Changes in Stratospheric Dynamics Simulated by the ECEarth Model From CMIP5 to CMIP6

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS003756

Changes in Stratospheric Dynamics Simulated by the EC-Earth Model From CMIP5 to CMIP6

F. Serva[1], B. Christiansen[2], P. Davini[3], J. von Hardenberg[3,4], G. van den Oord[5], T. J. Reerink[6], K. Wyser[7], and S. Yang[2]

[1]Consiglio Nazionale delle Ricerche, Institute of Marine Sciences (CNR-ISMAR), Rome, Italy, [2]Danish Meteorological Institute (DMI), Copenhagen, Denmark, [3]Consiglio Nazionale delle Ricerche, Institute of Atmospheric Sciences and Climate (CNR-ISAC), Torino, Italy, [4]Politecnico di Torino, Torino, Italy, [5]Netherlands eScience Center, Amsterdam, The Netherlands, [6]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands, [7]Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

Key Points:

- We evaluate the stratospheric dynamics of CMIP5 and CMIP6 versions of the EC-Earth model
- Simulation of the stratosphere is in better agreement with reanalyses in the EC-Earth CMIP6 configuration
- Biases in the upper atmosphere of the CMIP6 model version are quantified with DynVarMIP diagnostics

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

F. Serva,
federico.serva@terrarum.eu

Citation:

Serva, F., Christiansen, B., Davini, P., von Hardenberg, J., van den Oord, G., Reerink, T. J., et al. (2024). Changes in stratospheric dynamics simulated by the EC-Earth model from CMIP5 to CMIP6. *Journal of Advances in Modeling Earth Systems*, *16*, e2023MS003756. https://doi.org/10.1029/2023MS003756

Received 6 APR 2023
Accepted 13 DEC 2023

Author Contributions:

Conceptualization: F. Serva, B. Christiansen, S. Yang
Data curation: F. Serva, P. Davini, J. von Hardenberg, K. Wyser
Formal analysis: F. Serva, B. Christiansen
Investigation: F. Serva, B. Christiansen
Methodology: F. Serva, B. Christiansen, S. Yang
Software: F. Serva, B. Christiansen, P. Davini, J. von Hardenberg, G. van den Oord, T. J. Reerink, K. Wyser, S. Yang
Validation: F. Serva

Abstract The simulated stratospheric dynamics have been improved compared to previous generations in many climate models taking part in the Coupled Model Intercomparison Project Phase 6 (CMIP6). This was achieved by going from low to high-top configurations, that is, increasing the atmospheric vertical resolution, raising the model lid height and including parameterization schemes, such as non-orographic gravity wave drag (GWD), to simulate small-scale processes. This also applies to the EC-Earth model, for which comprehensive analysis is now possible thanks to outputs archived within the Dynamics and Variability Model Intercomparison Project (DynVarMIP). Here we analyze atmosphere-only simulations and evaluate the stratospheric dynamics of low-top CMIP5 and high-top CMIP6 configurations of EC-Earth against the MERRA-2 and ERA5 reanalyses. This allows us to identify and interpret model biases in the atmospheric component and to highlight deficiencies needed to be addressed in future model versions. We find substantial improvements in the simulation of the stratosphere in the CMIP6 configuration, both in the tropical and extratropical regions. The zonal wind momentum budget in the tropical stratosphere is much improved in the high-top configuration. Seasonal variability of both models is qualitatively similar to the reanalysis, but for the low-top model this is likely due to compensating errors, as its intraseasonal variability is less realistic. The non-orographic GWD forcing is weak in the high-top model and this limits the realism of the tropical stratosphere variability. We point to persisting issues in the stratosphere of EC-Earth that should be tackled to improve hemispheric-scale circulation and variability.

Plain Language Summary A realistic representation of the stratosphere in climate models is important since changes in this layer of the atmosphere have significant impacts on surface weather and climate. Current models are able to simulate this region but they often struggle to represent it realistically, mostly due to computational limitations and overly simplified process representation. In this work we document how the simulation of the stratosphere has improved in the EC-Earth model between phase 5 and 6 of the Coupled Model Intercomparison Project (CMIP). This analysis is facilitated by the availability of outputs prepared for these simulations and two recent reanalysis products, which we use as reference. We find substantial changes between the two model versions, with the CMIP6 configuration in better agreement with reanalysis results. The extended diagnostics help us by indicating possible developments to improve the fidelity of the model.

