

# Process-based evaluation of stratosphere-troposphere coupling in CMIP6 and AMIP6 models

Dillon Elsbury<sup>1,2</sup>, Zac Lawrence<sup>1,3</sup>, Amy Butler<sup>2</sup>

<sup>1</sup> Cooperative Institute for Research In Environmental Sciences

<sup>2</sup> NOAA Chemical Sciences Laboratory

<sup>3</sup> NOAA Physical Sciences Laboratory

## Abstract

This study evaluates stratosphere-troposphere coupling (STC) in reanalysis and pairs of climate models running with either prescribed 1979-2014 ocean boundary conditions (AMIP6) or with ocean-atmosphere coupling (CMIP6, historical). Motivated by evidence that differences between modeled and real-world ocean boundary conditions correlate with remote parts of the large-scale circulation, even the stratosphere, we focus on seasons with active two-way coupling between the troposphere and stratosphere: boreal winter in the North Hemisphere and austral spring in the Southern Hemisphere.

Relative to AMIP6, CMIP6 models produce colder sea surface temperatures (SSTs) over the tropical-subtropical Pacific and Arctic, but warmer SSTs over the tropical East Pacific, East Atlantic, and Southern Ocean. Regressions across models show that these regional SST biases correlate with more equatorward shifted subtropical jets and accelerated polar vortices in CMIP6 relative to AMIP6. The too-strong and too-cold CMIP6 polar vortices relative to AMIP6 and ERA5 reanalysis are further linked to overly weak climatological stationary waves in both hemispheres and issues simulating upward wave-mean flow interactions in the Southern Hemisphere spring. On the other hand, both AMIP6 and CMIP6 produce early season planetary waves that are too weak in either hemisphere compared to ERA5, but then become too strong later in the season, paralleling delayed sudden stratospheric warming timing and biases in annular mode persistence seasonality. AMIP6 better captures El Niño-Southern Oscillation (ENSO) teleconnections to the North and South Pacific; however both AMIP6 and CMIP6 still struggle to simulate ENSO-polar vortex interactions, suggesting that the source of these model biases is not solely due to poorly simulated SSTs. Similarly, the Quasi-Biennial Oscillation (QBO) morphology is comparable in AMIP6 and CMIP6. Larger QBO amplitudes and widths coincide with stronger high-latitude QBO impacts in CMIP6 models, but less so in AMIP6 models.

## Introduction

Changes in ocean boundary conditions stimulate two-way communication between the troposphere and the stratosphere. A key pathway for this coupling is through ocean-driven shifts in convective and latent heating, which can alter where and how planetary waves propagate horizontally and vertically, impacting the stratosphere and its connection to the troposphere. Stratosphere-troposphere coupling (STC) influences large-scale circulation, climate variability, and predictability from intra-seasonal to multi-decadal timescales (Butler et al. 2019, Scaife et al. 2022). Given its importance, accurately representing STC in climate models is crucial for understanding historical climate and its evolution. Herein, we evaluate how Coupled Model Intercomparison Project (CMIP) models reproduce aspects of the 1979-2014 large-scale circulation relevant to STC when run with interactive versus prescribed ocean boundary conditions.

The CMIP *historical* simulation (CMIP6) produces a synthetic climate record from 1850 to 2014 by branching from a preindustrial control state (perpetual 1850 forcing) and prescribing time-varying solar variability, volcanic aerosols,

and changes in atmospheric composition from anthropogenic greenhouse gasses (GHGs) and aerosols (Eyring et al. 2016). A key feature of CMIP6 is its coupled ocean-atmosphere configuration. In contrast, Atmospheric Model Intercomparison Project style simulations (Gates 1999), referred to as *AMIP6*, use the same forcings but prescribe observed sea surface temperature (SST) and sea ice concentration (SIC) from 1 January 1979 to 31 December 2014.

This study considers how prescribing the observed evolution of ocean boundary conditions changes CMIP6 models' representation of large-scale circulation, with a focus on STC. Comparing pairs of CMIP6 and AMIP6 simulations from the same models minimizes differences in model-specific parameterizations and internal dynamics, allowing the circulation differences to be attributed, broadly, to the different ocean configurations. CMIP6 versus AMIP6 comparisons have shown that prescribing SSTs in AMIP6 improve ENSO's extratropical teleconnections (Fang et al. 2024; Beniche et al. 2025), an effect that could enhance the realism of ENSO's stratospheric impacts (Domeisen et al. 2019; Manzini et al. 2025). The different SSTs may also be consequential for other parts of the stratosphere, for instance the tropical winds comprising the QBO (e.g., Zhou et al. 2024; Kawatani et al. 2025). However, CMIP6 versus AMIP6 comparisons have limitations. Complicating attribution, differences in circulation between CMIP6 and AMIP6 could arise from either the presence, or lack, of ocean-atmosphere coupling, and/or the impact of prescribing real-world SSTs that are unrepresentative of a model's internal climate system (e.g., Simpson et al. 2020). Given these drawbacks, we also compare CMIP6 and AMIP6 with ERA5 reanalysis (Hersbach et al. 2020), though this approach also has pros and cons. CMIP6 and ERA5 are more aligned in representing a coupled ocean-atmosphere system, while AMIP6's ocean boundary conditions are more similar to ERA5's (cf. Fig S1-S2). However, the models and ERA5 represent distinct climate systems, with basic-state differences, some unrelated to the ocean, that could drive differences in STC. The benefit of the paired CMIP6 versus AMIP6 comparison is that it better controls for these basic-state differences. Given the strengths and weaknesses of these methods to evaluate STC, we focus on comparing CMIP6 versus AMIP6, but also compare each with ERA5, with a focus on identifying how SST and SIC biases in models are related to large-scale circulation.

Model SST and SIC biases are already known to be intertwined with remote parts of the atmosphere. However, their influence on the vast network of STC pathways and mechanisms remains less documented. Rao et al. (2022) established that models with overly cold subtropical North Pacific Oceans tend to have too strong boreal winter polar vortices, and Feng et al. (2024) showed that overly cold subtropical South Pacific Oceans relate with too weak austral winter polar vortices. The high-latitude circulation has been characterized using Annular Modes (AMs), the dominant pattern of extratropical climate variability in either hemisphere (Thompson et al. 2000), which are represented marginally better by AMIP6 (Lee et al. 2021). Regional SST biases like the cold equatorial Pacific bias (Li and Xie 2014) impact ENSO's teleconnections (Bayr et al. 2024) and tropical-subtropical SST biases affect the storm-tracks (Priestley et al. 2023).

Impacts of sea ice biases are also expected to have far-reaching connections with the large-scale circulation (e.g., Smith et al. 2022). CMIP6 models generally capture the seasonality of historical Arctic and Antarctica sea ice (Shu et al. 2020; Watts et al. 2021). However, in the Arctic, they underestimate the sensitivity of observed sea ice decline to increasing GHGs or increasing global surface air temperature (Notz and SIMIP Community 2020). Model reductions of historical sea ice extent in the Barents Sea are too modest; this is an influential region on the strength of the boreal winter polar vortex, which weakens in some models when forced with sea ice loss alone (Sun et al. 2015; Smith et al. 2022, Pan et al. 2023). Around Antarctica, simulated sea ice remains too sparse during September and February (months of maximum and minimum extent, respectively), reflecting a bias towards negative sea ice trends that contrast with the relatively weak observed trends (Roach et al. 2020).

Our analysis focuses on 1979-2014 boreal winter and austral spring, since two-way coupling between the extratropical stratosphere and troposphere is pronounced during these seasons. Section 2 details the use of CMIP6 and AMIP6 simulations and Sections 3.1 and 3.2 document their boundary conditions and climatological states. Subsequent sections describe and compare the CMIP6 and AMIP6 zonal-mean zonal-winds, upward and downward STC, Annular Modes, major Sudden Stratospheric Warmings (SSWs), and select boreal winter teleconnections forced by ENSO and the QBO. We present our discussion and conclusion in Section 4.

## Methodology

All analyses focus on the 1979-2014 time period, the period of AMIP6 simulations. ERA5 reanalysis serves as the verification data set of real-world atmospheric variability against which the models are evaluated (Hersbach et al. 2020). Various “biases” are mentioned and its meaning varies with context: it can describe CMIP6 differences relative to AMIP6, or CMIP6 and AMIP6 differences relative to ERA5. The variables used and their CMIP variable names are temperature (*ta*), geopotential height (*zg*), zonal (east-west) wind (*ua*), meridional (north-south) wind (*va*), and surface air temperature (*tas*). We utilize both monthly and daily means, as clarified within each analysis. Since the main focus of this study is on large-scale circulation, all data is regressed onto a common 2.5° horizontal grid using bilinear interpolation.

The models are listed in Table S1 in supporting information. Every figure only compares models for which we have at least one CMIP6 and AMIP6 ensemble member providing all variables required for that analysis. Because the number of models used per figure is conditioned on availability of one to several variables in CMIP6 and AMIP6, the number of models per figure can fluctuate, but is at least 32 for monthly (*Amon*) and 26 for daily (*Eday + day*) analyses. We allow a model’s number of members to differ between CMIP6 and AMIP6 within an analysis, on the premise that standardizing to a common lower number of members disqualifies data. Table S1 lists model-specific configurations, including the altitude of each model top, whether or not it is a high-top model (> 80 kilometers), number of vertical levels, and horizontal resolution, all of which impact aspects of STC and the representation of climate modes (Charlton-Perez et al. 2013; Hurwitz et al. 2014; Lee et al. 2021; Lawrence et al. 2022; Garfinkel et al. 2025).

A few statistical approaches help benchmark model performance. To pinpoint pervasive model discrepancies, we identify spatial grid points where a majority of model-means (>80%) agree on the sign of a CMIP6 minus AMIP6 difference. For example, this picks out oceanic regions where surface air temperature differs systematically between CMIP6 and AMIP6 (Fig. 1a, 1d). To establish processes that models simulate well compared to reanalysis, 95th percentile confidence intervals of ERA5’s 1979-2014 interval variability are obtained by resampling observed seasons with replacement 10,000 times. This determines whether a model’s depiction of a process lies within the bounds of observed internal variability. Some analyses involve computing multi-model mean (MMM) mean correlations. In these instances, each model correlation coefficient is transformed using the Fisher Z formula, means are computed, and an inverse transformation is applied to obtain correlation coefficient averages.

Some common definitions are used across analyses. The polar vortex is defined as the zonal-mean zonal-wind at 10 hPa and area-averaged over 55°N-65°N/S. Polar cap averages refer to means of geopotential height poleward of 65°N/S. Planetary waves are defined two ways, by the amplitude of the eddy (deviation from zonal-mean) geopotential height along an isobar (e.g., 10 or 200 hPa), and by the eddy heat flux, the product of eddy meridional-wind and eddy temperature (Andrews et al. 1987). Annular modes (AMs) are defined following Baldwin and Thompson (2009) for each pressure level and hemisphere (poleward 20° latitude) using daily zonal-mean geopotential anomalies. Empirical orthogonal functions (EOFs) are derived as in Gerber et al. (2010). For each day

and pressure level, we subtract the global-mean geopotential at each latitude, deseasonalize using a centered 15-day running mean, detrend using a linear trend, apply  $\sqrt{\cos(\text{latitude})}$  weighting to account for meridian convergence towards the pole, and scale the EOFs by the square root of their eigenvalues.

Many of the STC diagnostics used in this study are publicly available through the NOAA Model Diagnostic Task Force (Maloney et al. 2019), accessible at <https://github.com/NOAA-GFDL/MDTF-diagnostics/tree/main/diagnostics>. These diagnostics were originally developed to evaluate a single climate model simulation against reanalysis, but for the purposes of this inter-model study, we have adapted them to operate across multiple models. Within the repository, diagnostics associated with this study can be identified by the prefix “`stc_`”, and they generate single-model versions of the results shown in Figures 5–10 and Figure 16. Figures 12–15 and Figure 17 are also derived, though more distantly, from this diagnostics package, while Figures 1–4, and 11, are not part of the package.

### 3. Results

#### 3.1 Climatological circulation differences between AMIP6 and CMIP6

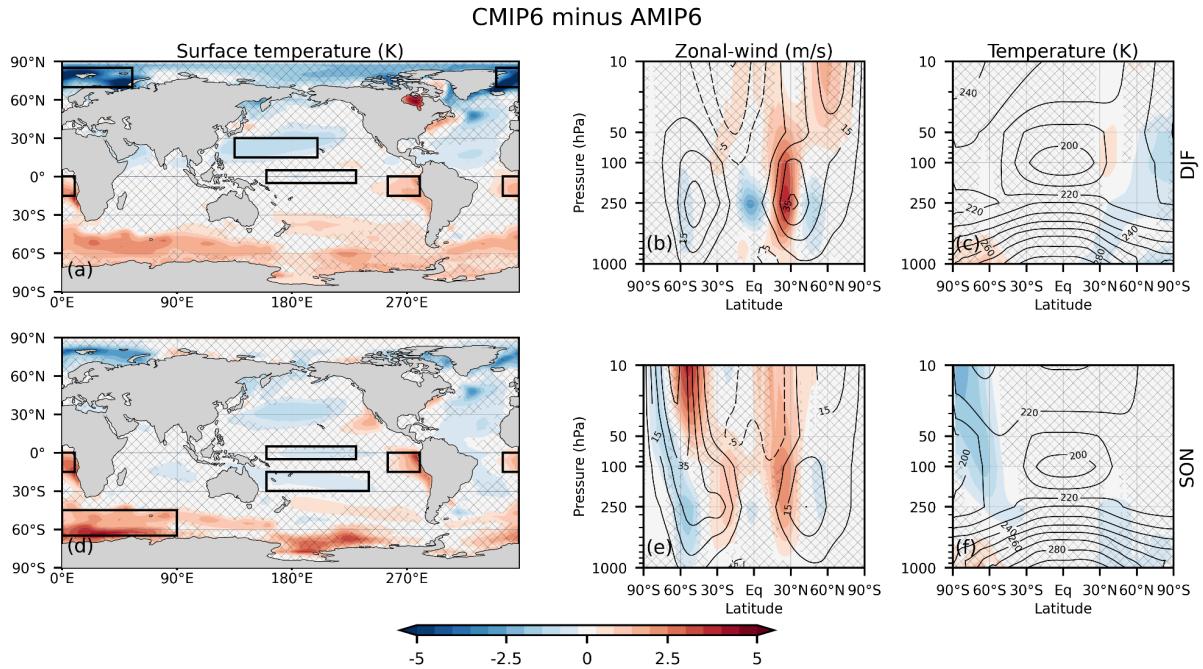


Figure 1: Shaded CMIP6 minus AMIP6 differences in monthly surface air temperature (left), zonal-mean zonal-wind (middle), and zonal-mean temperature (right) across 32 models. Rows one and two show DIF and SON, respectively. For the zonal-wind and temperature subplots, climatological AMIP6 quantities are overlaid in black contours. For all subplots, non-stippled areas signify where at least 26 of 32 models (~81%) agree on the sign of the CMIP6 minus AMIP6 differences. Black boxes on the surface air temperature plot show key regions used in the Figure 3 regressions.

Figure 1 compares the monthly climatologies of surface air temperature, zonal-mean zonal-wind, and zonal-mean temperature across 32 models in boreal winter (Fig. 1, top) and austral spring (Fig. 1, bottom). Because AMIP6 uses prescribed Merged SST based on UK MetOffice HadISST and NCEP OI2 ocean boundary conditions, the surface air temperatures in AMIP6 evolve similarly (though are not identical) to ERA5 over the 1979–2014 period (not shown). Figure 1 is produced without detrending so that mean state differences embedded in CMIP6 and AMIP6 are preserved. Since modeled and observed surface air temperature trends can differ significantly (Wills et al. 2022), they are briefly quantified here. For the period 1979–2014, AMIP6 models produce global-mean annual-mean surface air temperature increases between 0.43 and 0.59 K, comparable to the 0.59 K warming in ERA5. In contrast,

CMIP6 models show a broader range (0.55–1.34 K). Because CMIP6 simulations branch off of their preindustrial controls starting in 1850, these experiments effectively begin from 1979 base states that differ from AMIP6, which may explain the trend differences.

Climatologically, CMIP6 is warmer over the eastern Pacific and Atlantic in either season (Fig. 1a, 1d) and exhibits a “cold tongue bias” over the equatorial central-western Pacific, relative to AMIP6. The cold tongue bias is not consistently detected in boreal winter, but is robust in austral spring. CMIP6 is also colder than AMIP6 over the subtropical North Pacific and Atlantic in a majority of models. The North Atlantic to Barents-Kara region of the Arctic remains too cold during boreal winter in CMIP6 (Fig. 1a), while the Southern Ocean is too warm in austral spring (Fig. 1d). A too cold Arctic winter is a persistent bias in coupled climate models (Davy and Outten 2020). The too warm Southern Ocean bias may reflect deficiencies in model cloud representation (Gao et al. 2024) or incomplete coupling between atmosphere circulation and Ekman transport (Hu et al. 2024).

The CMIP6 subtropical jet is shifted equatorward during boreal winter (Fig. 1b) and austral spring (Fig. 1e) compared to AMIP6. The eddy-driven jets, located at higher-latitudes and characterized by stronger lower tropospheric winds, are equator-shifted in CMIP6. Figure 2 further illustrates these zonal-mean circulation differences by comparing maximum jet velocities and their corresponding latitudes between CMIP6 (blue) and AMIP6 (orange) (cf. Fig. 1 of Butchart et al. 2011).

