

Teleconnections of the Quasi-Biennial Oscillation in a multi-model ensemble of QBO-resolving models

James A. Anstey¹  | Isla R. Simpson² | Jadwiga H. Richter² | Hiroaki Naoe³ | Masakazu Taguchi⁴ | Federico Serva⁵ | Lesley J. Gray⁶ | Neal Butchart⁷ | Kevin Hamilton^{8,9} | Scott Osprey⁶ | Omar Bellprat¹⁰ | Peter Braesicke¹¹ | Andrew C. Bushell¹² | Chiara Cagnazzo⁵ | Chih-Chieh Chen² | Hye-Yeong Chun¹³ | Rolando R. Garcia² | Laura Holt¹⁴ | Yoshio Kawatani¹⁵ | Tobias Kerzenmacher¹¹ | Young-Ha Kim¹⁶ | Francois Lott¹⁷ | Charles McLandress^{1,18} | John Scinocca¹ | Timothy N. Stockdale¹⁹ | Stefan Versick¹¹ | Shingo Watanabe¹⁵ | Kohei Yoshida³ | Seiji Yukimoto³

¹Canadian Centre for Climate Modelling and Analysis (CCCma), Victoria, Canada

²National Center for Atmospheric Research (NCAR), Boulder, Colorado

³Meteorological Research Institute (MRI), Tsukuba, Japan

⁴Department of Earth Science, Aichi University of Education, Kariya, Japan

⁵Institute of Marine Sciences, National Research Council (ISMAR-CNR), Rome, Italy

⁶National Centre for Atmospheric Science (NCAS), University of Oxford, Oxford, UK

⁷Met Office Hadley Centre (MOHC), Exeter, UK

⁸Department of Atmospheric Sciences, University of Hawaii, Honolulu, Hawaii

⁹International Pacific Research Center (IPRC), Honolulu, Hawaii

¹⁰Barcelona Supercomputing Center (BSC), Barcelona, Spain

¹¹Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany

¹²Met Office, Exeter, UK

¹³Yonsei University, Seoul, South Korea

Abstract

The Quasi-biennial Oscillation (QBO) dominates the interannual variability of the tropical stratosphere and influences other regions of the atmosphere. The high predictability of the QBO implies that its teleconnections could lead to increased skill of seasonal and decadal forecasts provided the relevant mechanisms are accurately represented in models. Here modelling and sampling uncertainties of QBO teleconnections are examined using a multi-model ensemble of QBO-resolving atmospheric general circulation models that have carried out a set of coordinated experiments as part of the Stratosphere-troposphere Processes And their Role in Climate (SPARC) QBO initiative (QBOi). During Northern Hemisphere winter, the stratospheric polar vortex in most of these models strengthens when the QBO near 50 hPa is westerly and weakens when it is easterly, consistent with, but weaker than, the observed response. These weak responses are likely due to model errors, such as systematically weak QBO amplitudes near 50 hPa, affecting the teleconnection. The teleconnection to the North Atlantic Oscillation is less well captured overall, but of similar strength to the observed signal in the few models that do show it. The models do not show clear evidence of a QBO teleconnection to the Northern Hemisphere Pacific-sector subtropical jet.

KEYWORDS

North Atlantic Oscillation, polar vortex, Quasi-Biennial Oscillation, seasonal forecasting, stratosphere, stratosphere–troposphere coupling, teleconnection

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 Crown copyright. Quarterly Journal of the Royal Meteorological Society published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society. This article is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Reproduced with the permission of the Minister of Environment and Climate Change Canada.

¹⁴NorthWest Research Associates
(NWRA), Boulder, Colorado

¹⁵Japan Agency for Marine–Earth Science
and Technology (JAMSTEC), Yokohama,
Japan

¹⁶Institut für Atmosphäre und Umwelt,
Goethe-Universität, Frankfurt am Main,
Germany

¹⁷Laboratoire de Météorologie Dynamique
(LMD), Paris, France

¹⁸University of Toronto, Toronto, Ontario,
Canada

¹⁹European Centre for Medium-Range
Weather Forecasts (ECMWF), Reading,
UK

Correspondence

J. A. Anstey, Canadian Centre for Climate
Modelling and Analysis, Environment
and Climate Change Canada, Victoria,
British Columbia, Canada.
Email: james.anstey@canada.ca

Funding information

National Science Foundation (NSF) US
Department of Energy, Grant/Award
Numbers: 1852977, 1844590

1 | INTRODUCTION

The Quasi-Biennial Oscillation (QBO) of tropical stratospheric circulation is notable as a very long period variation that is generated by internal atmospheric processes. The oscillation is quasi-regular (periods have been observed to range between ~20 and 36 months) and models usually can produce skilful forecasts for lead times of 3 years or more (Scaife *et al.*, 2014b). Realizing the potential predictability of the QBO with comprehensive numerical model forecast systems is limited by uncertainties in modelling the QBO (Baldwin *et al.*, 2001; Butchart *et al.*, 2018). The QBO dominates the interannual variability of the lower tropical stratosphere but also influences other regions of the atmosphere. Due to the high predictability of the QBO itself, this influence may provide a source of additional seasonal and decadal predictability in those regions (Boer and Hamilton, 2008; Marshall and Scaife, 2009; Scaife *et al.*, 2014a; Marshall *et al.*, 2017; Garfinkel *et al.*, 2018; O'Reilly *et al.*, 2019; Nardi *et al.*, 2020).

The observed influence of the QBO on the Northern Hemisphere (NH) winter stratospheric polar vortex, often referred to as the Holton–Tan effect (Holton and

Tan, 1980), may be the most well-studied QBO teleconnection. Observations spanning the last ~65 years indicate that, when QBO winds in the lower tropical stratosphere at or near 50 hPa are used as an index of the QBO state, the NH stratospheric polar vortex during early to mid-winter (November–January) is strengthened when the QBO is westerly (QBO-W) and weakened when it is easterly (QBO-E). The difference in vortex strength between QBO-W and QBO-E phases peaks at roughly 10 m·s⁻¹ in the middle stratosphere (near 10 hPa) during January, with correlations between the QBO and vortex of ~0.5 for monthly-mean data, suggesting that the QBO phase is linearly associated with ~25% of the interannual variance of the vortex (Dunkerton and Baldwin, 1991). Coincident with this stratospheric response, the North Atlantic Oscillation (NAO) displays a preference, mainly in January, for its positive phase (poleward-shifted Atlantic jet) during QBO-W years and its negative phase (equatorward-shifted jet) during QBO-E years (Ebdon, 1975). In the Southern Hemisphere (SH) stratosphere, the late-winter final breakdown of the polar vortex is observed to occur slightly earlier when QBO winds at ~20 hPa are easterly and slightly later when they are westerly, which contrasts with the seasonal timing of the observed NH

response (early to mid-winter) and the QBO phase that best predicts it (~ 50 hPa). The stratospheric polar vortex teleconnection and its apparent downward extension into the troposphere has been the subject of numerous observational and modelling studies; Anstey and Shepherd (2014) give a review. The mechanism underlying QBO–vortex coupling is generally believed to be that the prevailing winds in the tropical stratosphere influence the propagation and dissipation of equatorward-propagating extratropical planetary waves, leading to high-latitude changes in these waves and consequently in the polar vortex (Holton and Tan, 1980; Dunkerton and Baldwin, 1991), but the precise details of how this occurs remain unclear.

Other QBO teleconnections examined in the literature include responses of the Pacific sector subtropical jet (Garfinkel and Hartmann, 2010; 2011a; 2011b; Seo *et al.*, 2013), NH winter storm tracks (Wang *et al.*, 2018a), tropical cyclones (Ho *et al.*, 2009; Camargo and Sobel, 2010), seasonal-mean tropical convection (Giorgetta *et al.*, 1999; Collimore *et al.*, 2003; Liess and Geller, 2012; Gray *et al.*, 2018), convection associated with the Madden–Julian Oscillation (MJO; Yoo and Son, 2016; Son *et al.*, 2017), and modulation of MJO teleconnections (Wang *et al.*, 2018b; Feng and Lin, 2019) including transport of air out of the tropics by atmospheric rivers (Mundhenk *et al.*, 2018; Toms *et al.*, 2020). Since the QBO influence coexists with other sources of interannual variability such as ENSO (e.g., Domeisen *et al.*, 2019), it can be difficult to assess the robustness of any QBO teleconnections using the relatively short observed record. Multiple linear regression can be applied to separate the influence of various predictors (e.g., Gray *et al.*, 2018) but results are potentially sensitive to the choice of predictors, co-variation between predictors (aliasing), and sample size (record length). Reanalysis records extend back to the mid-twentieth century for reanalysis systems that assimilate radiosonde and satellite data, and to the late 19th century for surface-input reanalyses (Fujiwara *et al.*, 2017). Brönnimann *et al.* (2016) surveyed QBO teleconnections using a reconstructed QBO index going back to 1900 (Brönnimann *et al.*, 2007) and found a weakened NAO response during the first half of the century coinciding with a strengthened European surface temperature response, but *vice versa* during the latter half of the century. Since the reconstructed QBO index has larger uncertainty before 1953 than afterward, apparent variations in the teleconnection strength could indicate that the state of the QBO is not well known in the earlier period. Even when the state of the QBO is well known, uncertainty about mechanisms means that the best choice of QBO index for a given response is not known with high confidence. This motivates choosing a QBO index that maximizes an observed or modelled

response, such as QBO wind at 50 hPa (Gray *et al.*, 2018), but the freedom to vary the predictor index is often not explicitly accounted for when statistical significance is calculated, which could lead to overconfident estimates of significance. The difficulty of separating out competing influences within a short observational record, compounded by uncertainty about underlying mechanisms, both contribute to uncertainty in diagnosis of QBO teleconnections. Uncertainty in the strength of diagnosed teleconnections also arises due simply to sampling variability – that is, chaotic behaviour of the atmosphere (and climate system overall). The NH stratospheric polar vortex, for example, can have large interannual variability even in the absence of “external” influences such as the QBO or ENSO.

In this study we examine the NH winter responses of the stratospheric polar vortex, the NAO, and the tropospheric subtropical jet to the QBO. Reanalyses are used to diagnose observed responses and the QBOi multi-model ensemble of QBO-resolving models is used to examine their uncertainty due to both internal variability and model formulation. One aspect of modelling uncertainty is that the characteristics of QBOs simulated in the different models can vary widely, as documented by other studies in this Special Section that examine the QBOi multi-model ensemble (Smith *et al.*, 2019; Bushell *et al.*, 2020; Holt *et al.*, 2020; Richter *et al.*, 2020; Stockdale *et al.*, 2020). Modelling uncertainties in the representation of QBO teleconnections may be affected by errors in the QBO itself and by errors in aspects of the circulation that respond to the QBO such as the stratospheric polar vortex or tropical precipitation. An “ensemble of opportunity” such as the QBOi ensemble samples a variety of imperfect models which may nevertheless capture some essential features of the teleconnections. A similar approach has been used to study the QBO teleconnection to the NH winter stratosphere using those QBO-resolving models that are available in the CMIP5 and CMIP6 ensembles (Rao *et al.*, 2020). The QBOi multi-model ensemble was created with a primary goal of understanding the modelling uncertainties of the QBO itself, with analysis of teleconnections a secondary goal. Ensembles of opportunity such as these (CMIP5, CMIP6, QBOi) are valuable tools for assessing teleconnections because they sample modelling uncertainty under consistent forcings (coordinated experiments) and the widest available range of models is included by default.

The article is structured as follows. Section 2 describes the model and reanalysis datasets used, and the methods used to define QBO phase and test the statistical significance of the results. Section 3 examines the stratospheric NH winter response to the QBO, including vortex strength, Sudden Stratospheric Warmings (SSWs)

and the Eliassen–Palm (EP) flux. Section 4 examines the tropospheric NH winter response of the NAO and subtropical jet. Conclusions are given in Section 5.

2 | METHODS

2.1 | Models and reanalyses

We use models from the QBOi multi-model ensemble that carried out the two present-day QBOi experiments following the QBOi protocol as defined in Butchart *et al.* (2018). Both experiments were performed with atmospheric general circulation models (AGCMs) using prescribed sea-surface temperature and sea ice (SST/SI). The first, Experiment 1, prescribes monthly time-varying SST/SI conditions and other external forcings (e.g., carbon dioxide concentrations) based on observations for the 1979–2008 period. The second, Experiment 2, is a timeslice experiment that uses climatological boundary SST/SI conditions and forcings based on the 1988–2007 period. The key difference between Experiments 1 and 2 (Exp1 and Exp2 hereafter) is that in Exp2 all interannual variability arises from internal atmospheric dynamics, whereas in Exp1 the boundary conditions are also a source of interannual variability. Hence Exp1 potentially includes El Niño–Southern Oscillation (ENSO) effects while Exp2 excludes them. For further information on the QBOi experiments and model characteristics, the reader is referred to Butchart *et al.* (2018). It should be noted that the configuration of each QBOi model used here is that described in Butchart *et al.* (2018), and could differ from the standard version of the model (if there is one) due to the model being optimized to produce a spontaneous QBO (as requested by the QBOi protocol). We use all available ensemble members for every model that carried out Exp1 and Exp2. The eleven models from Butchart *et al.* (2018) for which Exp2 was available were 60LCAM5, AGCM3-CMAM, ECHAM5sh¹, EMAC, LMDz6, MIROC-AGCM-LL, MIROC-ESM, MRI-ESM2, UMGA7, UMGA7gws, and CESM1(WACCM5-110L). For Exp1, these same models and also HadGEM2-A and HadGEM2-AC were available. For brevity, AGCM3-CMAM and CESM1(WACCM5-110L) are referred to herein as CMAM and WACCM, respectively. One additional model run is used: CMAMmodOGWD is a version of the CMAM differing from the one described in Butchart *et al.* (2018) only in the value of one parameter in

the orographic gravity wave drag parametrization². This change improves the representation of the NH winter stratospheric polar vortex, but the QBO remains virtually identical to that in the standard QBOi version of CMAM. A 450-year integration of Exp2 was done with this model.

Observed teleconnections are quantified using multiple reanalysis datasets of varying record lengths. Throughout the article we refer to teleconnections diagnosed from reanalyses as the “observed” response, with the understanding that reanalysis products in fact represent a blend of observational data and forecast model output (Fujiwara *et al.*, 2017). Data are obtained from the S-RIP multi-reanalysis zonal-mean dataset (Martineau, 2017) or directly from the reanalysis centre websites. We use the following datasets covering the indicated periods: ERA-Interim (1979–2018), ERA-40 (September 1957 to August 2002), ERA-20C (1900–2010), MERRA-2 (1980–2018), and JRA-55 (1958–2018). The ERA-Interim, JRA-55, and MERRA-2 reanalyses are available up to almost present day. JRA-55 provides the longest available record (since 1958, roughly 60 years) of a modern reanalysis product. ERA-20C provides the longest available record (1900–2010, 111 years) but is a surface-input reanalysis, although its forecast model uses a high model lid (0.01 hPa) and hence represents the stratosphere.

The ERA-20C forecast model exhibits a QBO, but since no tropical radiosonde observations are assimilated, the ERA-20C QBO fluctuates in strength over the record and is not in phase with the observed QBO. To examine QBO teleconnections using the ERA-20C reanalysis, we use the Brönnimann *et al.* (2007) QBO reconstruction which extends back to 1900. This reconstruction is based on the QBO imprint on the signature of the solar semi-diurnal tide in surface pressure (Hamilton, 1983), corroborated by available balloon wind observations. The uncertainty of this reconstruction increases as the record goes further back, being particularly large prior to 1920 (Brönnimann *et al.*, 2007). Using more recently digitized pre-1953 balloon observations to validate the reconstruction, Brönnimann *et al.* (2016) found only 60% agreement between the reconstructed pre-1953 QBO winds and the observations, and suggested that phase onset times could be in error by 3–4 months. We nevertheless include some results for the whole ERA-20C period, but with the caveat that it is unclear how much reliable information about QBO phase is available prior to 1953.

¹An error in the oceanic boundary conditions in the ECHAM5sh QBOi simulations was discovered after revisions of this article were complete, but was determined to have negligible impact on the ECHAM5sh results shown herein.