1. Introduction

A realistic representation of stratospheric processes is currently recognized as an important benchmark for comprehensive climate models, since the relevance of the coupling between the stratosphere and the troposphere is well established (Baldwin & Dunkerton, 2001; Christiansen, 2001; Haynes et al., 2021; Thompson et al., 2005). In particular, signatures of stratosphere-troposphere coupling have been identified in boreal winter weather (Butler et al., 2019), decadal oceanic variability in the Atlantic (Reichler et al., 2012), as well as in the occurrence of extreme events globally (Domeisen & Butler, 2020). It is therefore important to include a realistic simulation of the stratosphere in comprehensive climate simulations.

Atmospheric general circulation models (AGCMs) can be classified as low or high-top depending on the model lid height, with the boundary between the two classes nominally set at 1 hPa or 48 km (Charlton-Perez et al., 2013). High-top models generally also benefit from a finer vertical discretization and more physically

Visualization: B. Christiansen
Writing – original draft: F. Serva, B. Christiansen, P. Davini, J. von Hardenberg, G. van den Oord, T. J. Reerink, K. Wyser, S. Yang

consistent parameterizations for dynamical and radiative processes in the upper atmosphere (Geller et al., 2013; Keeble et al., 2021). The Coupled Model Intercomparison Project (CMIP) Phase 6 (Eyring et al., 2016) includes more models with a well-resolved stratosphere compared to previous phases (Richter et al., 2020). Despite the uncertainty in projected changes, a multimodel ensemble approach allows us to better understand the variability of the stratosphere as well as its response to forcings (Garcia, 2021). To further shed light onto stratospheric processes, CMIP6 included the Dynamics and Variability model Intercomparison Project (Gerber & Manzini, 2016, DynVarMIP), an initiative which specifically aimed for a set of additional diagnostics useful for stratospheric studies. Some contributing modeling centers benefitted from the experience from previous intercomparison exercises, such as the Chemistry-Climate Model intercomparison, (Morgenstern et al., 2017, CCMi), but additional technical efforts required for DynVarMIP limited wider participation so far. At the time of writing, less than 10 CMIP6 models provided the outputs required for the analysis presented here.

In this work we document the changes between CMIP phases of the atmospheric component of EC-Earth in the configuration used for most experiments, by considering atmosphere-only simulations and taking advantage of the implementation of the DynVarMIP diagnostics. While Earth system models are needed to estimate the full effects of climate forcings, analyses of simulations following the Atmospheric Model Intercomparison Project (AMIP) protocol (Gates et al., 1999) are useful to quantify the impacts of atmospheric model developments (Garcia & Richter, 2019; Lu et al., 2020; Richter et al., 2014). We also describe how DynVarMIP additional variables can be derived from recent reanalysis data.

Model improvements carried out since CMIP5 and the vertical extension of the domain increased the realism of the EC-Earth variability in the stratosphere (Döscher et al., 2022), but assessments of the modeled stratospheric dynamics have been fairly limited. Previous works (Palmeiro et al., 2020, 2022) focused on the climatological and seasonal variability of recent versions of the EC-Earth model in the northern hemisphere (NH) winter stratosphere under fixed boundary conditions. They found that the wave forcing of the extratropical boreal stratosphere and the representation of the equatorial stratospheric variability are improved in the high-top version compared to the low-top configuration. Here we expand upon earlier works by considering both extratropical sectors and the tropics up to the stratopause, and also evaluate the intraseasonal variability in the polar regions, as it did not receive much attention before. Furthermore, analysis of the DynVarMIP diagnostics helps us understand model biases which should be accounted for when analyzing CMIP projections, and how they could be reduced by means of further developments.

This paper is organized as follows: in Section 2, we introduce the model and reanalysis data; in Section 3.1 we discuss the model climatological state and large-scale biases; in Section 3.2 we focus on the equatorial stratosphere variability; in Section 3.3 we discuss the intraseasonal variability of in the extratropical stratosphere. Finally in Section 4 we provide concluding remarks and perspectives for further research.