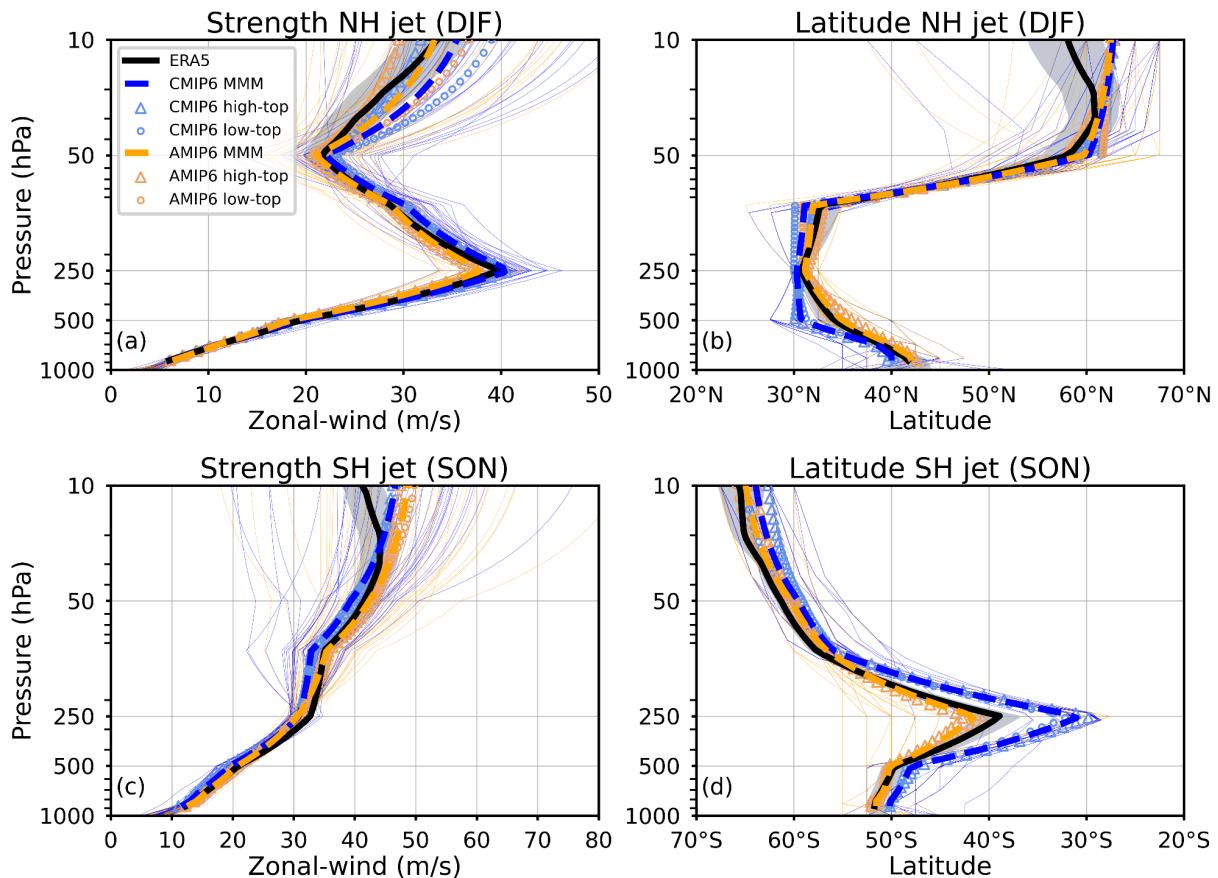


Figure 2: Maximum velocity of the monthly zonal-mean zonal-wind between 20° and 90°N/S at each isobar (left) and their corresponding latitudes (right). Top shows NH (DJF) and bottom shows SH (SON). (a, c) show jet speed maximums and (b, d) show latitudes. ERA5 quantities and their confidence intervals are shown by black

dashes and gray shading, respectively. CMIP6 and AMIP6 MMMs are shown by blue and orange dashes, respectively. Thin lines show individual model (41) responses. High-top (20) and low-top (21) model responses are shown as unfilled triangles and circles, respectively.

In boreal winter in ERA5 (Fig. 2a, b; black dashes), the NH zonal-mean jet strength peaks at 250 hPa in the subtropics, with a secondary maxima in the stratospheric polar vortex. The CMIP6 and AMIP6 (blue and orange dashes, respectively) MMMs of jet strength and latitude lie close to ERA5, however individual model responses can fall far outside of ERA5's confidence intervals. The CMIP6 subtropical jet is too strong and equatorward relative to ERA5/AMIP6. The CMIP6 polar vortex is also stronger than in AMIP6/ERA5 (Fig. 1b, 2a), coinciding with colder polar lower stratospheric temperatures in a majority of the CMIP6 models (Fig. 1c). AMIP6 reduces the vortex bias, but even its high-latitude wind above 50 hPa is stronger than ERA5 (Fig. 2a). Additionally, CMIP6 and AMIP6 shift the 10 hPa vortex poleward relative to ERA5 (Fig. 2b), suggesting these biases may arise from model deficiencies of stratospheric origin.

In austral spring, CMIP6 and AMIP6 SH jet strength mimics ERA5 in the lower atmosphere, but diverges above 20 hPa where both model polar vortices are too fast (Fig. 2c). Paradoxically, despite a colder polar stratosphere in CMIP6 than AMIP6 (Fig. 1f), 10-30 hPa jet maximum speeds are stronger in AMIP6, which may be due to differences in the computed latitudes of the jet maximums (Fig. 2d). The entire austral spring jet column in CMIP6 is shifted equatorward relative to AMIP6/ERA5, though the bias maximizes at 250 hPa, where the jet is almost 10 degrees further equatorward (Fig. 2d). The tendency of the SH jet stream to be too far equatorward is noted in previous studies (Simpson et al. 2016, Bracegirdle et al. 2020) and was shown to be of slightly larger magnitude in CMIP5 (Bracegirdle et al. 2013). Notably, here CMIP6 also shows an equatorward bias, but AMIP6 instead shows a small poleward bias relative to ERA5. 24 AMIP6 models are shifted poleward, while only 17 are shifted equatorward, relative to ERA5. This poleward bias appears to arise in AMIP6 because these models underestimate the subtropical jet maxima at 250 hPa.

When models are grouped by model top, high-top models show NH polar vortex amplitudes within ERA5's sampling variability, while low-top models produce a vortex that is too strong in CMIP6 and AMIP6 (Fig. 1a), in agreement with previous studies (e.g. Charlton-Perez et al. 2013). High-top models also more accurately capture NH jet latitude between 50–250 hPa (Fig. 1b). In the SH, the vortex amplitude bias is reduced in high-top models (Fig. 1c), though low-top models better match ERA5 in vortex latitude placement (Fig. 1d).

In summary, the climatological circulations of CMIP6 and AMIP6 differ in boreal winter and austral spring. CMIP6 runs tend to show amplified subtropical jet equatorward flanks and stronger polar vortices relative to AMIP6, consistent with colder lower stratospheric 60°N/S temperatures in a majority of models. These polar stratospheric biases are also evident if CMIP6 and AMIP6 are compared to ERA5 in a manner like Figure 1 (Fig. S1, Fig. S2), underscoring that the models have basic-state issues separate from the atmosphere-ocean coupling impacting STC. Between AMIP6 and CMIP6, AMIP6 runs show a wider distribution of stratospheric polar vortex velocities in both hemispheres. The zonal-mean circulation differences underscore the sensitivity of large-scale circulation to the ocean boundary conditions in CMIP6 and AMIP6, which are considered more in the next section.

### 3.2 Linking biases in ocean boundary conditions to differences in climatological circulation

To explore how the CMIP6 minus AMIP6 regional surface air temperature anomalies (Fig. 1a, 1d) are associated with the circulation differences (Fig. 1b, 1e), we regress across models the climatological zonal-wind differences from each model at every latitude-isobar grid point onto their regionally averaged temperature differences in black boxes on Figures 1a and 1d. To account for differences in trends amongst models, each model's surface air

temperature differences are divided by its 1979–2014 global-mean surface air temperature change. The regression coefficients are then scaled by their respective regionally averaged temperature anomaly to obtain an estimate of the expected zonal-mean zonal-wind anomaly associated with a given regional temperature imbalance between CMIP6 and AMIP6 (Fig. 3). The top and bottom of Figure 3 show boreal winter and austral spring, respectively, and are compared with overlaid black contours, which show CMIP6 minus AMIP6 climatological differences from Figures 1b and 1e.

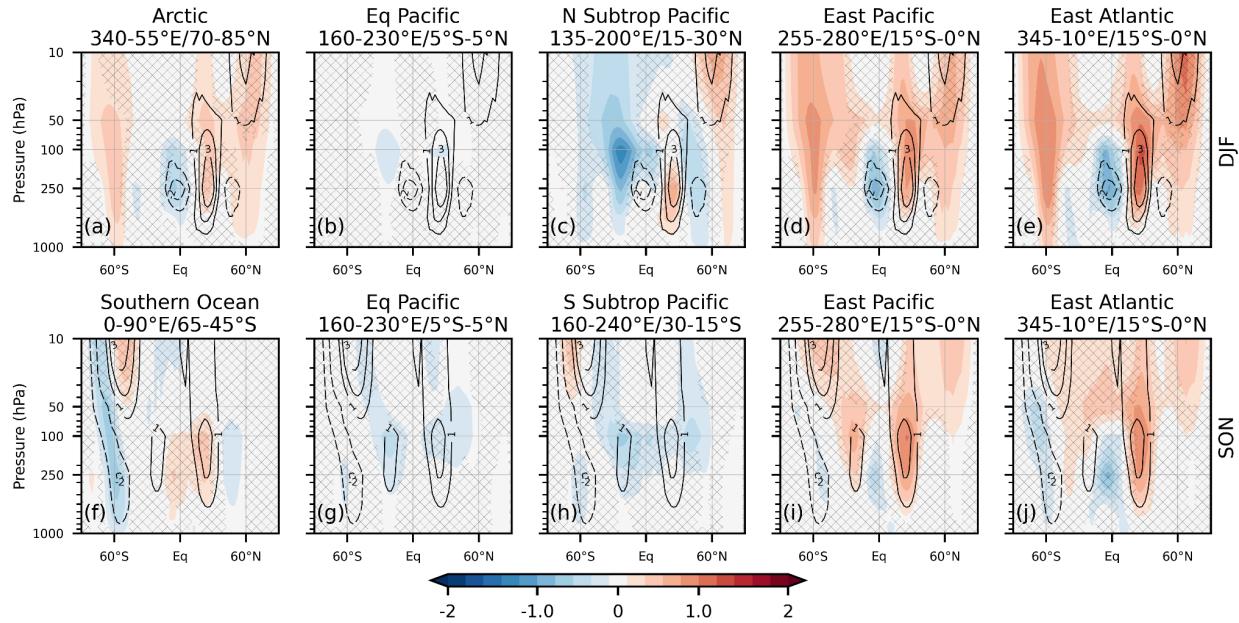


Figure 3: Across model regressions of DJF (top) and SON (bottom) CMIP6 minus AMIP6 zonal-mean zonal-wind onto regional averages of CMIP6 minus AMIP6 surface air temperature taken over the domain indicated in each title. Shading shows regressed zonal-wind. Statistically significant regression coefficients ( $p < 0.05$ ) are not hatched. Overlaid black contours show the CMIP6 minus AMIP6 zonal-mean zonal-wind anomalies that are shaded in Figure 1b and 1e.

Reductions in Barents-Kara sea-ice can weaken the polar vortex (Xu et al. 2024), suggesting that the pronounced CMIP6 cold anomaly in this region, indicative of more sea-ice (Fig. 1a), might explain the accelerated NH polar vortex relative to AMIP6 (Fig. 1b). Our analysis indicates a relation between sea ice and the stratospheric vortex of the correct sign, although the relationship is weak and not statistically significant above 100 hPa (Fig. 3a). The Arctic regression bears a close resemblance to the East Pacific, and East Atlantic regressions, hinting at a common dynamical link (Fig. 3a, 3d-e).

The equatorial Pacific cold tongue bias in CMIP6 (Fig. 1), acting effectively as a central Pacific La Niña-like base state, might be expected to accelerate the CMIP6 NH polar vortex via established ENSO pathways (Domeisen et al. 2019). However, the central/west Pacific cold bias shows little association with high-latitude zonal-winds and instead corresponds to a minor weakening of the subtropical jet in both hemispheres (Fig. 3b), in contrast with the CMIP6 minus AMIP6 subtropical jet acceleration (Fig. 1b).

The subtropical North Pacific cold bias, tropical eastern Pacific warm bias, and tropical eastern Atlantic warm bias in CMIP6 (Fig. 1a) are associated with an intensified equatorward flank of the NH subtropical jet (Fig. 3c-e), mirroring the CMIP6 jet acceleration relative to AMIP6 (Fig. 1b). Additionally, the subtropical North Pacific regression resembles the CMIP6 minus AMIP eddy-driven jet differences (Fig. 3c). Overall, these three regions, along with the Arctic response, exhibit the strongest wind anomalies. They all include an accelerated stratospheric polar vortex in

CMIP6, although none of the polar stratospheric regression coefficients are statistically significant above 50 hPa in the NH (Fig. 3a, 3c-e)

A similar general picture emerges for the SH during austral spring. Anomalously low spring Antarctic sea-ice can weaken the polar vortex (Wang et al. 2019; Rea et al. 2024), implying that CMIP6 Southern Ocean warming (Fig. 1d), indicative of less sea-ice, might explain the modest deceleration of the SH high-latitude wind relative to AMIP6 (Fig. 1e). The regression indicates a relation between warm Southern Ocean and weaker 50-1000 hPa high-latitude zonal-wind (Fig. 3f), however, there is no robust change in the polar stratospheric zonal-wind above 50 hPa.

In CMIP6, the cold biases in the equatorial and subtropical South Pacific exhibit minimal correlation with high-latitude SH zonal-wind differences relative to AMIP6 (Fig. 3g-h). Instead, these biases align with a slight weakening of the subtropical jet in both hemispheres, a pattern that contrasts with the subtropical jet acceleration seen in CMIP6 minus AMIP6 (Fig. 1e). In contrast, the CMIP6 East Pacific and Atlantic warm biases are related to the intensified equatorward subtropical jet flanks in both hemispheres and are also suggestive of NH polar vortex strengthening during SON (Fig. 3i-j)

In summary, the regressions point to a close relationship between biases in CMIP6 subtropical jets and biases in tropical-subtropical ocean boundary conditions (though we note that causality cannot be assessed). However, connections to the too strong and cold polar vortices remain more elusive. To further investigate differences in AMIP6 and CMIP STC processes, we next explore biases in large-scale atmospheric waves and their coupling to the mean flow.

### **3.3 Effect of ocean boundary conditions on upward troposphere to stratosphere coupling**

Vertically and horizontally propagating quasi-stationary planetary waves drive variability in the extratropical circulation. These waves originate from interactions between the zonal-mean zonal-wind, orography, diabatic heating, oceanic horizontal heat fluxes and land-sea contrasts (White et al. 2017; Garfinkel et al. 2020). Tropical convective heating modulates planetary waves (Scaife et al. 2017), enabling them to communicate the influence of tropical climate variability to high-latitudes (Yamazaki et al. 2020; Manzini et al. 2024). In boreal winter and austral spring, these waves can reach the stratosphere and occasionally break (McIntyre and Palmer 1983), depositing momentum and slowing down the polar stratospheric zonal-winds in events called sudden stratospheric warmings/SSWs (Matsuno 1971; Perlitz and Harnik 2004).

SSWs are characterized by rapid increases in polar stratospheric temperatures and a reversal of the daily 10 hPa westerly winds at 60°N from westerlies to easterlies (Charlton and Polvani 2007; Butler and Gerber 2018). They occur roughly every other winter, but multiple are possible during a given season (Lee et al. 2025). These events are primarily driven by upward propagating planetary waves decelerating the zonal-mean zonal-wind (Matsuno 1971; Birner and Albers 2017; Martineau et al. 2018). SSWs can drive significant and persistent surface impacts across the northern hemisphere. Canonically, this impact resembles the negative phase of the North Atlantic Oscillation, which shifts the storm-track and associated weather impacts south (Butler et al. 2017; Ayarzagüena et al. 2018; Maycock et al. 2020), although whether or not this response develops depends on the magnitude of stratospheric disturbance (Karpechko et al. 2017), the morphology of the SSW wave forcing (White et al. 2019), and the ocean and tropospheric boundary conditions present during the SSW (Domeisen et al. 2020; Dai and Hitchcock 2021).

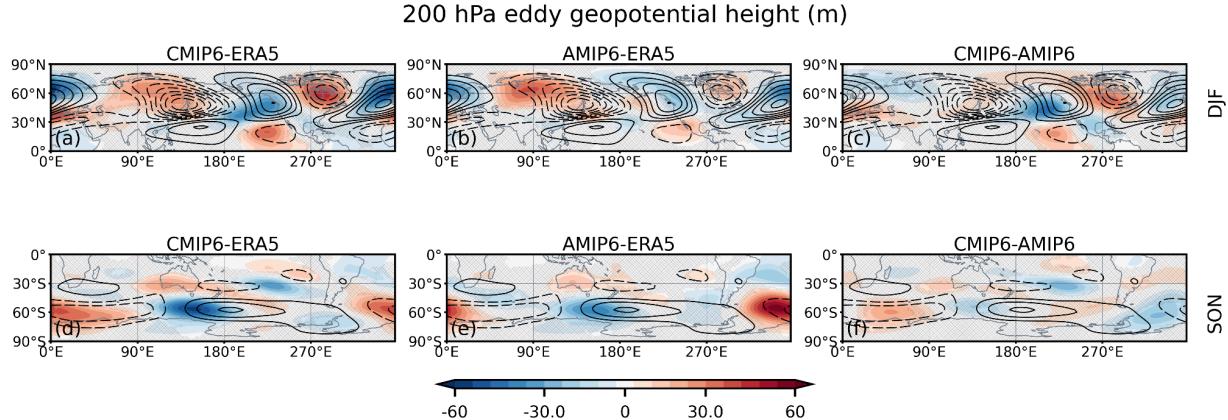


Figure 4: Differences in monthly 200 hPa eddy geopotential height between 44 models and ERA5 and between CMIP6 and AMIP6. (a-c) Top rows show DJF and (d-f) bottom row shows SON. Climatological ERA5 quantities are overlaid in black contours with intervals of  $\pm 40$  meters. Non-hatched areas signify where at least 35 of 44 models (80%) of models agree on the sign of the anomalies.

Section 3.1 showed that CMIP6 models produce polar vortices that are too strong and cold relative to AMIP6 (Fig. 1 and 2). Some possible explanations involve vertical planetary wave propagation: perhaps models simulate reduced amplitude upward-propagating planetary waves or misrepresent wave-mean flow interactions. Pertaining to option one, we use the eddy geopotential height at 200 hPa (Fig. 4), as this altitude shows the configuration of the large-scale wave entering the stratosphere from the troposphere. Extratropical CMIP6 eddy geopotential heights are consistently weaker than ERA5 (Fig. 4a). The counterclockwise rotating East Asia Trough and the clockwise rotating Alaskan Ridge, which are co-located with the strongest upward eddy heat flux climatologically (Plumb 1985; Palmeiro et al. 2023), are weaker in CMIP6 models compared to ERA5 during boreal winter (Fig. 4a). Height biases improve in AMIP6, particularly over oceans, but persist relative to ERA5 (Fig. 4b-c). Similarly, in austral spring, AMIP6 reduces some CMIP6 SH extratropical biases, but high-latitude biases still persist relative to ERA5 (Fig. 4d-f). At 60°S, the climatological eddy exhibits a zonal-wavenumber-1 structure, having one large trough and ridge that each occupy roughly half of the longitudes. ERA5's climatological ridge is opposed by the models, particularly in CMIP6 (Fig. 4d-f).