²The inverse critical Froude number, which affects the breaking altitudes of parametrized gravity waves (Sigmond and Scinocca, 2010), was 0.375 in the standard QBOi version of CMAM (Butchart *et al.*, 2018) but changed to 0.22 in CMAMmodOGWD. The other key parameter discussed in Sigmond and Scinocca (2010), the factor G that scales the launch momentum flux, was 1.0 in both versions of CMAM.

2.2 | Defining QBO phase

Previous studies have found that different QBO indices can maximize the response of different teleconnections (Baldwin and Dunkerton, 1998; Anstey *et al.*, 2010; Gray *et al.*, 2018). For this reason we do not use a single definition of QBO phase (e.g., 50 hPa equatorial wind) for all analyses presented in this article. For example, Baldwin and Dunkerton (1998) showed the observed NH winter stratospheric vortex response correlates most strongly with QBO winds at ~ 40 hPa, while the SH winter stratospheric vortex response correlates most strongly with QBO winds at ~ 20 hPa. Since we aim to determine if observed teleconnections are manifest in the model runs, the most straightforward model–observation comparison is to apply the same QBO phase definitions to the models that are optimal for observed teleconnections. Where this is done (e.g., Section 3.2), a single-level phase definition is used with the two categories denoted as QBO-W (westerly) and QBO-E (easterly).

An advantage of this straightforward approach is its simplicity, but two potential drawbacks should be noted. The first is that QBO teleconnections are subject to large sampling variability, as will be shown in Sections 3 and 4. Since the observed record is not very long (the longest modern reanalysis product is ~ 60 years; Section 2.1), and there could be aliasing of the response to other influences (such as ENSO teleconnections), the optimal QBO phase for any observed QBO teleconnection is uncertain. Secondly, model error could cause a model's optimal QBO phase to differ from the true optimal phase (i.e., the observed optimal phase, if it could be reliably determined from the observations). This could be due to deficiencies in the QBO as well as any other aspects of the model that might affect a teleconnection (e.g., its climatological winter polar vortex strength). Hence it is also useful to take an exploratory approach by examining responses to all QBO phases so as to determine the optimal index, if any, for a given QBO teleconnection in each model.

The sensitivity to different QBO phase definitions is examined using a QBO phase angle, ψ , defined by the two leading empirical orthogonal functions (EOFs) of tropical stratospheric monthly-mean zonal-mean wind, which typically capture upward of 90% of the variance (Wallace *et al.*, 1993; Baldwin and Dunkerton, 1998). The calculation follows Wallace *et al.* (1993)³, with the range $[0,1]$ representing a full QBO cycle. Additionally, ψ for each model is shifted by a constant value so that $\psi = 0$ corresponds to the

50 hPa easterly QBO phase onset in each model. This has no effect on the results (only differences between ψ values are physically meaningful) but ensures that similar values in the $[0,1]$ range correspond to similar vertical structure of winds in each model, making it easier to compare results across models. The motivation for using ψ rather than the wind at selected altitudes is that the EOFs are determined using data at all altitudes in the QBO region; hence ψ is a metric of QBO phase which accounts for the whole vertical structure of the oscillation.

The tropical wind used to define the QBO is the 2°S – 2°N averaged zonal-mean zonal wind over an appropriate time frame for a given analysis, e.g., December–February (DJF). Averages over latitude are area-weighted (i.e., weighted by the cosine of latitude). Further details are given in the relevant sections if necessary.

2.3 | Statistical tests

To assess the uncertainties in correlation coefficients (Section 3.2), we use a bootstrapping with replacement technique. It is assumed that each year is independent which, as will be discussed in Section 3.2, provides a reasonable assessment of uncertainties in simulations that are long enough to test this assumption. Consider the correlation between January QBO wind and January polar vortex strength as an example. For a model that provides N years of simulation for a given experiment after pooling all ensemble members together, we randomly re-sample from these N years (with replacement) another set of N years and re-calculate the correlation coefficient using this re-sampled set of years. This procedure is repeated 1,000 times, and the uncertainty range is given as the 5th to 95th percentile range of this distribution of bootstrapped correlation coefficients. Where this 5–95% percentile range does not encompass zero, the correlation coefficient is considered to be statistically significant. This is equivalent to significance at the 90% level by a two-sided test. In the bar plots (Figures 5, 11 and 13 below) we also show the minimum to maximum range of the bootstrapped distribution. To assess the extent to which observed correlations are consistent with the model statistics, we perform a similar re-sampling procedure to that described above, but instead of re-sampling from the model using N years, we re-sample an equivalent number of years as that in the observational record (e.g., 60 years for JRA-55). This quantifies the probability of obtaining the observed correlation from the model's climate but with an equivalent sample size to that in the observational record. For this assessment, we make use only of models and experiments with 100 years or more of data available.

³Pressure levels in the 10–70 hPa range are used to calculate the EOFs, as in Wallace *et al.* (1993). For the JRA-55 data used here, the available levels are 70, 50, 30, 20, and 10 hPa. For QBOi models, the available levels are 70, 60, 50, 40, 30, 20, 15, and 10 hPa (Butchart *et al.*, 2018).

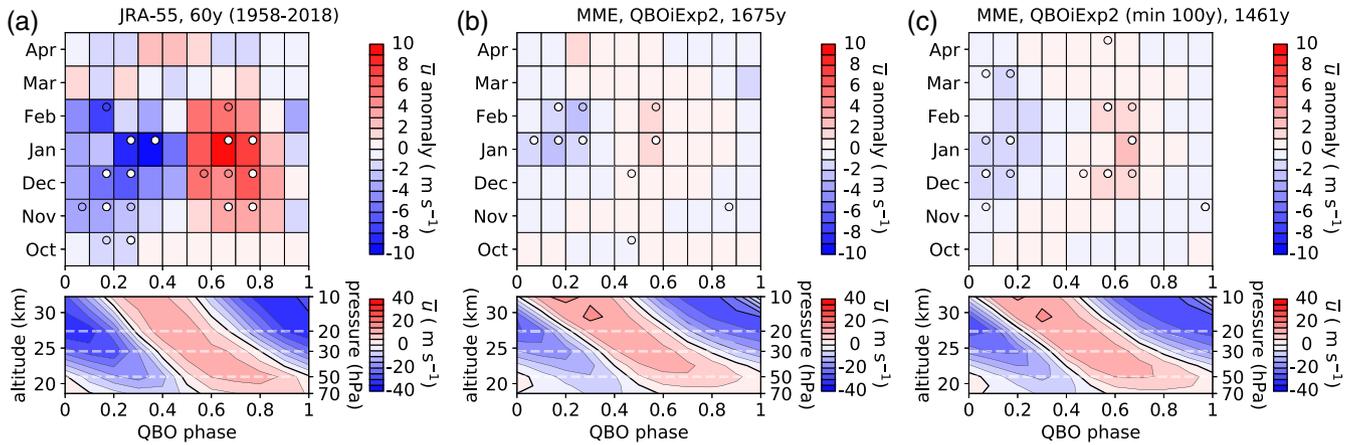


FIGURE 1 (a) Top panel: Seasonal variation for October–April of observed NH vortex response to QBO phase in the JRA-55 reanalysis, for the 60 NH winters from 1958/1959 to 2017/2018. The vortex metric is the deseasonalized monthly-mean zonal-mean zonal wind, \bar{u} , at 10 hPa, 60°N. Open circles and white-filled circles indicate statistical significance at the 90% and 95% levels, respectively (further details in text). Bottom panel: Vertical structure of equatorial \bar{u} (2°S–2°N) corresponding to a given value of QBO phase angle, ψ , shown by equatorial \bar{u} profiles composited for times when ψ takes the values on the x-axis. Dashed white horizontal lines indicate the 50, 30 and 20 hPa levels. (b) Mean over the multi-model ensemble (MME) of the diagnostic shown in (a) for all QBOi models that performed Exp2. (c) As (b), but excluding models that provided less than 100 years of simulation [Colour figure can be viewed at wileyonlinelibrary.com]

3 | WINTER STRATOSPHERIC RESPONSE

The NH vortex response to QBO phase is observed to peak in midwinter, around January, but is also present in early winter. We first consider whether the models show a similar intraseasonal timing in their responses (Section 3.1). The midwinter response is then examined in more detail, addressing whether systematic differences between the models and observations are detectable (Section 3.2). Since the vortex response may be associated with the occurrence of sudden stratospheric warmings (SSWs; e.g., Dunkerton *et al.*, 1988; Andrews *et al.*, 2019) we look for evidence for this association in the QBOi models (Section 3.3). QBO signatures in extratropical planetary wave forcing are examined in Section 3.4.

3.1 | Seasonally varying response

The response of the NH stratospheric polar vortex to the phase of the QBO over the October–April period is shown in Figure 1 for reanalysis and for the multi-model ensemble for Exp2, the present-day timeslice experiment. The JRA-55 reanalysis is used to characterize the observed response (Figure 1a), covering 60 NH winters from 1958/1959 to 2017/2018⁴. JRA-55 is shown because it provides the longest record of any modern reanalysis product

(Fujiwara *et al.*, 2017) over the era when the QBO winds are well observed and extensive radiosonde coverage in the NH constrains the polar vortex state. The monthly-mean zonal-mean zonal wind, \bar{u} , at 10 hPa and 60°N is used to characterize the vortex state. Similar results using this vortex metric are found for other reanalyses; results are more sensitive to record length (sample size) than to choice of reanalysis (Hitchcock, 2019; Anstey *et al.*, 2021). QBO phase is characterized by the phase angle ψ (Section 2.2). Each bin (i.e., box) in the top panel of Figure 1a shows the composited deseasonalized \bar{u} , denoted $\Delta\bar{u}$, calculated using all times that occur (1) within a bin of width $\Delta\psi = 0.3$ centred on the ψ value indicated on the horizontal axis, and (2) at the indicated month of the seasonal cycle. The choice of $\Delta\psi = 0.3$ for the bin width represents a compromise between achieving an adequate sample size on the one hand, and resolution in QBO phase space on the other (i.e., distinguishing between different equatorial wind states). The horizontal bin spacing, $\Delta\psi = 0.1$, is chosen to be smaller than the bin width so as to explore the sensitivity of the results to different definitions of QBO phase, which is akin to searching for values of ψ that optimize the vortex response (Baldwin and Dunkerton, 1998).

In JRA-55 the NH polar vortex during November–February shows a clear alignment with the phase of the QBO (Figure 1a, top panel). For roughly half of the range of ψ the composited $\Delta\bar{u}$ is easterly ($\psi \sim 0$ –0.5) and for the other half it is westerly ($\psi \sim 0.5$ –0.9), indicating weakened

⁴The tropical wind state during the 2015/2016 NH winter from roughly January onward was atypical due to the QBO disruption (Newman *et al.*,

2016; Osprey *et al.*, 2016; Coy *et al.*, 2017), but excluding this one winter makes essentially no difference to the results shown in Figure 1a.

and strengthened vortex states, respectively. The bottom panel of Figure 1a shows equatorial \bar{u} composited for each value of ψ . For this panel the equatorial wind is composited over a range $\Delta\psi = 0.1$, the horizontal bin spacing, centred on the value of ψ shown on the horizontal axis, since using the larger bin width of the top panel would excessively smooth out the QBO wind profiles. Comparing the top and bottom panels indicates that the bins centred on $\psi \sim 0.2\text{--}0.4$ (for easterly anomalies) and $\psi \sim 0.6\text{--}0.8$ (westerly anomalies) correspond to optimizing the NH midwinter (December–February) response by using the 50 hPa QBO wind to define QBO phase, as expected from previous studies (e.g., Holton and Tan, 1980; Baldwin and Dunkerton, 1998; Gray *et al.*, 2018). The magnitude of the strongest anomalies, which occur in midwinter, is $\Delta\bar{u} \sim 8\text{--}10 \text{ m}\cdot\text{s}^{-1}$, about 20–30% of the climatological vortex strength⁵. The vortex response begins to emerge in early winter (October–November) and strengthens in midwinter (December–February). No clear response is evident during March. In April there is a hint of a response: weak westerly anomalies occur for $\psi \sim 0.3\text{--}0.6$, and weak easterly anomalies for other ψ , suggesting a QBO influence on the timing of final vortex breakdown although it is not statistically significant. Comparison with the bottom panel shows that the westerly anomalies align with the QBO wind near 20 hPa, unlike the November–February NH response but reminiscent of the QBO modulation of the final breakdown of the SH winter vortex (Baldwin and Dunkerton, 1998).

Due to the large interannual variability of the NH polar vortex, it is essential to quantify the likelihood that the anomalies shown in Figure 1a could occur by chance rather than some physical mechanism related to the QBO. Statistical significance for each bin (box) in Figure 1a is calculated using the one-sample *t*-test (von Storch and Zwiers, 1999, Section 6.6.6), for which the test statistic is $t = \Delta\bar{u}/(\sigma/\sqrt{n})$ where σ is the standard deviation of deseasonalized \bar{u} (computed separately for each month of the seasonal cycle) and n the number of times falling within the bin. Significant results at the 90 and 95% levels (two-sided test) are shown by open circles and white-filled circles, respectively. For the November–February period there are a number of ψ values yielding statistically

significant $\Delta\bar{u}$ values, indicating that these $\Delta\bar{u}$ values would be unlikely if there were no physical mechanism linking the QBO and NH polar vortex. The April response for most ψ is only slightly weaker than the November response but is not significant⁶.

The corresponding response for each QBOi model that performed Exp2 is shown in Figure 2, and Figure 1b shows the multi-model ensemble (MME) mean of these responses. All models are given equal weight in the multi-model mean. The ensemble-mean vortex response is seen to be weaker than the reanalysis response (Figure 1a), but qualitatively similar: in NH midwinter, the vortex is stronger when the equatorial winds near 50 hPa are westerly and weaker when they are easterly. The strongest ensemble-mean responses occur in January–February. Statistical significance is calculated using the one-sample *t*-test as in Figure 1a, except that σ is replaced with the inter-model standard deviation of $\Delta\bar{u}$ and n is the number of models ($n = 12$). (A bootstrap test using 5,000 resamples with replacement gave essentially the same results for significance as the *t*-test.) The number of significant responses in January–February is similar to that in the reanalysis (albeit with weaker $\Delta\bar{u}$ for the MME), but unlike the reanalysis the models mostly fail to show significant responses in November–December. Similarly to the reanalysis, there is little evidence of a response in March–April. The equivalent of Figure 1b for Exp1 shows similar features as Exp2, although with slightly more of a response in December and slightly less in February (not shown).

Comparison of the individual model responses shown in Figure 2 indicates that many models exhibit qualitatively similar behaviour to that seen in reanalysis: equatorial wind in the lowermost tropical stratosphere, 30 or 50 hPa, usually aligns with the vortex response (a more detailed examination of this similarity is given in Section 3.2). However the individual models mostly show weaker responses than the reanalysis, both in terms of absolute magnitude and fewer statistically significant $\Delta\bar{u}$ values. For example, UMG7 and WACCM (Figure 2j,l) show response patterns similar to, but weaker than, JRA-55 (Figure 1a). But other models also show vortex responses that align well with ~ 50 hPa equatorial wind even if their response patterns look less similar overall to the reanalysis – for example, 60LCAM5, CMAM-modOGWD, ECHAM5sh, EMAC, and MIROC-AGCM-LL.

Nevertheless, the visual impression from Figure 2 is that the response pattern – how $\Delta\bar{u}$ varies intraseasonally and with ψ for each model – differs widely among models. Some of this variation, perhaps most of it, is

⁵Note that $\Delta\bar{u}$ is the anomaly of the wind from climatology, not a composite difference between westerly and easterly QBO phases. A difference between opposite QBO phases would be approximately twice this value, $15\text{--}20 \text{ m}\cdot\text{s}^{-1}$. This is larger than values of $\sim 10 \text{ m}\cdot\text{s}^{-1}$ at 10 hPa, 60°N typically reported in the QBO literature (e.g., figure 4a of Anstey and Shepherd, 2014) because $\Delta\psi = 0.3$ has been used, that is, each column makes use of $\sim 1/3$ of the data. Using $\Delta\psi = 0.5$ in Figure 1a gives peak anomalies of $\sim 5 \text{ m}\cdot\text{s}^{-1}$ (not shown), consistent with a difference between opposite QBO phases of $\sim 10 \text{ m}\cdot\text{s}^{-1}$ when each composite group contains half of the available data.

