*Supplement of*

# QBOi El Niño–Southern Oscillation experiments: overview of the experimental design and ENSO modulation of the QBO

**Yoshio Kawatani et al.**

*Correspondence to:* Yoshio Kawatani (kawatani@ees.hokudai.ac.jp)

This document provides supplementary analyses and figures that support the main article. It consists of five sections (S1–S5).

**Section S1: Strength of ENSO SSTs used in the experiments**

We chose to use amplified anomalies to maximize the ENSO signal in the QBO response and clarify the underlying mechanisms. In this section, we explain our rationale and the procedure in more detail.

The SST anomalies used in Kawatani et al. (2019; hereafter K2019) represent a "moderate" ENSO based on observations from all El Niño/La Niña SST anomalies. ENSO SST composites were constructed using data from 1950-2016, based on the Japan Meteorological Agency (JMA) ENSO indices. For each calendar month, El Niño and La Niña events were identified according to the JMA definition. Monthly SST anomalies were then weighted by the corresponding NINO.3 index and averaged to create monthly composite SST anomalies.

This process resulted in "moderate" composite ENSO SST anomalies, as illustrated by the January El Niño example. Seventeen January El Niño events were identified between 1950-2016, with NINO.3 anomalies ranging from 0.4K to 3.2K. The resulting composite NINO.3 SST anomaly for January was 1.92K, representative of what would generally be regarded as a moderate El Niño event. In the observational record, the highest NINO.3 anomaly value in a January El Niño event is +3.5 K (Table S1).

To ensure a clear and robust QBO response in our model experiments, we amplified these composite SST anomalies. El Niño anomalies were multiplied by 1.8 and La Niña anomalies by 1.4. These factors were chosen to approximate the peak SST anomalies observed during the strongest observed El Niño and La Niña events, as illustrated in Figs. 1c,d. This approach allowed us to better isolate the impact of ENSO on the QBO. We emphasize that SST anomalies employed are not 'unrealistically' large in the sense that actual anomalies of this magnitude are observed on occasion.

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| El Niño months | 17 | 15 | 11 | 13 | 14 | 18 | 18 | 17 | 18 | 18 | 18 | 17 |
| Max | 3.2 | 2.6 | 2.1 | 1.8 | 2.1 | 2.0 | 2.5 | 2.9 | 3.0 | 3.3 | 3.6 | 3.5 |
| La Niña months | 16 | 15 | 13 | 14 | 12 | 12 | 15 | 15 | 16 | 16 | 16 | 16 |
| Min | -1.8 | -1.5 | -1.0 | -1.3 | -1.4 | -2.0 | -1.6 | -1.6 | -1.3 | -1.6 | -1.7 | -1.8 |

**Table S1.** The number of El Niño and La Niña months during 1950-2016. Max and Min indicate maximum NINO3 anomalies for El Niño and minimum for La Niña (unit: K).

**Section S2: Method of Taguchi (2010)**

Here we present a figure (S1) illustrating our QBO phase definition, which follows Taguchi's (2010) method.
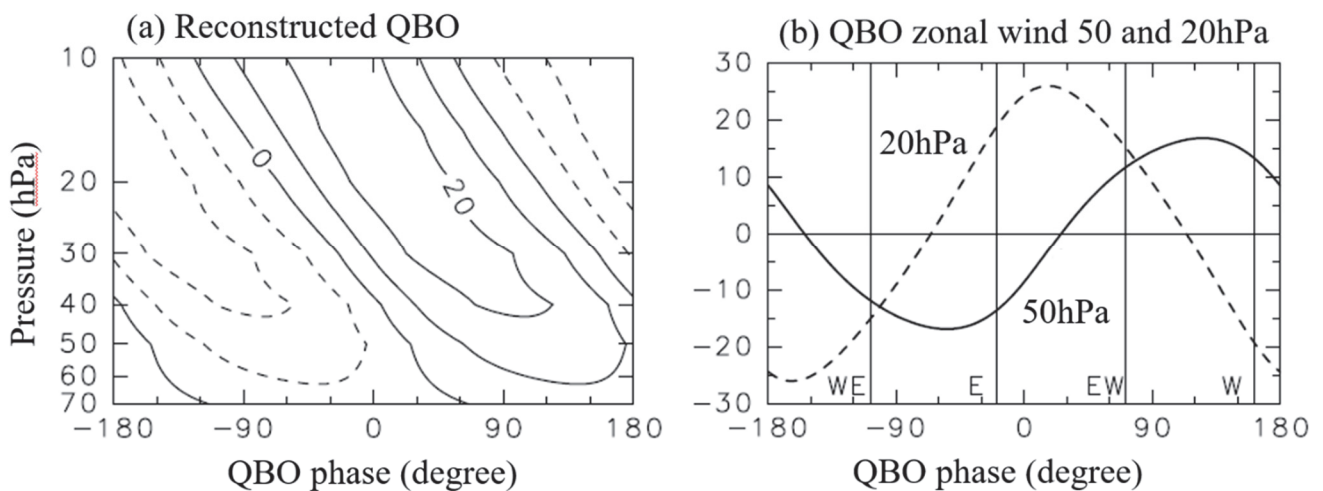


**Figure S1. The definitions of WE, E, EW and W phases. (a) Reconstructed QBO zonal wind. (d) Reconstructed wind at 50 (solid line) and 20 (dashed line) hPa.**

