JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD032362

Progress in Simulating the Quasi-Biennial Oscillation in CMIP Models

Jadwiga H. Richter[1], **James A. Anstey**[2], **Neal Butchart**[3], **Yoshio Kawatani**[4], **Gerald A. Meehl**[1], **Scott Osprey**[5,6], **and Isla R. Simpson**[1]

[1]Climate and Global Dynamics, National Center for Atmospheric Research, Boulder, CO, USA, [2]Canadian Centre for Climate Modelling and Analysis (CCCma), Environment and Climate Change Canada, Victoria, British Columbia, Canada, [3]Met Office Hadley Centre, Exeter, UK, [4]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan, [5]National Centre for Atmospheric Science, Leeds, UK, [6]Department of Physics, University of Oxford, Oxford, UK

Key Points:
- The number of models with a QBO has tripled from CMIP5 to CMIP6
- The mean period of the QBO is well represented in CMIP5 and CMIP6 models
- CMIP5 and CMIP6 models generally are deficient in representing the QBO amplitude at all levels below 20 hPa

Correspondence to:
J. H. Richter,
jrichter@ucar.edu

Citation:
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 Geophysical Research: Atmospheres*, *125*, e2019JD032362. https://doi.org/10.1029/2019JD032362

Received 30 DEC 2019
Accepted 23 MAR 2020
Accepted article online 7 APR 2020

Author Contributions
Conceptualization: Jadwiga H. Richter, Neal Butchart, Yoshio Kawatani, Scott Osprey
Data curation: Yoshio Kawatani, Gerald A. Meehl, Scott Osprey, Isla R. Simpson
Methodology: Jadwiga H. Richter, Neal Butchart
Software: James A. Anstey
Writing - Original Draft: Jadwiga H. Richter, Gerald A. Meehl
Formal Analysis: James A. Anstey
Writing - review & editing: James A. Anstey, Scott Osprey, Isla R. Simpson

Abstract The quasi-biennial oscillation (QBO) of the zonal mean zonal wind is the primary mode of variability in the tropical lower stratosphere. The QBO is characterized by alternating easterly westerly shear layers that descend down from ~10 to 100 hPa. The QBO is also seen in lower stratospheric tropical temperature, water vapor, and ozone and affects tropospheric variability through various teleconnections. We examine here the progress in simulating the QBO in the Coupled Model Intercomparison Project (CMIP) models, more specifically in CMIP3, CMIP5, and CMIP6 models. We show that the number of models that are able to simulate the QBO has increased from 0 in CMIP3, to 5 in CMIP5, to 15 in CMIP6. While the number of models with an internally generated QBO has tripled from CMIP5 to CMIP6, the fidelity of the simulation averaged over the CMIP models has not improved. We show that CMIP5 and CMIP6 models represent the QBO period and latitudinal extent quite well; however the QBO amplitude is shifted upwards relative to observations resulting in large underestimation of QBO amplitude at all levels below 20 hPa. The underestimation of QBO amplitude in the lowermost stratosphere and lack of variations downward to the tropopause and below will likely impact the quality of teleconnections seen in the current generation Earth system models.

Plain Language Summary The quasi-biennial oscillation (QBO) is an important mode of variability in the tropical lower stratosphere and impacts variability in the stratosphere and troposphere. We show that the number of climate or Earth system models being able to simulate the QBO in Coupled Model Intercomparison Project (CMIP) models has increased by a factor of 3 from CMIP5 to CMIP6. However, the quality of the simulation of the QBO has not improved. Overall, models simulate the period of the QBO well but underestimate the QBO amplitude at all levels below 20 hPa.

1. Introduction

The QBO is a prominent feature of the equatorial lower stratosphere and upper troposphere, with alternating easterly and westerly zonal mean wind shear zones descending from ~10 to 100 hPa. The period of the observed QBO ranges from about 20 to 32 months with an average of ~28 months (Baldwin et al., 2001; Bushell et al., 2020). The QBO impacts the variability of many parts of the Earth system. It directly influences the lower tropical stratospheric and upper tropospheric temperature and chemical constituents such as water vapor, ozone, and methane (Randel & Wu, 1996; Randel et al., 1998). In addition, the QBO impacts the stratosphere and troposphere through teleconnections. The QBO affects the variability of the stratospheric polar vortex through the so-called Holton-Tan effect (Holton & Tan, 1980), which may also be modulated by the solar cycle (Anstey & Shepherd, 2014; Labitzke, 1987; Labitzke & van Loon, 1988). The QBO effects on the polar vortex manifest themselves in the troposphere as changes in the variability of the North Atlantic Oscillation (NAO), with the QBO West-East difference resembling the positive phase of the NAO (Anstey & Shepherd, 2014). The QBO also affects extratropical storm tracks (Wang et al., 2018) and the Pacific subtropical jet (Garfinkel & Hartmann, 2010). In the tropics, studies have demonstrated the influence of the QBO on convection (Collimore et al., 2003; Liess & Geller, 2012), tropical cyclone tracks in the western North Pacific (Ho et al., 2009), and recently, several studies have found that in observations, the QBO modulates the amplitude of the Madden-Julian oscillation (MJO) in boreal winter, and this influence is larger than that from El Niño–Southern Oscillation (ENSO) (Son et al., 2017; Yoo & Son, 2016).

Coupled Model Intercomparison Project (CMIP) is run by the World Climate Research Programme (WCRP) Working Group on Coupled Modelling (WGCM) with the central goal of advancing scientific understanding of the Earth system. Since 1995, CMIP has coordinated climate model experiments involving multiple international modeling teams worldwide and has developed in phases (Meehl et al., 1997, 2007; Taylor et al., 2012). As the field of Earth system modeling has matured, so has the number of models involved with CMIP, ranging from 20 models in CMIP1 and CMIP2 to 23 models in CMIP3, to 42 models in CMIP5, and to over 60 models in the sixth phase, CMIP6, now in progress (Eyring et al., 2016). CMIP has led to better understanding past, present, and future climate change and variability within multimodel frameworks defining common experiment protocols, forcings, and outputs.

From its inception, CMIP was formulated to utilize the current state-of-the-art global coupled climate models developed by international modeling groups for scientific research. The development of these models is a balance between model complexity, model resolution (horizontal and vertical), and available computing resources, to allow multicentury simulations. Over time, modeling groups have strived to include added complexity and resolution, including in the stratosphere, as stratospheric dynamics were viewed as making important contributions to the climate system simulations. At the time of CMIP3, only a few so-called high-top Earth system models (ESMs) appeared with added vertical resolution to better resolve stratospheric dynamics and were being run as experimental versions. For example, HadGEM1 did not simulate a QBO in CMIP3 but did in a subsequent 60-level configuration (Osprey et al., 2010; Schenzinger et al., 2017). In CMIP5, only five models were able to simulate a QBO, which in large part motivated the Stratosphere Processes and their Role in Climate (SPARC) QBO initiative (QBOi) to improve the simulation of tropical stratospheric variability in general circulation models (GCMs) and ESMs (Butchart et al., 2018). As part of QBOi, simulations of the QBO were examined in 13 atmosphere-only models run under present-day climate (Bushell et al., 2020), and 11 models' QBO response to doubling and quadrupling CO_2 was examined by Richter et al. (2020). In addition, the proximal forcing of the QBO was examined by Holt et al. (2019). Modeling centers participating in QBOi were attentive to the representation of their models stratosphere, and the simulation of the QBO in particular. In most models, a parameterization of nonorographic gravity waves (GWs) is required in order to simulate the QBO. Among QBOi models, parameterized GWs provide 40% to 80% of the momentum forcing required to drive the QBO, and these parameterizations are highly tuned in order to obtain the correct QBO characteristics (Bushell et al., 2020; Butchart et al., 2018).

It is only with the generation of ESMs now being run in CMIP6 that "high-top" atmospheric models are more of a standard feature as faster supercomputers allow increased complexity and vertical resolution, with the latter being crucial for the simulation of the QBO (Geller et al., 2016; Richter et al., 2014). Additionally, the importance of stratospheric processes, including the QBO, to credible simulations of Earth system variability and skillful near-term climate predictions has been demonstrated in numerous studies (Anstey & Shepherd, 2014; Scaife, Arribas, et al., 2014; Scaife, Athanassiadou, et al., 2014; Wang et al., 2018). Presently, there are sufficient QBO-simulating ESMs to perform multimodel comparisons within the context of CMIP. Here, we examine the key modeled characteristics of the QBO within CMIP6, comparing those derived for CMIP5.