2. Material and Methods

We consider two versions of the EC-Earth model, namely v2.3 (Hazeleger et al., 2012) used to carry out the CMIP5 experiments (For more details, also see https://ec-earth.org/wp-content/uploads/sites/10/2022/09/ec-earth-science-and-implementation-plan-2014.pdf (last accessed, March 2023)) and v3.3 (Döscher et al., 2022) adopted in CMIP6. The EC-Earth model is developed by an European consortium, with the Integrated Forecast System (IFS) atmospheric model of the European Center for Medium Range Weather Forecasts (ECMWF) as the atmospheric component.

Relevant changes in the atmosphere between the CMIP5 (based on IFS cy31r1) and CMIP6 (based on IFS cy36r4) versions are a finer resolution of the spectral triangular grid (T159 and T255 with linear truncation, approximately 125 and 80 km respectively) and more model levels (respectively 62 with top at 5 hPa and 91 with top at 0.01 hPa). The L62 configuration has only one level above 10 hPa and 10 levels above 100 hPa, while these numbers increase to 18 and 38 for L91 (see this page (https://confluence.ecmwf.int/display/UDOC/Model+level+definitions (last accessed, August 2023)) for details on the vertical grids). The CMIP5 version of EC-Earth (v2.3, hereafter EC-Earth2) can be classified as a low-top model while the CMIP6 (v3.3, indicated as EC-Earth3) belongs to the high-top model category (Charlton-Perez et al., 2013).

Physical parameterizations relevant for the stratosphere were also changed between the two model versions. Both use the simplified GCM version of the Rapid Radiative Transfer Model scheme (Mlawer et al., 1997), which was

updated between the two IFS cycles, including improvements in the treatment of cloud effects on radiation and the shortwave radiation scheme (Morcrette et al., 2008). Furthermore, the microphysics of the CMIP6 model was updated, with a more complex representation of cloud condensates and interaction between aerosols and liquid clouds (Wyser et al., 2020). Starting with v3.1, a parameterization for non-orographic gravity wave drag (NOGWD) based on Orr et al. (2010) was included in EC-Earth, which is important for stratospheric dynamics (Davini et al., 2017). The parameterization for OGWD is instead unchanged between the two model versions, and Rayleigh friction is applied for levels above the middle stratosphere.

We note that these are the standard EC-Earth atmospheric configurations used for running experiments in CMIP, and that outputs for alternative versions, such as the vertically extended EC-Earth2 considered by Manzini et al. (2014), are not publicly available. Other options are available for both CMIP5 and CMIP6 versions, including coupling with interactive ocean and sea ice, land and chemistry: however, in this work we will focus our attention on AMIP simulations to focus on atmospheric dynamics. In these experiments sea ice concentration and sea surface temperatures (SSTs) are prescribed from observations, and other forcings are also set according to the common AMIP protocol.

For CMIP6, priority 1 DynVarMIP variables have been published for realization r10i1p1f1 (1979–2017) which is our main focus, but we also consider five additional ensemble members (1, 3, 4, 7, 9) for which daily data on 19 pressure levels are available. Unless otherwise stated, the results shown are for r10i1p1f1. For EC-Earth2, a single realization (1977–2008) not available from the CMIP5 repository, but archived online (Serva & Wyser, 2022), is analyzed. For this experiment, vorticity and divergence were stored on model levels, but vertical pressure tendency was not. Therefore the latter is calculated with the kinematic method (Martin, 2006; O'Brien, 1970), producing good results above orography (not shown).