**Section S3: El Niño minus La Niña differences of climatological annual mean, zonal mean temperature and zonal wind.**

Figure S2 shows the climatological annual mean, zonal mean temperature in the El Niño (purple contours) and La Niña (green contours) runs, along with their El Niño minus La Niña differences (color shading) in ERA5 and in the individual models. Over ~30°S–30°N, the troposphere is warmer, and the stratosphere is cooler in El Niño compared to La Niña runs (cf. Domeisen et al., 2019) for all the models and in the ERA5 data. In both the troposphere and stratosphere, the model El Niño-La Niña differences near the equator tend to be spread somewhat wider meridionally than indicated by ERA5. In the troposphere, the ERA5 composite displays clearly defined regions poleward of ~30° latitude where the El Niño conditions are cooler than the La Niña conditions, and these regions are not as clear in the model results except for EC-EARTH. One possible reason is the use of amplified ENSO SSTs in the model experiments. In their results using MIROC-AGCM with imposed ENSO SST anomalies based on just more recent ENSOs (and without the extra amplification used in the present study), Kawatani et al (2019; Fig. 7) found these cooler differences, although not as cool as indicted by the ERA5 results (see Fig. 7 in Kawatani et al., 2019), but this feature is not found in the present experiment with MIROC-AGCM.

Warm differences in the mid-to-high-latitude stratosphere are found in all models, which are also qualitatively similar to those in ERA5. However, the vertical and meridional extent of these warm anomalies differ among the models. LMDz shows significantly cooler El Niño-La Niña anomalies in the Southern Hemisphere stratosphere, where MRI and CESM1 simulated significant warm differences. The cooler El Niño-La Niña anomalies in the troposphere around 60°N-90°N in ERA5 are not found in any of the models.

Figure S3 is the same as Fig. S2 but for zonal mean zonal wind. The models successfully reproduce the overall pattern of the El Niño-La Niña differences in the zonal mean zonal wind observed in ERA5. In the equatorial troposphere, weak westerly differences around 900 hPa and easterly differences around 500–100 hPa are found in both ERA5 and the models, except for CESM1, which simulates smaller easterly differences over the central Pacific and stronger westerly differences at other longitudes (Fig. 5). In the low-to-mid latitudes, the zero contours of the zonal wind are higher in El Niño than in La Niña (see purple and green lines, respectively) in ERA5, and relatively large westerly differences from the lower troposphere to ~30 hPa are observed outside the equatorial region. All models simulate an upward shift of the zero wind contours in the El Niño run consistent with these westerly differences, but their vertical patterns differ significantly among models. Specifically, significantly deeper westerly differences are found up to 5 hPa in GISS, LMDz, and MIROC-ESM. In the poleward regions, weak easterly differences are found from the surface to the stratosphere in ERA5. These easterly differences also vary among the models, being largest in GISS and smallest in EMAC. LMDz, MRI, and CESM1 also show larger easterly anomalies compared to ERA5.