2. Models and Observations

The primary quantity analyzed throughout this study is monthly mean zonal mean zonal mean wind and temperature for the time period of 1979 to 2014 from CMIP models. We combine output from historical simulations for years 1979–2005 with years 2006–2014 of the Representative Concentration Pathway 8.5 (RCP8.5) simulations for 35 CMIP5 models. For CMIP6, we use output from historical simulations from 1979 to 2014 for the 30 models that were available at the time of the preparation of this manuscript. Since none of the CMIP3 models had an internally generated QBO, we do not examine these simulations in detail; however, we still use output from historical simulations from 22 models, years 1979 to 1999, to compare the tropical stratospheric variability across the CMIP3 ensemble. All CMIP models used in this study are listed in Table 1. We use the first available ensemble member for all models, which gives all models equal weight in the multimodel mean. QBO characteristics are very consistent across individual ensemble members. We use ERA-Interim reanalysis (ERAI, Dee et al., 2011) as the observational baseline. CMIP6 output is available on eight stratospheric pressure levels: 1, 5, 10, 20, 30, 50, 70, and 100 hPa. We interpolate all other model output and reanalysis to these levels for consistency.

For reference, we also show the amplitude of the QBO from Experiment 1 of QBOi (Bushell et al., 2020; Butchart et al., 2018). Thirteen models carried out this experiment (60LCAM5, AGCM3-CMAM,

Table 1
Summary of Models Used From CMIP3, CMIP5, and CMIP6

#	CMIP3	CMIP5	CMIP6
1	BCCR-BCM2-0	ACCESS1-0	**AWI-CM-1-1-MR**
2	CCMA-CGCM3-1	ACCESS1-3	**BCC-CSM2-MR**
3	CCMA-CGCM3-1-T63	bcc-csm1-1	BCC-ESM1
4	CNRM-CM3	bcc-csm1-1-m	CAMS-CSM1-0
5	CSIRO-MK3-0	BNU-ESM	CanESM5
6	GFDL-CM2-0	CanESM2	CESM2
7	GFDL-CM2-1	CCSM4	**CESM2-WACCM**
8	GISS-AOM	CESM1-CAM5	**CNRM-CM6-1**
9	GISS-MODEL-E-H	CESM1-WACCM	**CNRM-ESM2-1**
10	GISS-MODEL-E-R	CMCC-CM	**E3SM-1-0**
11	IAP-FGOALS1-0-G	**CMCC-CMS**	**EC-Earth3**
12	INGV-ECHAM4	CNRM-CM5	**EC-Earth3-Veg**
13	INMCM3-0	CSIRO-Mk3-6-0	FGOALS-g3
14	IPSL-CM4	FGOALS-g2	GFDL-CM4
15	MIROC3-2-HIRES	FIO-ESM	**GFDL-ESM4**
16	MIROC3-2-MEDRES	GFDL-CM3	GISS-E2-1-G
17	MPI-ECHAM5	GFDL-ESM2G	GISS-E2-1-G-CC
18	MRI-CGCM2-3-2A	GFDL-ESM2 M	GISS-E2-1-H
19	NCAR-CCSM3-0	GISS-E2-H	**HadGEM3-GC31-LL**
20	NCAR-PCM1	GISS-E2-R	**IPSL-CM6A-LR**
21	UKMO-HadCM3	HadGEM2-AO	MCM-UA-1-0
22	UKMO-HadGEM1	**HadGEM2-CC**	**MIROC6**
23		HadGEM2-ES	MIROC-ES2L
24		inmcm4	**MPI-ESM1-2-HR**
25		IPSL-CM5A-LR	**MRI-ESM2-0**
26		IPSL-CM5A-MR	NESM3
27		IPSL-CM5B-LR	NorCPM1
28		MIROC5	NorESM2-LM
29		**MIROC-ESM**	SAM0-UNICON
30		**MIROC-ESM-CHEM**	**UKESM1-0-LL**
31		MPI-ESM-LR	
32		**MPI-ESM-MR**	
33		MRI-CGCM3	
34		NorESM1-M	
35		NorESM1-ME	

Note. Models that are able to simulate the QBO are marked in bold.

CESM1 (WACCM5-110L), ECHAM5sh, EMAC, HadGEM2-A, HadGEM2-AC, LMDz6, MIROC-AGCM-LL, MIROC-ESM, MRI-ESM2, UMGA7, and UMGA7gws), which consisted of a simulation with observed prescribed sea surface temperatures (SSTs) and sea ice as well as external forcings for the period between January 1979 and February 2009. Details of the QBOi models and experiments can be found in Butchart et al. (2018) and Bushell et al. (2020).

3. QBO Metrics

The observed QBO, with the exception of the interrupted cycle in year 2016, which is not being considered here, is a very regular oscillation, and various commonly used methodologies to arrive at basic QBO metrics, such as QBO period and amplitude, result in the same answer. However, due to the deficiencies in the

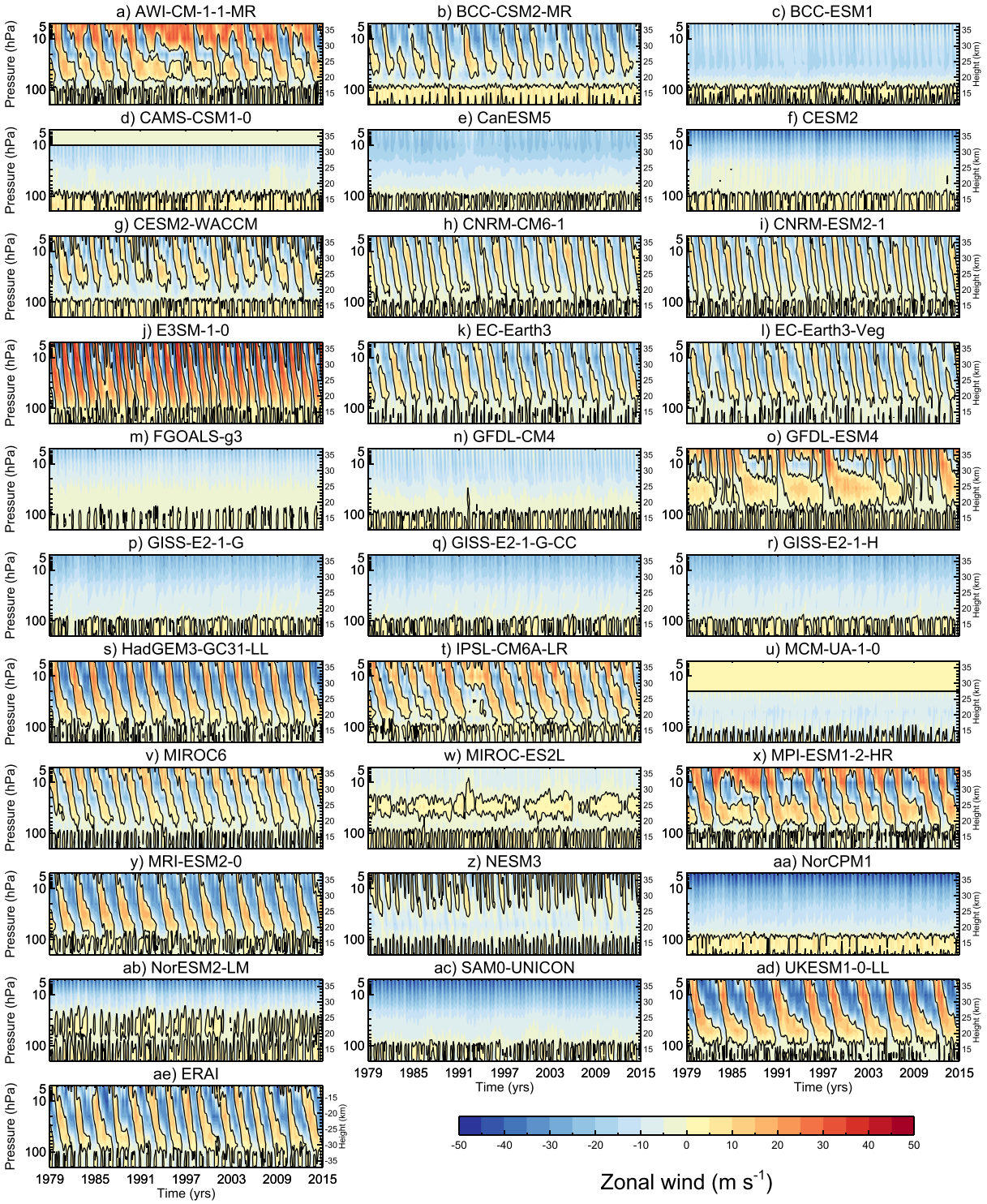

Figure 1. Zonal mean zonal wind averaged between 5°S and 5°N as function of time and pressure for CMIP6 models and ERAI reanalysis in m s^{-1}. Model names are specified in panel titles.

