.

Ern, M., Diallo, M. A., Khordakova, D., Krisch, I., Preusse, P., Reitebuch, O., Ungermann, J., and Riese, M.: The quasi-biennial oscillation (QBO) and global-scale tropical waves in Aeolus wind observations, radiosonde data, and reanalyses, Atmos. Chem. Phys., 23, 9549–9583, https://doi.org/10.5194/acp-23-9549-2023, 2023.

Fan, Y., Huang, C. M., Zhang, S. D., Huang, K. M., and Gong, Y.: Long-term study of quasi-16-day waves based on ERA5 reanalysis data and EOS MLS observations from 2005 to 2020, J. Geophys. Res-Space Phys., 127, e2021JA030030, https://doi.org/10.1029/2021JA030030, 2022.

Fornberg, B.: Classroom note: Calculation of weights in finite difference formulas. SIAM review, 40(3), 685-691, 1998.

Holt, L. A., Lott, F., Garcia, R. R., Kiladis, G. N., Cheng, Y. M., Anstey, J. A., Braesicke, P., Bushell, A. C., Butchart, N., Cagnazzo, C., Chen, C. C., Chun, H.-Y., Kawatani, Y., Kerzenmacher, T., Kim, Y.-H., McLandress, C., Naoe, H., Osprey, S., Richter, J. H., Scaife, A. A., Scinocca, J., Serva, F., Versick, S., Watanabe, S., Yoshida, K., and Yukimoto, S.: An evaluation of tropical waves and wave forcing of the QBO in the QBOi models, Q. J. Roy. Meteorol. Soc., 148, 1541–1567, https://doi.org/10.1002/qj.3827, 2022.

Lear, E. J., Wright, C. J., Hindley, N. P., Polichtchouk, I., and Hoffmann, L.: Comparing gravity waves in a kilometer-scale run of the IFS to airs satellite observations and ERA5, J. Geophys. Res.-Atmos., 129, e2023JD040097, https://doi.org/10.1029/2023JD040097, 2024.