Different distributions of the background zonal mean zonal wind could alter wave propagation properties and thus modify the stratospheric Brewer–Dobson circulation (BDC). The westerly El Niño minus La Niña differences in the mid-latitude lower stratosphere are somewhat similar to what is seen in global warming versus present-day climate simulations (e.g., Kawatani et al., 2011).
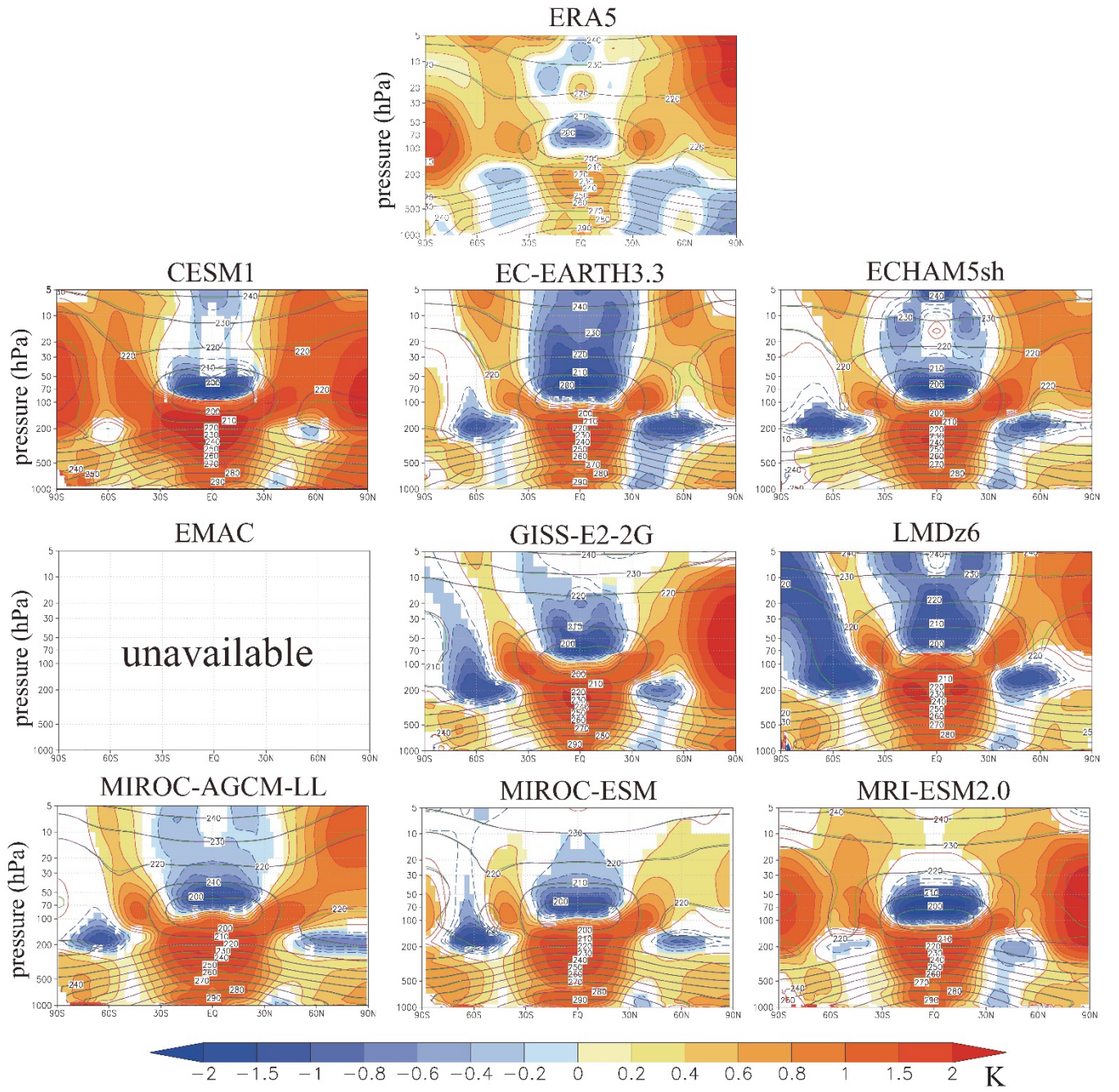
**Figure S2.** Contour lines show the climatological annual mean, zonal mean temperature for (purple) El Niño conditions and (green) La Niña conditions for (center top) ERA5 and (others) QBOi models. Contour intervals are 10 K for mean temperature. Color shading shows the El Niño minus La Niña differences and are only plotted where the magnitudes exceed 0.1 K for ERA5 and where the differences are judged different from zero with statistical confidence ≥ 95% for the model results. Color intervals are ±0.2, 0.4, 0.6, 1, 1.5, and 3 K.
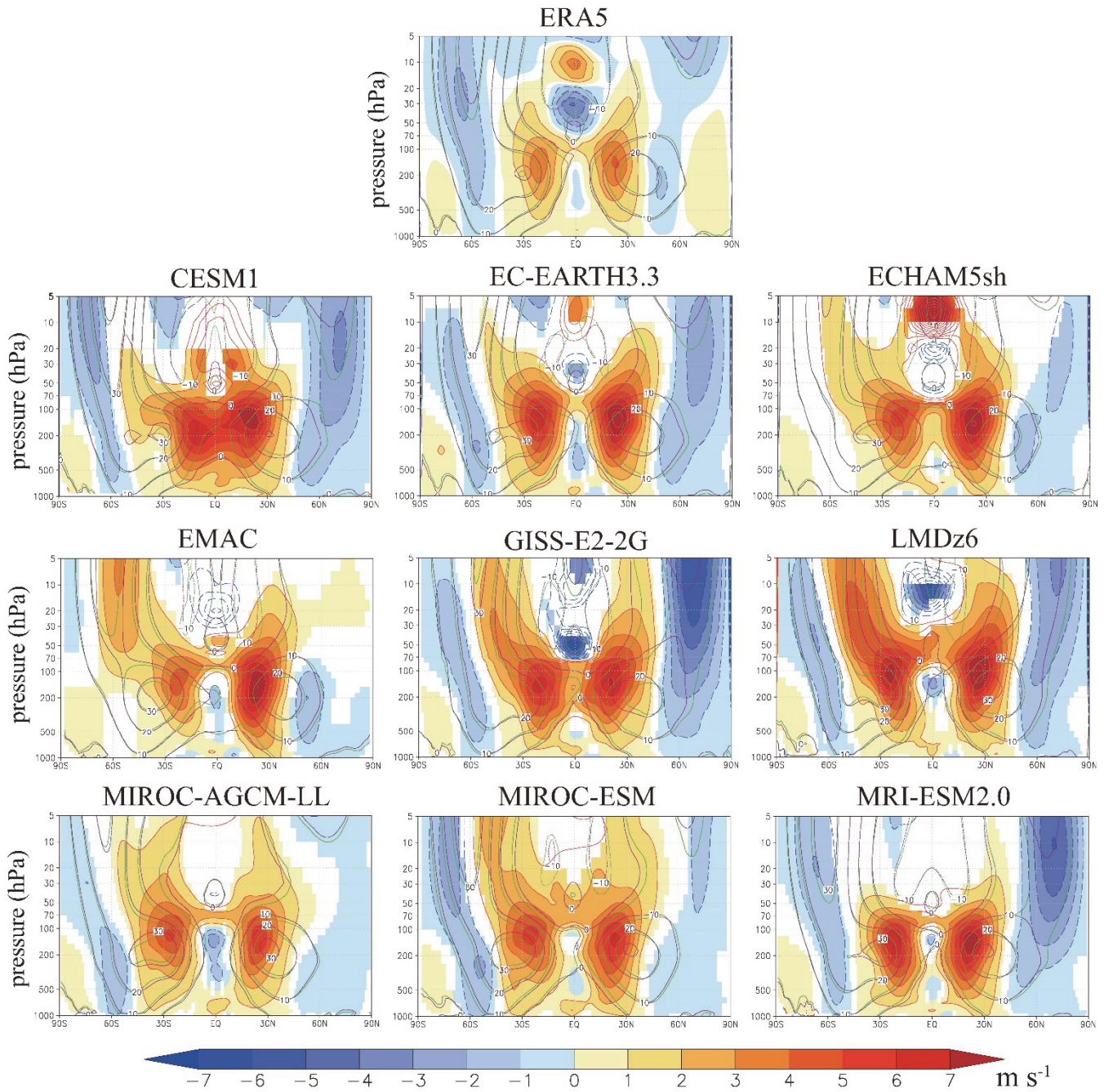
## ERA5