simulation of the QBO in models, different methods of arriving at QBO characteristics sometimes lead to different values, especially in a warming climate (Bushell et al., 2020; Richter et al., 2020). Schenzinger et al. (2017) defined metrics of the QBO for use in climate models and applied them to CMIP5 models and models submitted to Phase 2 of the Chemistry-Climate Model Validation Activity (SPARC, 2010). We use the same metrics here to evaluate QBOs in CMIP models, as well as several additional ones that have been demonstrated to be particularly useful in determining properties of less regular QBOs by Bushell et al. (2020) and by Richter et al. (2020). Schenzinger et al. (2017) defined over 20 metrics of the QBO. Here, we focus only on key metrics that capture the most important properties of the QBO and are of interest to a wide audience. We focus here on QBO metrics derived from the zonal mean zonal wind, as the QBO signature is clearest in this quantity. We show QBO amplitude also in zonal mean temperature; however, this quantity is influenced by multiple factors in the lower stratosphere, such as changes in the seasonal cycle and volcanic eruptions; hence, the QBO signal cannot be as clearly isolated as in the zonal mean zonal wind.

3.1. Definition of the QBO

Figure 1 shows the 5°S to 5°N averaged zonal mean zonal wind in the stratosphere and upper troposphere for CMIP6 models. A similar illustration is shown in Figure 1 of Butchart et al. (2018) for CMIP5 models. Before QBO metrics can be defined, a choice needs to be made with regard to which models we consider to have a QBO. In observations, the QBO is characterized by alternating descending easterly and westerly shear zones, between ~10 and 100 hPa, with an average period of 28 months. Visual inspection of Figure 1 shows that tropical stratospheric zonal mean wind oscillations exist in many models, but their characteristics greatly vary. For example, the time between individual oscillation cycles is on average as short as 16 months in E3SM-1-0 and as long as 70 months in GFDL-ESM. In some models, easterly and westerly layers descend regularly (e.g., EC-Earth3 and HadGEM3-GC31-LL), and in some, the westerlies often stall (e.g., in AWI-CM-1-1-MR and MPI-ESM1-2-HR). Some models show alternating easterlies/westerlies; however, the vertical range and amplitude of the oscillations are much smaller than in observation (e.g., NESM3 and NorESM2-LM). Here, we consider a model to have a QBO if the 5°S to 5°N averaged zonal mean zonal wind time series between 100 and 10 hPa exhibits alternating easterlies and westerlies, and the standard deviation of the deseasonalized time series is at least 20% of that seen in observations at all levels between 70 and 10 hPa. The latter criterion eliminates models that have a very weak amplitude oscillation that does not resemble the observed QBO, such as NESM3 and NorESM2-LM, from being classified as simulating a QBO. Models that simulate the QBO according to the above criteria are marked in bold in Table 1. The majority of the models marked as having a QBO easily exceed the 20% amplitude criterion. The only exception to that is the GFDL-ESM4 model, which would not be classified as having a QBO if the amplitude criterion was raised to 25% of the observed value.

3.2. QBO Period

The period is one of the key characteristics of the QBO. We derive it here using two methods:

- QBO transition times (TT): Transitions from westerlies to easterlies and easterlies to westerlies are identified in the 5°S to 5°N averaged zonal mean zonal wind time series at 10 hPa. A period of each QBO cycle is then the difference in time between every other phase change. Using the distribution of QBO periods derived in this manner, the minimum, maximum, and mean periods are calculated. This calculation follows Schenzinger et al. (2017), except here we pick the reference level of 10 hPa for all models, irrespective of whether the amplitude maximum is at this level or not. Since most models have a maximum amplitude near 10 hPa, this choice does not really affect the calculation results, and using a consistent level across models allows for a more fair comparison of various methods of period and amplitude estimation.
- Fast Fourier transform (FFT): FFT analysis is performed on a 5°S to 5°N averaged zonal mean zonal wind deaseasonalized time series at 10 hPa, which was padded at the end with zeros to 10 times the length of original time series to increase the spectral resolution. The dominant QBO period is identified as the period at which the FFT spectrum peaks. The exact same method was applied to QBOi models by Richter et al. (2020). FFT spectra for ERAI and models with oscillations resembling ERAI typically have one large well-defined peak, and in those cases, the FFT-derived QBO period is very similar to that derived using the TT method. Models that have less regular tropical oscillations tend to show multiple spectral peaks of similar amplitude, and for those models, the FFT- and the TT-derived periods differ.

3.3. QBO Amplitude

QBO amplitude is another key characteristic of the QBO. We calculate it here using the following two methods:

- TT: The calculation is closely related to the methodology for TT period calculation described in the previous section. The westerly (easterly) amplitude associated with each QBO cycle is calculated by taking the maximum (minimum) value of the 5°S to 5°N averaged zonal mean zonal wind time series at 10 hPa in between the QBO phase changes. Mean westerly (easterly) amplitude is arrived at by averaging values from individual cycles. The total QBO amplitude is then estimated by adding half of the mean easterly amplitude to half of the mean westerly amplitude. This methodology also follows Schenzinger et al. (2017) with the exception of the calculation being done always at 10 hPa.
- Dunkerton and Delisi (DD): This calculation follows Dunkerton and Delisi (1985) who showed that when most of the variability in the monthly mean equatorial winds results from the QBO, the mean amplitude of the oscillation can be approximated by $\sqrt{2}\sigma$, where σ is the standard deviation of the deseasonalized time series of the monthly mean zonal mean wind (in our case, the 5°S to 5°N averaged zonal mean zonal wind). This calculation allows the mean QBO amplitude to be easily calculated as a function of latitude and height, even when QBOs are not regular and when cycles cannot be easily distinguished. Bushell et al. (2020) and Richter et al. (2020) showed that the DD-derived amplitude agrees well with the mean amplitude obtained from the TT method for simulations with well-defined QBOs. In addition, this diagnostic is a useful measure of tropical variability, even when a well-defined oscillation is not present.

3.4. Spatial Extent

Lastly, we evaluate the spatial structure of the QBO by calculating its latitudinal and vertical extent and the lowest level to which it reaches. The derivation of these metrics mainly follows the methodology presented by Schenzinger et al. (2017). First, the QBO zonal wind amplitude latitude-height structure, $A_u(\phi, z)$, is derived by summing the squares of the amplitudes for the Fourier harmonics corresponding of periods between the minimum and maximum QBO periods (derived from the TT method) and dividing that quantity by one half the field standard deviation at each spatial location. Subsequently, the following three quantities are derived from A_u:

- Vertical extent: is defined to be the full width at half maximum of the vertical profile of 5°S–5°N averaged A_u.
- Lowest level: is defined as the lowermost pressure level at which the amplitude falls to 10% of the maximum value of 5°S–5°N averaged A_u.
- Latitudinal extent: is defined to be the full width at half maximum of the latitudinal profile of A_u at 10 hPa (rather than at the level of maximum amplitude, as is done in Schenzinger et al., 2017).

Linear interpolation is used in the above calculations to arrive at values in between the coarse CMIP6 levels and in between the latitude grid of the data.