⁶Background variability is larger in April than November ($\sigma = 8.2 \text{ m}\cdot\text{s}^{-1}$ versus $6.5 \text{ m}\cdot\text{s}^{-1}$). April $\Delta\bar{u}$ for the $\psi = 0.4\text{--}0.5$ bin has 87% significance and all other ψ give much lower significance.

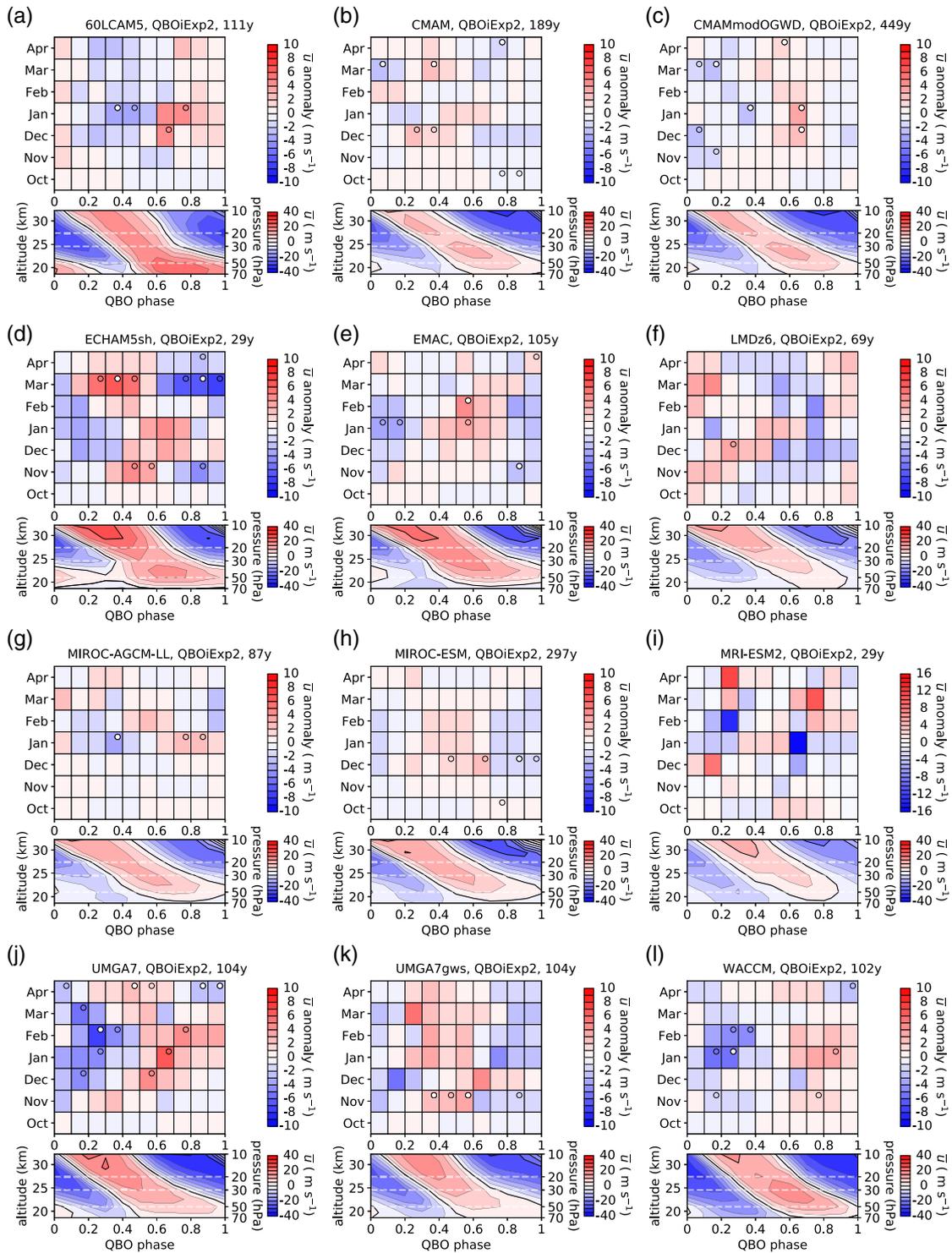


FIGURE 2 As Figure 1a, but for each individual model that performed Exp2, combining all available Exp2 ensemble members. (Note the larger colour bar range for MRI-ESM2) [Colour figure can be viewed at wileyonlinelibrary.com]

due to random sampling variability, since few of the $\Delta \bar{u}$ anomalies are statistically significant. The MME-mean of Figure 1b was constructed by assigning equal weight to all models, irrespective of the sample size for that model. Since the responses in Figure 2 for models with smaller samples are more likely to be contaminated

by random sampling variability, Figure 1c shows the MME-mean response computed using only models that provided at least 100 years of Exp2 data. This yields an increased number of statistically significant $\Delta \bar{u}$ values, most notably in December, and suggests a slightly better overall agreement between the reanalysis and models,

even though the MME-mean response remains weak compared to reanalysis (Figure 1a). Some indication of a March response also emerges, suggesting that the intraseasonal timing of the vortex response to the QBO is slightly later in the models (December–March) than reanalysis (November–February).

Figure 2 invites some speculation as to why models exhibit varying response patterns (even while remembering that an appreciable fraction of the differences may simply be random noise). The lower panels of each plot in Figure 2 show the different morphologies of the QBOs in the models, which are examined in more detail in Bushell *et al.* (2020). Comparison with the QBO as represented in JRA-55 (Figure 1a, lower panel) indicates that most model QBOs have weaker amplitude than JRA-55 at the lowest altitudes (e.g., 50 hPa), and also that they often tilt more strongly in the altitude–phase plane than JRA-55. Both of these features are evident in the MME-mean QBO composites (Figure 1b,c). The consistency across the model ensemble in showing some response to QBO winds near 50 hPa (Figure 1b,c) suggests this is a causal relation (otherwise one would not expect it to emerge from the noise of the MME). If so, then weak QBO amplitude near 50 hPa should weaken the QBO teleconnection to the vortex. The contrast between UMGA7 and UMGA7gws responses (Figure 2j,k, top panels) supports this interpretation since these two models differ only in their parametrized non-orographic gravity wave drag (Butchart *et al.*, 2018), which resulted in weaker QBO amplitude near 50 hPa in UMGA7gws (Figure 2j,k, bottom panels), but essentially no difference in climatological vortex strength (not shown). The stronger phase tilt in the models suggests that vertically deep QBO phases may be less common in the models than in reality. If the vortex responds more strongly to deeper equatorial wind anomalies (Dunkerton *et al.*, 1988; Gray *et al.*, 2003; Andrews *et al.*, 2019), then this could also weaken the QBO teleconnection to the vortex. However it is important to remember that many factors besides the QBO differ among the models. Comparison of the two CMAM results, Figures 2b and 2c, illustrates the impact of a single change in the model that did not affect the QBO (as seen by comparing the lower panels of each plot). As noted in Section 2.1, CMAM-modOGWD has a more realistic NH stratospheric polar vortex due to a single parameter change in the orographic gravity wave drag (OGWD) parametrization. This version of CMAM exhibits a midwinter (December–February) response that is slightly stronger, more significant, and in better agreement with the JRA-55 response pattern (Figure 1a) than the other CMAM. This suggests that inter-model differences in climatological vortex properties are another factor contributing to inter-model differences in the QBO–vortex teleconnection.

3.2 | Midwinter response

The QBO-composited results of Section 3.1 suggest that the models underestimate the vortex response to the QBO in comparison to reanalyses. Moreover, the statistical significance of the vortex anomalies in Figures 1 and 2 is generally low for the models, even for larger samples (~100 years or more for some models). This suggests that model errors, whether in the representation of the QBO itself (e.g., its amplitude near 50 hPa) or the extratropical stratosphere (e.g., the climatological strength of the vortex and the SSW frequency), may be adversely affecting the models' ability to represent the observed Holton–Tan response. However, the observed and modelled responses may have large uncertainty due to sampling variability, which must be taken into account when comparing them.

Figure 3 shows the correlation coefficient in January between vortex strength and equatorial (2°S – 2°N) wind at different altitudes for the JRA-55 and ERA-Interim reanalyses, and all models that performed Exp2. In the reanalyses the correlations are strongest for the 50 hPa equatorial wind, being roughly 0.45 and 0.5 in JRA-55 and ERA-Interim, respectively, with slightly weaker positive correlations at the adjacent 70 and 30 hPa levels. Opposite-signed correlations almost as strong as those at 50 hPa occur near 10 hPa, consistent with the vertical structure of the QBO. The uncertainty range (grey shading) shows the 5–95% range of correlation coefficients derived from bootstrap resampling (Section 2.3). Although this confidence interval clearly excludes zero at several altitudes (by the largest margin at 50 hPa) it also demonstrates a large uncertainty in the strength of the correlation. At 50 hPa, the range extends from roughly 0.3 to 0.6 for JRA-55 and 0.2 to 0.7 for ERA-Interim (the longer JRA-55 record provides a narrower confidence interval). The models similarly show a distribution of plausible correlation values, with smaller uncertainty for larger samples. Most models show a statistically significant correlation at some altitudes (i.e., the uncertainty range excludes zero, as indicated by red stars) but the altitudes of peak correlation differ among models. While this could mean that the QBO phase inducing the strongest vortex response differs among models, it is also clear from the uncertainties shown in Figure 3 that there is substantial overlap between many of the model distributions. This overlap implies that, in most cases, inter-model differences in the strength and vertical distribution of the correlation cannot be easily distinguished.

Although the models generally show weaker correlations than the reanalysis in Figure 3, the general shape of the vertical profiles is similar between the reanalysis and most models: a positive peak appears at the lower altitudes of the QBO (in the vicinity of 50 hPa), and there is

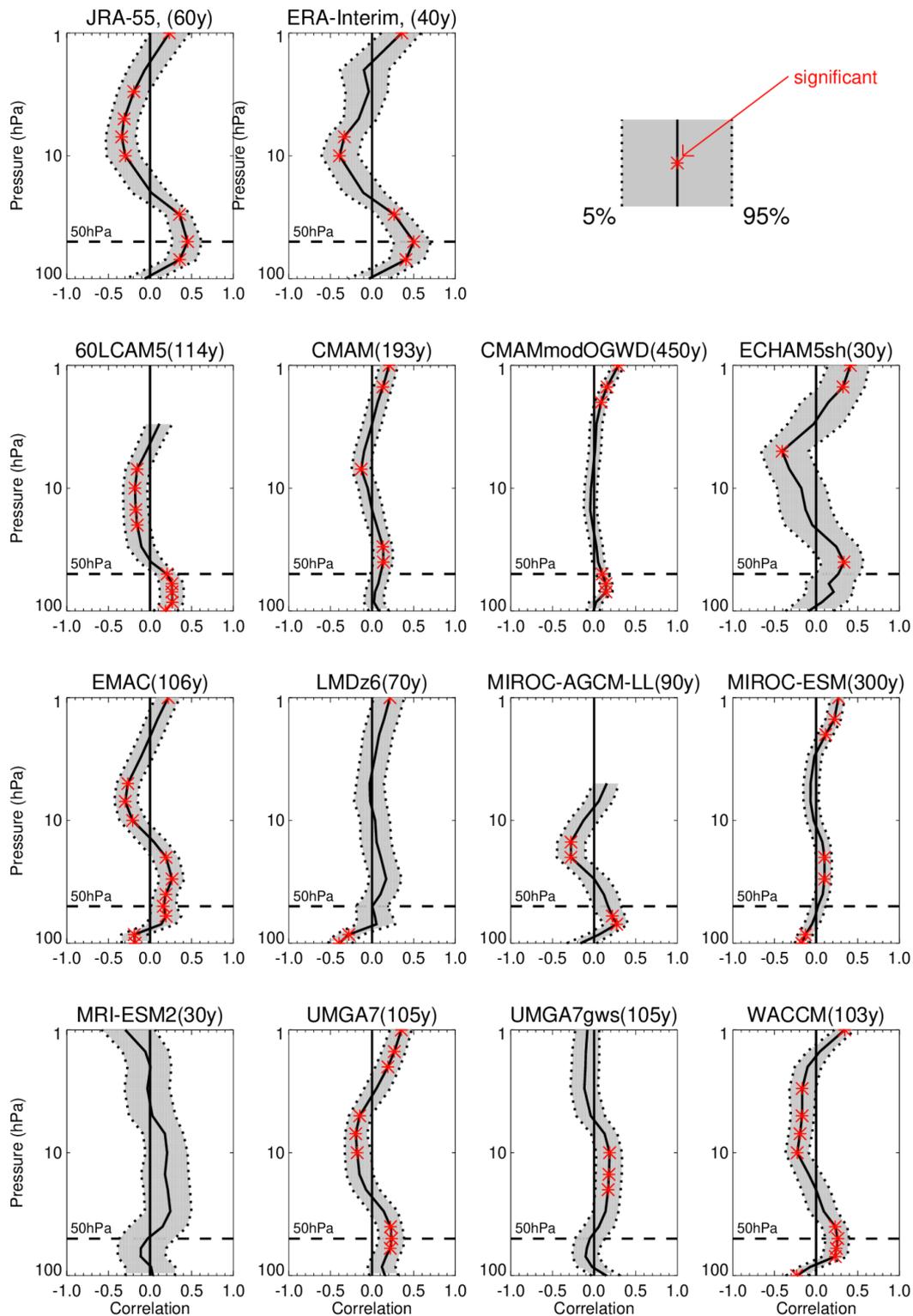
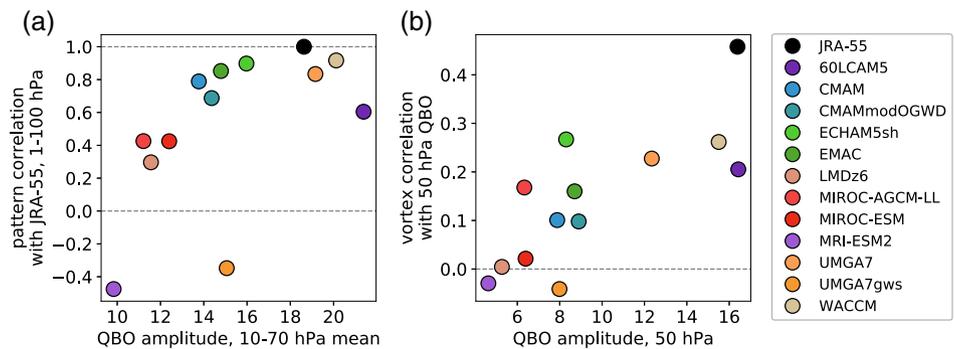


FIGURE 3 January correlation between vortex strength (monthly-mean \bar{u} at 10 hPa, 60°N) and equatorial wind at different altitudes (monthly-mean \bar{u} at 2°S–2°N) for all models that performed Exp2, and for the JRA-55 and ERA-Interim reanalyses. Uncertainty is shown in grey shading by the 5–95% range of correlations derived from bootstrap resampling (Section 2.3). A red star is shown if the uncertainty range excludes zero (i.e., if the null hypothesis can be rejected at the 90% level for a two-sided significance test). Note that the vertical resolution of pressure levels data provided for the QBOi models is higher than that available for the reanalysis. Also note that slight year count differences compared to Figure 2 are due to whether Exp2 was provided as a single ensemble member or multiple ensemble members that were combined (each year used in Figure 2 is a complete NH October–April period) [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 4 (a) Pattern correlation over 1–100 hPa between JRA-55 and model vertical profiles of January QBO–vortex correlation as shown in Figure 3, versus QBO amplitude averaged over 10–70 hPa. (b) January QBO–vortex correlation using 50 hPa QBO, versus QBO amplitude at 50 hPa. QBO amplitude is defined following Dunkerton and Delisi (1985) [Colour figure can be viewed at wileyonlinelibrary.com]



often a negative peak higher up (as expected due to the vertical structure of the QBO). Since previous studies show that a variety of tropical wind perturbations can induce a high-latitude response (Anstey and Shepherd, 2014 give an overview of these results), and given the inter-model variation in the models' QBOs (Figure 2), it is not a foregone conclusion that the model responses in Figure 3 should resemble the reanalysis. The similarity can be quantified by computing the pattern correlation between the JRA-55 vertical profile and model profiles. Figure 4a shows that the pattern correlation is positive for ten of the twelve models, and 0.6 or higher for seven of them, indicating that the vertical profile of QBO–vortex correlation in most models has a similar shape to that seen in reanalysis. This indicates that many models at least qualitatively reproduce the observed midwinter vortex response. Figure 4a also shows that the similarity tends to be greater for models with larger QBO amplitude in the 10–70 hPa layer, where most models underestimate the amplitude (Bushell *et al.*, 2020). A similar relationship is not evident if the 10–70 hPa layer is replaced with the 1–10 hPa layer in Figure 4a (not shown), suggesting that the lower-level QBO amplitude is important for the QBO–vortex teleconnection. Models with weaker 50 hPa QBO amplitude show weaker correlation in January between the 50 hPa QBO wind and the polar vortex (Figure 4b), consistent with the hypothesis that unrealistically weak low-level QBO amplitude can weaken the teleconnection.