Figure S3. As in Fig. S2 but for zonal mean zonal wind. Contour intervals are 10 m s⁻¹. Color shading included where the magnitude of El Niño minus La Niña differences exceed 0.2 m s⁻¹ for ERA5 and where the differences are judged different from zero with statistical confidence ≥ 95% for the model results.  Color intervals are ±1, 2, 3, 4, 5, 6 and 7 m s⁻¹.

**Section S4: Contrasting wave forcing and residual mean meridional circulations in El Niño and La Niña**

Figure S4 shows the differences between El Niño and La Niña in ERA5 and in each model, focusing on annual mean TEM related quantities. Specifically, the figure presents El Niño-La Niña differences in the resolved EP-flux divergence, zonal forcing due to parameterized orographic and to parameterized non-orographic gravity waves (note that only parameterized total zonal forcing differences are shown for ERA5), and the stream function of the residual mean circulation. In the left three columns, black contour lines indicate the zonal mean zonal wind in each El Niño run. Positive wave forcing values correspond to stronger eastward wave driving of the mean flow, while negative values indicate stronger westward driving in El Niño compared to La Niña. Note also that positive and negative residual mean circulation streamfunction values represent clockwise and counterclockwise circulations, respectively.

4

90        First, we focus on results from the MIROC-AGCM-LL, which was also used in the K2019 study (but with more moderate imposed ENSO SST anomalies, as explained in the previous section). Figure 8 of K2019 showed the EP flux vectors for El Niño conditions, along with the differences between El Niño and La Niña simulations. Notably this showed that resolved waves transporting westward mean momentum propagate upward around 30°N–70°N and 30°S–70°S and also propagate equatorward in low latitudes in the El Niño run. This equatorward propagation weakens considerably around the zero lines of

95        the mean zonal wind. There is also strong westward wave forcing due to orographic GWP in the mid-latitude upper-troposphere and lower-stratosphere (UTLS) region in both hemispheres, where westerlies are weak. This presumably plays a role in decelerating the upper part of the subtropical westerly jet.

          These characteristics of the EP flux vectors are very similar to those found in the present MIROC-AGCM experiments (not shown).  The El Niño-La Niña differences in EP-flux divergence seen in the present Fig. S4 (and the earlier Kawatani et

100      al., 2019, Fig. 8) features strong westward forcing above ~70-100 hPa and eastward forcing below these levels, showing that more resolved waves transporting westward momentum preferentially propagate into the lower stratosphere during El Niño. This is consistent with the upward shift of zero lines of zonal wind in El Niño compared to La Niña (see purple and green contours in Fig. S3). Westward forcing differences due to orographic GWP occur in the mid-latitudes around 30–80 hPa in both the Northern and Southern Hemispheres, whereas eastward forcing anomalies are seen below the westward forcing

105      anomalies, particularly in the Northern Hemisphere. K2019 noted that surface winds over major mid-latitude mountain regions change very little between the El Niño and La Niña simulations, and the change in orographic GWP may be mainly due to altered mean flow filtering. Shifting the zero line of the zonal wind also allows deeper penetration of orographic waves and spreads the westward forcing from the waves to higher altitudes.

          The overall characteristics of resolved wave forcing and orographic GWP differences in the off-equatorial UTLS region

110      are qualitatively similar among models, except orographic GWP in LMDz. In GISS, LMDz, and MRI, relatively large westward forcing anomalies by resolved waves extending from the lower to the upper stratosphere coincide with large easterly mean wind differences around 60°N (Fig. S3).

          The residual mean circulation differences (4th column) show positive and negative anomalies in the Northern Hemisphere and Southern Hemisphere, respectively. Consistent with wave forcing differences by both resolved and parameterized waves,

115      MIROC-AGCM and MIROC-ESM display differences of the residual mean circulation that are mainly found in the shallow branch of the BDC, while LMDz and MRI exhibit larger differences and that extend to the deep branch of the BDC.