4. Results

4.1. QBO Period

According to the criteria discussed in section 3.1, 5 CMIP5 and 15 CMIP6 models have an internally generated QBO. Hence, the number of models with a QBO increased by a factor of 3 since CMIP5. Figure 2a compares the mean, minimum, and maximum QBO periods for these models. For the vast majority of models, the dominant QBO period derived using the FFT method (colored bars) is very similar to the mean period derived from the TT method (plus signs). However, for models for which the QBO is not very regular (e.g., CMCC-CMS in CMIP5 and GFDL-ESM4 in CMIP6), there is a discrepancy between the two methods. Overall, there is a much greater range of simulated QBO periods in the CMIP6 ensemble as compared to the CMIP5 ensemble. Deficiencies in the simulation of the QBO period in some models can also be seen by comparing the minimum and maximum period values derived using the TT method (open triangles). In observations, QBO cycles range from 22 to 33 months in length, whereas in CMIP5 models, the maximum QBO period reaches ~70 months (CMCC-CMS), and in CMIP6, some models have cycles as short as ~5 months and as long as 75 months (GFDL-ESM4). GFDL-ESM4 is also the model with amplitude very close to not meeting the QBO criteria. However, other CMIP6 models also show a very broad range of QBO periods, such as AWI-CM-1-1-MR, BCC-CSM2-MR, and MPI-ESM2-0. In the ensemble mean, the FFT-derived QBO period is 28.5 months for CMIP5 and 30.5 months for CMIP6 models, which are similar to

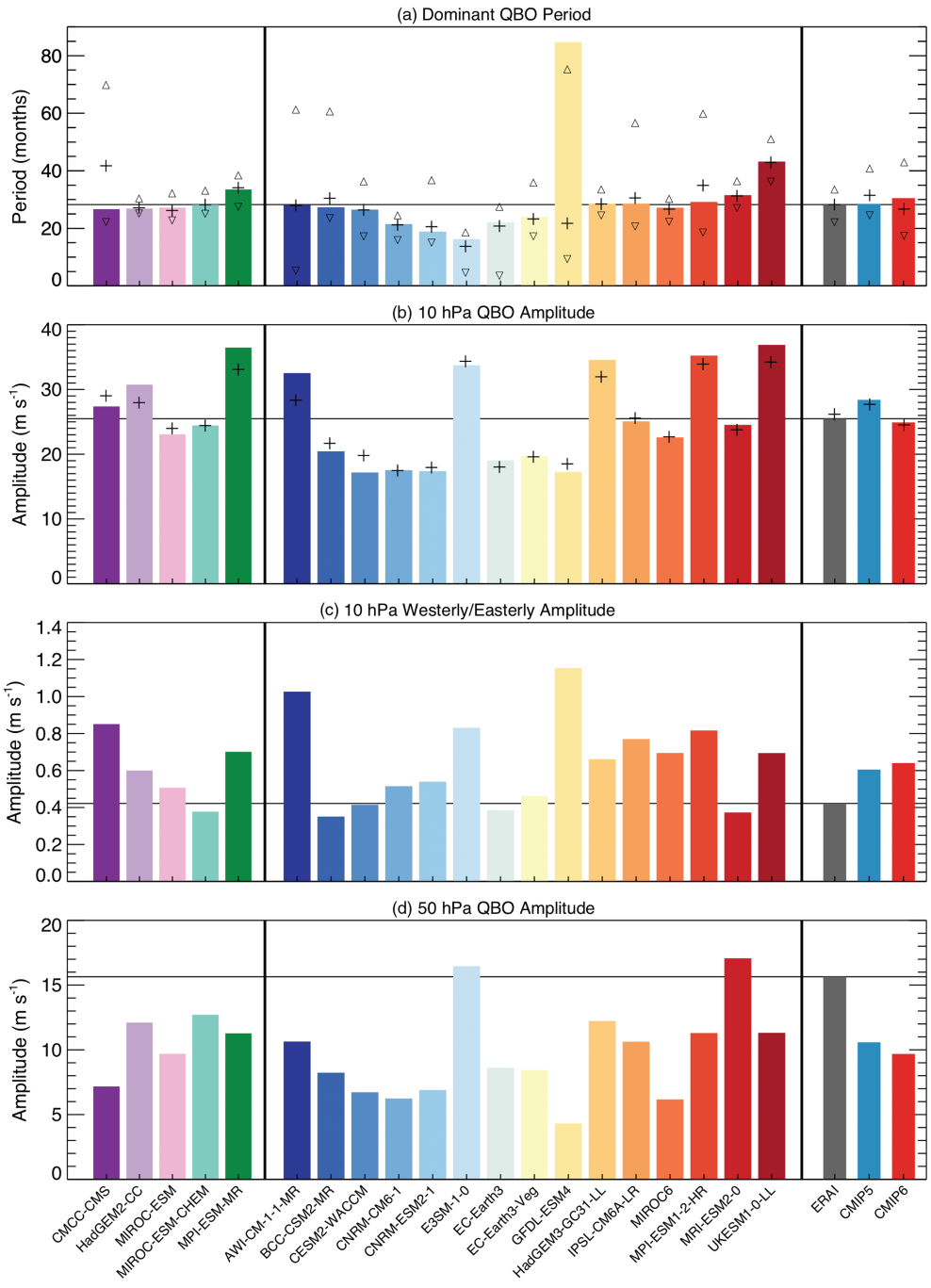

Figure 2. Dominant (a) QBO periods, (b) 10-hPa QBO amplitude, (c) 10-hPa ratio of westerly to easterly QBO amplitude, and (d) 50-hPa QBO amplitude for individual CMIP5 models (leftmost five models), individual CMIP6 models (navy through dark red), ERAI (gray), CMIP5 ensemble mean (blue), and CMIP6 ensemble mean (red). Ensemble means only include models with a QBO. Colors of bars have no particular meaning; they are there for ease of recognizing an individual model across different panels. In panel (a), colored bars correspond to QBO periods derived using the FFT method, and pluses and triangles correspond to the mean and min/max QBO period derived using the TT method, respectively. In panel (b), colored bars denote the 10-hPa amplitude derived using the DD method, whereas pluses denote the amplitude derived from the TT method. In panel (c), westerly and easterly QBO amplitude is derived from the TT method, and in panel (d), the QBO amplitude is derived using the DD method.

the observational QBO period of 28.3 months. Similarly, the TT-derived QBO mean period is 28.3 months for ERAI and 31.5 and 26.7 months for CMIP5 and CMIP6, respectively, showing that both CMIP5 and CMIP6 models, on average, represent the QBO period well. However, the differences between the mean maximum and mean minimum QBO periods are 16 months for CMIP5 and 26 months for CMIP6 models, which are ~1.5 times and over 2 times that of the observed difference of 11 months, showing that there is more variability in QBO periods in many models as compared to observations. Higher variability in QBO periods in CMIP models as compared to observations is also evident in the standard deviation (across QBO cycles) of the TT-derived periods, which is 5.5 and 6.6 months averaged over the CMIP5 and CMIP6 ensemble, respectively, and only 3.5 months for ERAI (not shown).

4.2. QBO Amplitude

In observations, the DD-derived QBO amplitude peaks at the equator at 20 hPa as illustrated in Figure 3a. Note that Bushell et al. (2020) reported that the ERAI QBO amplitude peaks at 15 hPa; however, the level of 15 hPa does not exist in the CMIP6 archive, and we reduced ERAI to the CMIP6 levels; hence, the peak is at 20 hPa instead of 15 hPa. The latitude/height structure of the QBO for CMIP5 and CMIP6 models with a QBO (Figure 3b) shows that in general, the QBO in models peaks at 10 hPa. The exception to that is MIROC-ESM-CHEM in the CMIP5 ensemble, which peaks at 15 hPa (Schenzinger et al., 2017), and MRI-ESM2-0 in CMIP6, which peaks at 20 hPa (not shown). Figures 3c and 3d further demonstrate that relative to ERAI, the QBO amplitude is generally strongly underestimated below 20 hPa and overestimated above 10 hPa in CMIP5 and CMIP6 models. In other words, the QBO amplitude of the CMIP models is shifted upward relative to observations. The amplitude underestimation below 20 hPa is slightly stronger in the CMIP6 ensemble. The difference between peak amplitude height between CMIP models and observations is distorted somewhat due to lack of model output at 15 hPa. In addition, comparison of Figures 3a–3d shows that the QBO amplitude decreases more quickly away from the equator in CMIP models as compared to observations, resulting in a narrower QBO. The underestimation of QBO width was also found by Bushell et al. (2020) for QBOi models.