To quantify the effects of model development we mainly use the MERRA-2 reanalysis (Gelaro et al., 2017), which is based on a model with a realistic simulation of the stratosphere (Molod et al., 2015) and lower mesosphere (Kawatani et al., 2020). The Goddard Earth Observing System (GEOS) forecast model of MERRA-2 has approximate latitude-longitude resolution of $0.5° \times 0.625°$ and 72 hybrid levels from the surface to 0.01 hPa. Here we consider MERRA-2 6-hourly data for the period 1980–2021 on model levels. While most comparisons are against MERRA-2, as it is independent from EC-Earth (i.e., not based on the IFS), we include the ECMWF ERA5 (Hersbach et al., 2020), which has been shown to be equivalent to other reanalyses and benefits from higher vertical resolution (Long et al., 2017; Wright & Hindley, 2018). We note that the use of the IFS in ERA5 and EC-Earth may be the reason for better agreement between them, but in any case considering multiple reanalyses is useful when evaluating DynVarMIP diagnostics, as reliability of derived variables is hard to assess. Unless otherwise stated, we compare reanalyses and simulations for the common period 1980–2008. The ERA5 version of IFS has 137 model levels, an approximate latitude-longitude resolution of $0.25° \times 0.25°$ and outputs are available hourly. We retrieved 6-hourly ERA5 data on model levels for the period 1979–2021 with a reduced resolution of $0.5° \times 0.5°$ to compute the required variables. Differences among reanalyses depend on their model formulation (Diallo et al., 2021) but also on the assimilated observations, and particularly so above 1 hPa due to data sparsity (Kawatani et al., 2020).

As discussed above, we take advantage of DynVarMIP diagnostic variables to examine model biases. For a comprehensive evaluation, we derive the transformed Eulerian mean (TEM) momentum terms for both models and the reanalyses, including zonal mean tendencies from flux divergence and residual circulation. A circulation-based framework, highlighting the wave-driven residual circulation (Andrews et al., 1987), is useful to interpret climatological model biases in the stratosphere (Kobayashi & Iwasaki, 2016). Following Gerber and Manzini (2016), all quantities are computed using native horizontal and vertical model grids and 6-hourly frequency data. Values at pressure levels below orography are extrapolated. The calculation requires seven zonal mean fields: dynamics terms (zonal, meridional and vertical wind, temperature, i.e. u, v, w, T) and zonal mean fluxes ($\overline{u'v'}, \overline{u'w'}, \overline{v'T'}$, zonal means are indicated with overbars and departures from zonal means with primes). Derivatives are calculated with finite differences and constants are set following Gerber and Manzini (2016). More details on the implementation of the TEM diagnostics are provided in the Appendix A. For the zonal wind tendency, non-orographic GWD is not included in EC-Earth2, both orographic and non-orographic terms are archived for EC-Earth3, while MERRA-2 provides net tendency from gravity wave schemes and ERA5 the tendency due to all parameterizations. Additional variables analyzed from models and reanalyses are daily sea level pressure and geopotential height.