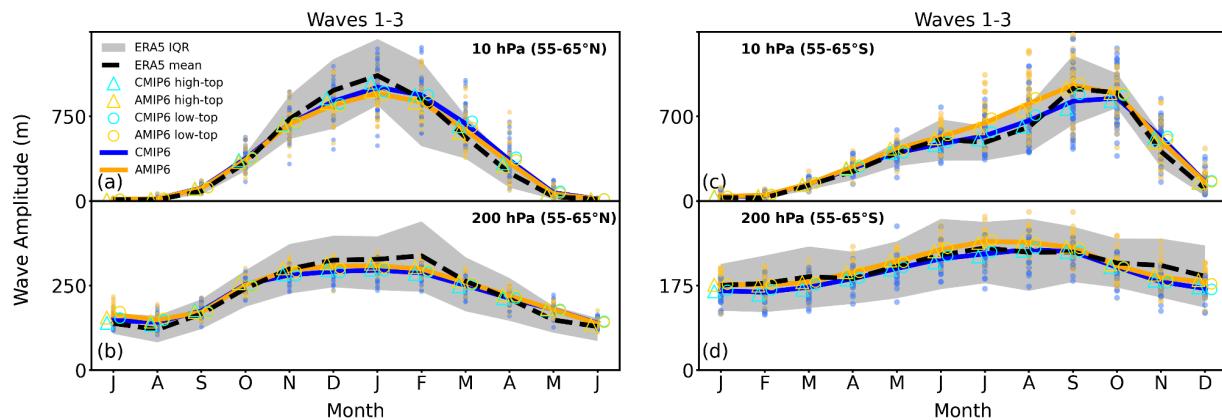


Figure 5: Amplitude versus time plots of monthly 55-65°N/S geopotential height filtered for zonal wavenumbers 1, 2, and 3 at 10 hPa (top) and 200 hPa (bottom) in 44 pairs of models. Black dashed lines show ERA5 means, gray shading shows its interquartile (25-75%) range, yellow and blue lines show CMIP6 and AMIP6 MMMs, respectively, and yellow/blue circles show model-means. Unfilled triangles and circles show amplitudes of high-top and low-top models, respectively.

Figure 5 further details the representation of planetary waves in models and ERA5, showing the amplitude of the high-latitude eddy geopotential height at 10 hPa (top) and 200 hPa (bottom) filtered for zonal wavenumbers-1, 2,

and 3; Supplementary Figure S3 shows wave-1, 2, and 3 separately. CMIP6 and AMIP6 MMMs fall within ERA5's interquartile range (IQR), indicating adequate planetary wave representation. At 10 hPa in the NH, relative to ERA5, planetary wave amplitudes are underestimated in NH early winter and overestimated in late winter (Fig. 5a). Similar timing biases arise at 200 hPa, with underestimations during DJF (consistent with Fig. 4a-b) and overestimations beginning after March (Fig. 5b). Models severely underestimate the upper bound of ERA5's IQR at 200 hPa (Fig. 5b). At 10 hPa in the SH, CMIP6 and AMIP6 match ERA5 well in SON, but overestimate amplitudes from June to August and again in November-December (Fig. 5c). Wave amplitudes at 200 hPa are reasonably captured, though weaker than ERA5 from October to December (Fig. 5d; see also 4d-e). High-top have larger early boreal winter amplitudes (Fig. 5a-b), whereas low-top models have larger austral winter/spring amplitudes (Fig. 5c-d). Relative to ERA5, NH 10 hPa amplitudes are more accurate in high-top models in DJFMA (Fig. 5a). High-top SH 10 amplitudes better mimic ERA5 during JJA, but degrade relative to low-top models during SON (Fig. 5b).

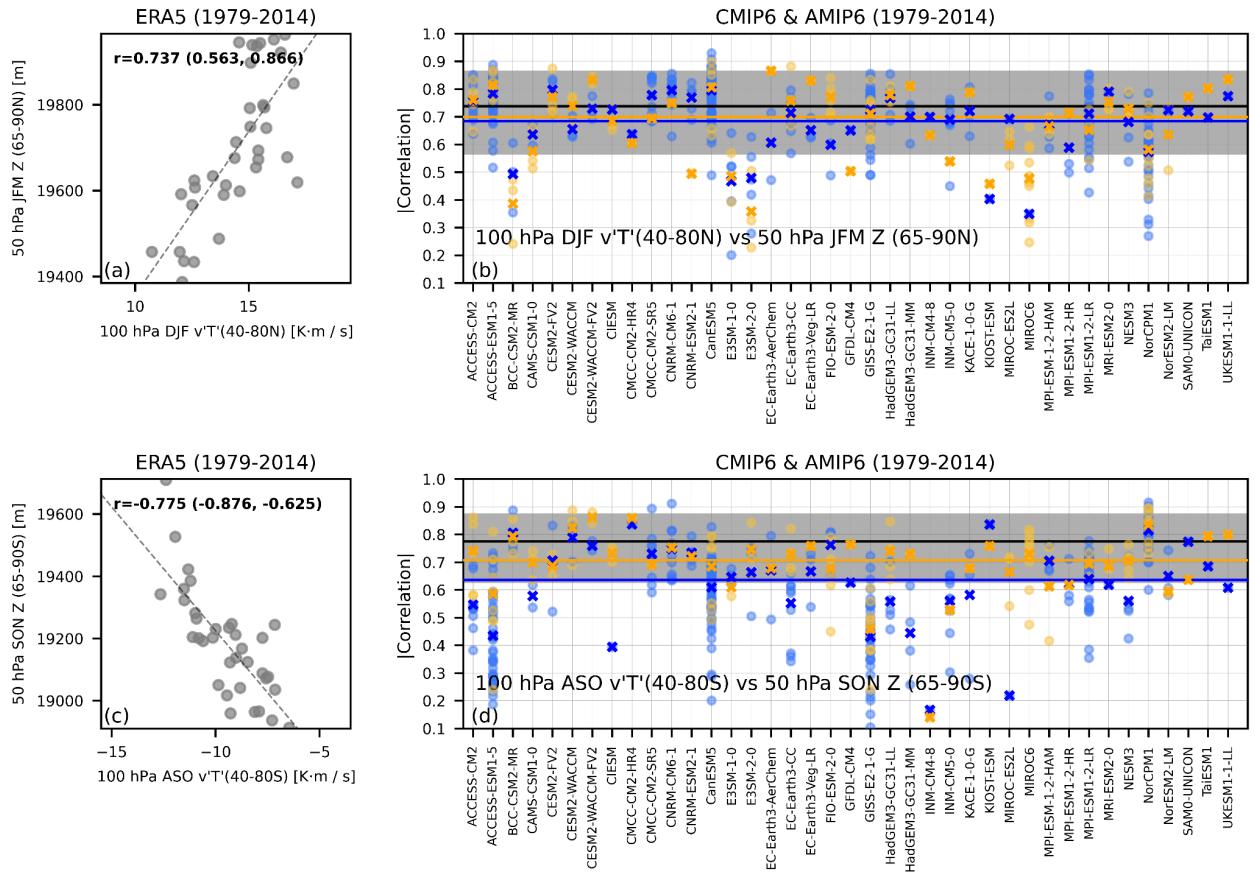


Figure 6: (a, c) Correlations between seasonally averaged monthly 100 hPa extratropical ( $40^{\circ}\text{N}$ - $80^{\circ}\text{N/S}$ ) eddy heat flux and 50 hPa polar cap ( $65^{\circ}\text{N}$ - $90^{\circ}\text{N/S}$ ) geopotential height, where the geopotential heights lag the eddy heat fluxes by one month. Correlation coefficients are printed in black with their 2.5/97.5 percentile confidence intervals, which are obtained by resampling the 1979-2014 records by replacement 10,000 times. (b) Correlations between the same fields as in (a), but from individual models (39) and their members. Blue and orange correspond to CMIP6 and AMIP6, respectively, where circles are individual members, crosses are model-means, and lines are MMMs. The gray shading on (b) and (d) shows ERA5's correlation confidence intervals.

In addition to reduced planetary wave amplitudes, perhaps the too cold CMIP6 polar stratospheres stem from deficient representation of upward troposphere-to-stratosphere wave-mean flow coupling. On seasonal timescales, enhanced eddy heat flux signifies stronger upward planetary wave propagation and is correlated with increases in extratropical stratospheric temperatures from wave induced secondary circulation effects (Dunkerton et al. 1981; Newman et al. 2001). Strong relationships linking midseason mid-latitude eddy heat flux (DJF and ASO) to late

season polar stratospheric circulation (JFM and SON) are apparent in ERA5 in the NH (Fig. 6a) and SH (Fig. 6c), respectively.

CMIP6 and AMIP6 MMM correlations of NH upward coupling are comparable to ERA5, despite some individual models significantly underestimating it (Fig. 6b). Rather than issues with upward coupling representation, other model deficiencies appear to be more important drivers of the colder boreal vortex in CMIP6 compared to AMIP6 (Fig. 1b). Conversely, in the SH, AMIP6 demonstrates a notable improvement in representing upward coupling compared to CMIP6, although both ensembles simulate weaker upward coupling than ERA5 (Fig. 6d). Consequently, the polar stratospheric cold bias observed in CMIP6 for the SH may result from insufficient upward wave coupling (Fig. 6d).

Figure 7 presents boreal winter major SSW statistics for CMIP6 and AMIP6 to further elucidate model sensitivity of polar vortex variability to differences in ocean boundary conditions. In agreement with CMIP6 showing colder and stronger climatological polar vortices relative to AMIP6 (Fig. 1), CMIP6 models show fewer SSWs per month (Fig. 7a) and per year (Fig. 7b) than AMIP6, though both underestimate SSW frequency between November and February and overestimate the frequency during March relative to ERA5. Although the multi-model mean CMIP6 and AMIP6 SSW frequencies are not statistically different (Fig. 7b), most individual models show higher frequency of SSWs in their AMIP configuration relative to CMIP.

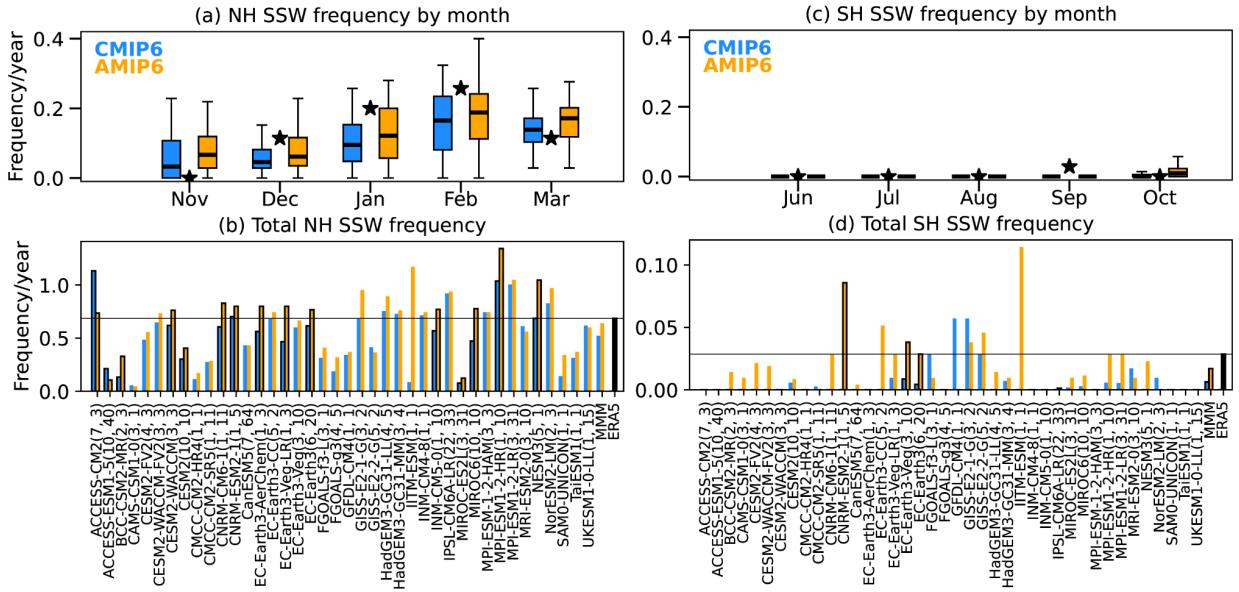


Figure 7: Frequency of SSWs by month in CMIP6 (blue) and AMIP6 (orange) over the 1979-2014 period. (a,c) Boxplots show the median and interquartile range and the whiskers denote the 95% sampling distributions, for (a) NH and (c) SH. Black stars show mean frequencies from ERA5. (b,d) Frequency of SSWs in simulations of individual models, where the parentheses show the number of AMIP6 and CMIP6 members, for (b) NH and (d) SH. Black contouring around bars of a model indicates statistically significant differences between its AMIP6 and CMIP6 simulations. Black bar and line indicate ERA5 frequency for comparison. Note that the y-axis range is the same for (a) and (c) to highlight NH/SH differences, but different between (b) and (d) so that AMIP/CMIP differences can be seen in (d).

In the SH, major SSWs are much more rare (Fig. 7c). There was only one observed major SSW that occurred in the 1979-2014 period (in September 2002). In general the models simulate a frequency of SH SSWs even smaller than observed, and this is especially true for CMIP6 simulations (Fig. 7c,d). The models also place a preference for these events in October, which is slightly later than observed (Fig. 7c). AMIP6 simulations consistently show higher SH

SSW frequency relative to CMIP6 (for which many models show no SH SSWs at all), which is likely because of improved upward wave coupling in AMIP6 (Fig. 6d).

In this section, we have assessed upward coupling between the troposphere and stratosphere and its impact on polar stratospheric variability (in the form of SSWs) in CMIP6 and AMIP6 models. Compared to ERA5, AMIP6 and especially CMIP6 planetary wave amplitudes are too weak during early to mid-boreal winter (Fig. 4, Fig. 5). We suggest this issue partly underlies the colder and less disturbed NH polar stratospheric polar vortex in both AMIP6 and CMIP6 because the models simulate boreal winter upward wave mean-flow interactions adequately (Fig. 6). This is in contrast to the SH, where CMIP6 upward coupling is weaker and SSW frequency is smaller than AMIP6 (and both are deficient compared to ERA5) (Fig. 6, Fig. 7). In the next section, we explore the quality of simulated downward coupling from the stratosphere to the troposphere.

### **3.4 Effect of ocean boundary conditions on downward stratosphere to troposphere coupling**

Changes in polar stratospheric winds can influence the tropospheric jet stream over timescales ranging from sub-seasonal to multi-decadal (Kidston et al. 2015). The covariation between the tropospheric and stratospheric circulation often resembles the Northern and Southern Annular Modes (NAM & SAM; Thompson and Wallace 2000; Lim et al. 2018), the primary modes of extratropical climate variability coupling the troposphere and stratosphere over a range of altitudes (Baldwin and Thompson 2009; Gerber et al. 2010; Simpson et al. 2020). They are maintained by wave-mean flow feedbacks in the troposphere (Lorenz and Hartmann 2001; 2003) and are influenced by climatological jet structure (Eichelberger and Hartmann 2007), transient phenomena like SSWs (Karpechko et al. 2017) and ENSO (Lim et al. 2013), and by changes in atmospheric composition (Gillett and Fyfe 2013). In the troposphere, a positive phase of the AM equates to a poleward shifting eddy-driven jet stream and concurrent changes in the storm-track and regional surface climate.

To get a basic idea of how models represent these phenomena, we compare polar cap averaged geopotential height at 50 hPa with the synchronous heights at 850 hPa. In NH during JFM (Fig. 8a) and SON in the SH (Fig. 8c), ERA5 exhibits a correlation of 0.65, illustrating the strong linkage between stratospheric and tropospheric circulation at high-latitudes. In the NH, AMIP6 and CMIP6 show weaker downward coupling than ERA5, though AMIP6 outperforms CMIP6 and both MMMs are successfully within the ERA5 bootstrapped sampling distribution (Fig. 8b). Surprisingly, there are some models for which no member falls within the reanalysis sampling distribution, suggesting some models do not capture downward coupling well at all. In the SH, the MMM correlations for both AMIP6 and CMIP6 closely match ERA5, indicating reasonable downward coupling strength.

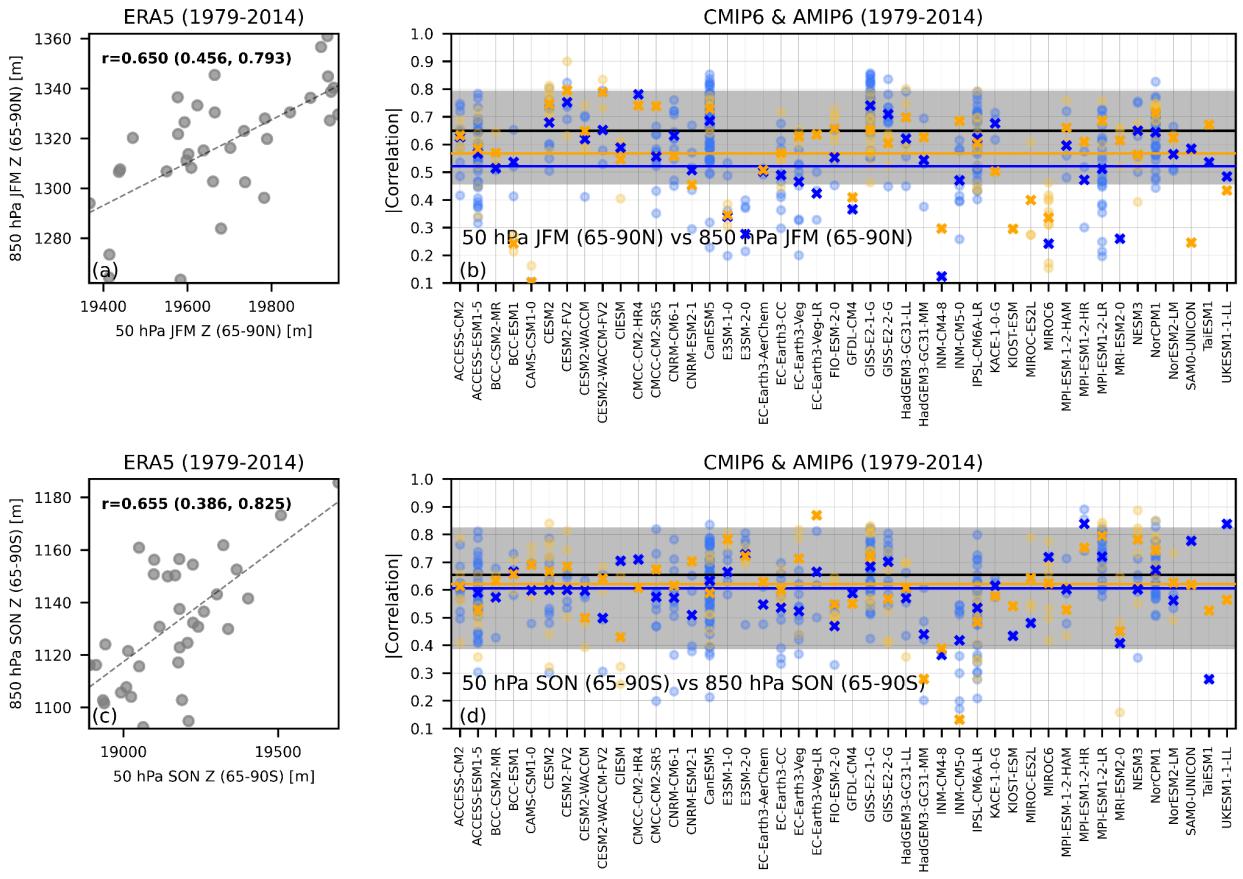


Figure 8: Correlations between seasonally averaged 50 hPa polar cap ( $65^{\circ}\text{N}$ - $90^{\circ}\text{N/S}$ ) geopotential heights and synchronous heights at 850 hPa during JFM in the NH (top) and during austral spring in the SH (bottom). Correlation coefficients are printed in black with their 2.5/97.5 percentile confidence intervals, which are obtained by resampling the 1979-2014 records by replacement 10,000 times. (b) Correlations between the same fields as in (a), but from individual models (39) and their members. Blue and orange correspond to CMIP6 and AMIP6, respectively, circles are individual members, crosses are model-means, and lines are MMMs. (b,d) Gray shading shows ERA5's bootstrapped sampling distribution of correlations.