Figure 2b compares the QBO amplitude at 10 hPa derived using the DD and TT methods for individual CMIP models. The two measures of amplitude shown yield similar values, however discrepancies of up to 4 $m\,s^{-1}$ exist. The QBO amplitude derived using DD and TT methods would match if the QBO time series resembled a sine wave, but since it does not (in either observations or the models), the shape of the oscillation will affect whether the DD-derived amplitude is smaller or larger that the TT-derived amplitude. Figure 2b shows that there is a large range of values of the QBO amplitude at 10 hPa among CMIP models, especially among CMIP6 models. However, in a multimodel ensemble mean, the CMIP5 and CMIP6 10-hPa QBO amplitudes are similar to that found in ERAI. Although the mean QBO amplitude is similar to observations, the split between easterly and westerly amplitude is not. Figure 2c shows that at 10 hPa in observations, the westerly QBO amplitude is about 40% of the easterly amplitude. Most models however overestimate the westerly amplitude and underestimate the easterly amplitude (not shown) resulting in a multimodel mean westerly to easterly amplitude ratio of ~0.6. As the QBO amplitudes using the TT method are derived from the raw (and not deseasonalized zonal mean zonal wind), another interpretation of this finding is that models tend to carry a mean westerly bias in the tropical stratospheric zonal mean zonal wind, which is confirmed by comparing the annually averaged tropical zonal mean zonal wind to observations (not shown).

Although the QBO amplitude peaks near 10 hPa, numerous studies have shown that the connection of the QBO to tropospheric variability is strongly correlated with QBO winds in the lowermost stratosphere. Teleconnections of the QBO to the polar vortex are most often examined at 50 hPa (e.g., Anstey & Shepherd, 2014). Studies showing the influence of the QBO on the MJO (e.g., Son et al., 2017; Yoo & Son, 2016) also use the 50-hPa QBO index. Figure 2d shows that the 50-hPa QBO amplitude is underestimated in all CMIP5 models and in all but two CMIP6 models (E3SM-1-0 and MRI-ESM2-2-HR). In an ensemble average, 50-hPa QBO amplitude is 68% of the observed value in CMIP5 models and only 62% of the observed value in CMIP6 models. Similar deficiencies in QBO amplitude in GCMs are found at 70 hPa (not shown).

Figures 3e and 3f summarize our findings regarding tropical variability in CMIP models. For models with QBOs, the DD method was demonstrated above to be a very good proxy of the QBO amplitude (Figure 3e). Applied to all models, with and without QBOs, the DD amplitude is a measure of tropical variability (Figure 3f). Figure 3e shows that below 30 hPa, the CMIP5 and CMIP6 ensembles are indistinguishable from each other as well as from the QBOi models. Between 10 and 30 hPa, the CMIP5 (CMIP6) QBO amplitudes

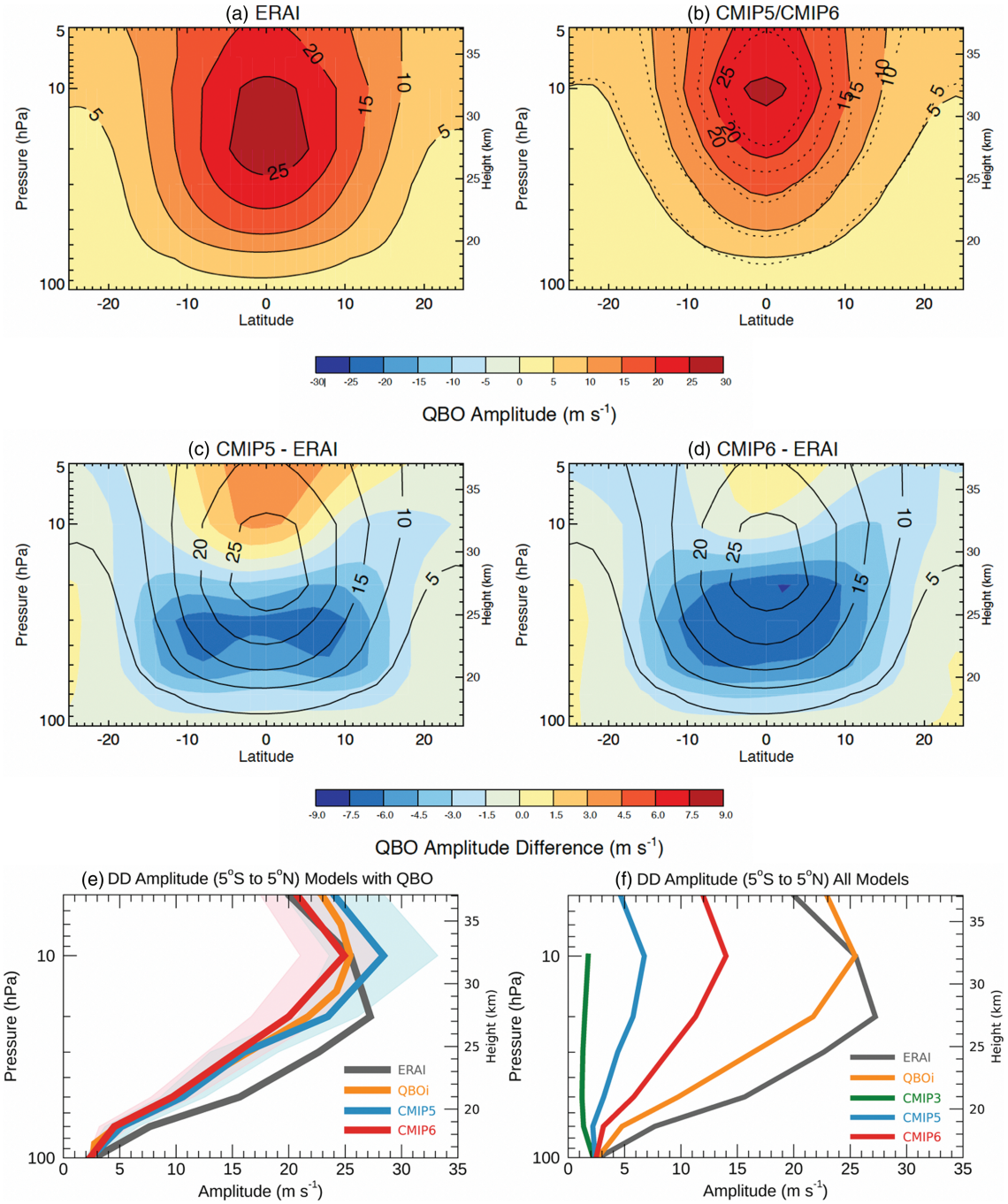

Figure 3. DD amplitude for (a) ERAI, (b) CMIP6 (shading and solid line) and CMIP5 (dotted line) ensembles, (c) CMIP5 ensemble—ERAI (shading), and (d) CMIP6 ensemble—ERAI (shading). Solid contours in panels (c) and (d) show the ERAI DD amplitude. Panel (e) shows the vertical profile of DD amplitude averaged between 5°S and 5°N for ERAI (black), QBOi models (orange), CMIP5 models with a QBO (blue), and CMIP6 models with a QBO (red). Blue (pink) shading represents the ±2 standard error (2 times the multimodel standard deviation divided by $\sqrt{n}$, where n is the number of models) for the CMIP5 (CMIP6) ensemble. Gray shading is the region where pink and blue shadings overlap. Panel (f) shows the same solid lines as in panel (e) but with DD amplitude calculated for all models in the CMIP5 and CMIP6 ensembles. In addition the CMIP3 ensemble is added in green.