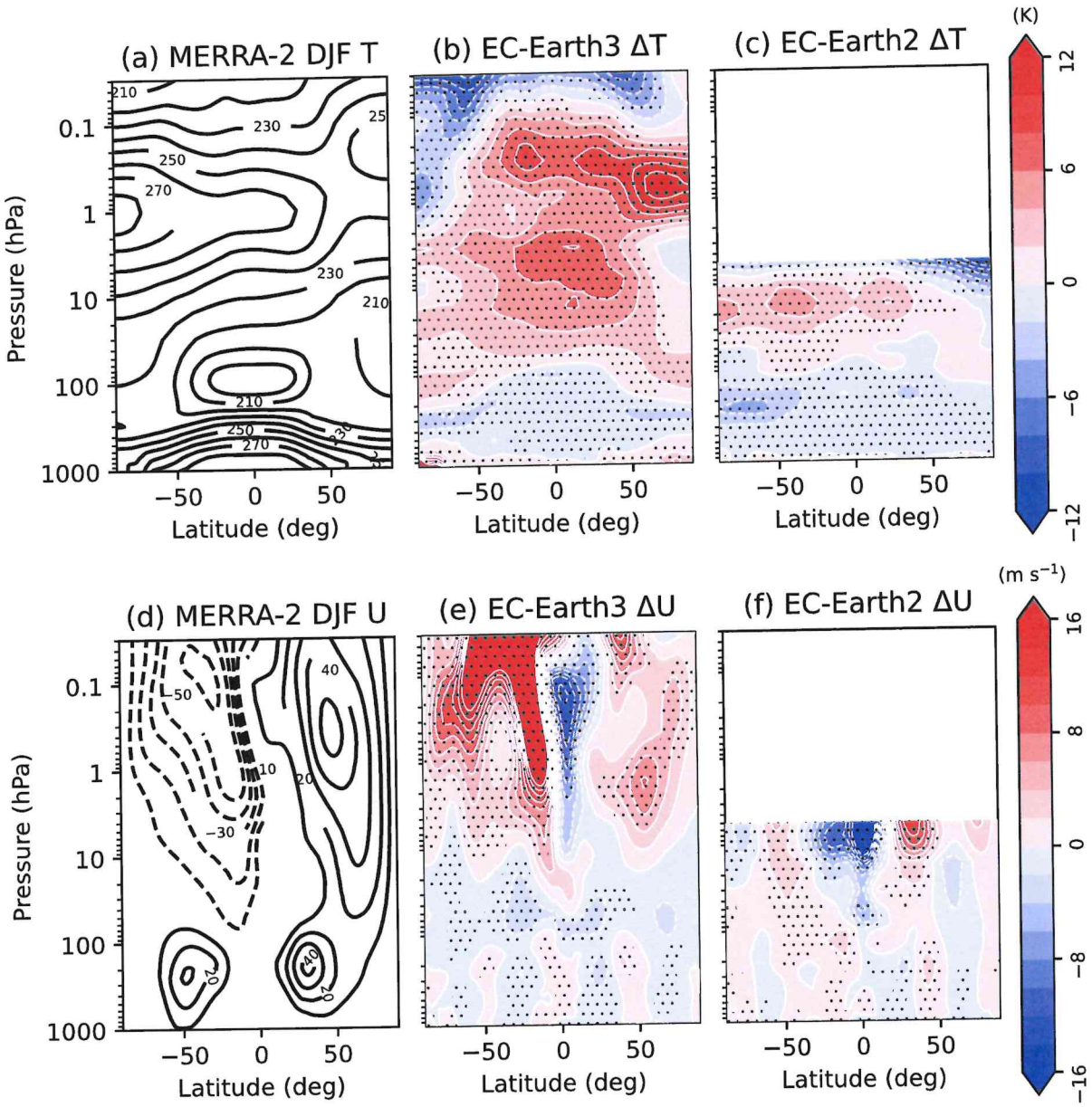

Figure 1. Zonal mean temperature and zonal wind as a function of latitude and pressure in DJF for (a, d) MERRA-2 and differences (b, e) EC-Earth3–MERRA-2 and (c,f) EC-Earth2–MERRA-2. Stippling denotes significance at the 5% level according to a *t*-test.

3. Results and Discussions

3.1. Global Zonal Mean Conditions

We start the model evaluation by considering zonal mean averages of key quantities, such as zonal wind and temperature (Figure 1). Following Palmeiro et al. (2022), we start the discussion from the boreal winter (December-January-February, DJF) climatology. Plots for the June-July-August (JJA) season are reported in Figure S1 in Supporting Information S1. To compare with previous works, the degrees of freedom for the statistical significance are calculated from the number of years. The decorrelation time scale exceeds 1 year at high latitudes and in the upper stratosphere, and reducing the number of degrees of freedom has little effect (not shown). We have verified that the climatology of r10i1p1f1 does not differ from the selected 5-member ensemble mean of EC-Earth3, with small regions of significant differences in the winter stratosphere (not shown), likely due to sampling uncertainty.