Figure 9 examines how well models represent the annular modes and their downward impacts in each hemisphere. AM “persistence” equals the e-folding timescale of the AM index autocorrelation function, which is smoothed over a 181-day window using a Gaussian filter like in Simpson et al. (2011). The left columns show that, at 700 hPa, tropospheric NAM and SAM are most persistent during November-December in ERA5 (Fig. 9a-b). In CMIP6 and AMIP6, persistence peaks later than in ERA5, with the NAM maximizing in late boreal winter (Fig. 9a). SAM persistence is delayed by a month and overestimated in models, although AMIP6 yields more realistic timescales (Fig. 9b).

The right panels of Figure 9 illustrate “annular mode predictability,” the squared correlation between the 50 hPa AM today (day 0) and the 30-day averaged 700 hPa AM over days 10-40 after. This obtains the fraction of variance of tropospheric AM index during days 10-40 that can be “predicted” using a persistence forecast based on today’s stratospheric AM index (Baldwin et al. 2003; Gerber et al. 2010). In ERA5, the 50 hPa NAM in late November/early December predicts 15% of the variance in the January 700 hPa NAM (Fig. 9c). Similarly, the late September/early October 50 hPa SAM predicts over 20% of the November 700 hPa SAM variance (Fig. 9d). In ERA5, predictability is partly driven by early season polar vortex variability, but in models, predictability is delayed in each hemisphere (Fig. 9c-d), reflecting delayed AM persistence and likely biases in stratospheric dynamics.

The delay in NAM persistence and predictability relative to observations may result from differences in the timing of stratospheric warmings between models and reanalysis. As noted in Figure 7a, CMIP6 and AMIP6 models simulate fewer SSWs than ERA5 between November and February, but a higher frequency during March. As SSWs are known drivers of NAM variability (Baldwin et al. 2021; Butler and Domeisen 2021), the delayed SSW seasonality can at least partly explain the late season NAM biases (Fig. 9a, 9c). Likewise, climate models have a commonly known bias of delayed final warmings in the SH relative to reanalysis (Wilcox and Charlton-Perez 2013), which may partly explain the delayed predictability of the tropospheric SAM (Fig. 9d).

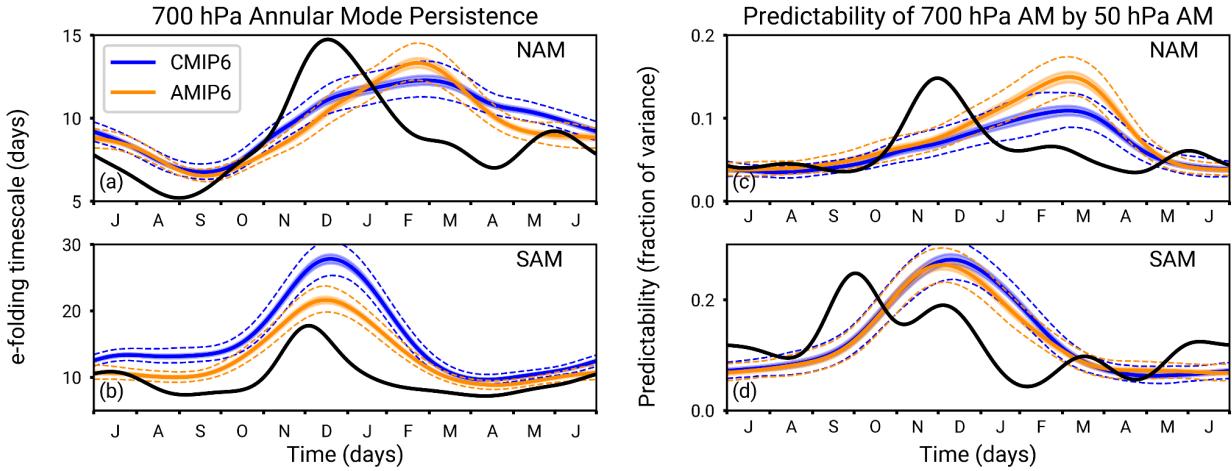


Figure 9: (a) and (b) Seasonal cycles of the NAM and SAM e-folding timescales derived using daily geopotential height data from 26 pairs of CMIP6 and AMIP6 models. Blue and yellow correspond to CMIP6 and AMIP6, respectively. (c) and (d) show the fraction of variance that the 50 hPa AM on a given day explains in the 700 hPa AM averaged from 10 to 40 days later. Plots show the median (thick colored lines), first to third quartiles (colored envelopes), and 0.5-99.5 percentile confidence intervals (dashed lines).

To summarize how the full chain of events from upward to downward vertical coupling is simulated in CMIP6 and AMIP6, Figure 10 relates the mid-season 100 hPa eddy heat flux between  $40^{\circ}$ - $80^{\circ}$  in each hemisphere with the polar cap geopotential height over a broad range of months and isobars. In the NH, positive correlations imply that upward propagating planetary waves (i.e., positive eddy heat fluxes) weaken the circumpolar zonal-wind (i.e., positive polar cap geopotential height anomalies). ERA5 boreal winter data reveals that upward planetary wave propagation is preceded by a late autumn tropospheric ridge and polar vortex strengthening that steepens the potential vorticity gradient (Lawrence and Manney 2020), providing favorable preconditioning for subsequent upward wave propagation. This culminates in late winter increases in polar cap geopotential height that start in the stratosphere and peak at the surface in March, indicating weakened circumpolar zonal-winds throughout the atmospheric column (Fig. 10a).

CMIP6 and AMIP6 capture some aspects of the ERA5 vertical coupling relationships, but with reduced amplitude in the lowermost stratosphere, during the middle and late seasons (Fig. 10b-c). They largely fail to include the early season tropospheric precursor, a shortcoming of multiple CMIP generations (cf. Fig. 9 of Furtado et al. 2015), though CMIP6 shows a more robust tropospheric precursor signal in early winter than AMIP6, suggesting these precursors may arise in part from different ocean configurations. As also shown in Figure 8b, AMIP6 and especially CMIP6 show a weaker downward effect from the stratosphere in late boreal winter compared to ERA5, and both largely miss the amplified coupling at the surface (e.g., Baldwin et al. 2024).

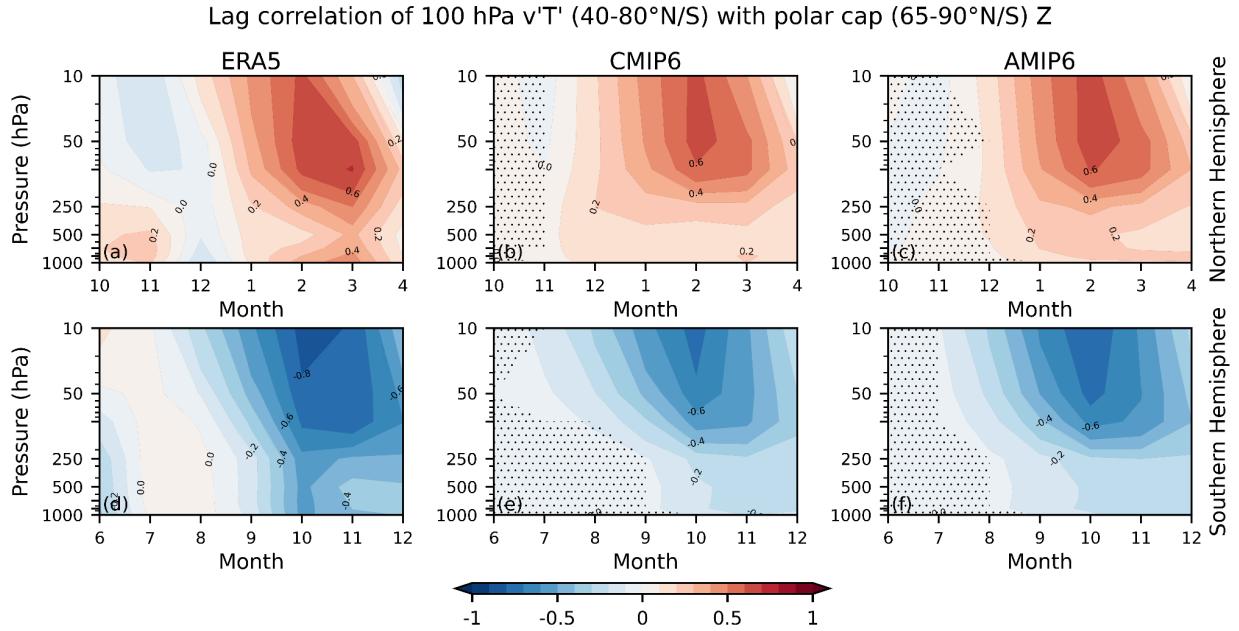


Figure 10: Lead-lag correlation contour plots assessing the relationship between 100 hPa 40°N-80°N/S eddy heat flux during DJF (top) and ASO (bottom) with polar cap geopotential heights for each month. (a) and (d) show ERA5, (b) and (e) show CMIP6 averages, and (c) and (f) show AMIP6 averages. Non-stippled areas on the model subplots show where at least 31 of 39 (>80%) models agree on the sign of the correlation..

The seasonal evolution of vertical coupling in the SH is broadly similar (Fig. 10d); note, correlations are negative because negative eddy heat fluxes correspond to upward propagating waves in the SH. Differences between CMIP6 and AMIP6 are fairly minor (Fig. 10e-f). Both ensembles emulate ERA5, but with substantially reduced correlations in the stratosphere (also shown in Fig. 6d). Interestingly, though Fig. 8d suggests reasonable downward coupling of a stratospheric signal to the troposphere in both CMIP6 and AMIP6, Fig. 10e-f suggests that both ensembles fail to capture the strong ERA5 downward response in September to November to an early spring eddy heat flux. This implies that because the models fail to simulate adequate upward coupling to the stratosphere, the downward coupling is then also subsequently underestimated, even though the downward coupling itself is reasonably simulated.

In general, sections 3.1-3.4 suggest that AMIP6 boundary conditions make the large-scale circulation more consistent with ERA5; there are improvements to the zonal-mean zonal-wind, planetary wave amplitudes and mean-flow process, and aspects of vertical wave-mean flow coupling. Nonetheless, some issues are common to CMIP6 and AMIP6, like the delayed seasonal cycle peak in planetary waves, SSW timing, and AM predictability. The next section builds on these results by looking at CMIP6 and AMIP6 differences in ENSO and QBO teleconnection pathways to the high-latitude stratosphere.

### 3.5 Effect of ocean boundary conditions on ENSO teleconnections

The El Niño Southern Oscillation is a recurring climate pattern characterized by alternating warm (El Niño) and cold (La Niña) tropical Pacific SST anomalies over inter-annual timescales (Philander 1990). ENSO modulates global circulation, including high-latitude stratospheric dynamics. During El Niño, the boreal winter subtropical jet stream shifts equatorward, vertically propagating waves are enhanced, and the polar vortex strength weakens in mid to late winter, relative to ENSO-neutral conditions (Domeisen et al. 2019). However, the exact location of the tropical Pacific warm anomalies can modify the stratospheric response (Hurwitz et al. 2014; Calvo et al. 2017). In contrast,

La Niña is associated with a boreal polar vortex acceleration, although observed responses exhibit strong internal variability (Iza et al. 2016). ENSO's extratropical teleconnections and their seasonal effects are also shaped by stratospheric downward coupling (Butler et al. 2014, Polvani et al. 2017, Oehrlein et al. 2019). In the SH, ENSO phase inversely relates to the SAM, peaking in late austral spring to early austral summer (Lim et al. 2013). Analogous to the Pacific North American teleconnection pattern in the NH (Manzini et al. 2024), the tropical Pacific links to the SH high-latitudes via a Pacific-South America Rossby wave train (Fang et al. 2024), whose signal can be modulated by the austral polar vortex (Osman et al. 2022).

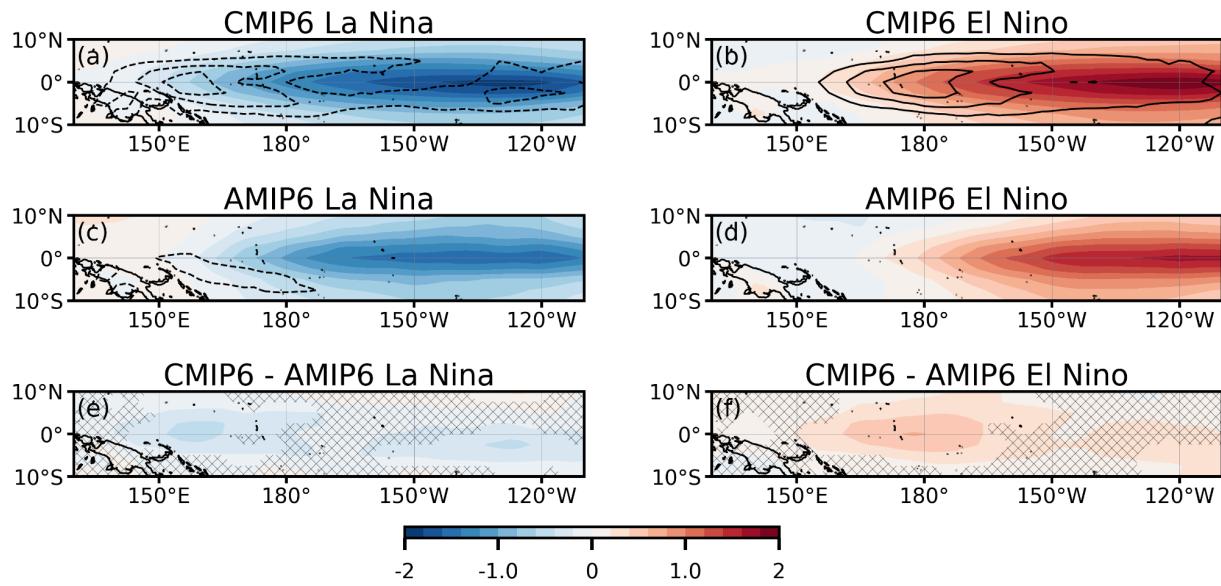


Figure 11: Shaded September-February surface air temperature anomalies by ENSO phase for CMIP6 (top), AMIP6 (middle), and their difference (bottom). Black contours show differences relative to ERA5 (model minus ERA5) with intervals of +/- 0.2 Kelvin. Non-stippled regions in the bottom row show where 27 of 34 (>80%) models agree on the sign of the CMIP6 minus AMIP6 surface air temperature differences.

CMIP6 models capture the broad tropical Pacific features of both ENSO phases, as illustrated by their surface air temperature composites (Fig. 11). These are multi-model averages of all cases when the Nino3.4 September to February averaged intra-member standardized surface air temperature anomalies exceed +/-1 standard deviation. Although the model patterns are generally realistic, CMIP6 consistently produces temperature anomalies during either ENSO phase that extend too far west (Fig. 11a-d), that exceed ERA5's amplitude (black contours, Fig. 11a-b), and that are too large relative to AMIP6, consistently between 150°E and 180°E (Fig. 11e-f).

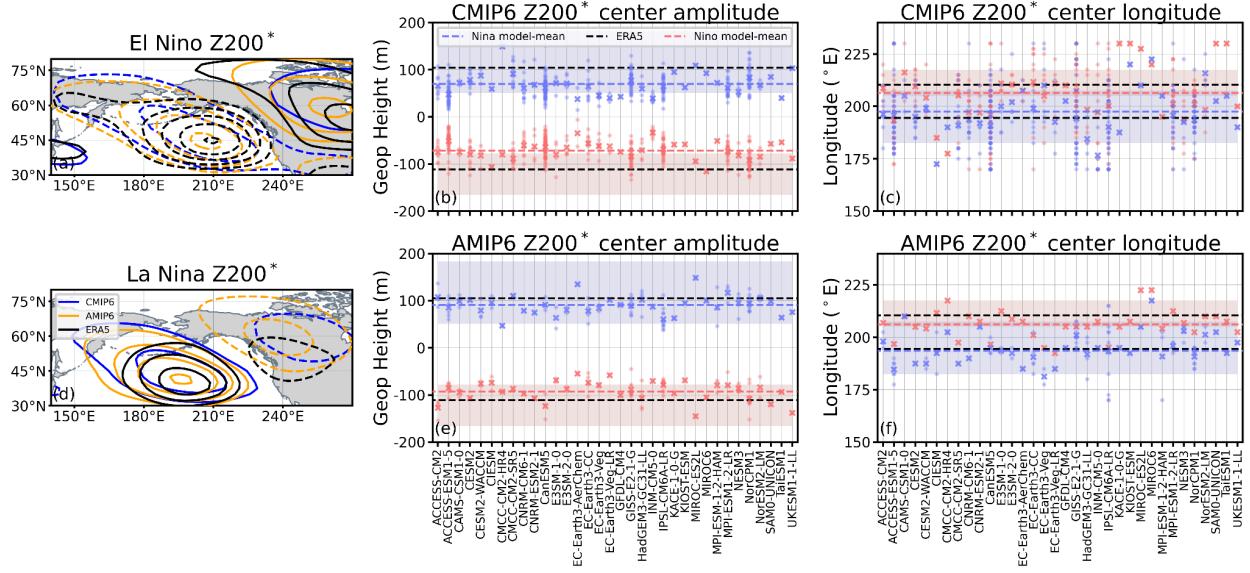


Figure 12: NH 200 hPa eddy geopotential height,  $Z200^*$ , in CMIP6, AMIP6, and ERA5 during DJF. (a, D) The maps show ensemble mean  $Z200^*$  ( $\pm 20$  m) during El Niño (top) and La Niña (bottom) with CMIP6 in blue, AMIP6 in orange, and ERA5 in black. (b, e) Center and (c, f) right columns show amplitude and longitude, respectively, of maximum  $Z200^*$  anomaly within  $170\text{--}230^\circ\text{E}$ ,  $30\text{--}60^\circ\text{N}$  in CMIP6 and AMIP6. Red and blue crosses are El Niño and La Niña model-means and red and blue dots are members. Dashed red and blue horizontal lines are MMMs for El Niño and La Niña, respectively, and dashed black lines are ERA5 means, which are enclosed by shaded sampling distributions (bounds = min/max across (6) El Niño years and (5) La Niña years).