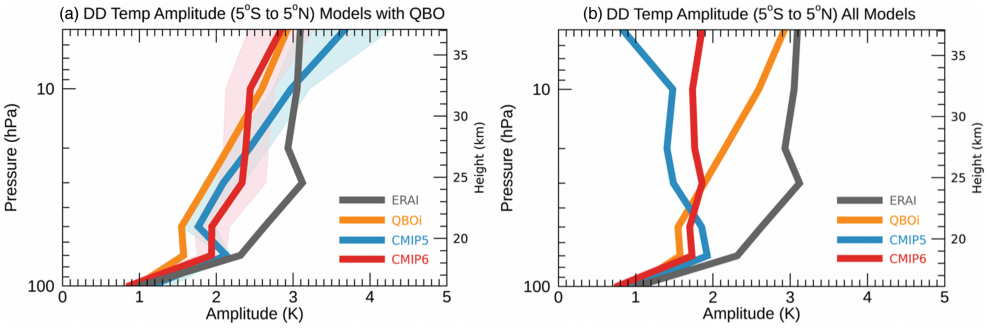

Figure 4. Same as Figures 3e and 3f but for temperature.

are somewhat higher (lower) than those of QBOi models; however, the distributions across models (indicated by the standard error shading) are overlapping, indicating that the CMIP5 and CMIP6 values are not significantly different. It is clear from Figure 3e that ERAI QBO amplitude lies outside of the CMIP5 and CMIP6 distribution at and below 20 hPa, showing that CMIP5 and CMIP6 models strongly underestimate the QBO amplitude in the lower stratosphere. Since all of the simulations considered in QBOi were done with prescribed SSTs, and CMIP5 and CMIP6 are coupled ocean-atmosphere simulations, the similarity of the amplitudes between these ensembles also implies that coupling to the ocean is not a likely cause of deficiencies in the QBO amplitude in CMIP models. Although no progress has been made in representing the QBO amplitude in CMIP models from CMIP5 to CMIP6, the tropical variability across the CMIP ensembles has quadrupled from CMIP3 to CMIP5 models and has nearly doubled from CMIP5 to CMIP6 models due to the greater fraction of models representing the QBO (Figure 3f).

As several of the teleconnections of the QBO to the troposphere are hypothesized to depend on the temperature perturbations associated with the QBO, Figures 4a and 4b show the QBO amplitude in temperature derived using the DD method, analogously to the DD amplitude in zonal mean wind shown in Figures 3e and 3f. For ERAI, variability within the range of typical QBO periods (as defined by a Fourier window of 20–40 months) accounts for approximately 80–90% of the DD amplitude for zonal-mean zonal wind in the 10- to 70-hPa layer, corroborating that the DD amplitude is a useful measure of QBO amplitude. This fraction is similar for zonal mean temperature near the altitudes where QBO amplitude peaks (20 and 30 hPa), but smaller at lower altitudes (65% at 70 hPa and 50% at 100 hPa), indicating that other variability besides the QBO contributes appreciably to the DD temperature amplitude at altitudes in the lower tropical stratosphere (not shown). One of the largest factors that impacts temperature perturbations below 50 hPa is volcanic eruptions, such as the eruption of Mt. Pinatubo in 1991. Figure 4 shows that the amplitude of the QBO signal in temperature above 70 hPa is much smaller in CMIP models than observed, consistent with what was shown in the zonal mean zonal wind in Figures 3e and 3f. However, below 70 hPa, the differences between the modeled and observed DD temperature amplitude are not as large, as only a small portion of that signal is coming from the QBO. Comparison of Figure 4b to Figure 4a shows that in all models, the DD amplitude vastly differs between models with a QBO and all models only for altitudes above 50 hPa, further showing that the temperature signal is clearly affected by the QBO in the middle to upper stratosphere. In addition, Figure 4a also shows that the temperature perturbations in the lowermost stratosphere is lower in QBOi models than CMIP6 models. This is due to the fact that only about half of the QBOi models include the effects of volcanoes and hence do not include the strong warming perturbations that are associated with their eruptions, which decreases the DD amplitude in temperature by 0.2 to 0.3 K in the lowermost tropical stratosphere (not shown).

4.3. QBO Spatial Extent

The QBO spatial extent characteristics, shown in Figure 5, are largely a reflection of and quantification of what was already observed in the latitude/height QBO amplitude figures. The latitudinal extent of the QBO at 10 hPa is underestimated by most models; however, the difference from observations is rather small (Figure 5a). Only two models, GFDL-ESM4 and IPSL-CM6A-LR, underestimate the QBO latitudinal extent by more than 10%. In a multimodel average, CMIP5 models underestimate the QBO latitude extent at 10 hPa by ~4% and CMIP6 models by ~8%.

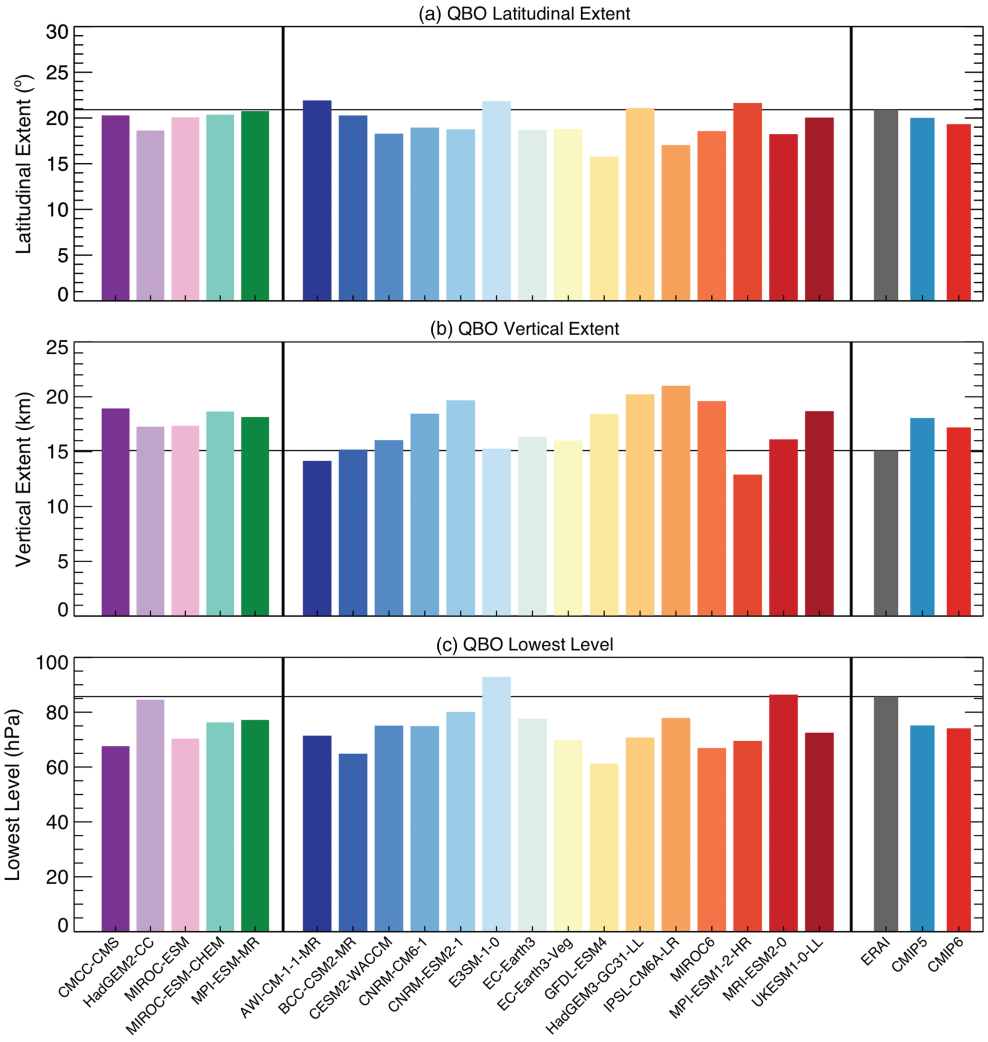

Figure 5. QBO (a) latitudinal extent (°) at 10 hPa, (b) vertical extent in log-pressure altitude (km), and (c) lowest level (pressure level at which amplitude drops to 10% of maximum amplitude). Model labels are the same as in Figure 2.

The vertical extent, or the depth of the QBO, is overestimated by all CMIP5 models and about half of the CMIP6 models; 25% of the CMIP6 models overestimate the vertical extent by over 25%, resulting in 15% overestimation in QBO vertical extent of the CMIP6 ensemble and 20% overestimation for the CMIP5 ensemble. This bias in QBO depth is manifested as the higher than observed QBO amplitudes above 10 hPa, especially in CMIP5 models (Figure 3c), or an upward shift of the QBO. The QBO lowest level, shown in Figure 5c, shows that despite the larger depth, the QBO does not reach far enough down into the lower stratosphere for a vast majority of CMIP5 and CMIP6 models. The exceptions to that are HadGEM2-CC in CMIP5 and E3SM-1-0 and MRI-ESM-2-0 out of the CMIP6 models. In a multimodel average, the discrepancy between the QBO lowest level in CMIP5 and CMIP6 models and observations is about 10 hPa or 12%, or roughly 1 km in log-pressure altitude.