To argue that model error leads to unrealistically weak QBO–vortex correlations, the comparison between models and reanalyses must account for the large uncertainty in the correlation values (Figure 3). This is examined in Figure 5a,b, which shows the correlation coefficient in January between vortex strength and 50 hPa equatorial wind for multiple reanalyses and for both Exp1 and Exp2 in the QBOi models. The observed response varies with the reanalysis and period chosen, but all reanalyses show that, for any of the periods containing only years from 1958 onward, the correlation is positive and statistically significant (5–95% range excludes zero). Its value ranges from

roughly 0.3 to 0.5, and all of the 5–95% confidence intervals overlap. For the longest record, 1900–2017 for ERA-20C + ERA-Interim, the correlation is weak (~ 0.1) and not significant, and its confidence interval does not overlap those of the most modern reanalyses (leftmost three bars). As noted in Section 2.1, for ERA-20C the Brönnimann *et al.* (2007) QBO reconstruction has been used to represent QBO phase and its reliability prior to 1953 is unclear. The weak correlation may simply reflect that QBO phase is not well known for this period. It is also possible that ERA-20C, which only assimilates surface data, does not provide a reliable representation of the stratospheric polar vortex at 10 hPa.

Correlations in the models are weaker than the observed correlations of ~ 0.5 in the most recent reanalyses, ranging from near zero to about 0.3. Over all models for both Exp1 and Exp2, six out of 25 have correlations that are statistically significant. In many cases the same model shows fairly different correlation values for Exp1 and Exp2, which could be caused by interference with ENSO in Exp1 but also could be due to random sampling variations, as indicated by the uncertainty ranges. Nevertheless, the correlations in the models, in both experiments, are predominantly positive. This supports a real physical connection between the QBO and the vortex in the models, since it would be unlikely to obtain the same sign of correlation over most of the ensemble due purely to chance. The number of significant correlations (six out of 25) exceeds the number which would be expected due to randomness to lie outside a 90% confidence interval (i.e., two or three).

The weak model correlations suggest errors in their representation of QBO–vortex coupling, but there is also substantial overlap between the uncertainty ranges of the models with each other and with reanalyses. To assess whether the models differ from the observations, we ask: how likely would it be to draw the observed correlation randomly from the model distributions? Figure 5c shows the correlations for Exp1 and Exp2 in the models with sample sizes close to 100 years or larger, with the uncertainty range now showing the distribution of correlations

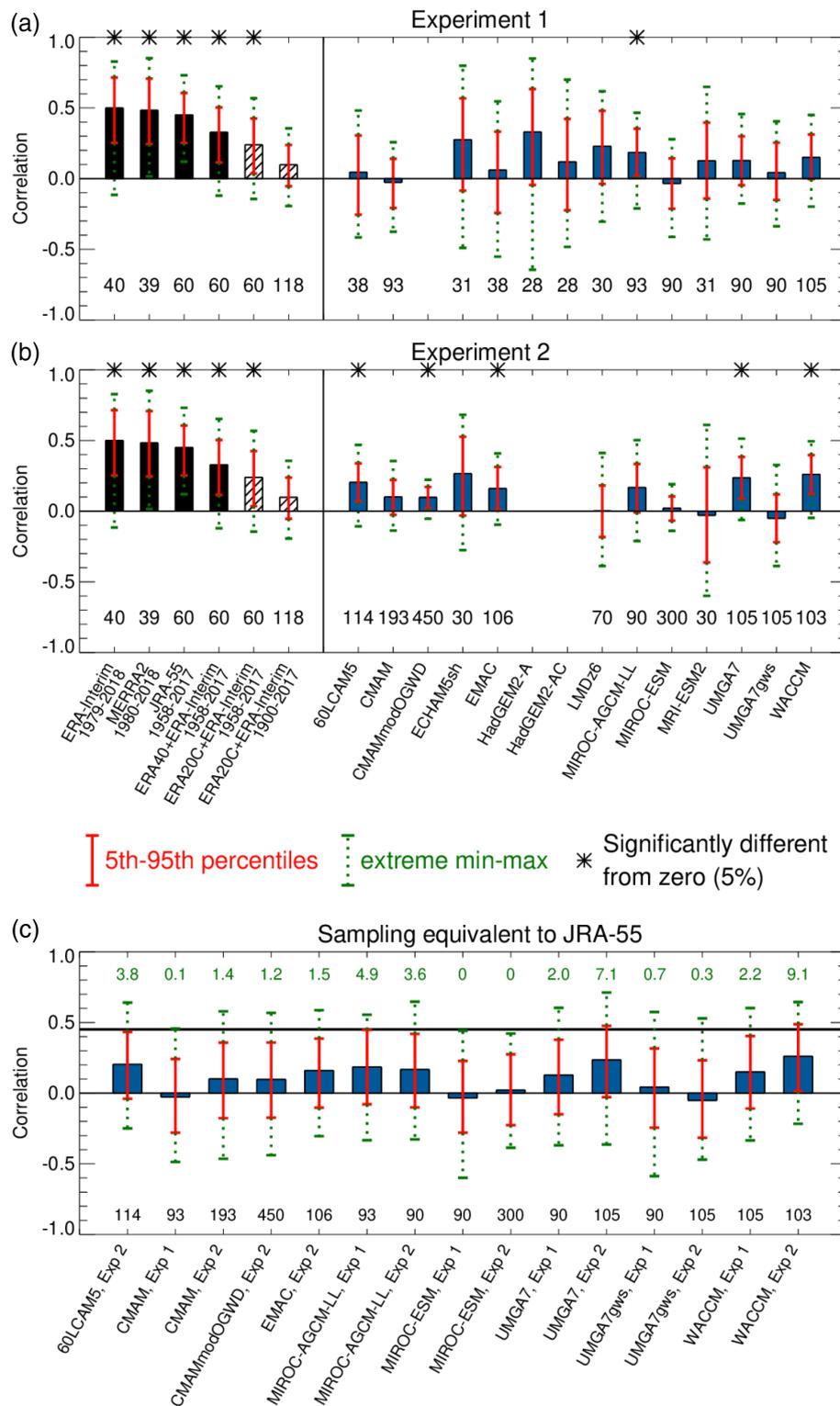


FIGURE 5 January correlation between vortex strength (monthly-mean \bar{u} at 10 hPa, 60°N) and 50 hPa equatorial wind (monthly-mean \bar{u} at 2°S–2°N). (a) Black bars at left show various reanalysis datasets over various time periods (correlations using only surface-constrained ERA20C reanalysis are hatched to symbolize that less faith should be placed in those estimates). Blue bars at right show Exp1 for each of the models. Numbers at bottom denote the number of Januaries available in each dataset, red (green) ranges show the 5–95% (min–max) range on the correlation estimated using bootstrapping (with replacement). Asterisks denote correlations that are significantly different from zero at the 10% level by a two-sided test (i.e., where the 5–95th percentile range does not encompass zero), or equivalently at the 5% level by a one-sided test. (b) is as (a) but the blue bars show Exp2. Reanalysis results are identical to those in (a) and are repeated here for ease of comparison with the models. (c) Correlations from (a, b) for model runs where 90 or more Januaries are available, but the red/green bars show the range of values that can be obtained by sub-sampling with an equivalent number of years to the JRA-55 dataset (60 years). The actual JRA-55 correlation is shown by the black horizontal line, and green numbers depict the likelihood (percentage) of obtaining a correlation this large as a random sample from the model distributions [Colour figure can be viewed at wileyonlinelibrary.com]

that could be obtained from each if it were sampled like JRA-55, that is, if 60-year samples were drawn randomly from the model distributions. The actual JRA-55 correlation of ~ 0.45 is seen to lie on the fringes of most of these distributions. In only two of the models (UMGA7 and WACCM) is there more than a 5% chance of sampling the observed correlation from a 60 year sample from the

model distribution. The multi-model mean of the probability of sampling a correlation as large as observed is only 2.5%, strongly suggesting that model error affects the representation of the midwinter vortex response to the QBO in most of the QBOi models.

In Figure 5c only one (out of 15) of the model correlations is deemed significant when 60-year samples are

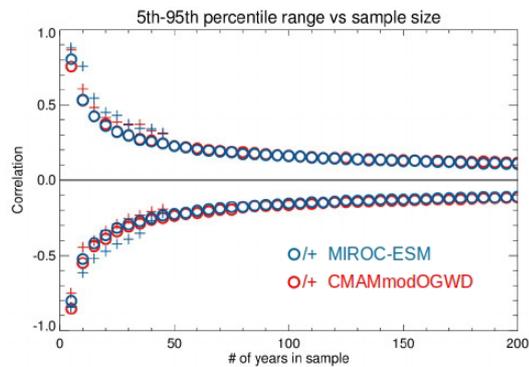


FIGURE 6 The 5–95% ranges estimated from bootstrap samples, as a function of sample size. Uncertainty ranges have been calculated for Exp2 runs of MIROC-ESM and CMAMmodOGWD as these each have a large number of years. Values shown are relative to the mean correlation for each model such that they are centred on zero. Circles show the 5–95% range for 5000 bootstrap samples calculated without autocorrelation, that is, each year is treated as independent and individual years are chosen at random. Crosses show the 5–95% range for each overlapping chunk within the model simulation, that is, this includes the possible effects of autocorrelation. The uncertainty range including autocorrelation is only shown up to a sample length of 50 years. Beyond this the estimate can be uncertain due to the lack of available independent samples [Colour figure can be viewed at wileyonlinelibrary.com]

drawn from the model distributions, but Figures 5a,b show that six of these are significant when the full sample is used for each model. How many years of model integration should be required for robust detection of a Holton–Tan response? Figure 6 shows the 5–95% range of correlation coefficients obtained by resampling from two models with very large sample sizes for Exp2, MIROC-ESM (300 years) and CMAMmodOGWD (450 years). The uncertainty range rapidly diminishes with increasing sample size over the 5–30 year range, but is beginning to stabilize at roughly 50 years. For samples larger than 100 years the range decreases very slowly. This suggests that model integrations of about 50 years should be considered a minimum, and roughly 100 years optimal, for assessing the QBO response. However it is possible that larger samples could be desirable for separating the influence of multiple factors, such as ENSO or the 11-year solar cycle. Exp2 probably represents the simplest possible case for evaluating QBO–vortex coupling in an AGCM, since by construction it only includes internal atmospheric variability.

3.3 | Sudden stratospheric warmings

A large part of the interannual variability of the NH stratospheric polar vortex is associated with SSWs (Charlton and Polvani, 2007; Butler *et al.*, 2017). The observed NH

vortex response to the QBO is not only associated with SSWs since there is a highly significant monthly-mean response in November (when SSWs are rare) but from December onward any QBO influence on the occurrence of SSWs (Dunkerton *et al.*, 1988; Andrews *et al.*, 2019) would project onto the monthly-mean responses which were shown in Sections 3.1 and 3.2. Here we define SSWs as reversals (i.e., transition to easterlies) of the daily-mean zonal-mean zonal wind at 10 hPa, 60°N following Butler *et al.* (2017)⁷. QBO influence on SSW frequency is examined by classifying each November–March (NDJFM) period as QBO-W or QBO-E based on whether the NDJFM-mean zonal-mean zonal wind at 50 hPa, 2°S–2°N is westerly or easterly.

Figure 7 shows SSW frequency in the QBOi models for Exp2 and in the JRA-55 reanalysis. The observed climatological frequency in the JRA-55 record is 0.62 (37 SSWs in 60 years). There are a similar number of SSWs during QBO-E and QBO-W phases (18 and 19 respectively; solid bars in Figure 7), but since there are fewer QBO-E years (23) than QBO-W years (37) the likelihood of an SSW occurring during a QBO-E phase is higher (0.78) than during a QBO-W phase (0.51; hatched bars in Figure 7). This is consistent with the expectation that tropical easterlies favour a more disturbed vortex (McIntyre, 1982).

The models in Figure 7 are sorted in order of increasing climatological SSW frequency, which ranges from an SSW on average occurring every four years (0.25) to one or more each year (1.1); the reanalysis lies in the middle of this distribution. A resampling-based estimate (10,000 resamples) gives a 95% confidence interval of [0.45, 0.79] for the ERA-Interim SSW frequency, and five models have climatological SSW frequencies within this interval – 60LCAM5 (0.5), WACCM (0.63), MRI-ESM2 (0.75), MIROC-ESM (0.76), and marginally LMDz6 (0.45). Six out of the eleven models considered here do not have a realistic simulation of the frequency of SSWs, which could be affecting the quality of their QBO–vortex teleconnections. Of the models that have a realistic representation of SSWs, WACCM and MRI-ESM2 show that SSWs are more likely during QBO-E years than QBO-W years (frequency differences of 0.33 and 0.31, respectively). The next largest difference in SSW frequency occurs for UMGA7 (0.29), also with SSWs more likely during QBO-E, but the occurrence of SSWs in this model is much larger than in observations and hence processes responsible for SSWs may not be well

⁷The event date of a SSW is defined to occur when daily-mean \bar{u} at 10 hPa and 60°N first changes from westerly to easterly between November and March. The wind must return to westerly for 20 consecutive days between events. If the wind does not return to westerly for at least 10 consecutive days before 30 April, then the warming is a final warming and is not included.