          These characteristics found in the models are qualitatively similar to those in ERA5 (note again that ERA5 provides only sum of parameterized zonal forcing).
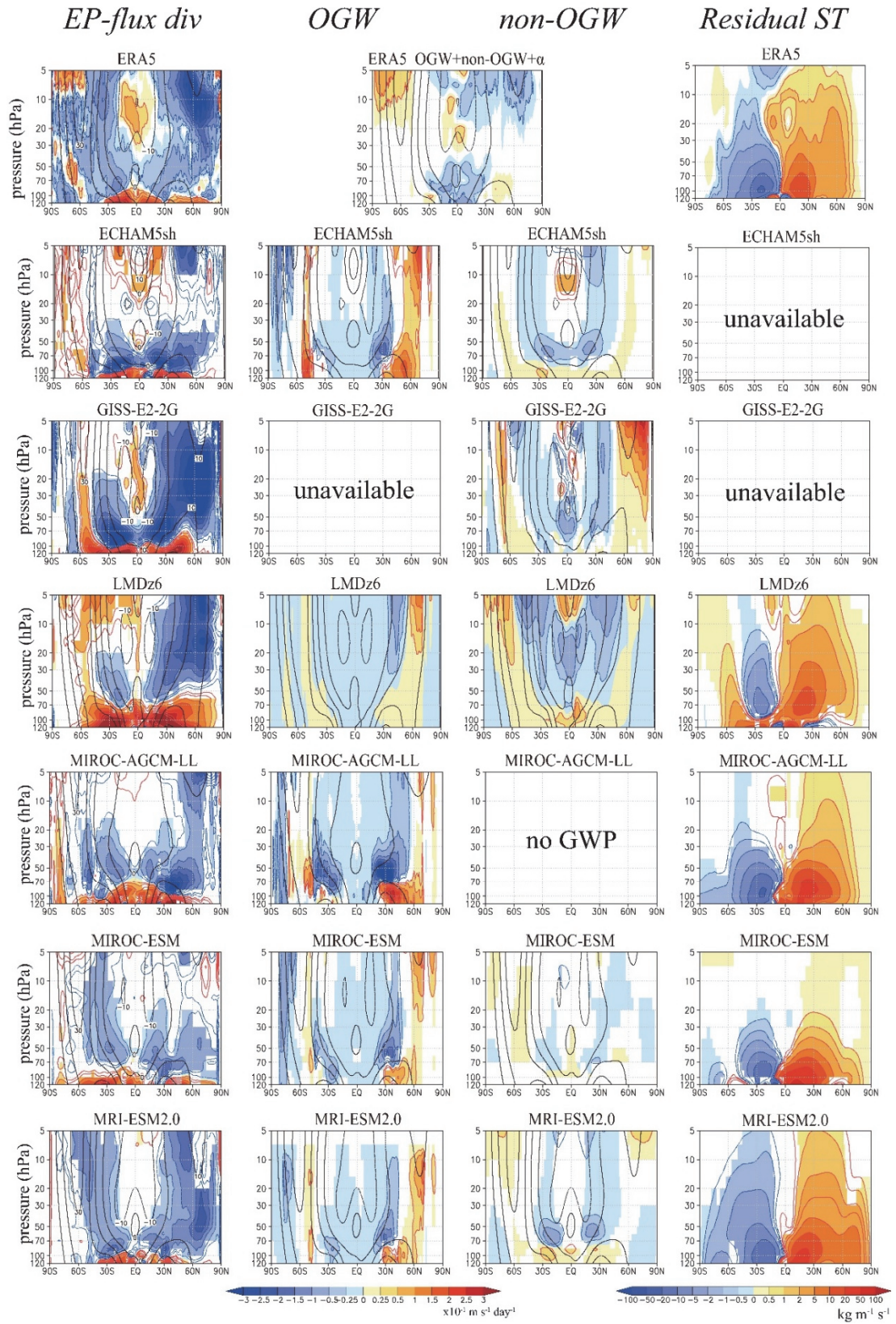
**Figure S4.** The color shading shows El Niño minus La Niña annual mean differences of: (left column) EP-flux divergence, and zonal momentum forcing due to (2nd column) orographic and (3rd column) non-orographic GWP, and (4th column) residual mean circulation streamfunction. For the first 3 columns the mean zonal wind is contoured with intervals of 10 m s$^{-1}$. Color intervals are at ±0.25, ±0.5, ±1.0, ±1.5, ±2.0, ±2.5, and ±3.0 × 10$^{-1}$ m s$^{-1}$ day$^{-1}$ for the first 3 columns and ±0.5, ±1.0, ±2.0, ±5.0, ±10, ±20, ±50, and ±100 kg m$^{-1}$ s$^{-1}$ for the 4th column. Top panels show corresponding results from ERA5. ERA5 includes unresolved zonal parameterized forcing, corresponding to the sum of the 2nd and 3rd column plus additional forcing. For ERA5, differences exceeding 0.1 × 10$^{-1}$ m s$^{-1}$ day$^{-1}$ in zonal forcing and 0.2 kg m$^{-1}$ s$^{-1}$ in the stream function are shaded. In the model results, color shading is shown only where differences in El Niño minus La Niña values are statistically significant at a confidence level ≥ 95%.

6

**Section S5: ENSO modulation of the QBO in the QBOi Experiment 1**

130    A complementary approach is provided by QBOi Experiment 1 (Exp1), an AMIP-type simulation conducted under the QBOi protocol. Exp1 used observed SSTs, sea ice, and radiative forcings from January 1979 to February 2009 (Butchart et al., 2018). The integration period for each ensemble member is 31 years. ECHAM5sh, LMDz, and MRI-ESM2.0 contributed one ensemble member each, while MIROC-AGCM, MIROC-ESM, and CESM1 each contributed three ensemble members. El Niño and La Niña months were identified using the ENSO definition provided by the Japan Meteorological Agency (JMA),
135    consistent with the method used for the observational analyses. It should be noted that the number of El Niño and La Niña samples in Exp1 is considerably smaller than in our new prescribed ENSO experiments.

Figure S5 shows the distribution of QBO phase progression rates from Exp1, classified by QBO phase and season. Figure S6 shows the corresponding distribution of QBO amplitude. Observational results, previously presented in Figs. 5 and 6 of the main text, are here repeated to facilitate comparison. These Exp1 results can be compared with those from the ENSO
140    experiments. To summarize the differences, Figure S7 presents simple averages of QBO amplitude and phase progression rate across all QBO phases and seasons, for each model and experiment type, along with the corresponding observational values.