5. Summary and Discussion

We have examined here the key characteristics of the QBO in coupled historical simulations, 1979 to 2014, in CMIP models. The number of models with an internally generated QBO rose from 0 in CMIP3, to 5 in CMIP5, to 15 in CMIP6. Despite the tripling of the number of models that are able to simulate the QBO, the quality of the simulation of the QBO from CMIP5 to CMIP6 has not changed. Most of the CMIP5 models capture the correct mean QBO period. With only a few exceptions, CMIP6 models also represent well the mean QBO period; however, the range of simulated periods is twice as large as in observations.

Both CMIP5 and CMIP6 models accurately represent the QBO amplitude near 10 hPa, which is the level of maximum QBO amplitude for most models. CMIP models also do a reasonable job representing the QBO latitudinal extent at 10 hPa. In a multimodel average, CMIP5 (CMIP6) models underestimate the QBO amplitude at 10 hPa by only ~4(8)%. However, the majority of models significantly underestimate the QBO amplitude at and below 20 hPa, and overestimate it above 10 hPa, especially in the CMIP5 models. In other words, the QBO appears to be shifted upwards in many models relative to what is seen in observations. The underestimation of QBO amplitude in the lower stratosphere is large. At 50 hPa, the QBO amplitude in CMIP5 models is on average 68% of the observed value, and it is only 62% of the observed value in the CMIP6 ensemble. Similar biases are seen at 70 hPa. The depth, or vertical extent, of the QBO is on average higher than observed in CMIP5 and CMIP6 models; however, the lowest level that the QBO reaches is at a higher altitude than in observations 10-hPa difference, pointing again to an upward shift of the oscillation in the models and particular deficit in the representation of the QBO in the lowermost levels.

Shortcomings in the QBO amplitude below 20 hPa and the inadequate reach into the lowermost stratosphere of the QBOs in CMIP models may have consequences for those QBO teleconnections that have been found to correlate strongly with QBO winds near 50 hPa (Anstey & Shepherd, 2014; Son et al., 2017; Wang et al., 2018; Yoo & Son, 2016). In addition, if the QBO in zonal mean zonal wind does not reach far down enough, the temperature anomalies associated with the QBO, which are one-fourth cycle out of phase with the wind anomalies, will likely be weak in the lowermost stratosphere. As temperature anomalies are one of the mechanisms in which the QBO is thought to influence tropospheric variability, these connections might also be weak. Hence, although more models than ever are simulating the QBO, its influence on the tropospheric variability in GCMs may still be elusive. The influence of the QBO on the stratospheric polar vortex and the MJO in CMIP6 models are actively being investigated and will be reported on in separate publications.

There are several factors that are likely responsible for the deficiencies in the representation of the QBO in climate models. First, vertical resolution of ~500 m has been demonstrated to be needed to adequately represent the QBO. For example, Richter et al. (2014) have shown that increased vertical resolution increased the generation of and forcing of the QBO by Kelvin and mixed GWs. Geller et al. (2016) have shown that vertical resolution is necessary to resolve wave-mean flow interactions that cause the shear layers to descend down to 100 hPa. However, many CMIP models have vertical resolution coarser than 500 m, and it is not clear how vertical resolution finer than 500 m would affect the QBO and momentum deposition from Kelvin and mixed Rossby GWs. CMIP6 model, CESM2-WACCM, produces an internally generated QBO; however, it is deficient in amplitude in the lower stratosphere due to relatively poor vertical resolution. A much better QBO was obtained in a previous version of this model, CESM1-WACCM, when the vertical resolution was doubled in the troposphere and lower stratosphere (Garcia & Richter, 2019). Holt et al. (2019) has demonstrated a clear relationship between the wave forcing from resolved waves in the lower stratosphere and vertical resolution in QBOi models. Similar dependence of lower stratospheric wave forcing on horizontal resolution was not found. Holt et al. (2019) also showed that even models that have ample wave momentum flux entering the lower stratosphere seem to underestimate the contribution of these waves to the QBO driving in the upper stratosphere, which suggests that factors such as vertical resolution and numerical dissipation might influence the effectiveness of Kelvin and mixed Rossby GWs in driving the QBO.

Since the resolved wave forcing of the QBO is underestimated in most models, GCMs typically depend on parameterizations of nonorographic GW drag to account for the rest of the forcing (Bushell et al., 2020; Holt et al., 2019). The majority of models utilize GW parameterizations that use a fixed GW source spectrum that is the same in the tropics and extratropics and is not explicitly linked to spatiotemporally varying tropospheric sources. In reality, the GW spectrum in the extratropics is different than that in the tropics; hence, this spectrum is a tuned compromise, likely inadequately representing the phase speed spectrum of GWs generated by tropical convection. In particular, an inadequate amount of momentum flux from waves with low-phase speeds, less than 10 m s^{-1}, will impede the downward propagation of the shear zones. A few modeling centers in recent years have changed their GW parameterizations to be linked to tropospheric sources in the hope that they would be more realistic and respond to tropospheric variability and climate change (Bushell et al., 2015; Lott et al., 2012; Richter et al., 2010). However, Richter et al. (2020) have shown that there is no convergence among those parameterizations in how the tropical GW spectrum, and hence the QBO, will change in a warming climate. In addition, these parameterizations were designed to more accurately represent the spectrum of waves that is not resolved in climate models, but in reality, they also need

to compensate for errors due to inadequate representation of Kelvin and mixed Rossby GWs, zonal mean wind biases, numerical diffusion, and inadequate vertical resolution.

Another factor that could contribute to the inadequate representation of the QBO in CMIP models is the process by which CMIP models are developed. The majority of resources are typically invested in developing the tropospheric physics and tropospheric climate, and the stratosphere is often tuned last, shortly before the deadline of model delivery and with limited computational resources. This could explain why there is such a wide range of mean QBO periods among CMIP6 models as compared to QBOi models for example, but does not explain the deficiencies in QBO amplitude, as QBOi models showed to have the same biases (Bushell et al., 2020) and those modeling centers self-selected to participate in QBOi and typically spent ample time tuning the stratospheric simulation.

Lastly, in this manuscript, we focused on the simulation of the QBO in historical simulations. These have a larger range of SSTs than simulations with prescribed observed SSTs, a factor that could affect the forcing of Kelvin and mixed Rossby GWs. In models with source-dependent GW parameterizations, this could also alter the GW spectrum. Simulations with prescribed 20th-century SSTs were not available yet for all the models considered here for CMIP6 (and were not available for CMIP5 models either); hence, we cannot quantify this influence at this time. However, averaged over CMIP5 and CMIP6, the QBO near-equatorial amplitude is virtually indistinguishable from QBOi models throughout most of the stratosphere, and QBOi models all used prescribed observed SSTs. Hence, it is unlikely that the variable SSTs are a large factor contributing to the biases in the representation of the QBO seen in the CMIP models.

In summary, while there has been a large increase in the number of models with the QBO among CMIP models in the last 20 years, the fidelity of the simulation of the QBO among the QBO-resolving models, in a multimodel mean, has not improved. The presence of 50% of models with the QBO in the CMIP6 ensemble could help answer the question of how much does the QBO matter for representing tropospheric variability correctly; however, the biases in the QBO amplitude in the lower stratosphere are likely a limiting factor in assessing this adequately. Several studies of QBO teleconnections in CMIP6 models are already underway to answer this question.

Acknowledgments
The ERAI reanalysis used in this study is available from the website (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim). CMIP3, CMIP5, and CMIP6 data are freely available from the Earth System Grid Federation (ESGF; https://esgf-node.llnl.gov/). We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the ESGF for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. Output from QBOi experiments is available via the JASMIN scientific data analysis environment by requesting access to the QBOi group workspace at CEDA (https://www.ceda.ac.uk/services/). Data format and storage are described in the supporting information of Butchart et al. (2018). This work was supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) via National Science Foundation IA 1844590. Y. K. was supported by the Japan Society for the Promotion of Science KAKENHI Grants JP15KK0178, JP17K18816, and JP18H01286 and by the Environment Research and Technology Development Fund (2-1904) of the Environmental Restoration and Conservation Agency of Japan. N. B. was supported by the Met Office Hadley Centre Programme funded by BEIS and Defra. We thank two anonymous reviewers for their comments that helped to shape the final version of the manuscript.