Both ENSO phases in CMIP6 exhibit a westward extension of tropical Pacific SSTs that shifts the associated tropical-extratropical wave trains westward (Bayr et al. 2019; Bayr et al. 2024; Beniche et al. 2025). During SH December–February, CMIP6 models produce westward shifted Pacific–South America wave trains, a discrepancy that AMIP6 corrects (Fang et al. 2024). Repeating the analysis of Fang et al. (2024), but for the NH, we assess the amplitude and longitude of the North Pacific “Low,” or “trough,” an integral component of the Pacific North American teleconnection modulated by ENSO. El Niño and La Niña events have opposite observed atmospheric impacts on the Low, namely a deepening (ridging) of the low during El Niño (La Niña).

Figure 12 shows maps on the left of 200 hPa eddy geopotential height during El Niño and La Niña. The middle panels display the maximum amplitude of the North Pacific Low anomaly, while the right panel shows its corresponding longitude. For consistency with Hurwitz et al. (2014), the anomaly is defined over  $170\text{--}230^\circ\text{E}$  and  $30\text{--}60^\circ\text{N}$ , and is shown separately for La Niña (blue) and El Niño (red). For comparison, MMMs are shown as solid blue and red lines, model means as crosses, members as circles, ERA5 means in black dashes, and shading is a min-max sampling distribution across observed El Niño and La Niña seasons.

Overall, CMIP6 models emulate the El Niño teleconnection relative to ERA5, though they underestimate the anomalous trough’s max amplitude (Fig. 12a, 12b) and shift its center westward (Fig. 12c). The La Niña ridge is also represented with fidelity, though its amplitude is weaker than ERA5 (Fig. 12d, 12b, 12c). In contrast with CMIP6, AMIP6 demonstrates reduced inter-model variation in center amplitude and longitude of the North Pacific response, indicating a more consistent representation of the teleconnection (Fig. 12e, 12f).

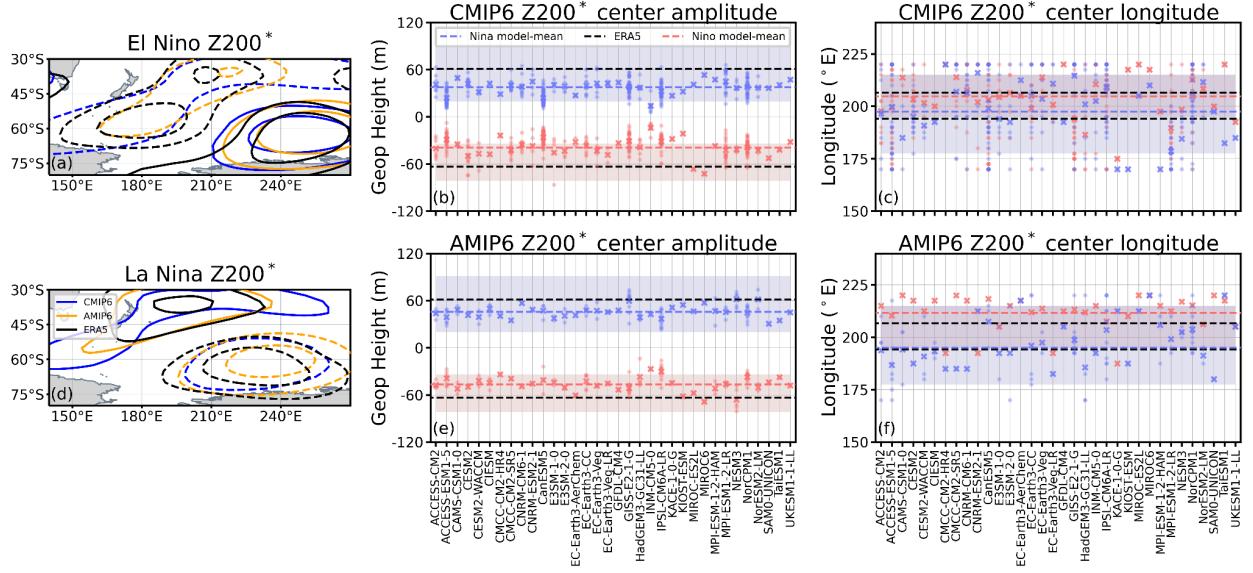


Figure 13: As in Figure 12, but for Southern hemisphere 200 hPa eddy geopotential height,  $Z200^*$ , in CMIP6, AMIP6, and ERA5 during SON. Amplitude (middle) and longitude (right) statistics are computed within  $170\text{--}220^\circ\text{E}$ ,  $45\text{--}30^\circ\text{S}$ , consistent with Hurwitz et al. (2014).

During austral spring, a wave-train of alternating increases and decreases in geopotential height emanates from the equatorial Pacific into the extratropical SH in response to El Niño and La Niña (Osman et al. 2022). We analyze the El Niño trough and La Niña ridge that form between  $170\text{--}220^\circ\text{E}$  and  $45\text{--}30^\circ\text{S}$ . CMIP6 models capture the teleconnection in either ENSO phase (Fig. 13a, 13d), but their El Niño and La Niña amplitudes fall at the lower end of ERA5's sampling distribution (Fig. 13b, 13e). AMIP6 improves, but does not entirely resolve the overly weak amplitude bias (Fig. 13e). Additionally, AMIP6 reduces the spread in center longitudes relative to CMIP6, but is still biased relative to ERA5 for El Niño (Fig. 13f). The concentration of some model-mean longitude responses along the  $170^\circ\text{E}$  and  $220^\circ\text{E}$  in Figures 13c and 13f is due to the  $170\text{--}220^\circ\text{E}$  domain being too narrow for the variety of wave-trains generated by the models.

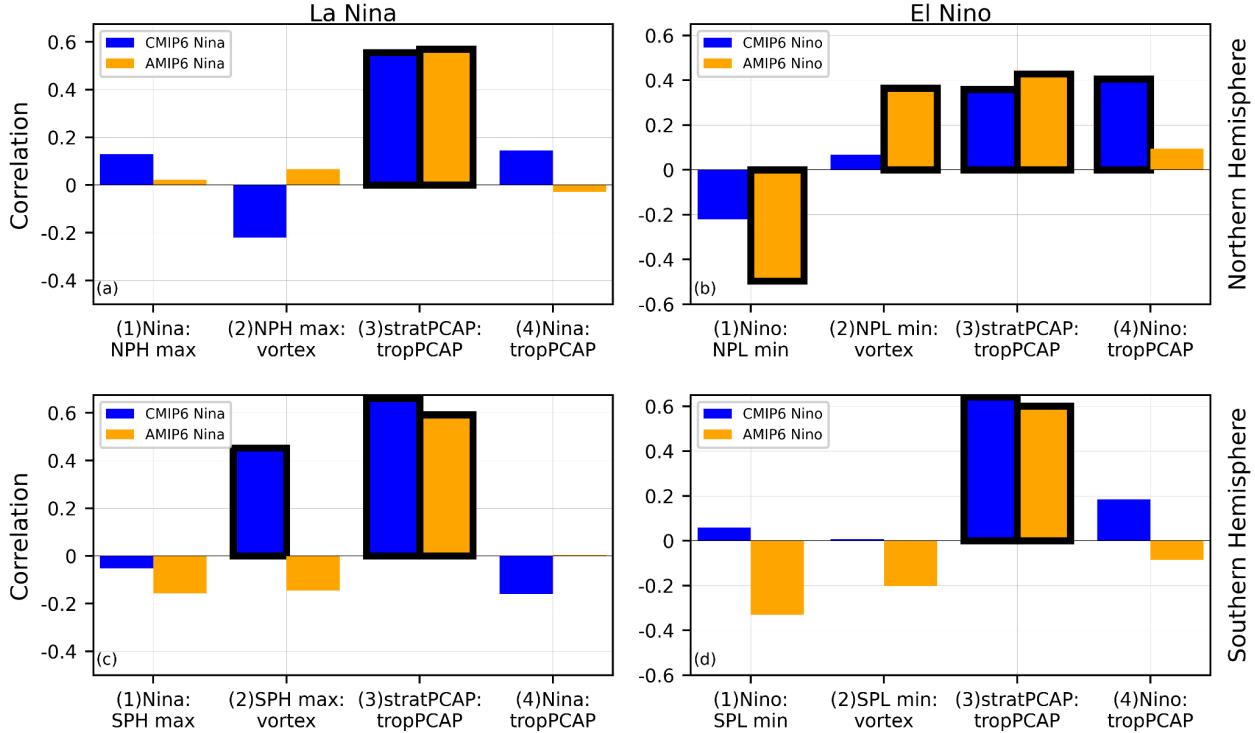


Figure 14: In blue and orange, respectively, CMIP6 and AMIP6 ensemble-mean ENSO teleconnections to the NH (top) and SH (bottom) high-latitude circulations. Four teleconnections are shown per subplot, defined as across model-mean correlations; from left to right, (1) Nino3.4 surface air temperature correlation with North (South) Pacific 2200' geopotential height min/max amplitudes, (2) North (South) Pacific 2200' geopotential height min/max amplitude correlation with NH (SH) zonal-wind at 10 hPa, 55°N–65°N/S, (3) correlations between polar cap geopotential height at 10 hPa and 850 hPa, and (4) correlation between Nino3.4 surface air temperature and 850 hPa polar cap geopotential height. Black contours enclose statistically significant ( $p < 0.05$ ) correlations. ENSO surface air temperature and Pacific sector (same as Fig. 12 and 13) geopotential heights quantities are computed during DJF and SON whereas polar zonal-wind and polar cap geopotential heights are lagged by one month, JFM and OND, respectively.

The North Pacific Low's amplitude and location affect the strength of upward planetary wave propagation into the stratosphere (Smith and Kushner 2012). El Niño deepens the low and intensifies upward propagation of wavenumber-1, whereas La Niña has opposite effects (Palmeiro et al. 2023). Consequently, differences in the North Pacific teleconnection between CMIP6 and AMIP6 or between models may lead to distinct polar circulation responses. For instance, models with a strong North Pacific response to ENSO may exhibit a stronger polar cap response than models with a weak Pacific teleconnection. To quantify the strength of these across model relationships, we form composite averages of each model's La Niña and El Niño oceanic and atmospheric response, and then compute correlations across models between different pathways of the ENSO-to-high-latitude teleconnection. We consider four pathways, shown left to right in each subplot: (1) from ENSO (tas) to extratropical Pacific 200 hPa circulation (zg), (2) from extratropical Pacific 200 hPa circulation response (zg) to polar vortex response (ua), (3) from 10 hPa stratospheric polar cap response (zg) to the same in troposphere (zg) at 850 hPa, and (4) from ENSO (tas) to the tropospheric polar cap response (zg) at 850 hPa. Teleconnection strength across the CMIP6 and AMIP6 ensembles, represented by the bars in Figure 14, is measured as the correlation across model-means.

A robust part of the ENSO teleconnection is pathway (3), the synchronous relationship between stratospheric and tropospheric polar cap circulation. Any polar stratospheric circulation response to ENSO in models is tightly related to a similar circulation response in the troposphere irrespective of CMIP6 or AMIP6, ENSO phase, or hemisphere, (Fig. 14a-d). Aside from pathway (3), the boreal winter La Niña teleconnections are not robust (Fig. 14a). Pertaining to pathway (1), CMIP6 and AMIP6 La Niñas exhibit a weak relationship between the amplitude of a model's La Niña

and the amplitude of its anomalous North Pacific High (Fig. 14a). In contrast, stronger El Niños produce deeper North Pacific Lows (Fig. 14b) in CMIP6 ( $r = -0.22$ ,  $p = 0.21$ ) and especially in AMIP6 models ( $r = -0.50$ ,  $p = 0.00$ ). These deeper Lows are positively correlated with weaker 10 hPa 55°N-65°N zonal-mean zonal-wind (pathway 2), which can couple down to the troposphere and increase polar cap geopotential heights (Fig. 14b). Interestingly, CMIP6 models have a stronger El Niño to tropospheric polar cap relationship than in AMIP6 models (pathway 4), which may be consistent with the similarity between the East Pacific and Arctic across model regression in Figure 3.

As with the NH, the La Niña teleconnection pathways in the SH during austral spring are weak (Fig. 14c). CMIP6 models exhibit a positive correlation between the strength of their anomalous extratropical ridges (Fig. 13a) and the speed of their polar vortices, but no other pathways are statistically significant (Fig. 14c). The El Niño SH pathways also lack robustness, implying weak overall relationships of ENSO with the austral spring high-latitude circulation (Fig. 14d).

In Figure 15, we relate the 10 hPa polar vortices in either hemisphere (55°N-65°N/S) with the concurrent monthly Nino3.4 surface air temperatures taken across all 1979-2014 seasons. ERA5 reveals a modest positive correlation in early boreal winter that shifts to an inverse relationship in mid to late winter (Fig. 15a); note these correlations are quite small in part because of non-linearities introduced by SSWs; see Domeisen et al. 2019.

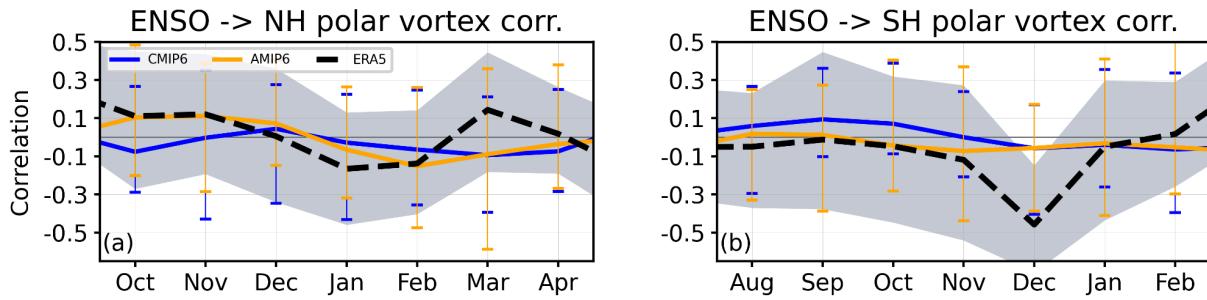


Figure 15: Ensemble-mean CMIP6 (blue) and AMIP6 (orange) correlations between monthly Nino3.4 surface air temperature and 10 hPa 55°N-65°N/S zonal-wind in the NH (left) and SH (right) across all years, 1979-2014, and 34 models. Whisker bounds are the minimum/maximum correlation across models in the CMIP6 and AMIP6 ensembles. Dashed black lines show ERA5 correlations and gray shading encloses their 2.5/97.5 percentile confidence intervals.

CMIP6 models demonstrate weak ENSO-polar vortex coupling in boreal winter. While they capture the inverse relationship in mid to late winter, the signal persists too late into the season, consistent with Hu et al. (2023). In contrast, AMIP6 correlations better resemble ERA5, although the peak inverse correlation occurs in February instead of ERA5's January peak. In the SH, ERA5 indicates an inverse relationship between ENSO and austral polar vortex strength, especially during December, mimicking the ENSO-SAM teleconnection seasonality and sign (Lim et al. 2013). However, both CMIP6 and AMIP6 struggle to simulate this teleconnection accurately.

To sum up, CMIP6 models produce overly strong and west-shifted ENSO surface air temperature anomalies relative to ERA5, which accompanies biases in tropical-extratropical planetary waves in both hemispheres. AMIP6 boundary conditions help correct the North and South Pacific teleconnections, nonetheless, both CMIP6 and AMIP6 encounter difficulties simulating the southern hemisphere polar vortex teleconnection.

### 3.5 Effect of ocean boundary conditions on the Quasi-Biennial Oscillation and its teleconnections

The Quasi-Biennial Oscillation (QBO) is characterized by alternating easterly and westerly wind regimes descending through the tropical stratosphere roughly every 28-months (Baldwin et al. 2001). This cycle results from

interactions between the tropical zonal-mean zonal-wind and various waves, including small-scale gravity, Kelvin, mixed-Rossby gravity, and inertia-gravity waves (Pahlavan et al. 2021). Although more models now generate QBOs (Richter et al. 2020), differences in model resolutions, convective parameterizations, and non-orographic gravity wave forcing lead to varied QBO representation (Holt et al. 2021; Garfinkel et al. 2022; Bushell et al. 2022). For instance, the QBO explains 80% of the observed tropical stratospheric zonal-wind variance at 20 hPa (Pascoe et al. 2005), but models place its center of action near 10 hPa (Richter et al. 2020).

The QBO influences the extratropical circulation in both hemispheres. In early boreal winter, westerly (and easterly) QBO winds at 30-50 hPa relate to accelerated (and decelerated) polar vortex circulation via modifications to the large-scale background for planetary wave propagation (Garfinkel et al. 2012; Watson and Gray 2014; Lu et al. 2020). In austral spring, the southern vortex varies in phase with the 20 hPa QBO winds (Yamashita et al. 2018; Rao et al. 2023). Previous studies indicate that stronger (Garfinkel et al. 2011; Elsbury et al. 2021; Anstey et al. 2022) and wider (Hurwitz et al. 2011; Hansen et al. 2013) QBOs enhance these high-latitude teleconnections, although limited research exists for the austral vortex.