The phase progression rate exhibits greater consistency between experiments than does the QBO amplitude. Most models, with the exception of ECHAM5, show a greater number of segments with faster QBO descent during El Niño than during La Niña in both experiments. This indicates that enhanced QBO phase progression under El Niño conditions is a robust feature
145    across models and forcing types.

The amplitude differences in Exp1 exhibit substantial inter-model variability, similar to the ENSO experiments. For instance, LMDz shows consistently stronger amplitude during El Niño across all segments in the ENSO experiments and in most segments in Exp1. In contrast, ECHAM5sh shows weaker amplitude during El Niño in all segments of the ENSO experiments but displays a less coherent pattern in Exp1. This likely reflects unrealistic QBO behavior in the El Niño
150    simulation of ECHAM5sh (Fig. 2), where the EOF phase-space pattern deviates from a well-defined ellipse. MIROC-ESM shows generally weaker amplitude during La Niña in both experiments. MIROC-AGCM contains segments with stronger amplitude during El Niño, though these are typically not statistically significant. In most models, the number of segments with significant amplitude differences is smaller in Exp1 than in the ENSO experiments.

When averaged across all QBO phases and seasons, all models except ECHAM5 exhibit faster QBO descent during El
155    Niño in both experiments. In LMDz and MIROC-AGCM, this acceleration is stronger in the ENSO experiments. In EMAC, it is more prominent in Exp1. MIROC-ESM shows inconsistent behavior between the two experiments, with a slightly larger difference in Exp1 segment averages. MRI-ESM2.0 and CESM1 show minimal differences between the two experiments.

Regarding QBO amplitude, the observation shows slightly stronger QBO amplitude during La Niña. Among the models, EMAC and LMDz show stronger amplitude during El Niño in both experiment types, while MIROC-ESM and CESM1
160    consistently show weaker amplitude during El Niño. EMAC, LMDz, and CESM1 exhibit larger El Niño–La Niña amplitude differences in the ENSO experiments than in Exp1, whereas MIROC-ESM shows the opposite tendency, with slightly larger differences in Exp1, although the overall magnitude of the difference remains small. MIROC-AGCM and MRI show very small amplitude differences under both conditions. ECHAM shows a reversal in the sign of the amplitude difference, but interpretation is complicated by its unrealistic QBO structure during El Niño. Overall, the amplitude response shows substantial
165    model dependence. No consistent pattern is found across the ensemble, and the degree of sensitivity to SST amplitude varies considerably.

It should also be noted that differences between Exp1 and the ENSO experiments reflect not only SST amplitude but also fundamental aspects of experimental design. Exp1 incorporates year-to-year variability in sea surface temperatures as well as external forcings such as volcanic eruptions and solar fluctuations, and it includes a much smaller number of El Niño and La
170    Niña months. These factors introduce sampling uncertainty and complicate direct comparisons with the perpetual ENSO experiments, in which ENSO signals are isolated under fixed boundary conditions and longer integration periods.

7

Against this background, the present results do not imply that the modulation of the QBO by ENSO forcing scales linearly with the magnitude of SST anomalies. While the tendency for faster QBO descent during El Niño appears robust across models and experiment types, the magnitude of the phase progression difference does not exhibit a consistent linear relationship with SST amplitude. In K2019, both MIROC-AGCM and MIROC-ESM demonstrated that stronger SST anomalies were associated with larger QBO period differences between El Niño and La Niña, suggesting a degree of sensitivity. However, this result could not be confirmed in Exp1, where MIROC-ESM failed to reproduce a similarly robust signal. This inconsistency highlights the limitation of inferring SST sensitivity from relatively short-duration simulations with limited samples.