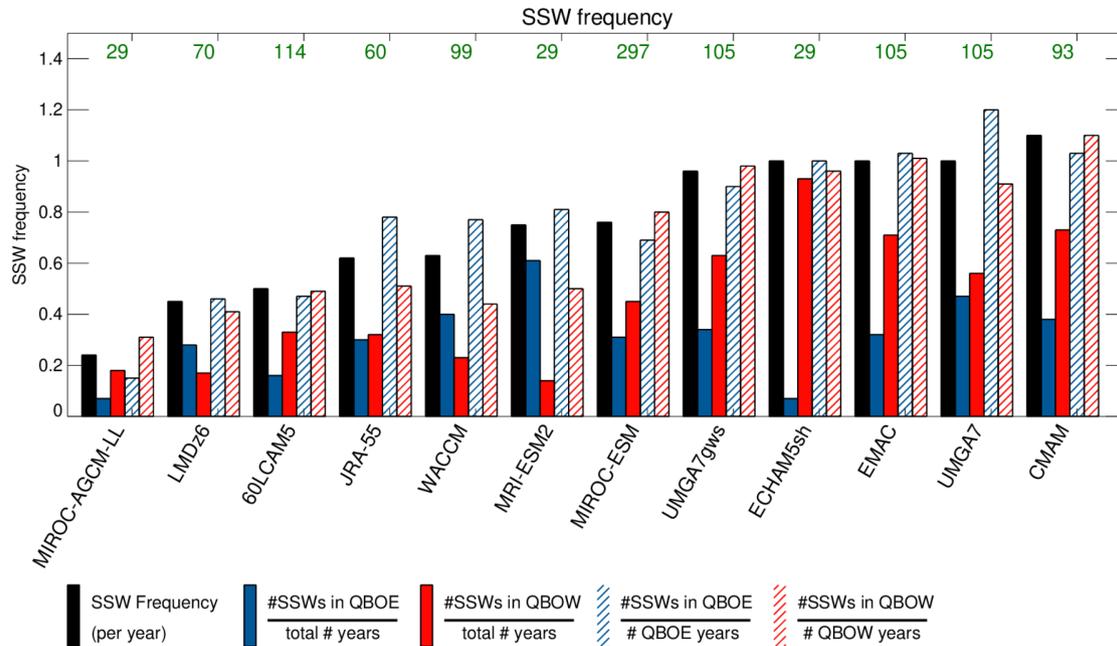


FIGURE 7 SSW frequencies during the November–March (NDJFM) period in Exp2 and JRA-55. SSWs are defined following Butler *et al.* (2017). QBO phase is defined by the sign of 50 hPa, 2°S–2°N NDJFM-mean zonal-mean zonal wind. Green numbers at top give the number of winters for each model. Models and reanalysis are sorted from lowest to highest climatological SSW frequency [Colour figure can be viewed at wileyonlinelibrary.com]

represented. The statistical significance of the E-minus-W SSW frequency differences can be assessed by resampling the climatological SSW distribution of a model to randomly draw (10,000 times) two groups the size of the E and W groups and compute the difference in SSW frequency between these two groups, yielding a distribution of SSW frequencies under the null hypothesis of no QBO influence on SSW frequency. The resulting 95% confidence intervals indicate that only WACCM and UMGA7 have SSW frequency differences that are significant at the 5% level (two-sided test). The 0.31 SSW frequency difference for MRI-ESM2, while similar to the WACCM and UMGA7 differences, is not statistically significant because of the small sample size.

The two models showing significantly more SSWs during QBO-E than QBO-W in Exp2, WACCM and UMGA7, are two of the models that showed significant correlations for the January-mean vortex winds (Figure 5, and using 50 hPa equatorial wind to define the QBO in both cases). While SSWs are distinct events, their imprint on the monthly-mean winds should lead to some similarity between the two diagnostics. Figure 8 shows the relation between SSW frequency and DJF-mean zonal-mean zonal wind at 10 hPa, 60°N in Exp2 and JRA-55. The high correlation (-0.785) between SSW frequency and mean wind is consistent with the expectation that high climatological SSW frequency is associated with a climatologically weak polar vortex and *vice versa*. This figure also indicates that

the simulation of the mean and variability of the polar vortex is quite different from observations in most models, which could be partly a reason why the QBO–polar vortex teleconnections are weaker than in observations. The models with large SSW frequency differences between QBO-E and W groups – WACCM, UMGA7, and MRI-ESM2 – have DJF-mean winds near the middle of the distribution and close to JRA-55. (UMGA7 is near the middle of the mean wind distribution even though it has unrealistically high climatological SSW frequency.) This is consistent with the occurrence of SSWs being influenced by the QBO when the vortex is in an intermediate state, rather than when it is extremely disturbed or extremely quiescent, which is a robust finding from mechanistic modelling studies (Anstey and Shepherd, 2014). However a realistic DJF-mean vortex strength seems not to be a sufficient condition for strong QBO influence on SSW occurrence, since other models near the middle of the distribution in Figure 8 show no appreciable difference in SSW frequency, at least when the response is defined using 50 hPa QBO phase. This is perhaps not surprising since other model errors (such as in the QBO itself) might impact the Holton–Tan relation. Nevertheless the results indicate that the QBO can exert an influence on the occurrence of SSW events, which in turn may lead to associated impacts on extratropical surface climate due to the downward influence of SSWs (e.g., Baldwin and Dunkerton, 2001).

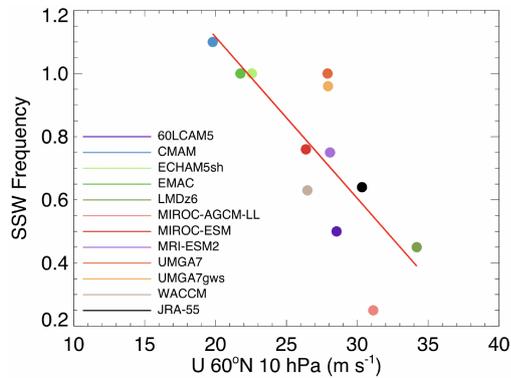


FIGURE 8 SSW frequency calculated between November and March (as in Figure 7) versus climatological December–February (DJF) zonal-mean zonal wind at 10 hPa, 60°N for QBOi models Exp2 and JRA-55 reanalysis. Red line shows a linear fit through the model data. The Pearson correlation coefficient of the fitted line is -0.785 [Colour figure can be viewed at wileyonlinelibrary.com]

3.4 | Wave forcing

Planetary-scale Rossby waves that propagate from the troposphere to stratosphere drive the polar vortex away from its quiescent radiative equilibrium state, sometimes leading to SSWs. For the QBO to influence the polar vortex state it must somehow influence the propagation and dissipation of these waves. Holton and Tan (1980) hypothesized that the QBO-E phase would cause a poleward shift of the $\bar{u} = 0$ line, which is a critical surface for stationary waves (zonal phase speed $c=0$ with respect to the ground), leading to increased wave dissipation at high latitudes and hence a weakened polar vortex. Subsequent studies suggested that the QBO-induced mean meridional circulation could affect the background wind state, altering planetary wave propagation but with a shift of the $\bar{u} = 0$ line not being the main factor causing the vortex response (e.g., Garfinkel *et al.*, 2012). We cannot resolve this question here, but in this section we examine the differences in extratropical wave forcing between QBO phases to gain insight into the vortex responses described in Sections 3.1–3.3.

Figure 9 shows the correlation between daily 10 hPa \bar{u} and 50 hPa DJF-mean QBO wind as a function of latitude and time in the seasonal cycle, for the JRA-55 reanalysis and the models for Exp2. In October JRA-55 shows significant positive correlation at roughly 30°–70°N, which then shifts poleward over the season until vanishing in March. (Negative correlations in the Tropics for JRA-55 and all models are due to 10 hPa QBO wind being anticorrelated with 50 hPa QBO wind.) Many models similarly show positive low-latitude correlation early in winter, but in many cases this remains confined to lower latitudes.

The two models showing a pattern most similar to JRA-55 are WACCM and UMGA7, which were noted earlier for their similar appearance to observations in Figure 2 and significant January correlations in Figure 5. Some indication of poleward propagation is also seen in 60LCAM5 and MRI-ESM2, although statistically significant anomalies do not reach poleward of 60°N. The correlation between the same QBO index and upward EP-flux at 100 hPa is also shown in Figure 9 (cyan and magenta contours, shown only for the NH). JRA-55 shows a statistically significant November high-latitude reduction in upward EP-flux coinciding with increased vortex strength, as well as significant reductions in the midlatitudes/subtropics in October and December–March. This 100 hPa EP-flux response is less clear in the models overall, although again WACCM and UMGA7 are perhaps the most similar (but with weaker correlations than the reanalysis, as is the case also for the wind correlation). In general, many models show statistically significant changes in extratropical 100 hPa upward EP-flux at various times over the extended winter season, but the strength and timing of this correlation differs from that seen in JRA-55. Models often show a subtropical reduction in 100 hPa EP-flux over the course of winter, as well as a high-latitude reduction coinciding with a stronger vortex, but there is substantial inter-model variation in these features. Correlations with 100 hPa EP-flux in the Tropics likely represent equatorial waves modulated by the QBO signal near the tropopause.

DJF-mean QBO W-minus-E composite differences in EP-flux divergence are shown for Exp2 models in Figure 10. Unlike the 100 hPa upward EP-flux differences shown in Figure 9, changes in EP-flux divergence at higher altitudes are more consistent across models. In most cases a large region of anomalous divergence is seen at high altitudes in the Extratropics. Below this, in the Subtropics, a vertical dipole of anomalous convergence over divergence occurs in all models. This pattern indicates decreased EP-flux convergence on the poleward flank of 50 hPa QBO-W winds and increased convergence near 10 hPa, where a poleward-shifted $\bar{u} = 0$ line is expected (due to 10 hPa QBO winds being anticorrelated with 50 hPa QBO winds, as shown in Figure 9). In some cases the lower positive lobe of this dipole extends poleward to 60°N or beyond, most clearly in ECHAM5sh, MRI-ESM2, UMGA7, and WACCM. This is consistent with the positive January correlation shown in Figure 5 for all of these except MRI-ESM2, which in Figures 2h and 9 has the timing of its strongest polar response shifted to later in the winter. Since there is a dynamical balance between wave forcing and wind anomalies in the seasonal mean, consistency between the wind response and wave forcing response is expected and these

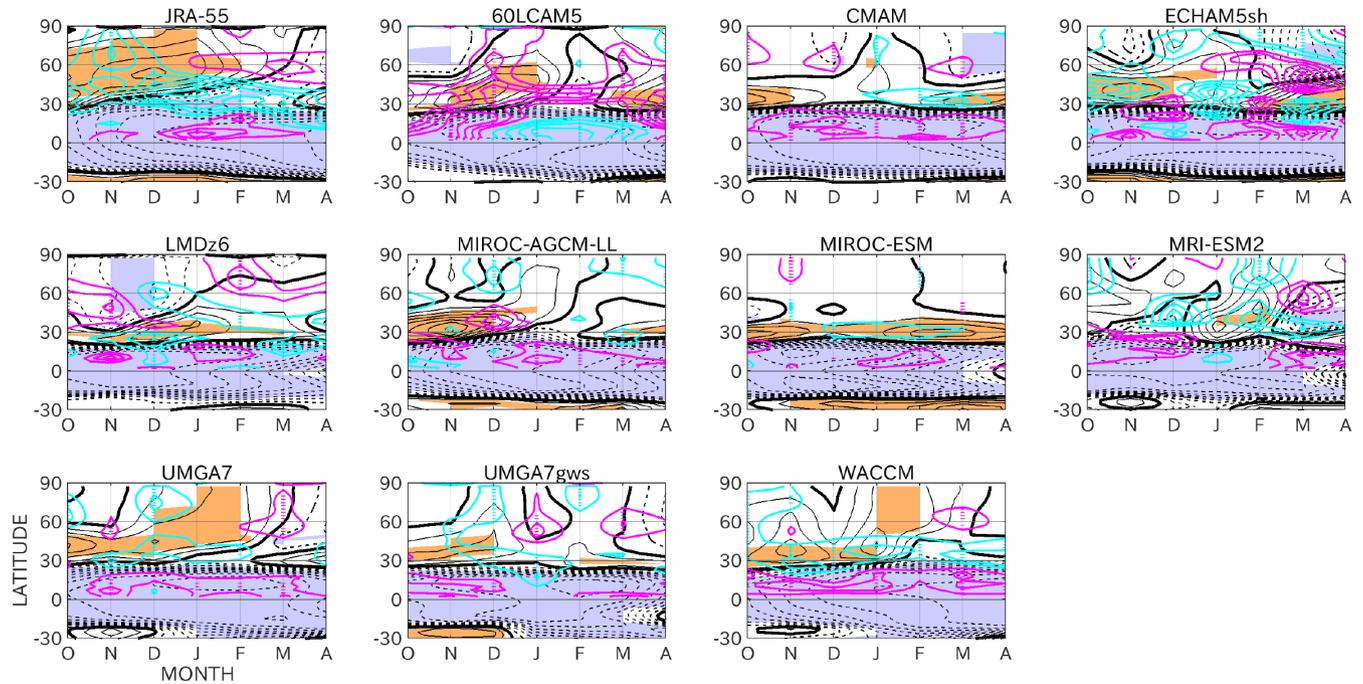


FIGURE 9 Latitude–time evolution of the correlation between 10 hPa \bar{u} and 50 hPa QBO index in JRA-55 reanalysis (1958–2018) and QBOi models for Exp2 (black line contours: solid positive, dashed negative, bold solid is zero, contour interval 0.1). Orange and purple shading denotes 90% statistical significance. Also shown are correlations with QBO index of monthly upward EP-flux at 100 hPa (coloured line contours, only shown for NH: magenta positive, cyan negative, zero omitted, contour interval 0.1). Magenta and cyan small horizontal bars denote 90% statistical significance. QBO index is defined as DJF-mean \bar{u} at 50 hPa, 2°S–2°N [Colour figure can be viewed at wileyonlinelibrary.com]

results do not explain how the wave response is induced by the prevailing tropical wind state. Nevertheless, the consistency in the models' stratospheric DJF-mean wave forcing responses indicates that some aspects of the wave response to the QBO are robust to model differences.

Despite the robustness of the subtropical EP-flux divergence response, the anomalous EP-flux vectors in Figure 10 show widely varying responses across the models, particularly at high latitudes. This strongly suggests that the high-latitude extension of QBO influence on stratospheric planetary waves is sensitive to inter-model differences. Differences in the stratospheric basic state (Figure 8) could affect the high-latitude stratospheric wave response, although cause and effect cannot be distinguished in the seasonal mean shown here. Tropospheric variability uncorrelated with the QBO could be one factor contributing to the inter-model spread in the correlation between the QBO and 100 hPa EP-flux (Figure 9), and it may be that that models do not adequately capture the relevant troposphere–stratosphere coupling mechanisms. A further possibility is that, given the varied zonal wind responses (Figures 2 and 9), upward EP-flux at 100 hPa is strongly influenced by the vortex state itself (Scott and Polvani, 2004).

4 | TROPOSPHERIC TELECONNECTIONS

Various QBO teleconnections to the troposphere have been described in the literature, as noted in Section 1, and there are a number of mechanisms by which QBO influence could be transmitted to the surface (figure 1 of Gray *et al.*, 2018 gives a schematic overview). In this section we examine two of these tropospheric teleconnections: the relation of the QBO to the NAO (Section 4.1) and to the Pacific subtropical jet (Section 4.2). QBO influence on seasonal-mean tropical precipitation during NH winter (December–February) and summer (June–August) was also examined, but results were inconclusive and further analysis is left for future work (Gray *et al.*, 2018 give the observed response). The apparent QBO influence on convective activity associated with the MJO (e.g., Son *et al.*, 2017) is also of interest but is not examined here because the output variables archived for the QBOi model runs are not suitable to characterize the MJO (Butchart *et al.*, 2018 give details). However, we refer readers to the recent study by Kim *et al.* (2020), which demonstrated that none of the CMIP6 models (i.e., similar generation to those used in QBOi) simulate the connection between QBO variability and MJO activity that is found in observations.