Ocean conditions further modulate the QBO and its teleconnections. For example, boreal polar vortex weakening is favored when easterly QBO coincides with La Niña (Kumar et al. 2022; Ma et al. 2023). SSTs can also affect QBO periodicity, with weaker evidence of effects on QBO amplitude (Taguchi et al. 2010; Zhou et al. 2024). Recent disruptions to the QBO underscore its sensitivity to extratropical planetary waves, which are influenced by ocean boundary conditions (Wang et al. 2023).

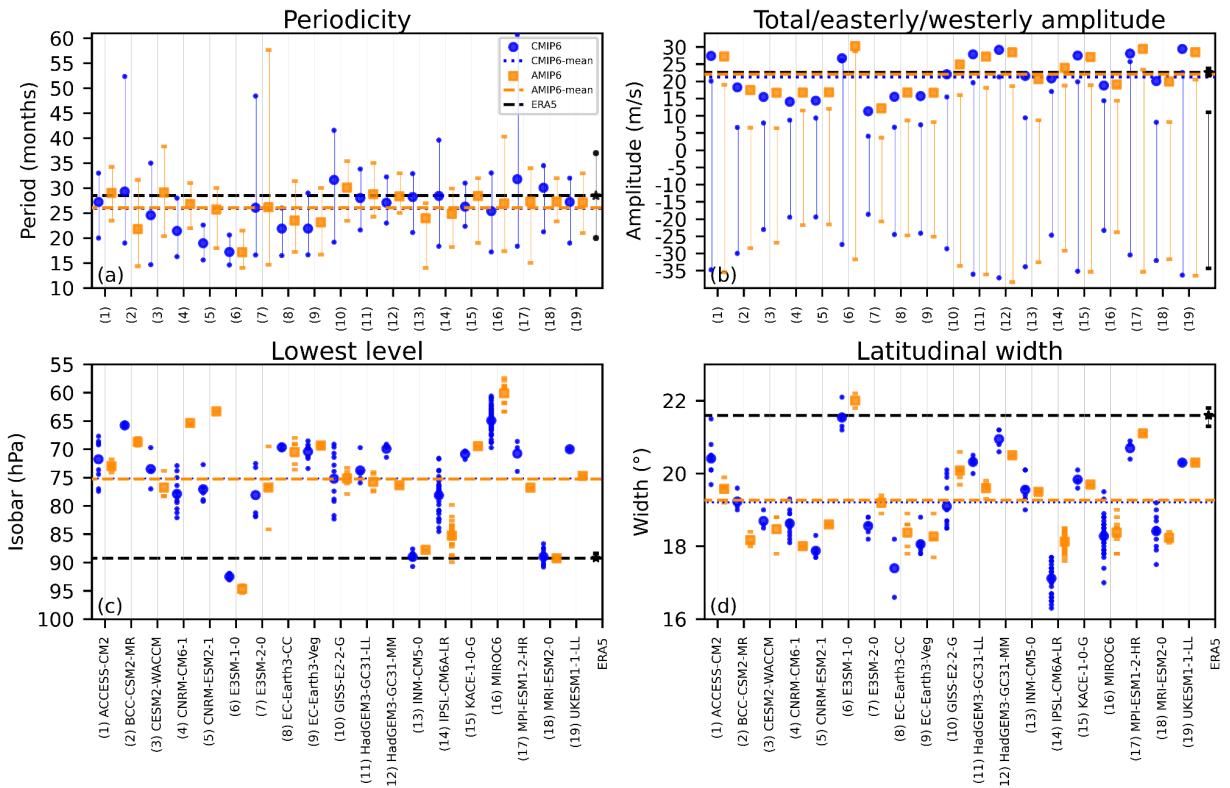


Figure 16: Periodicity, amplitude, lowest level of descent, and latitudinal width of the QBO in CMIP6 and AMIP6 models. Blue quantities show CMIP6, orange quantities show AMIP6. (a) mean periodicity for CMIP6 and AMIP6 models where whiskers extend to minimums and maximums taken across all members, CMIP6 mean is dotted, AMIP6 mean is dashed, and ERA5 is shown in black. (b) circles and squares show total (not mean) amplitudes (“TT amplitude” from Richter et al. 2020), tops of whiskers show westerly amplitudes, and bottoms of whiskers show easterly amplitudes. (c) lowest isobar the QBO descends to, the level at which the

QBO Fourier amplitudes falls to 10% of its maximum (Schenzinger et al. 2017). Smaller circles and squares show ensemble members. (d) latitudinal width, at a user defined height (10 hPa here), is the full width at half amplitude maximum of a Gaussian fit to the QBO Fourier amplitude (Schenzinger et al. 2017).

We evaluate the physical structure and frequency of model QBOs at 10 hPa (unless otherwise mentioned) using the Schenzinger et al. (2017) metrics. In this framework, QBO Fourier amplitude is defined as the fraction of QBO timescale zonal-wind variance (Pascoe et al. 2005), and following Richter et al. (2020), a simulation is deemed to have a QBO if its 5°S-5°N 10-100 hPa zonal-wind alternates between easterlies and westerlies and has a deseasonalized standard deviation  $\geq 20\%$  of the ERA5 10-70 hPa values. Only those models that meet the criteria for having a QBO are analyzed here; roughly 50% of models.. We build on Richter et al. (2020) by incorporating multiple ensemble members per model, allowing for comparison of model QBO internal variability with observed internal variability. The latter is defined as the 2.5-97.5 confidence intervals of the QBO metrics, derived by resampling observed cycles (a pair of westerlies and easterlies phases) with replacement 10,000 times and forming synthetic QBO Fourier amplitude structures from which synthetic spatial metrics are obtained. By this approach, observed internal variability in amplitude is 21.6 to 23.6 m/s, in latitude is 21.3 to 21.8 degrees latitude, and in lowest level is 88.4 to 89.2 hPa. Judging by their members, spatial metrics vary substantially more in models than ERA5 (Fig. 16), interpreted as the QBOs being more unstable.

The main finding is that there are not substantial differences between the CMIP6 and AMIP6 QBOs. The subset of models with QBO well estimate the ERA5 QBO periodicity (Fig. 16a) and the total amplitude (Fig. 16b). Model easterly and westerly amplitudes can vary widely from ERA5. The depth of mean QBO descent into the lower stratosphere is comparable between CMIP6 and AMIP6 and  $\sim 15$  hPa shallower than reanalysis (Bushell et al. 2022), though some models perform well (Fig. 16c). Model QBOs in both ensembles are consistently narrower than ERA5 (Fig. 16d), with widths in individual models strongly associated with QBO amplitude (inter-model correlations in CMIP6: 0.800; AMIP6: 0.787).

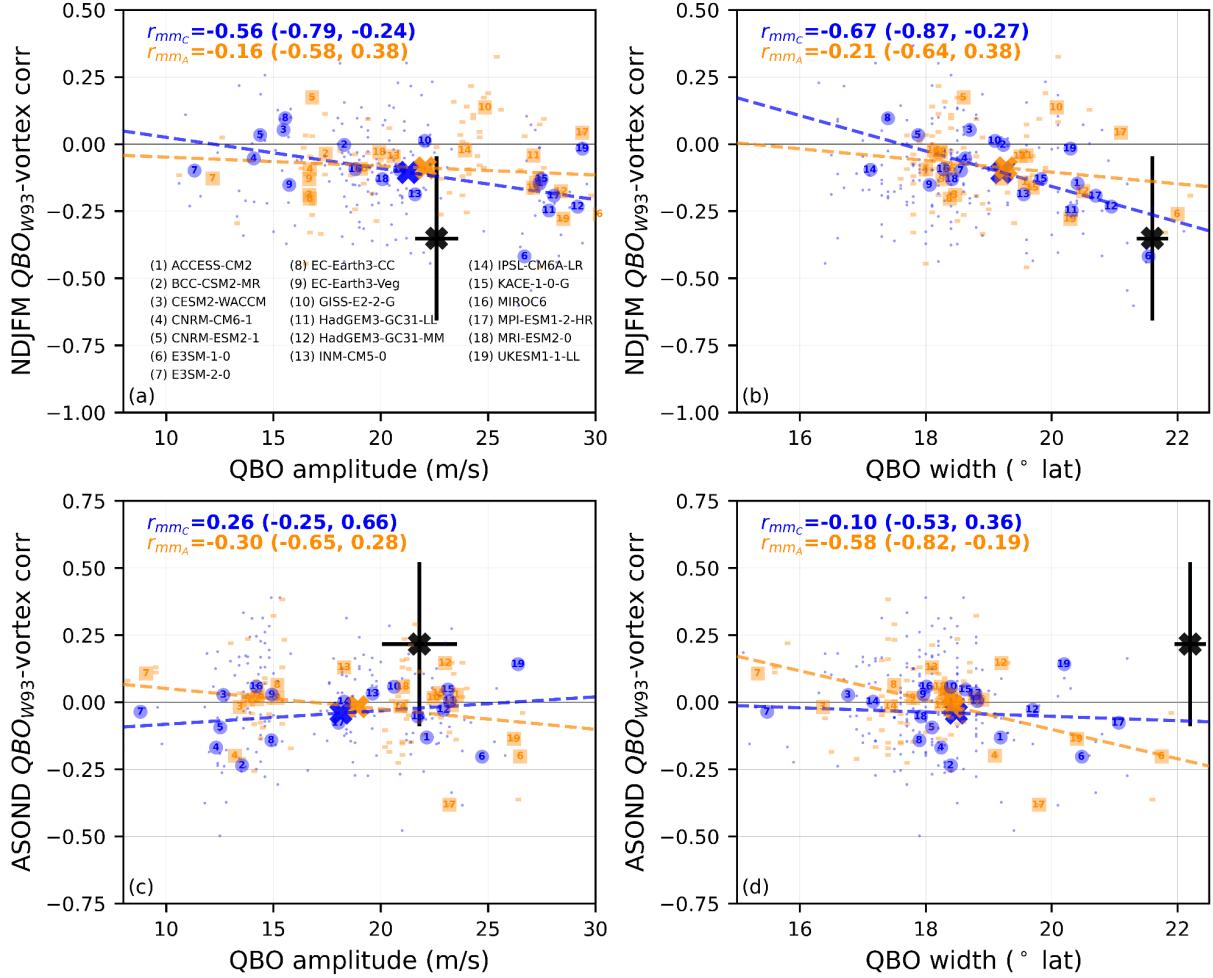


Figure 17: Relationship of QBO amplitudes (left) and widths (right) with the strength of the QBO’s high-latitude teleconnection to the boreal winter (top) and austral spring (bottom) polar vortex. Teleconnection strength is shown on y-axes and x-axes show QBO amplitude or width. The teleconnection is defined as the correlation between seasonally averaged (NH: NDJFM and SH: ASOND) 5°S-5°N EOF based QBO (NH: at 10 hPa; SH: at 20 hPa) and the concurrent 10 hPa polar vortex, the 55°-65°N/S zonal-mean zonal-wind. The QBO amplitude and width are computed at 10 hPa (20 hPa) for the NH (SH) case. Across model CMIP6 (AMIP6) correlations and their confidence intervals are printed in blue (orange); model mean quantities are shown as blue circles (orange squares) and individual members are little circles (squares). Big Xs show CMIP6 and AMIP6 ensemble-means and black quantities show ERA5 and its bootstrapped 95th percentile confidence intervals.

Despite differences in ocean boundary conditions, CMIP6 and AMIP6 yield similar QBO physical structures and frequencies. However, the presence of atmosphere-ocean coupling may influence the QBO’s high-latitude teleconnections. We quantify teleconnection strength as the correlation between tropical (5°S-5°N) QBO wind and the concurrent polar vortex wind (55°-65°N/S, 10 hPa) in boreal winter and austral spring. To more precisely obtain the portion of wind attributable to the QBO in the correlation, each simulation’s tropical 10-70hPa zonal-wind is projected onto Wallace et al. (1993) QBO empirical orthogonal functions derived using the Free University of Berlin record as in Gray et al. (2018). Although the QBO polar teleconnections vary intra-seasonally (Lu et al. 2020; Rao et al. 2023), we compute seasonally averaged correlations (NH: November-March; SH: August-December) to account for different intra-seasonal variability in observations and models conveyed by SSWs (Elsbury et al. 2024).

In ERA5, the 10 hPa QBO anticorrelates with the NH polar vortex ( $r = -0.35$ ), which is also captured, albeit weakly, by CMIP6 ( $r = -0.108$ ) and AMIP6 ( $r = -0.089$ ). Stratifying models by QBO amplitude and width reveals that stronger, wider QBOs yield intensified teleconnections with the polar vortex (Fig. 12a-b). This relationship is stronger in CMIP6 than AMIP6, suggesting that when the atmosphere and ocean are coupled, differences in the width and

amplitude of the QBO can more strongly affect the teleconnection, perhaps because the atmosphere itself is less constrained. .

For the austral spring teleconnection, the QBO metrics and vortex correlations are computed using 20 hPa QBO winds (rather than 10 hPa as for NH), which couple with the austral vortex more strongly (Rao et al. 2023). Models struggle overall to simulate the observed positive QBO-vortex teleconnection ( $r=0.216$ ), producing weak relationships in CMIP6 ( $r = -0.041$ ) and AMIP6 ( $r = -0.017$ ) (Fig. 14c-d). As for the NH, stronger amplitude QBOs in CMIP6 align better with the observed positive correlation (Fig. 14c), but the relationship reverses for QBO width for both AMIP6 and CMIP6 (Fig. 14d). Furthermore, given that some model members come close to the observed teleconnection strength, despite having much weaker and narrower QBOs, suggests that the amplitude and width of the QBO does not strongly constrain the simulated QBO teleconnection to the SH.

In summary, CMIP6 and AMIP6 show comparable QBO structure and periodicities. Both underestimate the width of the QBO and its descent into the lower stratosphere, which could explain why they generally do not capture the observed teleconnection of the QBO to the polar vortex in either hemisphere. However, models with larger QBO amplitude or width show improved teleconnections in boreal winter, and this is particularly true for CMIP6 models. The teleconnection with the austral spring polar vortex, however, does not generally show improvement for models with more realistic QBO amplitude or width.

#### 4. Discussion

This study compares the large-scale circulation and stratosphere-troposphere coupling in CMIP6 versus AMIP6 during dynamically active seasons (boreal winter and austral spring, 1979-2014). Motivated to understand the relationship of model ocean configuration with STC, we compare the fully coupled CMIP6 to the prescribed-ocean AMIP6, and each simulation with ERA5. Relative to ERA5, the representation of STC is modestly better when real-world SST/SIC are prescribed versus being internally generated. AMIP6 is more realistic than CMIP6 in positioning the latitude of the subtropical jet and polar vortex; it reduces biases in the amplitude and position of tropical-extratropical planetary waves; and it shows marginally better representations of upward and downward STC processes than CMIP6.

While AMIP6 offers modest improvements in certain areas, these gains may not outweigh the benefits of CMIP6's more realistic coupled atmosphere-ocean configuration. The choice between CMIP6 and AMIP6 for STC-related studies ultimately hinges on the sensitivity of the specific STC pathway to SST or SIC biases or to ocean-atmosphere coupling processes. For instance, AMIP6 appears particularly well-suited for ENSO teleconnection analyses due to reduced SST-driven biases, in agreement with Bayr et al. (2024), Fang et al. (2024), and Beniche et al. (2025). AMIP6 El Niños show stronger inter-model correlations with the polar vortex than their CMIP6 counterparts (Fig. 14b). Supporting this, Manzini et al. (2024) observed that El Niño induced vortex responses were inconsistent across CMIP6 models, even when North Pacific teleconnections were well represented. Figure 14b corroborates this finding, suggesting that AMIP6 may help resolve such discrepancies.

However, for some aspects of STC, prescribing observed ocean boundary conditions has negligible effects and does not justify forgoing CMIP6's relative realism. Both AMIP6 and CMIP6 simulate early-season planetary waves that are too weak in both hemispheres relative to ERA5, followed by an overamplification later in the season, mirroring the delayed timing of sudden stratospheric warmings and the seasonal persistence of annular modes. We also find little difference in QBO periodicity, amplitude, descent, and width between CMIP6 and AMIP6. The lack of difference is surprising and warrants more detailed analysis, given that an East Pacific El Niño, which matches the

sign and location of long-term tropical SST biases in CMIP6, hastens the QBO's downward propagation (Zhou et al. 2024; Kawatani et al. 2025).

## 5. Conclusions

CMIP6 sea surface temperature biases, including an overly cold equatorial Pacific (Li and Xie 2014), warm tropical East Pacific and East Atlantic (Wills et al. 2022), cold subtropics (Rao et al. 2022), excessive Arctic sea ice (Pan et al. 2023), and a warm Southern Ocean (Hu et al. 2024), are pervasive relative to AMIP6. These biases in turn correlate with climatological atmospheric circulation differences across models: CMIP6 shows robust (in a majority, >80%, of models) equator-shifted subtropical jets and overly cold polar lower stratospheres. These cold biases are consistent with (non-robust) accelerated boreal and austral polar vortices above 50 hPa (Fig. 2a, 2c). The vortex biases are also retained when CMIP6 and AMIP6 climatologies are computed relative to ERA5 (Fig. S1, S2).

Across-model regressions of boreal winter CMIP6 minus AMIP6 circulation differences onto CMIP6 minus AMIP6 surface air temperature biases reveal that the cold subtropical North Pacific is linked to an equatorward storm-track bias in boreal winter (Fig. 3c), echoing Priestly et al. (2023), while the East Pacific and Atlantic biases relate to an equatorward shifted subtropical jet whose signal extends upward into the stratosphere (Fig. 3d-e). The three aforementioned regions, plus the Arctic, also correlate with an accelerated NH polar vortex in CMIP6, though none of the regressions coefficients are statistically significant above 50 hPa (Fig. 3a, 3c-e). In the SH case, CMIP6's overly warm austral spring Southern Ocean is related to decelerated high-latitude 50-1000 hPa zonal-wind column relative to AMIP6 (Fig. 3f). The SH subtropical jet bias shows a relation to the East Pacific and Atlantic warm biases, more so than the cold equatorial and subtropical South Pacific biases (Fig. 3g-j).

The colder polar stratospheres in CMIP6 are partly attributable to vertical planetary wave shortcomings. Although AMIP6 improves aspects of the planetary wave field (i.e., the eddy geopotential height at 200 hPa and 10 hPa), such as its amplitude over the oceans, both AMIP6 and CMIP6 produce early season planetary waves that are too weak in either hemisphere, but then become too strong later in the season (Fig. 4, Fig. 5). Models also simulate too weak upward coupling between planetary waves and the polar vortex and too low SSW frequencies, especially in CMIP6 and in SH spring (Fig. 6, Fig. 7).