A more definitive assessment of the influence of SST amplitude on QBO modulation would require a dedicated set of sensitivity experiments, ideally involving long integrations (e.g., 100 years) under systematically varied SST anomaly strengths (e.g., ×1.0, ×1.5, ×2.0, ×2.5).
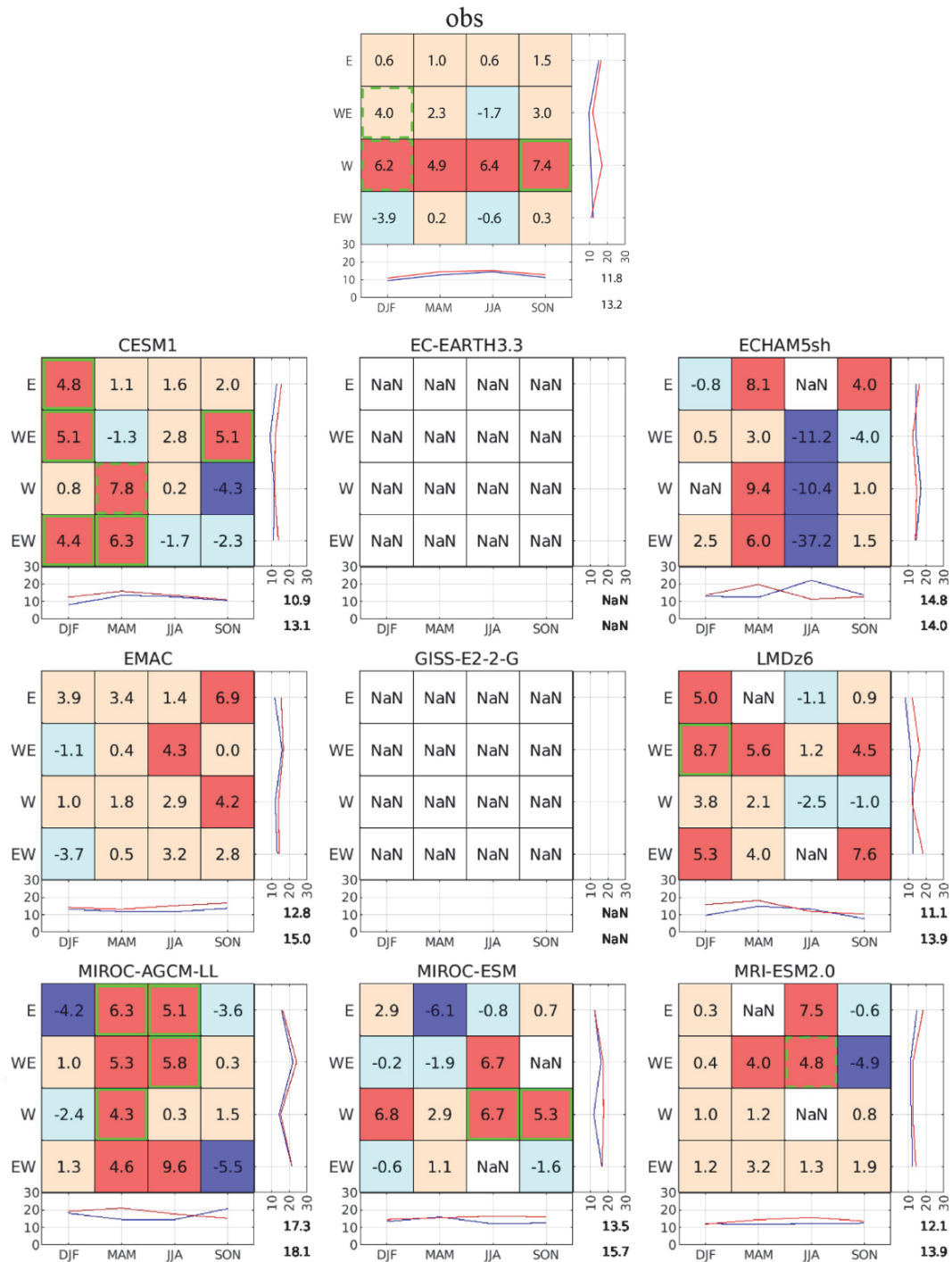