References

Anstey, J. A., & Shepherd, T. G. (2014). Review article high-latitude influence of the quasi-biennial oscillation. *Quarterly Journal of the Royal Meteorological Society*, *140*, 1–21. https://doi.org/10.1002/qj.2132

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., et al. (2001). The quasi-biennial oscillation. *Review of Geophysics*, *39*, 179–229.

Bushell, A. C., Anstey, J. A., Butchart, N., Kawatani, Y., Osprey, S., Richter, J. H., et al. (2020). Evaluation of the Quasi-Biennial Oscillation in global climate models for the SPARC QBO-initiative. *Quarterly Journal of the Royal Meteorological Society*, 1–31. https://doi.org/10.1002/qj.3765

Bushell, A. C., Butchart, N., Derbyshire, S. H., Jackson, D. R., Shutts, G. J., Vosper, S. B., & Webster, S. (2015). Parameterized gravity wave momentum fluxes from sources related to convection and large-scale precipitation processes in a global atmosphere model. *Journal of the Atmospheric Sciences*, *72*(11), 4349–4371. https://doi.org/10.1175/JAS-D-15-0022.1

Butchart, N., Anstey, J. A., Hamilton, K., Osprey, S., McLandress, C., Bushell, A. C., et al. (2018). Overview of experiment design and comparison of models participating in phase 1 of the SPARC quasi-biennial oscillation initiative (QBOi). *Geoscientific Model Development*, *11*(3), 1009–1032. https://doi.org/10.5194/gmd-11-1009-2018

Collimore, C. C., Martin, D. W., Hitchman, M. H., Huesmann, A., & Waliser, D. E. (2003). On the relationship between the QBO and tropical deep convection. *Journal of Climate*, *16*, 2552–2568.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *656*, 553–597.

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

Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., et al. (2016). Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, *7*(4), 813–830. https://doi.org/10.5194/esd-7-813-2016

Garcia, R. R., & Richter, J. H. (2019). On the momentum budget of the quasi-biennial oscillation in the whole atmosphere community climate model. *Journal of the Atmospheric Sciences*, *76*(1), 69–87. https://doi.org/10.1175/JAS-D-18-0088.1

Garfinkel, C. I., & Hartmann, D. L. (2010). Influence of the quasi-biennial oscillation on the North Pacific and El Niño teleconnections. *Journal of Geophysical Research*, *115*, D20116. https://doi.org/10.1029/2010JD014181

Geller, M. A., Zhou, T., Shindell, D., Ruedy, R., Aleinov, I., Nazarenko, L., et al. (2016). Modeling the QBO improvements resulting from higher-model vertical resolution. *Journal of Advances in Modeling Earth Systems*, *8*, 1092–1105. https://doi.org/10.1002/2016MS000699

Ho, C.-H., Kim, H.-S., Jeong, J.-H., & Son, S.-W. (2009). Influence of stratospheric quasi-biennial oscillation on tropical cyclone tracks in the western North Pacific. *Geophysical Research Letters*, *36*, L06702. https://doi.org/10.1029/2009GL037163

Holt, L., Lott, F., Garcia, R. R., Kiladis, G. N., Anstey, J. A., Braesicke, P., et al. (2019). An evaluation of tropical waves and wave forcing of the QBO in the QBOi models. *Quarterly Journal of the Royal Meteorological Society*, *49*, 75–104.

Holton, J. R., & Tan, H. (1980). The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sciences*, *37*, 2200–2208.

Labitzke, K. G. (1987). Sunspots, the QBO, and the stratospheric temperature in the north polar region. *Geophysical Research Letters*, *14*, 535–537.

Labitzke, K. G., & van Loon, H. (1988). Association between the 11-year solar cycle, the QBO and the atmosphere. Part I: The troposphere and stratosphere in the northern hemisphere in winter. *Journal of Atmospheric and Terrestrial Physics*, *50*, 197–206.

Liess, S., & Geller, M. A. (2012). On the relationship between QBO and distribution of tropical deep convection. *Journal of Geophysical Research*, *117*, D03108. https://doi.org/10.1029/2011JD016317

Lott, F., Guez, L., & Maury, P. (2012). A stochastic parameterization of non-orographic gravity waves: Formalism and impact on the equatorial stratosphere. *Geophysical Research Letters*, *39*, L06807. https://doi.org/10.1029/2012GL051001

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., & Stouffer, R. J. (1997). Intercomparison makes for a better climate model. *Eos, Transactions American Geophysical Union*, *78*(41), 445–451. https://doi.org/10.1029/97EO00276

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., et al. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*, *88*(9), 1383–1394. https://doi.org/10.1175/BAMS-88-9-1383

Osprey, S. M., Gray, L. J., Hardiman, S. C., Butchart, N., Bushell, A. C., & Hinton, T. J. (2010). The climatology of the middle atmosphere in a vertically extended version of the Met Office's climate model. Part II: Variability. *Journal of the Atmospheric Sciences*, *67*(11), 3637–3651. https://doi.org/10.1175/2010JAS3338.1

Randel, W. J., & Wu, F. (1996). Isolation of ozone QBO in SAGE II data by singular-value decomposition. *Journal of the Atmospheric Sciences*, *53*(17), 2546–2559.

Randel, W. J., Wu, F., Russell, J. M., Roche, A., & Waters, J. W. (1998). Seasonal cycles and QBO variations in stratospheric CH_4 and H_2O observed in UARS HALOE data. *Journal of the Atmospheric Sciences*, *55*(2), 163–185.

Richter, J. H., Bacmeister, J. T., & Solomon, A. (2014). On the simulation of the quasi-biennial oscillation in the Community Atmosphere Model, version 5. *Journal of Geophysical Research: Atmospheres*, *119*, 3045–3062. https://doi.org/10.1002/2013JD021122

Richter, J. H., Butchart, N., Kawatani, Y., Bushell, A. C., Holt, L., Serva, F., et al. (2020). Response of the Quasi-Biennial Oscillation to a warming climate in global climate models. *Quarterly Journal of the Royal Meteorological Society*, 1–29. https://doi.org/10.1002/qj.3749

Richter, J. H., Sassi, F., & Garcia, R. R. (2010). Towards a physically based gravity wave source parameterization in a general circulation model. *Journal of the Atmospheric Sciences*, *67*, 136–156.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*, 2514–2519. https://doi.org/10.1002/2014GL059637

Scaife, A. A., Athanassiadou, M., Andrews, M., Arribas, A., Baldwin, M. P., Dunstone, N., et al. (2014). Predictability of the quasi-biennial oscillation and its northern winter teleconnection on seasonal to decadal timescales. *Geophysical Research Letters*, *41*, 1752–1758. https://doi.org/10.1002/2013GL059160

Schenzinger, V., Osprey, S., Gray, L., & Butchart, N. (2017). Defining metrics of the quasi-biennial oscillation in global climate models. *Geoscientific Model Development*, *10*, 2157–2168. https://doi.org/10.5194/gmd-10-2157-2017

Son, S. W., Lim, Y., Yoo, C., Hendon, H. H., & Kim, J. (2017). Stratospheric control of the Madden-Julian oscillation. *Journal of Climate*, *30*(6), 1909–1922. https://doi.org/10.1175/JCLI-D-16-0620.1

SPARC (2010). SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models. In V. Eyring, T. Shepherd, & D. Waugh (Eds.), SPARC Report No. 5, WCRP-30/2010, WMO/TD – No. 40. Retrieved from www.sparc-climate.org/publications/sparc-reports/

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Wang, J., Kim, H.-M., & Chang, E. K. M. (2018). Interannual modulation of Northern Hemisphere winter storm tracks by the QBO. *Geophysical Research Letters*, *45*, 2786–2794. https://doi.org/10.1002/2017GL076929

Yoo, C., & Son, S. W. (2016). Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophysical Research Letters*, *43*, 1392–1398. https://doi.org/10.1002/2016GL067762