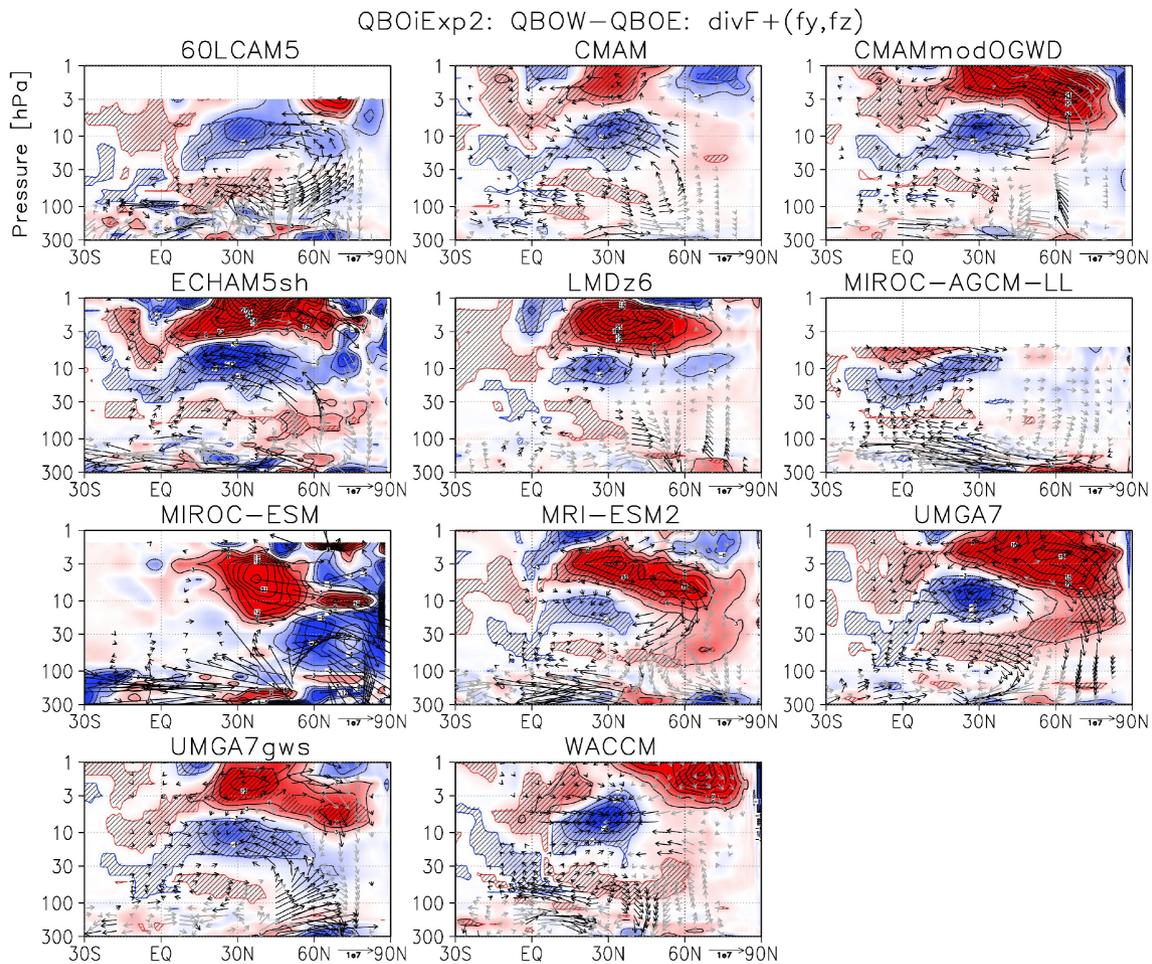


FIGURE 10 Composite differences (QBO-W minus QBO-E based on 50 hPa QBO wind) in EP-flux divergence and EP-flux vectors averaged for DJF, for Exp2 models. Statistical significance at the 95% level for EP-flux divergence differences is shown by cross-hatched areas. Black EP-flux vectors represent either poleward or upward EP-flux with a 90% statistical significance, and grey vectors with no significance. EP-flux is scaled by $\sqrt{1000/p}$ where p is the pressure in hPa. QBO index is defined as DJF-mean \bar{u} at 50 hPa, 2°S–2°N [Colour figure can be viewed at wileyonlinelibrary.com]

4.1 | North Atlantic Oscillation (NAO)

Following the same format as Figure 5, correlations in January between the NAO index and 50 hPa QBO wind are shown in Figure 11 for multiple reanalyses and for Exp1 and Exp2 models. The reanalyses all show modest positive correlation, which for two of the recent 60-year records is statistically significant. The lowest correlation, ERA-20C + ERA-Interim, may be affected by uncertainty in the reconstructed QBO phase as was noted in Section 3.2; the rest of the observed correlations range from roughly 0.2 to almost 0.3. Wide confidence intervals in all cases indicate large uncertainty in the observed value of the NAO–QBO correlation.

For Exp1, the models show no systematic relation between QBO phase and the NAO: correlations in Figure 11a are as often positive as negative (six models for each). In Exp1 it is possible that interannual

variability present in the prescribed SST and SI interferes with the NAO response to the QBO. Comparison of ENSO (Niño 3.4) and QBO indices (not shown) indicates that prominent aliasing between ENSO and 50 hPa QBO wind occurs for ECHAM5sh, LMDz6, HadGEM2-A, MRI-ESM2, UMGA7, and WACCM. However, these models do not show large differences between their Exp1 and Exp2 correlations, and confidence intervals for the Exp1–Exp2 pairs of each model have substantial overlap in all cases. This suggests that aliasing between the QBO and ENSO does not have a major impact on the model correlations shown in Figure 11a, although it is still possible that other interannual variability originating from the imposed SST and SI could interfere with the response.

For Exp2, Figure 11b shows that QBO–NAO correlations in January are more often positive than negative, and hence more consistent overall with the reanalysis

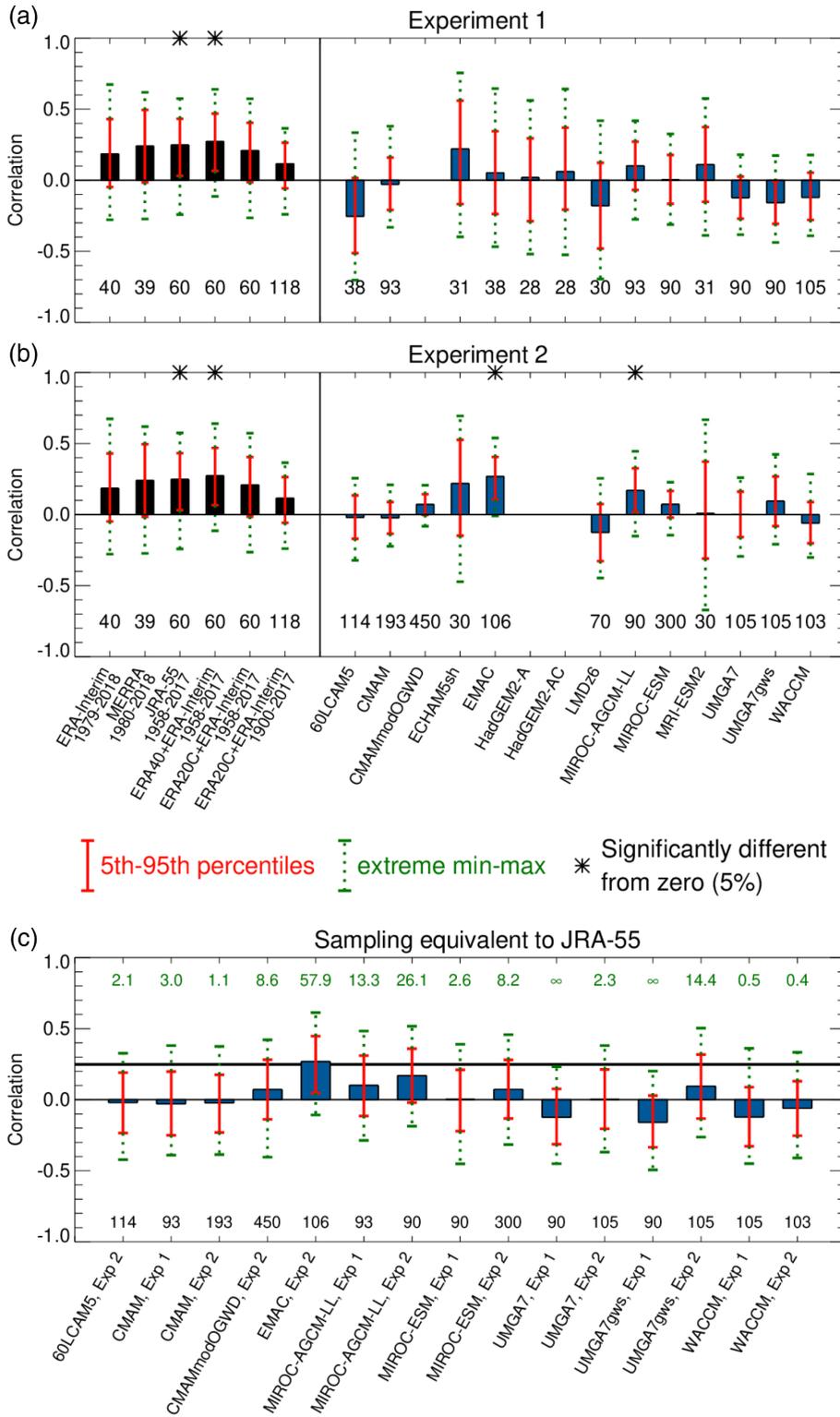


FIGURE 11 As Figure 5, but for the January correlation of 50 hPa QBO phase with the NAO. The NAO index is defined as the principal component time series of the first Empirical Orthogonal Function of monthly mean sea level pressure (MSLP) over the domain 20°–70°N, 90°W–40°E. ERA-20C boxes are not depicted as hatched due to greater confidence in the observational constraint provided by MSLP in ERA-20C than for the vortex strength shown in Figure 5 [Colour figure can be viewed at wileyonlinelibrary.com]

correlations. (Six models are positive, four negative, and the two remaining, although positive, are essentially zero.) Since Exp2 excludes non-atmospheric sources of interannual variability, the predominantly positive correlations in Figure 11b suggest there is a real QBO influence on the NAO. Two models, EMAC and MIROC-AGCM-LL, have statistically significant positive correlations, with samples

of ~100 years (although two out of twelve is only slightly higher than the number of significant correlations that would be expected by chance at the 10% significance level). Both of these models also have positive correlations for the stratospheric vortex in January (Figure 5b) which are close to being significant. This is consistent with the expectation that the NAO response and vortex response are

related, perhaps by downward influence from the stratospheric vortex, which is the “polar route” for QBO surface impacts as described by Gray *et al.* (2018). Figure 11b also shows that UMGA7 and WACCM, the two models with significant positive QBO–vortex correlations in Figure 5b, have essentially zero QBO–NAO correlation. However if Figure 5b is constructed using 30 hPa QBO phase instead of 50 hPa (not shown), correlations are similar for most models, including WACCM, but the UMGA7 QBO–NAO correlation increases to 0.17. Andrews *et al.* (2019) also found a positive NAO response to 30 hPa QBO phase using a version of the UM. This suggests that UMGA7 may also capture a “polar route” for QBO impact on the NAO. ECHAM5sh similarly shows positive QBO–vortex and QBO–NAO correlations, although neither are significant due to its short record.

Figure 11c evaluates the likelihood of drawing the observed JRA-55 correlation from the distributions of models with larger samples, as was done for the vortex (Figure 5c). The observed correlation, roughly 0.25, is marginal for many of the models – that is, it would be unlikely to draw this correlation by chance from the model distributions. This suggests model error degrades the representation of the QBO–NAO teleconnection in these models. However, Exp2 of EMAC and MIROC-AGCM-LL show substantially larger likelihoods of drawing the observed correlation; these models, according to this metric, are more consistent with the observed response. Using 30 hPa QBO phase (not shown) would also make UMGA7 more consistent with the observed response, as noted above. It is unclear why WACCM, which shows a significant positive QBO–vortex correlation in Figure 5b, has essentially zero QBO–NAO correlation.

Overall then, the QBOi multi-model ensemble provides some support for a physical link between the QBO and NAO, as indicated by predominantly positive correlations in Exp2, and the fact that drawing the observed correlation for 50 hPa QBO is likely in two of the Exp2 models. Model error may contribute to the weak correlations seen in many of the models, but it should be cautioned that the observed correlation is not large, and only barely significant in two of the reanalysis records.

4.2 | Subtropical jet

The Pacific subtropical jet has been shown to be modulated by QBO phase (Garfinkel and Hartmann, 2011b; Seo *et al.*, 2013). This response may be related to the “horseshoe-shaped” wind anomalies that extend downward from the QBO region in the Subtropics, which are likely associated with the QBO mean meridional circulation (Garfinkel and Hartmann, 2011a). Momentum

deposition by synoptic-scale waves on the upper flank of the subtropical jet, near the lowest QBO altitudes, may play a role (Naito, 2002). The subtropical jet response provides a “subtropical route” for QBO surface impacts (Gray *et al.*, 2018), which could be relevant to understanding QBO teleconnections to other phenomena that are affected by the subtropical jet, such as the NAO.

Figure 12 shows QBO-E (a–e) and QBO-W (f–j) composited anomalies from climatology, based on 50 hPa 2°S – 2°N averaged QBO wind, of the Pacific sector-mean (130° – 240°E) zonal wind averaged over October–March for the JRA-55 reanalysis and the multi-model means of Exp1 and Exp2. For JRA-55, peak sector-mean wind anomalies in the subtropical and midlatitude troposphere are small, less than $0.5\text{ m}\cdot\text{s}^{-1}$ (Figure 12a,f). At low latitudes it is important to account for the influence of ENSO, the main source of interannual variability in this region, which might confound any detection of QBO influence. The second column (Figure 12b,g) shows the same QBO composites if the ENSO response is first removed by linear regression⁸. Anomalies in the QBO region, the polar vortex, and the high-latitude troposphere are essentially unchanged. In the subtropical and midlatitude troposphere there is a strengthening of the QBO-E and QBO-W composited anomalies when ENSO is removed. However the overall pattern of the QBO-composited anomalies is not strongly sensitive to removal of ENSO. In either case the QBO response is fairly symmetric between QBO-E and QBO-W phases, is of small magnitude (less than about $0.5\text{ m}\cdot\text{s}^{-1}$), and extends deeply into the subtropical and midlatitude troposphere.

Figure 12c,h show the multi-model mean QBO-composite anomalies for Exp1. A very slight tropospheric response to QBO-E is evident, but even this vanishes when the ENSO response is removed by linear regression (Figure 12d,i). Similarly, in Exp2 (Figure 12e,j), which excludes ENSO effects by construction, almost no tropospheric response is evident in the multi-model mean. The downward “horseshoe-shaped” anomalies in the subtropical and midlatitude stratosphere resemble those in the reanalysis, but do not extend any appreciable depth below the tropopause. Hence in this seasonal (October–March) and multi-model mean the QBOi models show no evidence of a subtropical jet response.

Sampling variability of the October–March Pacific sector subtropical jet response, with ENSO removed by linear regression for the reanalysis and Exp1, is examined in Figure 13. The difference between QBO-E and QBO-W composites (E-minus-W) of the upper tropospheric

⁸The Niño 3.4 index is used as the ENSO predictor. It is defined as October–March averaged SST anomalies, linearly detrended and the averaged over the Niño 3.4 region (190° – 240°E , 5° – 5°N)

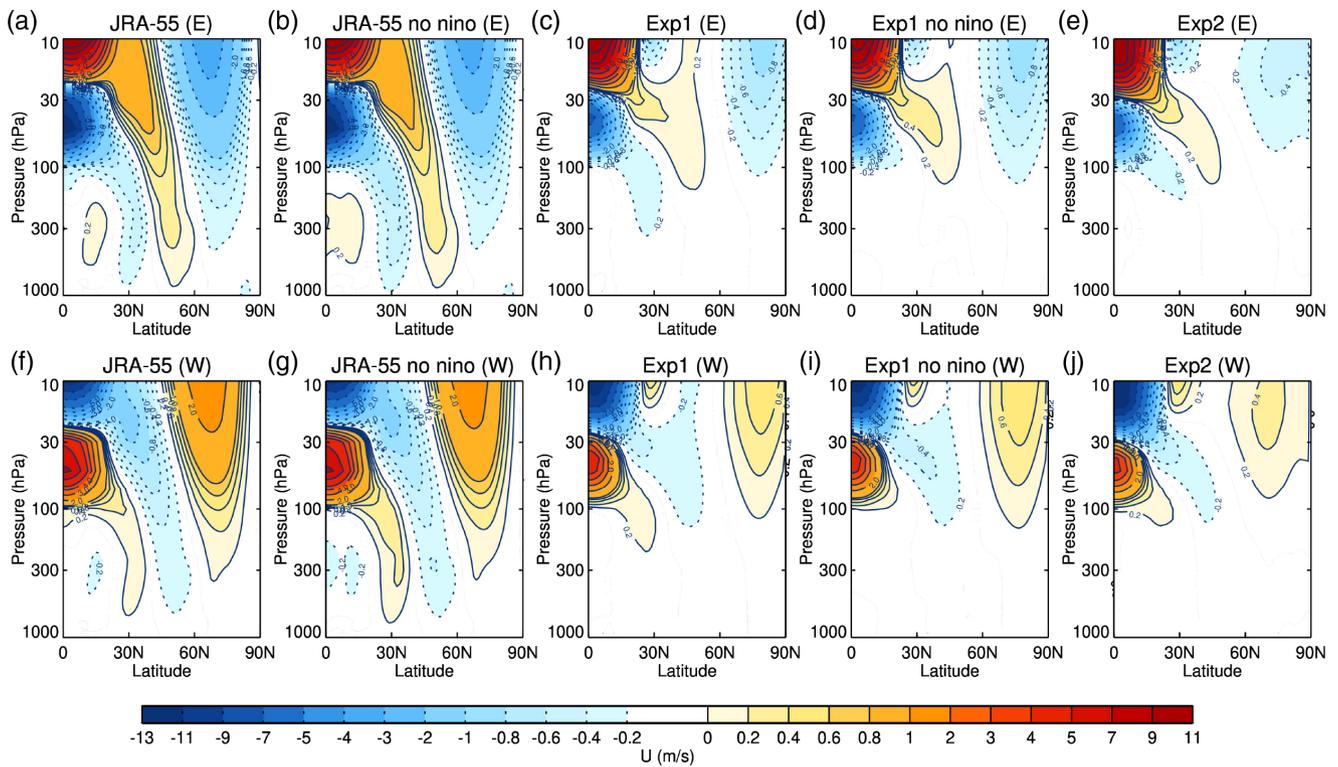


FIGURE 12 Response to QBO phase of October–March averaged Pacific sector-mean (130°–240°E) zonal wind. QBO phase is defined by the sign of 50 hPa, 2°S–2°N zonal-mean zonal wind averaged over the same months. (a–e) show QBO-E composites of the anomalies from climatology for (a) JRA-55, (b) is as (a) but with ENSO response removed, (c) multi-model mean of Exp1, (d) is as (c) but with ENSO response removed, and (e) multi-model mean of Exp2. ENSO response is removed by regressing out the Niño 3.4 contribution. (f–j) are as (a–e), but for QBO-W phase composites [Colour figure can be viewed at wileyonlinelibrary.com]