Regarding downward coupling, the strength of the coupling between polar stratospheric wind anomalies and the tropospheric circulation is captured reasonably well in models, if somewhat underestimated for CMIP6 in boreal winter (Fig. 8). Nonetheless, models delay the seasonal peak in annular mode persistence and predictability in both hemispheres (Fig. 9), paralleling the delayed planetary wave seasonality and preference for SSWs in March rather than January-February (Fig. 5, Fig. 7). These timing errors result in exaggerated downward coupling later in each season, affecting the timing of surface impacts. In general, while models can broadly depict the downward influence of the extratropical stratosphere on the troposphere, improving the seasonal cycle of planetary waves may correct multiple STC-related shortfalls. Curiously, while models show reasonable correlations of the tropospheric circulation to a given stratospheric perturbation (Fig. 8), their downward response is strongly underestimated in connection to a preceding lower stratospheric upward wave flux (Fig. 10). This suggests that because the upward coupling is too weak, the stratospheric perturbation is also too weak, reducing the downward effect. It would be worth investigating to what extent the missing early season tropospheric ridges in CMIP6 and AMIP6 (Fig. 10) diminish the stationary wave's early season amplitude (Fig. 5), potentially driving the planetary wave seasonal cycle bias and reduced upward wave coupling.

We also benchmark high-latitude teleconnections of ENSO and QBO in CMIP6 and AMIP6. CMIP6 simulates overly intense and westward shifted ENSO anomalies, which in turn biases the North Pacific Low response to ENSO; AMIP6 improves these biases (Fig. 11, Fig. 12, Fig. 13). However, neither ensemble produces the maximum observed NH ENSO-polar vortex relationship in January-February, with CMIP6 models simulating a lingering relationship into March-April and AMIP6 improving the seasonality (and amplitude) by a month (Fig. 15). Given that AMIP6 improves the ENSO teleconnections, we investigate if models with stronger ENSO anomalies or better Pacific teleconnections exhibit stronger polar stratospheric responses to ENSO or amplified coupling to the tropospheric polar cap circulation (Fig. 14). However, weak across model relationships are detected (Fig. 14a, 14b), consistent with Manzini et al. (2024), who found that vortex responses to ENSO are not guaranteed in models, even with a quality ENSO to North Pacific teleconnection.

For the SH during austral spring, AMIP6 improves the quality of the ENSO-South Pacific teleconnection relative to CMIP6 (Fig. 13), aligning with Fang et al. (2024), who reports a similar result for austral summer. Despite the improvement, both CMIP6 and AMIP6 struggle to reproduce the strong observed ENSO-vortex relationship during November-December (Fig. 15), pointing to deficiencies apart from the ocean conditions, such as with wave-mean flow interactions (Fig. 6, 7, 10). The across model relationships between the South Pacific teleconnection and the austral polar cap circulation are weak, even in AMIP6 (Fig. 14)

CMIP6 and AMIP6 models with a QBO reproduce well its observed periodicity and amplitude, but underestimate the lowest level of descent and its latitudinal width (Fig. 16). The models lack systematic differences between AMIP6 and CMIP6 QBOs. While both AMIP6 and CMIP6 struggle to capture observed QBO teleconnections to the boreal winter and austral spring polar vortex, stronger and wider QBOs do strengthen the boreal winter teleconnection, especially in CMIP6 (Fig. 17). Higher amplitude and wider QBOs do not appear to affect the austral spring teleconnection in any systematic fashion. While the differences between CMIP6 and AMIP6 QBO-polar vortex teleconnection relationships are not fully understood, a speculation is that the less constrained atmospheric circulation in CMIP6 might enable the width and amplitude of the QBO to more strongly modulate the QBO's middle stratospheric pathway to affect the vortex.

In conclusion, although CMIP6 and AMIP6 reproduce many aspects of 1979-2014 large-scale circulation and STC, systematic ocean boundary biases in CMIP6 are related to large-scale mean state circulation errors, including the representation of the tropospheric jet streams, stratospheric polar jets, planetary waves, and teleconnections. As we do not explore causality in the ocean-atmosphere interplay, it is plausible that atmospheric biases drive the SST/SIC biases, rather than the other way around. While we only consider the historical period in order to compare with AMIP6, we are curious to what extent these coupled ocean-atmosphere biases are pervasive in CMIP6 climate projections, perhaps undermining the fidelity of projected large-scale circulation in similar ways.

## References

- Ahmadi, R., and O. Alizadeh, 2023: The possible links between the Barents–Kara sea-ice area, Ural blocking, and the North Atlantic Oscillation. *Quart. J. Roy. Meteor. Soc.*, 149, 3357–3372, <https://doi.org/10.1002/qj.4543>.
- Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: Middle Atmosphere Dynamics. Academic Press, 489 pp.
- Anstey, J. A., and T. G. Shepherd, 2014: High-latitude influence of the quasi-biennial oscillation. *Quart. J. Roy. Meteor. Soc.*, 140 (678), 1–21. <https://doi.org/10.1002/qj.2132>.
- Anstey, J. A., and Coauthors, 2022: Teleconnections of the quasi-biennial oscillation in a multi-model ensemble of QBO-resolving models. *Quart. J. Roy. Meteor. Soc.*, 148 (744), 1568–1592, <https://doi.org/10.1002/qj.4048>
- Ayarzagüena, B., and Coauthors, 2018: Stratospheric connection to the abrupt end of the 2016/2017 Iberian drought. *Geophys. Res. Lett.*, 45 (22), 12–639. <https://doi.org/10.1029/2018GL079802>.
- Baldwin, M. P., and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, 294 (5542), 581–584. <https://doi.org/10.1126/science.1063315>.
- Baldwin, M. P., D. W. Thompson, E. F. Shuckburgh, W. A. Norton, and N. P. Gillett, 2003: Weather from the stratosphere? *Science*, 301 (5631), 317–319. <https://doi.org/10.1126/science.1085688>.
- Baldwin, M. P., and D. W. Thompson, 2009: A critical comparison of stratosphere–troposphere coupling indices. *Quart. J. Roy. Meteor. Soc.*, 135 (644), 1661–1672. <https://doi.org/10.1002/QJ.479>.
- Baldwin, M. P., and Coauthors, 2021: Sudden stratospheric warmings. *Rev. Geophys.*, 59 (1), e2020RG000708. <https://doi.org/10.1029/2020RG000708>.
- Baldwin, M. P., T. Birner, and B. Ayarzagüena, 2024: Tropospheric amplification of stratosphere–troposphere coupling. *Q. J. R. Meteorol. Soc.*, 150, 5188–5205, <https://doi.org/10.1002/qj.4701>.
- Bayr, T., D. I. Domeisen, and C. Wengel, 2019: The effect of the equatorial Pacific cold SST bias on simulated ENSO teleconnections to the North Pacific and California. *Climate Dyn.*, 53 (7), 3771–3789. <https://doi.org/10.1007/s00382-019-04746-9>.
- Bayr, T., J. F. Lübbecke, J. Vialard, and M. Latif, 2024: Equatorial Pacific cold tongue bias degrades simulation of ENSO asymmetry due to underestimation of strong eastern Pacific El Niños. *J. Climate*, 37 (23), 6167–6182. <https://doi.org/10.1175/JCLI-D-24-0071.1>.
- Beniche, M., J. Vialard, M. Lengaigne, and N. M. J. Hall, 2025: How well do AMIP6 and CMIP6 reproduce the specific Extreme El Niño teleconnections to North America? *J. Climate*, 130 (3), e2024JD041740. <https://doi.org/10.1029/2024JD041740>.
- Birner, T., and J. R. Albers, 2017: Sudden stratospheric warmings and anomalous upward wave activity flux. *SOLA*, 13 (Special Edition), 8–12. <https://doi.org/10.2151/sola.13A-002>.

Bracegirdle, T. J., E. Shuckburgh, J.-B. Sallee, Z. Wang, A. J. Meijers, N. Bruneau, T. Phillips, and L. J. Wilcox, 2013: Assessment of surface winds over the Atlantic, Indian, and Pacific Ocean sectors of the Southern Ocean in CMIP5 models: Historical bias, forcing response, and state dependence. *J. Geophys. Res. Atmos.*, 118, 547–562, <https://doi.org/10.1002/jgrd.50153>.

Bracegirdle, T. J., C. R. Holmes, J. S. Hosking, G. J. Marshall, M. Osman, M. Patterson, and T. Rackow, 2020: Improvements in circumpolar Southern Hemisphere extratropical atmospheric circulation in CMIP6 compared to CMIP5. *Earth Space Sci.*, 7, e2019EA001065, <https://doi.org/10.1029/2019EA001065>.

Bushell, A. C., J. A. Anstey, N. Butchart, Y. Kawatani, S. M. Osprey, J. H. Richter, F. Serva, P. Braesicke, C. Cagnazzo, C. C. Chen, and H. Y. Chun, 2020: Evaluation of the quasi-biennial oscillation in global climate models for the SPARC QBO-initiative. *Quart. J. Roy. Meteor. Soc.*, 148 (744), 1459–1489. <https://doi.org/10.1002/qj.3820>.

Butchart, N., and Coauthors, 2011: Multimodel climate and variability of the stratosphere. *J. Geophys. Res. Atmos.*, 116 (D5), D05102. <https://doi.org/10.1029/2010JD014995>.

Butler, A. H., L. M. Polvani, and C. Deser, 2014: Separating the stratospheric and tropospheric pathways of El Niño–Southern Oscillation teleconnections. *Environ. Res. Lett.*, 9 (2), 024014. <https://doi.org/10.1088/1748-9326/9/2/024014>.

Butler, A. H., and E. P. Gerber, 2018: Optimizing the definition of major sudden stratospheric warmings. *J. Climate*, 31, 233–254, <https://doi.org/10.1175/JCLI-D-17-0323.1>.

Butler, A. H., and D. I. V. Domeisen, 2021: The wave geometry of final stratospheric warming events. *Weather Clim. Dynam.*, 2 (2), 453–474. <https://doi.org/10.5194/wcd-2-453-2021>.

Calvo, N., and Coauthors, 2017: Northern hemisphere stratospheric pathway of different El Niño flavors in stratosphere-resolving CMIP5 models. *J. Climate*, 30 (12), 4351–4371. <https://doi.org/10.1175/JCLI-D-16-0132.1>.

Castanheira, J. M., and C. A. F. Marques, 2022: Biases of the barotropic atmospheric circulation variability in CMIP6 models. *J. Climate*, 35 (15), 5071–5085. <https://doi.org/10.1175/JCLI-D-21-0581.1>.

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Climate*, 20 (3), 449–469. <https://doi.org/10.1175/JCLI3996.1>.

Dai, Y., and P. Hitchcock, 2021: Understanding the basin asymmetry in surface response to sudden stratospheric warmings from an ocean–atmosphere coupled perspective. *J. Climate*, 34 (21), 8683–8698. <https://doi.org/10.1175/JCLI-D-21-0314.1>.

Davini, P., and F. d'Andrea, 2020: From CMIP3 to CMIP6: Northern Hemisphere atmospheric blocking simulation in present and future climate. *J. Climate*, 33, 10021–10038, <https://doi.org/10.1175/JCLI-D-19-0993.1>.

Davy, R., and S. Outten, 2020: The Arctic surface climate in CMIP6: Status and developments since CMIP5. *J. Climate*, 33, 8047–8068. <https://doi.org/10.1175/JCLI-D-19-0990.1>

Domeisen, D. I., C. M. Grams, and L. Papritz, 2020: The role of North Atlantic–European weather regimes in the

surface impact of sudden stratospheric warming events. *Weather Clim. Dynam.*, 1 (2), 373–388. <https://doi.org/10.5194/wcd-1-373-2020>.

Domeisen, D. I., C. I. Garfinkel, and A. H. Butler, 2019: The teleconnection of El Niño Southern Oscillation to the stratosphere. *Rev. Geophys.*, 57 (1), 5–47. <https://doi.org/10.1029/2018RG000596>.

Dunkerton, T., C. F. Hsu, and M. E. McIntyre, 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.*, 38 (4), 819–844. [https://doi.org/10.1175/1520-0469\(1981\)038<0819:SEALDF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<0819:SEALDF>2.0.CO;2).

Elsbury, D., Y. Peings, and G. Magnusdottir, 2021: CMIP6 models underestimate the Holton-Tan effect. *Geophys. Res. Lett.*, 48 (24), e2021GL094083. <https://doi.org/10.1029/2021GL094083>.

Elsbury, D., A. Butler, Y. Peings, and G. Magnusdottir, 2024: Sensitivity of easterly QBO's boreal winter teleconnections and surface impacts to SSWs. *J. Climate*, 37 (14), 3675–3688. <https://doi.org/10.1175/JCLI-D-23-0395.1>.

Eichelberger, S. J., and D. L. Hartmann, 2007: Zonal jet structure and the leading mode of variability. *J. Climate*, 20 (20), 5149–5163. <https://doi.org/10.1175/JCLI4279.1>.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, 9 (5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>.

Fang, Y., J. A. Screen, X. Hu, S. Lin, N. C. Williams, and S. Yang, 2024: CMIP6 models underestimate ENSO teleconnections in the Southern Hemisphere. *Geophys. Res. Lett.*, 51 (18), e2024GL110738. <https://doi.org/10.1029/2024GL110738>.

Feng, K., J. Rao, H. Chen, R. Ren, and D. Guo, 2024: Improvement of the simulated Southern Hemisphere stratospheric polar vortex across series of CMIPs. *Climate Dyn.*, 62 (6), 5605–5621. <https://doi.org/10.1007/s00382-024-07250-x>.

Furtado, J. C., J. L. Cohen, A. H. Butler, E. E. Riddle, and A. Kumar, 2015: Eurasian snow cover variability and links to winter climate in the CMIP5 models. *Climate Dyn.*, 45, 2591–2605. <https://doi.org/10.1007/s00382-015-2494-4>.

Gao, Z., S. Zhao, Q. Liu, S. M. Long, and S. Sun, 2024: Assessment of the Southern Ocean sea surface temperature biases in CMIP5 and CMIP6 models. *J. Ocean Univ. China*, 23 (5), 1135–1150. <https://doi.org/10.1007/s11802-024-5808-5>.

Garfinkel, C. I., and D. L. Hartmann, 2011: The influence of the quasi-biennial oscillation on the troposphere in winter in a hierarchy of models. Part I: Simplified dry GCMs. *J. Atmos. Sci.*, 68 (6), 1273–1289. <https://doi.org/10.1175/2011JAS3665.1>.

Garfinkel, C. I., T. A. Shaw, D. L. Hartmann, and D. W. Waugh, 2012: Does the Holton-Tan mechanism explain how the quasi-biennial oscillation modulates the Arctic polar vortex? *J. Atmos. Sci.*, 69 (5), 1713–1733. <https://doi.org/10.1175/JAS-D-11-0209.1>.

Garfinkel, C. I., I. White, E. P. Gerber, M. Jucker, and M. Erez, 2020: The building blocks of Northern Hemisphere wintertime stationary waves. *J. Climate*, 33 (13), 5611–5633. <https://doi.org/10.1175/JCLI-D-19-0181.1>.

Garfinkel, C. I., E. P. Gerber, O. Shamir, J. Rao, M. Jucker, I. White, and N. Paldor, 2022: A QBO cookbook: Sensitivity of the quasi-biennial oscillation to resolution, resolved waves, and parameterized gravity waves. *J. Adv. Model. Earth Syst.*, 14 (3), e2021MS002568. <https://doi.org/10.1029/2021MS002568>.

Garfinkel, C. I., and Coauthors, 2025: A process-based evaluation of biases in extratropical stratosphere–troposphere coupling in subseasonal forecast systems. *Weather Clim. Dyn.*, 6, 171–195, <https://doi.org/10.5194/wcd-6-171-2025>.

Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, 80 (1), 29–56.  
[https://doi.org/10.1175/1520-0477\(1999\)080<0029:AOOTRO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2).

Gerber, E. P., and Coauthors, 2010: Stratosphere-troposphere coupling and annular mode variability in chemistry-climate models. *J. Geophys. Res. Atmos.*, 115 (D3), D00M06. <https://doi.org/10.1029/2009JD013770>.

Gillett, N. P., and J. C. Fyfe, 2013: Annular mode changes in the CMIP5 simulations. *Geophys. Res. Lett.*, 40 (6), 1189–1193. <https://doi.org/10.1002/grl.50249>.

Gray, L. J., J. A. Anstey, Y. Kawatani, H. Lu, S. Osprey, and V. Schenzinger, 2018: Surface impacts of the quasi-biennial oscillation. *Atmos. Chem. Phys.*, 18 (11), 8227–8247. <https://doi.org/10.5194/acp-18-8227-2018>.

Gong, T., and D. Luo, 2017: Ural blocking as an amplifier of the Arctic sea ice decline in winter. *J. Climate*, 30, 2639–2654, <https://doi.org/10.1175/JCLI-D-16-0544.1>.

Hansen, F., K. Matthes, and L. J. Gray, 2013: Sensitivity of stratospheric dynamics and chemistry to QBO nudging width in the chemistry–climate model WACCM. *J. Geophys. Res. Atmos.*, 118 (18), 10,464–10,474.  
<https://doi.org/10.1002/jgrd.50812>.

Hartig, K., and E. Tziperman, 2024: Suppression of cold air outbreaks over the interior of North America in a warmer climate. *J. Climate*, 37 (17), 4541–4557. <https://doi.org/10.1175/JCLI-D-23-0477.1>.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, 146 (730), 1999–2049.  
<https://doi.org/10.1002/qj.3803>.

Hitchcock, P., and I. R. Simpson, 2014: The downward influence of stratospheric sudden warmings. *J. Atmos. Sci.*, 71 (10), 3856–3876. <https://doi.org/10.1175/JAS-D-14-0012.1>.

Hu, J., Y. Shen, J. Deng, Y. Jia, Z. Wang, and A. Li, 2023: Revisiting the influence of ENSO on the Arctic stratosphere in CMIP5 and CMIP6 models. *Atmosphere*, 14 (5), 785. <https://doi.org/10.3390/atmos14050785>.

Hurwitz, M. M., P. Braesicke, and J. A. Pyle, 2011: Sensitivity of the mid-winter Arctic stratosphere to QBO width in a simplified chemistry–climate model. *Atmos. Sci. Lett.*, 12 (3), 268–272. <https://doi.org/10.1002/asl.330>.

Hurwitz, M. M., N. Calvo, C. I. Garfinkel, A. H. Butler, S. Ineson, C. Cagnazzo, E. Manzini, and C. Peña-Ortiz, 2014: Extra-tropical atmospheric response to ENSO in the CMIP5 models. *Climate Dyn.*, 43, 3367–3376. <https://doi.org/10.1007/s00382-014-2110-z>.