**Figure S5: Same as Fig. 5 but for results for the QBOi experiment 1.**
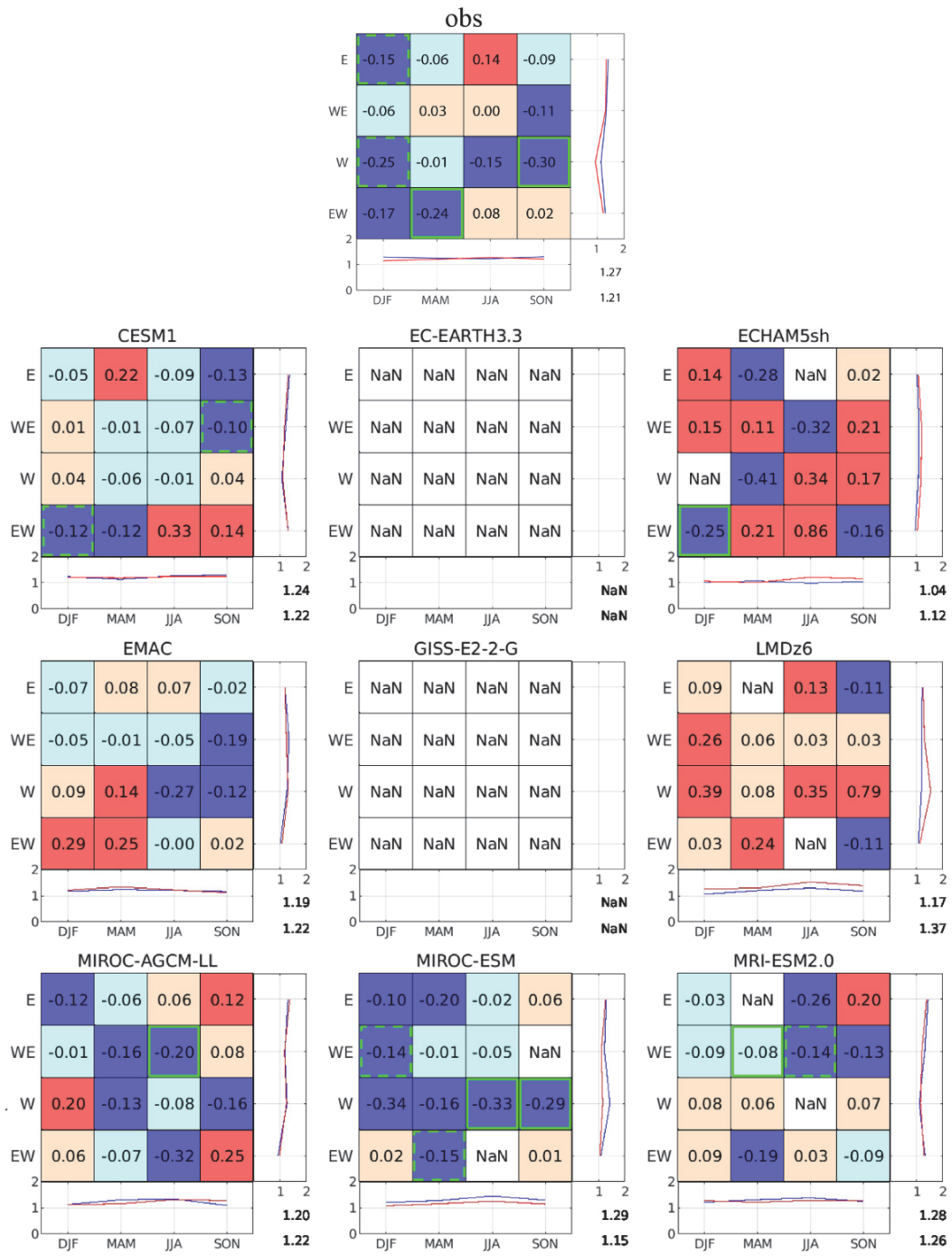
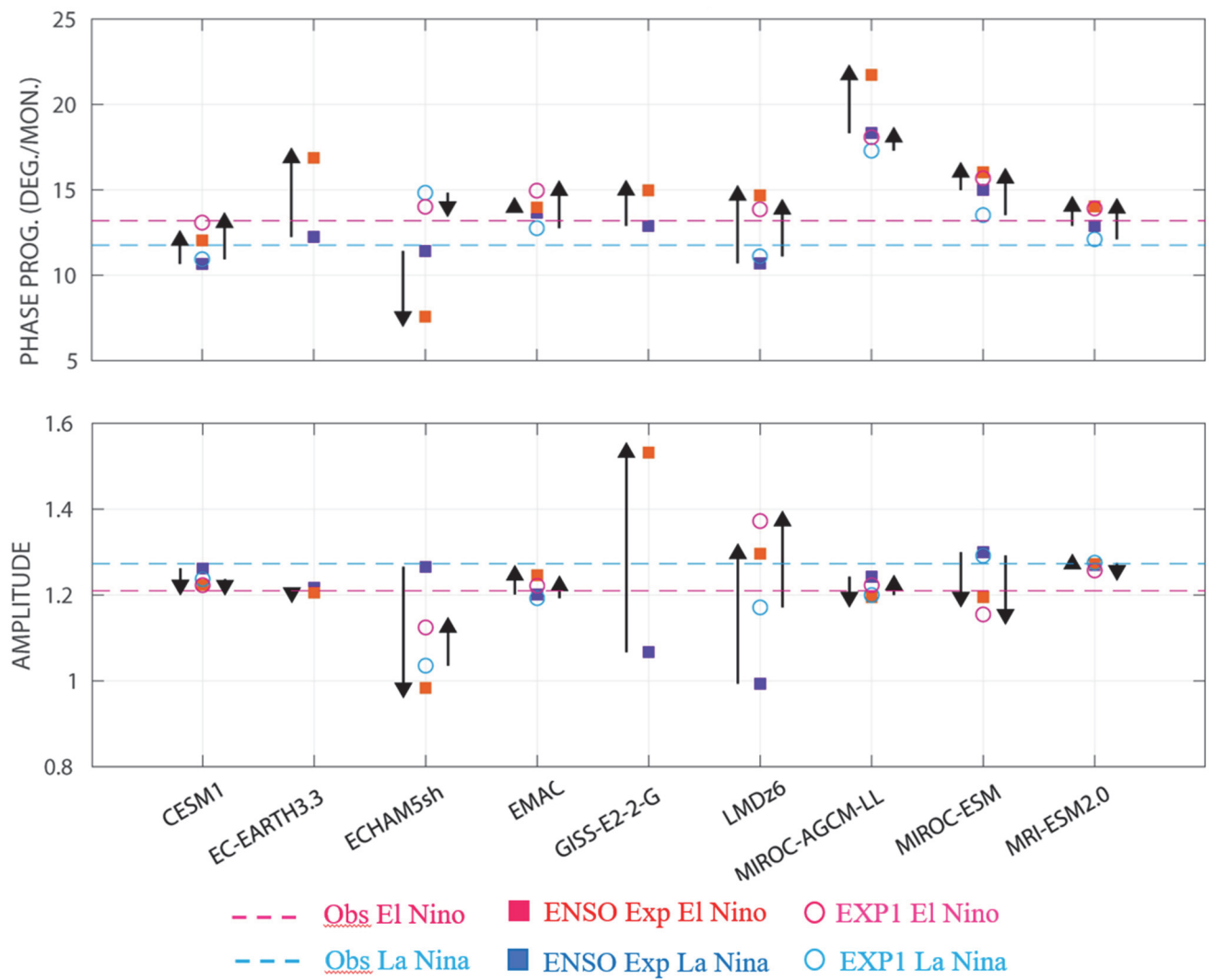**Figure S6: Same as Fig. 6 but for results for the QBOi experiment 1.**

**Figure S7: Mean QBO amplitude and phase progression rate in the ENSO experiments and Exp1, along with FUB observations. Dashed lines represent FUB observations. Filled squares indicate results from the ENSO experiments, while open circles represent Exp1. Upward-pointing vectors correspond to stronger amplitude or faster phase progression during El Niño compared to La Niña, and vice versa for downward-pointing vectors.**

**References cited only in the supplement**

Domeisen, D. I. V., Garfinkel, C. I. and Butler, A. H.: The teleconnection of El Niño Southern Oscillation to the stratosphere. Rev. Geophys., 57, 5-47, 2019.