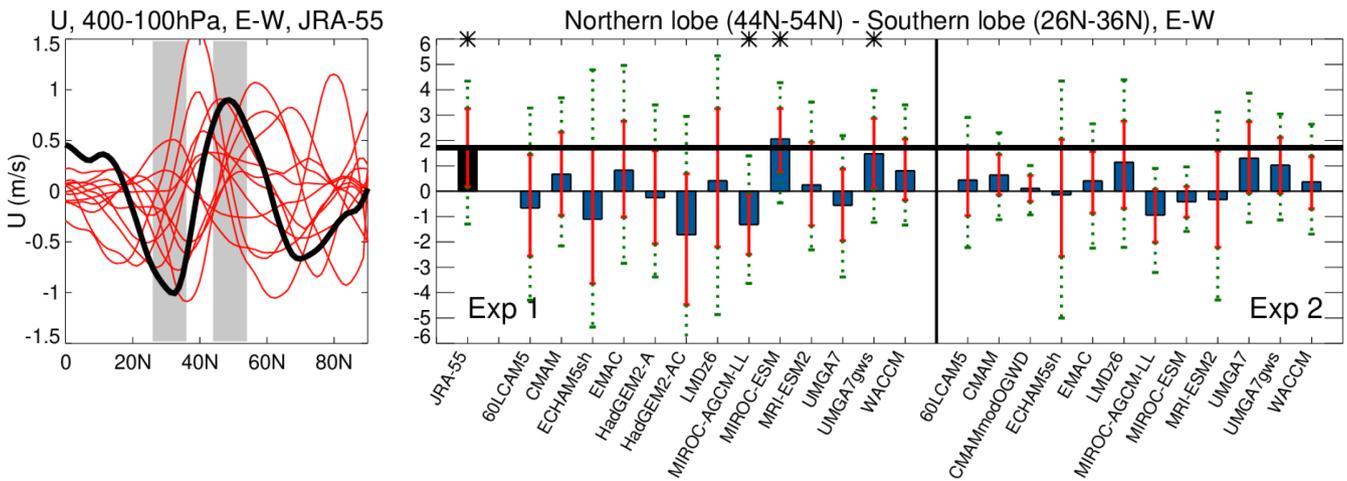


FIGURE 13 Anomalies from climatology of the pressure-weighted vertical average of zonal wind from 400 to 100 hPa averaged over the Pacific sector (130°–240°E) and averaged over October–March. (left) QBO E-minus-W difference (QBO phase defined as in Figure 12) for JRA-55 after regressing out the Niño 3.4 contribution (black) and Exp2 (red). (right) Difference between the wind anomalies averaged around the observed northern lobe (44°N–54°N) and the southern lobe (26°N–36°N) (averaging regions shown by the grey bars in the left panel). Left portion of the panel shows JRA-55 and Exp1 after removing the Niño 3.4 contribution, and right portion shows Exp2. Bars show the mean value (black for JRA-55, blue for the models). Red and green ranges are derived by resampling with replacement the number of easterly (westerly) years from within the easterly (westerly) distribution and recalculating the metric 1,000 times. Red range shows the 5–95% range and the green range shows the minimum and maximum values from these bootstrapped distributions (model sample sizes as in Figure 11). Asterisks indicate mean values that are significant as determined by where the 5–95% percentile range does not encompass zero. Bold black horizontal line shows the JRA-55 mean value for reference [Colour figure can be viewed at wileyonlinelibrary.com]

(100–400 hPa average) zonal wind anomaly is shown in order to maximize the signal. In the left panel of Figure 13, the JRA-55 response is a meridional dipole with peak values in midlatitudes and the subtropics, indicating a poleward (equatorward) Pacific sector jet shift under QBO-E (QBO-W). The difference between the northern and southern lobes of the response in JRA-55 is roughly $2 \text{ m}\cdot\text{s}^{-1}$ peak-to-peak. The models (red lines) are generally inconsistent with JRA-55 and with each other. Defining a metric of the subtropical jet response as the difference of the E-minus-W response between the northern and southern lobes, the right panel of Figure 13 shows this metric and its sampling variability for JRA-55 and the models. The JRA-55 response is statistically significant at the 10% level (two-sided test) although only marginally so. A few models show a response of similar magnitude to the JRA-55 response, but the sign of the response is not consistent across the models (15 positive, 10 negative). Three of the model responses are deemed statistically significant, which is roughly how many are expected by chance to lie outside a 90% confidence interval (i.e., two or three out of 25). Hence there is no evidence that the multi-model ensemble simulates a QBO teleconnection to the Pacific sector subtropical jet during October–March.

The lack of a response in the models suggests that model error impacts the teleconnection. One possible reason for this could be that, at the lowest altitudes of the QBO, the model QBOs are weaker and meridionally narrower than is observed, as is shown in Bushell *et al.* (2020) and can also be seen by comparing JRA-55 and model panels in Figure 12. If QBO amplitude and width in the lowermost tropical stratosphere are important for a subtropical response, this might explain why the models fail to show one. It is notable that the three models with significant Exp1 responses in Figure 13 (MIROC-AGCM-LL, MIROC-ESM and UMGA7gws) are three of the four models with highest grades (i.e., closest to reanalysis) for 50 hPa QBO meridional width in figure 11 of Bushell *et al.* (2020). However, the JRA-55 response falls within the 90% confidence interval of twelve out of 25 models in Figure 13. In many cases it would be likely to draw the JRA-55 response from model distributions that are consistent with no response, suggesting an alternate interpretation that the observed response is a statistical fluke. Further investigation into the potential effects of model biases on this teleconnection, as well as a longer observational record, may help to resolve this question.

5 | CONCLUSIONS

A multi-model ensemble of QBO-resolving models has been used to examine the robustness of several observed

QBO teleconnections during NH winter. The strength of these teleconnections (as defined by the chosen metrics for each case) varied across the models, suggesting that model error adversely impacts the representation of these teleconnections. Known errors in the representation of the QBO (Bushell *et al.*, 2020), as well as the potential effects of other model biases, make this interpretation plausible. In particular, the strength of the QBO teleconnection to the NH winter stratospheric polar vortex was shown to correlate with the amplitude of the QBO at 50 hPa, which is the altitude that shows the strongest correlation with the vortex in observations. Since weak QBO amplitude at lower levels is a systematic bias seen across multiple generations of models (Richter *et al.*, 2020), it is plausible that model representation of this teleconnection will not improve without correction of this bias.

As the sampling uncertainty of the chosen teleconnection metrics is large, it is difficult to distinguish model error from internal variability. The strength of the observed QBO teleconnections is also highly uncertain, implying that models showing weaker responses than observed are not necessarily unrealistic. Nevertheless, for both the stratospheric polar vortex and NAO responses, the low likelihood of obtaining the observed correlation from many of the models suggests that model error does indeed impact the representation of these teleconnections in the QBOi multi-model ensemble. This interpretation is most plausible in the case of the stratospheric polar vortex response (Holton–Tan effect), for which the observed teleconnection has high statistical significance. Based on statistical results from the QBOi models, it is recommended that model integrations of roughly 100 years are appropriate for examining the representation of the Holton–Tan effect in models. Fewer than 100 years is unlikely to provide robust characterization of a model's behaviour, but substantially more (exceeding 200 years) provides diminishing returns. However in more realistic climate models containing other sources of low-frequency variability besides the QBO (e.g., CMIP models), a larger sample might be required to separate QBO influence from other effects, particularly for noisier QBO teleconnections such as to the NAO or subtropical jet.

Internal variability will remain a major uncertainty in the characterization of teleconnections for the foreseeable future (i.e., until a considerably longer observational record is available), but the results presented here provide some evidence that model error impacts the representation of teleconnections in these models. This is useful because modelling uncertainties can be reduced through progress in representing relevant aspects of circulation such as the QBO and the stratospheric polar vortex, whereas internal variability is irreducible. The optimal

QBO index for the polar vortex response in many models lies in the vicinity of 50 hPa, as observed. The systematic weakness of the QBO–vortex coupling in the models might arise from the systematically weak QBO amplitudes near 50 hPa in the models (Bushell *et al.*, 2020), but other model errors could also play a role, such as climatological vortex biases (as indicated by the comparison between the two CMAM versions) or inadequate representation of stratosphere–troposphere coupling mechanisms. It is notable that in many models the optimal QBO level is slightly above 50 hPa, consistent with modelled QBOs being shifted upward relative to the real QBO (Bushell *et al.*, 2020), but in the case of the model with the largest low-level QBO amplitude (60LCAM5) the optimal level is shifted downward. This might indicate that the main causal level for the NH vortex response is the lowest altitude at which the QBO has sufficiently strong amplitude to modulate the behaviour of stratospheric planetary wave activity. One way to test this, or other hypotheses regarding QBO teleconnections, is to perform experiments using specified tropical wind states in a similar multi-model ensemble, thus controlling for one source of inter-model disagreement (i.e., the QBO itself) but otherwise sampling a range of modelling uncertainties. A better understanding of what aspects of the QBO are important for its teleconnections could, in turn, guide model development efforts to improve representation of the QBO in stratosphere-resolving climate and seasonal forecasting models.

ACKNOWLEDGEMENTS

We acknowledge the UK's Centre for Environmental Data Analysis (CEDA) for hosting the QBOi data archive. 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. NB was supported by the Met Office Hadley Centre Programme funded by BEIS and Defra. SO and LG were supported by the National Centre for Atmospheric Science and the Natural Environment Research Council (NE/P006779/1, NE/N018001/1). FL and SO were supported by the JPI-Climate/Belmont Forum 423 project GOTHAM (ANR-15-JCLI-0004-01, NERC NE/P006779/1). CC and FS were supported by the Copernicus Climate Change Service, funded by the EU and implemented by ECMWF. HN was supported by the Environment Research and Technology Development Fund

(S-12) of the Environmental Restoration and Conservation Agency of Japan and by Grant-in-Aid (JP18K03748, JP20H05171) for Science Research (KAKENHI) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. For the EMAC simulations, PB, TK, and SV acknowledge support by the state of Baden-Württemberg through bwHPC. YK was supported by the Japan Society for Promotion of Science (JSPS) KAKENHI (grant nos. JP15KK0178, JP17K18816, JP18H01286 and 19H05702) and by the Environment Research and Technology Development Fund (JPMEERF20192004) of the Environmental Restoration and Conservation Agency of Japan. YK and KH were supported by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) through its sponsorship of research at the International Pacific Research Center. SW and YK were partly supported by the “Integrated Research Program for Advancing Climate Models (TOUGOU program)” from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. The Earth Simulator was used for MIROC-ESM and MIROC-AGCM-LL simulations. We acknowledge the scientific guidance of the World Climate Research Programme (WCRP) for helping motivate this work, coordinated under the framework of the Stratosphere-troposphere Processes And their Role in Climate (SPARC) QBO initiative (QBOi) led by JA, NB, KH and SO.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

James A. Anstey  <https://orcid.org/0000-0001-6366-8647>

REFERENCES

- Andrews, M.B., Knight, J.R., Scaife, A.A., Lu, Y., Wu, T., Gray, L.J. and Schenzinger, V. (2019) Observed and simulated teleconnections between the stratospheric Quasi-Biennial Oscillation and Northern Hemisphere winter atmospheric circulation. *Journal of Geophysical Research: Atmospheres*, 124, 1219–1232. <https://doi.org/10.1029/2018JD029368>.
- Anstey, J.A. and Shepherd, T.G. (2014) High-latitude influence of the Quasi-Biennial Oscillation. *Quarterly Journal of the Royal Meteorological Society*, 140, 1–21.
- Anstey, J.A., Shepherd, T.G. and Scinocca, J.F. (2010) Influence of the Quasi-Biennial Oscillation on the extratropical winter stratosphere in an atmospheric general circulation model and in reanalysis data. *Journal of the Atmospheric Sciences*, 67, 1402–1419.
- Anstey, J.A., Gray, L.J., Fujiwara, M., Ivanciu, I., Kawatani, Y., Kiladis, G., Kim, Y.-H., Martineau, P., Schenzinger, V., Tegtmeier, S. and Wright, C. (2021). Quasi-Biennial Oscillation, Chapter 9 in SPARC S-RIP report. <https://www.sparc-climate.org/activities/reanalysis-intercomparison/>; accessed 5 May 2021.