Holt, L. A., F. Lott, R. R. Garcia, G. N. Kiladis, Y. M. Cheng, J. A. Anstey, P. Braesicke, A. C. Bushell, N. Butchart, C. Cagnazzo, and C. C. Chen, 2022: An evaluation of tropical waves and wave forcing of the QBO in the QBOi models. *Quart. J. Roy. Meteor. Soc.*, 148 (744), 1541–1567. <https://doi.org/10.1002/qj.3827>.

Iza, M., N. Calvo, and E. Manzini, 2016: The stratospheric pathway of La Niña. *J. Climate*, 29 (24), 8899–8914. <https://doi.org/10.1175/JCLI-D-16-0230.1>.

Karpechko, A. Y., P. Hitchcock, D. H. Peters, and A. Schneidereit, 2017: Predictability of downward propagation of major sudden stratospheric warmings. *Quart. J. Roy. Meteor. Soc.*, 143 (704), 1459–1470. <https://doi.org/10.1002/qj.3017>.

Kidston, J., A. A. Scaife, S. C. Hardiman, D. M. Mitchell, N. Butchart, M. P. Baldwin, and L. J. Gray, 2015: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geosci.*, 8 (6), 433–440. <https://doi.org/10.1038/NGEO2424>

Kumar, V., S. Yoden, and M. H. Hitchman, 2022: QBO and ENSO effects on the mean meridional circulation, polar vortex, subtropical westerly jets, and wave patterns during boreal winter. *J. Geophys. Res. Atmos.*, 127 (15), e2022JD036691. <https://doi.org/10.1029/2022JD036691>.

Labe, Z., Y. Peings, and G. Magnusdottir, 2019: The effect of QBO phase on the atmospheric response to projected Arctic sea ice loss in early winter. *Geophys. Res. Lett.*, 46 (13), 7663–7671. <https://doi.org/10.1029/2019GL083095>.

Lawrence, Z. D., and G. L. Manney, 2020: Does the Arctic stratospheric polar vortex exhibit signs of preconditioning prior to sudden stratospheric warmings? *J. Atmos. Sci.*, 77 (2), 611–632. <https://doi.org/10.1175/JAS-D-19-0168.1>.

Lawrence, Z. D., and Coauthors, 2022: Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems. *Weather Clim. Dyn. Discuss.*, 2022, 1–37, <https://doi.org/10.5194/wcd-2022-15>.

Lee, J., K. R. Sperber, P. J. Gleckler, K. E. Taylor, and C. J. Bonfils, 2021: Benchmarking performance changes in the simulation of extratropical modes of variability across CMIP generations. *J. Climate*, 34 (17), 6945–6969. <https://doi.org/10.1175/JCLI-D-20-0832.1>.

Lee, S. H., A. H. Butler, and G. L. Manney, 2025: Two major sudden stratospheric warmings during winter 2023/2024. *Weather*, 80 (2), 45–53. <https://doi.org/10.1002/wea.7656>.

Li, G., and S.-P. Xie, 2014: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *J. Climate*, 27 (4), 1765–1780. <https://doi.org/10.1175/JCLI-D-13-00337.1>.

Lim, E. P., H. H. Hendon, and H. Rashid, 2013: Seasonal predictability of the southern annular mode due to its association with ENSO. *J. Climate*, 26 (20), 8037–8054. <https://doi.org/10.1175/JCLI-D-13-00006.1>.

Lim, E. P., H. H. Hendon, and D. W. Thompson, 2018: Seasonal evolution of stratosphere-troposphere coupling in the Southern Hemisphere and implications for the predictability of surface climate. *J. Geophys. Res. Atmos.*, 123 (21), 12,002–12,016. <https://doi.org/10.1029/2018JD029321>.

Lorenz, D. J., and D. L. Hartmann, 2001: Eddy–zonal flow feedback in the Southern Hemisphere. *J. Atmos. Sci.*, 58 (21), 3312–3327. [https://doi.org/10.1175/1520-0469\(2001\)058<3312:EZFFIT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<3312:EZFFIT>2.0.CO;2).

Lorenz, D. J., and D. L. Hartmann, 2003: Eddy–zonal flow feedback in the Northern Hemisphere winter. *J. Climate*, 16 (8), 1212–1227. [https://doi.org/10.1175/1520-0442\(2003\)16<1212:EFFITN>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)16<1212:EFFITN>2.0.CO;2).

Lu, H., M. H. Hitchman, L. J. Gray, J. A. Anstey, and S. M. Osprey, 2020: On the role of Rossby wave breaking in the quasi-biennial modulation of the stratospheric polar vortex during boreal winter. *Quart. J. Roy. Meteor. Soc.*, 146 (729), 1939–1959. <https://doi.org/10.1002/qj.3775>.

Ma, T., W. Chen, X. An, C. I. Garfinkel, and Q. Cai, 2023: Nonlinear effects of the stratospheric Quasi-Biennial Oscillation and ENSO on the North Atlantic winter atmospheric circulation. *J. Geophys. Res. Atmos.*, 128

Manzini, E., B. Ayarzagüena, N. Calvo, and D. Matei, 2024: Nonlinearity and asymmetry of the ENSO stratospheric pathway to North Atlantic and Europe, revisited. *J. Geophys. Res. Atmos.*, 129 (2), e2023JD039992. <https://doi.org/10.1029/2023JD039992>.

Martineau, P., S.-W. Son, M. Taguchi, and A. H. Butler, 2018: A comparison of the momentum budget in reanalysis datasets during sudden stratospheric warming events. *Atmos. Chem. Phys.*, 18 (10), 7169–7187. <https://doi.org/10.5194/acp-18-7169-2018>.

Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, 28 (8), 1479–1494. [https://doi.org/10.1175/1520-0469\(1971\)028<1479:ADMOTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2).

Maycock, A. C., G. I. Masukwedza, P. Hitchcock, and I. R. Simpson, 2020: A regime perspective on the North Atlantic eddy-driven jet response to sudden stratospheric warmings. *J. Climate*, 33

McIntyre, M. E., and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature*, 305 (5935), 593–600. <https://doi.org/10.1038/305593a0>.

Mori, M., H. Tokinaga, Y. Kosaka, H. Nakamura, B. Taguchi, and H. Tatebe, 2024: The influence of extratropical ocean on the PNA teleconnection: Role of atmosphere-ocean coupling. *Geophys. Res. Lett.*, 51 (14), e2024GL110234. <https://doi.org/10.1029/2024GL110234>.

Newman, P. A., E. R. Nash, and J. E. Rosenfield, 2001: What controls the temperature of the Arctic stratosphere during the spring? *J. Geophys. Res. Atmos.*, 106 (D17), 19999–20010. <https://doi.org/10.1029/2000JD000061>.

Notz, D., and SIMIP Community, 2020: Arctic sea ice in CMIP6. *Geophys. Res. Lett.*, 47 (10), e2019GL086749. <https://doi.org/10.1029/2019GL086749>.

Oehrlein, J., G. Chiodo, and L. M. Polvani, 2019: Separating and quantifying the distinct impacts of El Niño and sudden stratospheric warmings on North Atlantic and Eurasian wintertime climate. *Atmos. Sci. Lett.*, 20 (7), e923. <https://doi.org/10.1002/asl.923>.

Osman, M., T. G. Shepherd, and C. S. Vera, 2022: The combined influence of the stratospheric polar vortex and ENSO on zonal asymmetries in the southern hemisphere upper tropospheric circulation during austral spring and summer. *Climate Dyn.*, 59 (9), 2949–2964. <https://doi.org/10.1007/s00382-022-06225-0>.

Pahlavan, H. A., Q. Fu, J. M. Wallace, and G. N. Kiladis, 2021: Revisiting the quasi-biennial oscillation as seen in ERA5. Part I: Description and momentum budget. *J. Atmos. Sci.*, 78 (3), 673–691. <https://doi.org/10.1175/JAS-D-20-0248.1>.

Palmeiro, F. M., J. García-Serrano, M. Rodrigo, M. Abalos, B. Christiansen, and S. Yang, 2023: Boreal winter stratospheric climatology in EC-EARTH: CMIP6 version. *Climate Dyn.*, 60 (3), 883–898. <https://doi.org/10.1007/s00382-022-06368-0>.

Pan, R., Q. Shu, Z. Song, S. Wang, Y. He, and F. Qiao, 2023: Simulations and projections of winter sea ice in the Barents Sea by CMIP6 climate models. *Adv. Atmos. Sci.*, 40 (12), 2318–2330. <https://doi.org/10.1007/s00376-023-2235-2>.

Pascoe, C. L., L. J. Gray, S. A. Crooks, M. N. Juckes, and M. P. Baldwin, 2005: The quasi-biennial oscillation: Analysis using ERA-40 data. *J. Geophys. Res. Atmos.*, 110 (D8). <https://doi.org/10.1029/2005JD006871>.

Peings, Y., 2019: Ural blocking as a driver of early-winter stratospheric warmings. *Geophys. Res. Lett.*, 46, 5460–5468, <https://doi.org/10.1029/2019GL082097>.

Perlitz, J., and N. Harnik, 2004: Downward coupling between the stratosphere and troposphere: The relative roles of wave and zonal mean processes. *J. Climate*, 17 (24), 4902–4909. <https://doi.org/10.1175/JCLI-3247.1>.

Philander, S. G., 1989: El Niño, La Niña, and the Southern Oscillation. *Int. Geophys. Ser.*, 46, X–289. <https://doi.org/10.1017/S0016756800015351>.

Plumb, R. A., 1985: On the three-dimensional propagation of stationary waves. *J. Atmos. Sci.*, 42 (3), 217–229. [https://doi.org/10.1175/1520-0469\(1985\)042<0217:OTTDPO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<0217:OTTDPO>2.0.CO;2).

Polvani, L. M., L. Sun, A. H. Butler, J. H. Richter, and C. Deser, 2017: Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia. *J. Climate*, 30 (6), 1959–1969. <https://doi.org/10.1175/JCLI-D-16-0277.1>.

Priestley, M. D., D. Ackerley, J. L. Catto, and K. I. Hodges, 2023: Drivers of biases in the CMIP6 extratropical storm tracks. Part I: Northern Hemisphere. *J. Climate*, 36 (5), 1451–1467. <https://doi.org/10.1175/JCLI-D-20-0976.1>.

Rao, J., C. I. Garfinkel, T. Wu, Y. Lu, and M. Chu, 2022: Mean state of the Northern Hemisphere stratospheric polar vortex in three generations of CMIP models. *J. Climate*, 35 (14), 4603–4625. <https://doi.org/10.1175/JCLI-D-21-0694.1>.

Rao, J., C. I. Garfinkel, R. Ren, T. Wu, and Y. Lu, 2023: Southern Hemisphere response to the quasi-biennial oscillation in the CMIP5/6 models. *J. Climate*, 36 (8), 2603–2623. <https://doi.org/10.1175/JCLI-D-22-0675.1>.

Rea, D., D. Elsbury, A. H. Butler, L. Sun, Y. Peings, and G. Magnusdottir, 2024: Interannual influence of Antarctic sea ice on Southern Hemisphere stratosphere–troposphere coupling. *Geophysical Research Letters*, 51, e2023GL107478, <https://doi.org/10.1029/2023GL107478>.

Richter, J. H., J. A. Anstey, N. Butchart, Y. Kawatani, G. A. Meehl, S. Osprey, and I. R. Simpson, 2020: Progress in simulating the quasi-biennial oscillation in CMIP models. *J. Geophys. Res. Atmos.*, 125 (8), e2019JD032362. <https://doi.org/10.1029/2019JD032362>.

Roach, L. A., J. Dörr, C. R. Holmes, F. Massonet, E. W. Blockley, D. Notz, T. Rackow, M. N. Raphael, S. P. O'Farrell, D. A. Bailey, and C. M. Bitz, 2020: Antarctic sea ice area in CMIP6. *Geophys. Res. Lett.*, 47 (9), e2019GL086729. <https://doi.org/10.1029/2019GL086729>.

Scaife, A. A., and Coauthors, 2017: Tropical rainfall, Rossby waves and regional winter climate predictions. *Quart. J. Roy. Meteor. Soc.*, 143 (702), 1–11. <https://doi.org/10.1002/qj.2910>.

Schenzinger, V., S. Osprey, L. Gray, and N. Butchart, 2017: Defining metrics of the quasi-biennial oscillation in global climate models. *Geosci. Model Dev.*, 10 (6), 2157–2168. <https://doi.org/10.5194/gmd-10-2157-2017>.

Shu, Q., Q. Wang, Z. Song, F. Qiao, J. Zhao, M. Chu, and X. Li, 2020: Assessment of sea ice extent in CMIP6 with comparison to observations and CMIP5. *Geophys. Res. Lett.*, 47 (9), e2020GL087965. <https://doi.org/10.1029/2020GL087965>.

Simpson, I. R., P. Hitchcock, T. G. Shepherd, and J. F. Scinocca, 2011: Stratospheric variability and tropospheric annular-mode timescales. *Geophys. Res. Lett.*, 38 (20), L20806. <https://doi.org/10.1029/2011GL049304>.

Simpson, I. R., and L. M. Polvani, 2016: Revisiting the relationship between jet position, forced response, and annular mode variability in the southern midlatitudes. *Geophys. Res. Lett.*, 43, 2896–2903, <https://doi.org/10.1002/2016GL067989>.

Simpson, I. R., J. Bacmeister, R. B. Neale, C. Hannay, A. Gettelman, R. R. Garcia, P. H. Lauritzen, D. R. Marsh, M. J. Mills, B. Medeiros, and J. H. Richter, 2020: An evaluation of the large-scale atmospheric circulation and its variability in CESM2 and other CMIP models. *J. Geophys. Res. Atmos.*, 125 (13), e2020JD032835. <https://doi.org/10.1029/2020JD032835>.

Smith, K. L., and P. J. Kushner, 2012: Linear interference and the initiation of extratropical stratosphere–troposphere interactions. *J. Geophys. Res. Atmos.*, 117, D13107, <https://doi.org/10.1029/2012JD017587>.

Smith, D. M., and Coauthors, 2022: Robust but weak winter atmospheric circulation response to future Arctic sea ice loss. *Nature Commun.*, 13 (1), 727. <https://doi.org/10.1038/s41467-022-28283-y>.

Taguchi, M., 2010: Observed connection of the stratospheric quasi-biennial oscillation with El Niño–Southern Oscillation in radiosonde data. *J. Geophys. Res. Atmos.*, 115 (D18). <https://doi.org/10.1029/2010JD014325>.

Thompson, D. W., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, 13 (5), 1000–1016. [https://doi.org/10.1175/1520-0442\(2000\)013<1000:AMITEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2).

Wallace, J. M., R. L. Panetta, and J. Estberg, 1993: Representation of the equatorial stratospheric quasi-biennial oscillation in EOF phase space. *J. Atmos. Sci.*, 50 (12), 1751–1762.  
[https://doi.org/10.1175/1520-0469\(1993\)050<1751:ROTESQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<1751:ROTESQ>2.0.CO;2).

Wang, G., H. H. Hendon, J. M. Arblaster, E. P. Lim, S. Abhik, and P. van Rensch, 2019: Compounding tropical and stratospheric forcing of the record low Antarctic sea-ice in 2016. *Nature Communications*, 10, 13,  
<https://doi.org/10.1038/s41467-018-07689-7>.

Wang, Y., J. Rao, Y. Lu, Z. Ju, J. Yang, and J. Luo, 2023: A revisit and comparison of the quasi-biennial oscillation (QBO) disruption events in 2015/16 and 2019/20. *Atmos. Res.*, 294, 106970.  
<https://doi.org/10.1016/j.atmosres.2023.106970>.

Watson, P. A., and L. J. Gray, 2014: How does the quasi-biennial oscillation affect the stratospheric polar vortex? *J. Atmos. Sci.*, 71 (1), 391–409. <https://doi.org/10.1175/JAS-D-13-096.1>.

Watts, M., W. Maslowski, Y. J. Lee, J. C. Kinney, and R. Osinski, 2021: A spatial evaluation of Arctic sea ice and regional limitations in CMIP6 historical simulations. *J. Climate*, 34 (15), 6399–6420.  
<https://doi.org/10.1175/JCLI-D-20-0491.1>.

White, R. H., D. S. Battisti, and G. H. Roe, 2017: Mongolian mountains matter most: Impacts of the latitude and height of Asian orography on Pacific wintertime atmospheric circulation. *J. Climate*, 30 (11), 4065–4082.  
<https://doi.org/10.1175/JCLI-D-16-0401.1>.

White, I., C. I. Garfinkel, E. P. Gerber, M. Jucker, V. Aquila, and L. D. Oman, 2019: The downward influence of sudden stratospheric warmings: Association with tropospheric precursors. *J. Climate*, 32 (1), 85–108.  
<https://doi.org/10.1175/JCLI-D-18-0053.1>.

Wilcox, L. J., and A. J. Charlton-Perez, 2013: Final warming of the Southern Hemisphere polar vortex in high- and low-top CMIP5 models. *J. Geophys. Res. Atmos.*, 118, 2535–2546, <https://doi.org/10.1002/jgrd.50254>.

Xu, M., J. A. Screen, W. Tian, J. Zhang, C. Zhang, and H. Yu, 2024: Influence of regional sea ice loss on the Arctic stratospheric polar vortex. *J. Geophys. Res. Atmos.*, 129 (15), e2023JD040571.  
<https://doi.org/10.1029/2023JD040571>.

Yamashita, Y., H. Naoe, M. Inoue, and M. Takahashi, 2018: Response of the Southern Hemisphere atmosphere to the stratospheric equatorial quasi-biennial oscillation (QBO) from winter to early summer. *J. Meteor. Soc. Japan*, 96 (6), 587–600. <https://doi.org/10.2151/jmsj.2018-057>.

Yamazaki, K., T. Nakamura, J. Ukita, and K. Hoshi, 2020: A tropospheric pathway of the stratospheric quasi-biennial oscillation (QBO) impact on the boreal winter polar vortex. *Atmos. Chem. Phys.*, 20 (8), 5111–5127.  
<https://doi.org/10.5194/acp-20-5111-2020>.

Zhou, T., K. J. DallaSanta, C. Orbe, D. H. Rind, J. A. Jonas, L. Nazarenko, G. A. Schmidt, and G. Russell, 2024: Exploring the ENSO modulation of the QBO periods with GISS E2.2 models. *Atmos. Chem. Phys.*, 24 (1), 509–532.  
<https://doi.org/10.5194/acp-24-509-2024>.