- Baldwin, M.P. and Dunkerton, T.J. (1998) Quasi-biennial modulation of the Southern Hemisphere stratospheric polar vortex. *Geophysical Research Letters*, 25, 3343–3346.
- Baldwin, M.P. and Dunkerton, T.J. (2001) Stratospheric harbingers of anomalous weather regimes. *Science*, 294, 581–584.
- Baldwin, M.P., Gray, L.J., Dunkerton, T.J., Hamilton, K., Haynes, P.H., Randel, W.J., Holton, J.R., Alexander, M.J., Hirota, I., Hironouchi, T., Jones, D.B.A., Kinnerson, J.S., Marquardt, C., Sato, K. and Takahashi, M. (2001) The Quasi-Biennial Oscillation. *Reviews of Geophysics*, 39, 179–229.
- Boer, G.J. and Hamilton, K. (2008) QBO influence on extratropical predictive skill. *Climate Dynamics*, 31, 987–1000.
- Brönnimann, S., Annis, J.L., Vogler, C. and Jones, P.D. (2007) Reconstructing the Quasi-Biennial Oscillation back to the early 1900s. *Geophysical Research Letters*, 34(22). <https://doi.org/10.1029/2007GL031354>.
- Brönnimann, S., Malik, A., Stickler, A., Wegmann, M., Raible, C.C., Muthers, S., Anet, J., Rozanov, E. and Schmutz, W. (2016) Multi-decadal variations of the effects of the Quasi-Biennial Oscillation on the climate system. *Atmospheric Chemistry and Physics*, 16, 15529–15543. <https://doi.org/10.5194/acp-16-15529-2016>.
- 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. (2020) Evaluation of the Quasi-Biennial Oscillation in global climate models for the SPARC QBO-initiative. *Quarterly Journal of the Royal Meteorological Society*, 147. <https://doi.org/10.1002/qj.3765>.
- Butchart, N., Anstey, J.A., Hamilton, K., Osprey, S.M., McLandress, C., Bushell, A.C., Kawatani, Y., Kim, Y.-H., Lott, F., Scinocca, J., Stockdale, T.N., Andrews, M., Bellprat, O., Braesicke, P., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Dobrynin, M., Garcia, R.R., Garcia-Serrano, J., Gray, L.J., Holt, L., Kerzenmacher, T., Naoe, H., Pohlmann, H., Richter, J.H., Scaife, A.A., Schenzinger, V., Serva, F., Versick, S., Watanabe, S., Yoshida, K. and Yukimoto, S. (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, 1009–1032. <https://doi.org/10.5194/gmd-11-1009-2018>.
- Butler, A.H., Sjöberg, J.P., Seidel, D.J. and Rosenlof, K.H. (2017) A sudden stratospheric warming compendium. *Earth System Science Data*, 9, 63–76. <https://doi.org/10.5194/essd-9-63-2017>.
- Camargo, S.J. and Sobel, A.H. (2010) Revisiting the influence of the Quasi-Biennial Oscillation on tropical cyclone activity. *Journal of Climate*, 23, 5810–5825. <https://doi.org/10.1175/2010JCLI3575.1>.
- Charlton, A.J. and Polvani, L.M. (2007) A new look at stratospheric sudden warmings. Part I: climatology and modeling benchmarks. *Journal of Climate*, 20, 449–469.
- Collimore, C.C., Martin, D.W., Hitchman, M.H., Huesmann, A. and Waliser, D.E. (2003) On the relationship between the QBO and tropical deep convection. *Journal of Climate*, 16, 2552–2568.
- Coy, L., Newman, P.A., Pawson, S. and Lait, L.R. (2017) Dynamics of the disrupted 2015/16 Quasi-Biennial Oscillation. *Journal of Climate*, 30, 5661–5674. <https://doi.org/10.1175/JCLI-D-16-0663.1>.
- Domeisen, D.I., Garfinkel, C.I. and Butler, A.H. (2019) The teleconnection of El Niño Southern Oscillation to the stratosphere. *Reviews of Geophysics*, 57, 5–47. <https://doi.org/10.1029/2018RG000596>.
- Dunkerton, T.J. and Baldwin, M.P. (1991) Quasi-biennial modulation of planetary-wave fluxes in the Northern Hemisphere winter. *Journal of the Atmospheric Sciences*, 48, 1043–1061.
- Dunkerton, T.J. and Delisi, D.P. (1985) Climatology of the equatorial lower stratosphere. *Journal of the Atmospheric Sciences*, 42, 376–396.
- Dunkerton, T.J., Delisi, D.P. and Baldwin, M.P. (1988) Distribution of major stratospheric warmings in relation to the Quasi-Biennial Oscillation. *Geophysical Research Letters*, 15, 136–139.
- Ebdon, R.A. (1975) The Quasi-Biennial Oscillation and its association with tropospheric circulation patterns. *Meteorological Magazine*, 104, 282–297.
- Feng, P.-N. and Lin, H. (2019) Modulation of the MJO-related teleconnections by the QBO. *Journal of Geophysical Research: Atmospheres*, 124, 12022–12033. <https://doi.org/10.1029/2019JD030878>.
- Fujiwara, M., Wright, J.S., Manney, G.L., Gray, L.J., Anstey, J.A., Birner, T., Davis, S., Gerber, E.P., Harvey, V.L., Hegglin, M.I., Homeyer, C.R., Knox, J.A., Krüger, K., Lambert, A., Long, C.S., Martineau, P., Molod, A., Monge-Sanz, B.M., Santee, M.L., Tegtmeier, S., Chabrillat, S., Tan, D.G.H., Jackson, D.R., Polavarapu, S., Compo, G.P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J.S. and Zou, C.-Z. (2017) Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, 17, 1417–1452. <https://doi.org/10.5194/acp-17-1417-2017>.
- Garfinkel, C.I. and Hartmann, D.L. (2010) The influence of the Quasi-Biennial Oscillation on the North Pacific and El Niño teleconnections. *Journal of Geophysical Research: Atmospheres*, 115(D20). <https://doi.org/10.1029/2010JD014181>.
- Garfinkel, C.I. and Hartmann, D.L. (2011a) The influence of the Quasi-Biennial Oscillation on the troposphere in winter in a hierarchy of models. Part I: simplified dry GCMs. *Journal of the Atmospheric Sciences*, 68, 1273–1289. <https://doi.org/10.1175/2011JAS3665.1>.
- Garfinkel, C.I. and Hartmann, D.L. (2011b) The influence of the Quasi-Biennial Oscillation on the troposphere in winter in a hierarchy of models. Part II: perpetual winter WACCM runs. *Journal of the Atmospheric Sciences*, 68, 2026–2041. <https://doi.org/10.1175/2011JAS3702.1>.
- Garfinkel, C.I., Shaw, T.A., Hartmann, D.L. and Waugh, D.W. (2012) Does the Holton–Tan mechanism explain how the Quasi-Biennial Oscillation modulates the Arctic polar vortex?. *Journal of the Atmospheric Sciences*, 69, 1713–1733.
- Garfinkel, C.I., Schwartz, C., Domeisen, D.I.V., Son, S.-W., Butler, A.H. and White, I.P. (2018) Extratropical atmospheric predictability from the Quasi-Biennial Oscillation in subseasonal forecast models. *Journal of Geophysical Research: Atmospheres*, 123, 7855–7866. <https://doi.org/10.1029/2018JD028724>.
- Giorgetta, M.A., Bengtsson, L. and Arpe, K. (1999) An investigation of QBO signals in the East Asian and Indian monsoon in GCM experiments. *Climate Dynamics*, 15, 435–450. <https://doi.org/10.1007/s003820050292>.
- Gray, L.J., Sparrow, S., Juckes, M., O’Neill, A. and Andrews, D.G. (2003) Flow regimes in the winter stratosphere of the Northern

- Hemisphere. *Quarterly Journal of the Royal Meteorological Society*, 129, 925–945.
- Gray, L.J., Anstey, J.A., Kawatani, Y., Lu, H., Osprey, S.M. and Schenzinger, V. (2018) Surface impacts of the Quasi-Biennial Oscillation. *Atmospheric Chemistry and Physics*, 18, 8227–8247. <https://doi.org/10.5194/acp-18-8227-2018>.
- Hamilton, K. (1983) Quasi-biennial and other long-period variations in the solar semidiurnal barometric oscillation: observations, theory and possible application to the problem of monitoring changes in global ozone. *Journal of the Atmospheric Sciences*, 40, 2432–2443. [https://doi.org/10.1175/1520-0469\(1983\)040<2432:QBAOLP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<2432:QBAOLP>2.0.CO;2).
- Hitchcock, P. (2019) On the value of reanalyses prior to 1979 for dynamical studies of stratosphere–troposphere coupling. *Atmospheric Chemistry and Physics*, 19, 2749–2764. <https://doi.org/10.5194/acp-19-2749-2019>.
- Ho, C.-H., Kim, H.-S., Jeong, J.-H. and Son, S.-W. (2009) Influence of stratospheric Quasi-Biennial Oscillation on tropical cyclone tracks in the western North Pacific. *Geophysical Research Letters*, 36. <https://doi.org/10.1029/2009GL037163>.
- 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.M., Richter, J.H., Scaife, A.A., Scinocca, J., Serva, F., Versick, S., Watanabe, S., Yoshida, K. and Yukimoto, S. (2020) An evaluation of tropical waves and wave forcing of the QBO in the QBOi models. *Quarterly Journal of the Royal Meteorological Society*, 147. <https://doi.org/10.1002/qj.3827>.
- Holton, J.R. and Tan, H.C. (1980) The influence of the equatorial Quasi-Biennial Oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sciences*, 37, 2200–2207.
- Kim, H., Caron, J.M., Richter, J.H. and Simpson, I.R. (2020) The lack of QBO–MJO connection in CMIP6 models. *Geophysical Research Letters*, 47(11). <https://doi.org/10.1029/2020gl087295>.
- Liess, S. and Geller, M.A. (2012) On the relationship between QBO and distribution of tropical deep convection. *Journal of Geophysical Research: Atmospheres*, 117(D3). <https://doi.org/10.1029/2011JD016317>.
- Marshall, A.G. and Scaife, A.A. (2009) Impact of the QBO on surface winter climate. *Journal of Geophysical Research: Atmospheres*, 114(D18). <https://doi.org/10.1029/2009JD011737>.
- Marshall, A.G., Hendon, H.H., Son, S.-W. and Lim, Y. (2017) Impact of the Quasi-Biennial Oscillation on predictability of the Madden–Julian Oscillation. *Climate Dynamics*, 49, 1365–1377. <https://doi.org/10.1007/s00382-016-3392-0>.
- Martineau, P. (2017) *S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels*. Chilton: Centre for Environmental Data Analysis. <https://doi.org/10.5285/b241a7f536a244749662360bd7839312>.
- McIntyre, M.E. (1982) How well do we understand the dynamics of stratospheric warmings? *Journal of the Meteorological Society of Japan*, 60, 37–65.
- Mundhenk, B.D., Barnes, E.A., Maloney, E.D. and Baggett, C.F. (2018) Skilful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian Oscillation and Quasi-Biennial Oscillation. *npj Climate and Atmospheric Science*, 1, 2397–3722. <https://doi.org/10.1038/s41612-017-0008-2>.
- Naito, Y. (2002) Planetary wave diagnostics on the QBO effects on the deceleration of the polar-night jet in the Southern Hemisphere. *Journal of the Meteorological Society of Japan*, 80, 985–995.
- Nardi, K.M., Baggett, C.F., Barnes, E.A., Maloney, E.D., Harnos, D.S. and Ciasto, L.M. (2020) Skilful all-season S2S prediction of US precipitation using the MJO and QBO. *Weather and Forecasting*, 35, 2179–2198. <https://doi.org/10.1175/WAF-D-19-0232.1>.
- Newman, P.A., Coy, L., Pawson, S. and Lait, L.R. (2016) The anomalous change in the QBO in 2015–2016. *Geophysical Research Letters*, 43, 8791–8797. <https://doi.org/10.1002/2016GL070373>.
- O’Reilly, C.H., Weisheimer, A., Woollings, T., Gray, L.J. and MacLeod, D. (2019) The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. *Quarterly Journal of the Royal Meteorological Society*, 145, 131–146. <https://doi.org/10.1002/qj.3413>.
- Osprey, S.M., Butchart, N., Knight, J.R., Scaife, A.A., Hamilton, K., Anstey, J.A., Schenzinger, V. and Zhang, C. (2016) An unexpected disruption of the atmospheric Quasi-Biennial Oscillation. *Science*, 353, 1424–1427. <https://doi.org/10.1126/science.aah4156>.
- Rao, J., Garfinkel, C.I. and White, I.P. (2020) Impact of the Quasi-Biennial Oscillation on the northern winter stratospheric polar vortex in CMIP5/6 models. *Journal of Climate*, 33, 4787–4813. <https://doi.org/10.1175/JCLI-D-19-0663.1>.
- Richter, J.H., Anstey, J.A., Butchart, N., Kawatani, Y., Meehl, G.A., Osprey, S.M. and Simpson, I.R. (2020) Progress in simulating the Quasi-Biennial Oscillation in CMIP models. *Journal of Geophysical Research: Atmospheres*, 125(8). <https://doi.org/10.1029/2019JD032362>.
- Richter, J.H., Butchart, N., Kawatani, Y., Bushell, A.C., Holt, L.A., Serva, F., Anstey, J., Simpson, I.R., Osprey, S.M., Hamilton, K., Braesicke, P., Cagnazzo, C., Chen, C.-C., Garcia, R.R., Gray, L.J., Kerzenmacher, T., Lott, F., McLandress, C., Naoe, H., Scinocca, J., Stockdale, T.N., Versick, S., Watanabe, S., Yoshida, K. and Yukimoto, S. (2020) Response of the Quasi-Biennial Oscillation to a warming climate in global climate models. *Quarterly Journal of the Royal Meteorological Society*, 147. <https://doi.org/10.1002/qj.3749>.
- Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D., Folland, C.K., Gordon, M., Hermanson, L., Knight, J.R., Lea, D.J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A.K., Smith, D., Vellinga, M., Wallace, E., Waters, J. and Williams, A. (2014a) Skilful 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., Dunstone, N., Knight, J., MacLachlan, C., Manzini, E., Muller, W.A., Pohlmann, H., Smith, D., Stockdale, T.N. and Williams, A. (2014b) 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>.
- Scott, R.K. and Polvani, L.M. (2004) Stratospheric control of upward wave flux near the tropopause. *Geophysical Research Letters*, 31(2). <https://doi.org/10.1029/2003GL017965>.
- Seo, J., Choi, W., Youn, D., Park, D.-S.R. and Kim, J.Y. (2013) Relationship between the stratospheric Quasi-Biennial Oscillation and the spring rainfall in the western North Pacific. *Geophysical Research Letters*, 40, 5949–5953. <https://doi.org/10.1002/2013GL058266>.

- Sigmond, M. and Scinocca, J.F. (2010) The influence of the basic state on the Northern Hemisphere circulation response to climate change. *Journal of Climate*, 23, 1434–1446.
- Smith, A.K., Holt, L.A., Garcia, R.R., Anstey, J.A., Serva, F., Butchart, N., Osprey, S.M., Bushell, A.C., Kawatani, Y., Kim, Y.-H., Lott, F., Braesicke, P., Cagnazzo, C., Chen, C.-C., Chun, H.-Y., Gray, L., Kerzenmacher, T., Naoe, H., Richter, J., Versick, S., Schenzinger, V., Watanabe, S. and Yoshida, K. (2019) The equatorial stratospheric semiannual oscillation and time-mean winds in QBOi models. *Quarterly Journal of the Royal Meteorological Society*, 147. <https://doi.org/10.1002/qj.3690>.
- Son, S.-W., Lim, Y., Yoo, C., Hendon, H.H. and Kim, J. (2017) Stratospheric control of the Madden–Julian Oscillation. *Journal of Climate*, 30, 1909–1922. <https://doi.org/10.1175/JCLI-D-16-0620.1>.
- Stockdale, T.N., Kim, Y.-H., Anstey, J.A., Palmeiro, F., Butchart, N., Scaife, A.A., Andrews, M., Bushell, A.C., Dobrynin, M., Garcia-Serrano, J., Hamilton, K., Kawatani, Y., Lott, F., McLandress, C., Naoe, H., Osprey, S.M., Pohlmann, H., Scinocca, J., Watanabe, S., Yoshida, K. and Yukimoto, S. (2020) Prediction of the Quasi-Biennial Oscillation with a multi-model ensemble of QBO-resolving models. *Quarterly Journal of the Royal Meteorological Society*, 147. <https://doi.org/10.1002/qj.3919>.
- Toms, B.A., Barnes, E.A., Maloney, E.D. and van den Heever, S.C. (2020) The global teleconnection signature of the Madden–Julian Oscillation and its modulation by the Quasi-Biennial Oscillation. *Journal of Geophysical Research: Atmospheres*, 125(7). <https://doi.org/10.1029/2020JD032653>.
- von Storch, H. and Zwiers, F.R. (1999) *Statistical Analysis in Climate Research*. Cambridge, UK: Cambridge University Press.
- Wallace, J.M., Panetta, R.L. and Estberg, J. (1993) Representation of the equatorial stratospheric Quasi-Biennial Oscillation in EOF phase space. *Journal of the Atmospheric Sciences*, 50, 1751–1762.
- Wang, J., Kim, H.-M. and Chang, E.K.M. (2018a) Interannual modulation of Northern Hemisphere winter storm tracks by the QBO. *Geophysical Research Letters*, 45, 2786–2794. <https://doi.org/10.1002/2017GL076929>.
- Wang, J., Kim, H.-M., Chang, E.K.M. and Son, S.-W. (2018b) Modulation of the MJO and North Pacific storm track relationship by the QBO. *Journal of Geophysical Research: Atmospheres*, 123, 3976–3992. <https://doi.org/10.1029/2017JD027977>.
- Yoo, C. and 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>.

How to cite this article: Anstey JA, Simpson IR, Richter JH, *et al.* Teleconnections of the Quasi-Biennial Oscillation in a multi-model ensemble of QBO-resolving models. *Q.J.R. Meteorol. Soc.* 2021;1–25. <https://doi.org/10.1002/qj.4048>