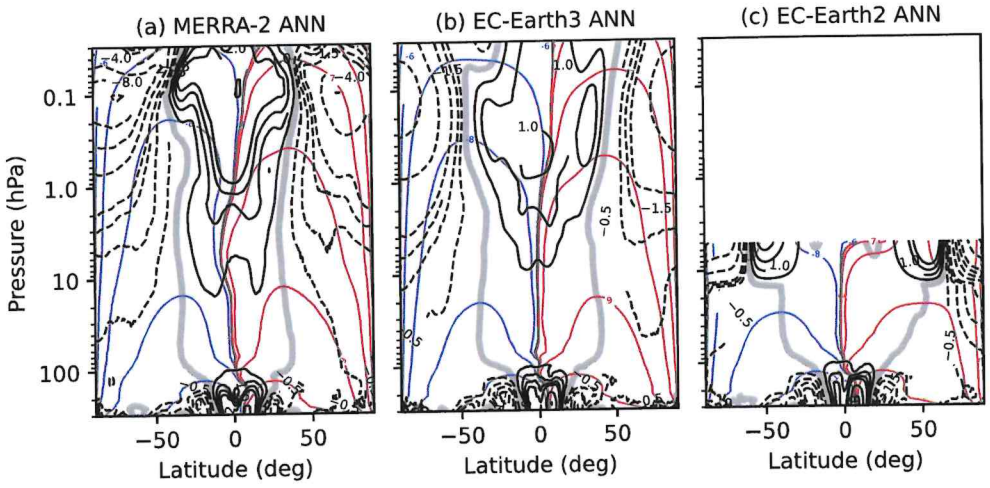

Figure 2. Annual mean residual circulation (blue and red contours, units kg s^{-1} and logarithmic levels) and residual vertical velocity (black contours, negative dashed, bold gray contour for the tropical pipe, units mm s^{-1}) for (a) MERRA-2, (b) EC-Earth3 and (c) EC-Earth2.

For zonal mean temperature in DJF (Figure 1, top row), both versions of EC-Earth show a cooler troposphere (between ~850 hPa and the tropopause) than MERRA-2, and while EC-Earth3 displays a warmer bias above, a cold bias is present over the Arctic in EC-Earth2. The meridional structure of the stratospheric temperature bias is different depending on the model domain extent. While in the lower stratosphere the magnitude of the bias is similar (~5K), the cold bias around 10 hPa in the winter stratosphere of EC-Earth2 is reduced in EC-Earth3. However, a warm bias is seen in EC-Earth3 in the winter mid-stratosphere (around 1 hPa). This warm bias may be related with ozone radiative processes as mentioned by Palmeiro et al. (2022), while a cold bias possibly related with the residual circulation characterize the summer upper stratosphere. Above 1 hPa dynamical processes may be more relevant as discussed by Orr et al. (2010) and further examined below. The pattern of the biases is in good agreement with Palmeiro et al. (2022), who analyzed the NH climatology for a centennial simulation and a different reanalysis reference.

In Figure 1 (bottom row) we also report the zonal mean zonal wind climatology and the model biases in DJF. The significant easterly bias in the tropics around 10 hPa of EC-Earth2 is absent in EC-Earth3, as its zonal wind variability is more realistic, but winds are too easterly in EC-Earth3 between 0.1 and 5 hPa. EC-Earth2 overestimates the strength of the polar night jet, especially closer to the model upper lid at 40°N. This bias is reduced in magnitude in EC-Earth3, which in turns has a significant westerly bias in the summer stratosphere above 10 hPa and southward of 10°S, linked to an overestimation of the intensity of the easterly jet. As discussed below and already reported in Döscher et al. (2022), the bias in the tropical upper stratosphere of EC-Earth3 is due to unrealistic seasonal variability. It should however be noted that reanalyses are not well constrained at high altitudes (Kawatani et al., 2020). Tropospheric zonal wind biases are relatively small in both model versions. Interestingly in both model versions the summer easterlies are weaker than in the reanalysis, partly due to numerical damping applied in upper model layers (Polichtchouk et al., 2017), but in EC-Earth3 also due to the temperature bias (Figure 1b) and to thermal wind balance.

The vertical component of the residual circulation ($\overline{w}^*$) and residual streamfunction are shown in Figure 2. In EC-Earth2 (Figure 2c) the annual mean width of the tropical upwelling region (Seviour et al., 2012, i.e., the zero contour line) gets latitudinally too large already above ~50 hPa, extending into the extratropics near the model top. Strong downward motions are seen poleward of 50° in both hemispheres, likely as an artifact of the model top placement. EC-Earth3 (Figure 2b) underestimates the upwelling strength in the tropical stratosphere (from 10 hPa), and residual circulation is shifted vertically (e.g., compare the height of the blue contours in the SH) with respect to reanalysis (Figure 2a), as found for other CMIP6 models (Abalos et al., 2021). The tropical pipe structure is more realistic than in EC-Earth2, but wider than in the reanalysis. Nevertheless, improvements from the CMIP5 to the CMIP6 versions of EC-Earth are evident. Consistent with other model studies (Kawatani

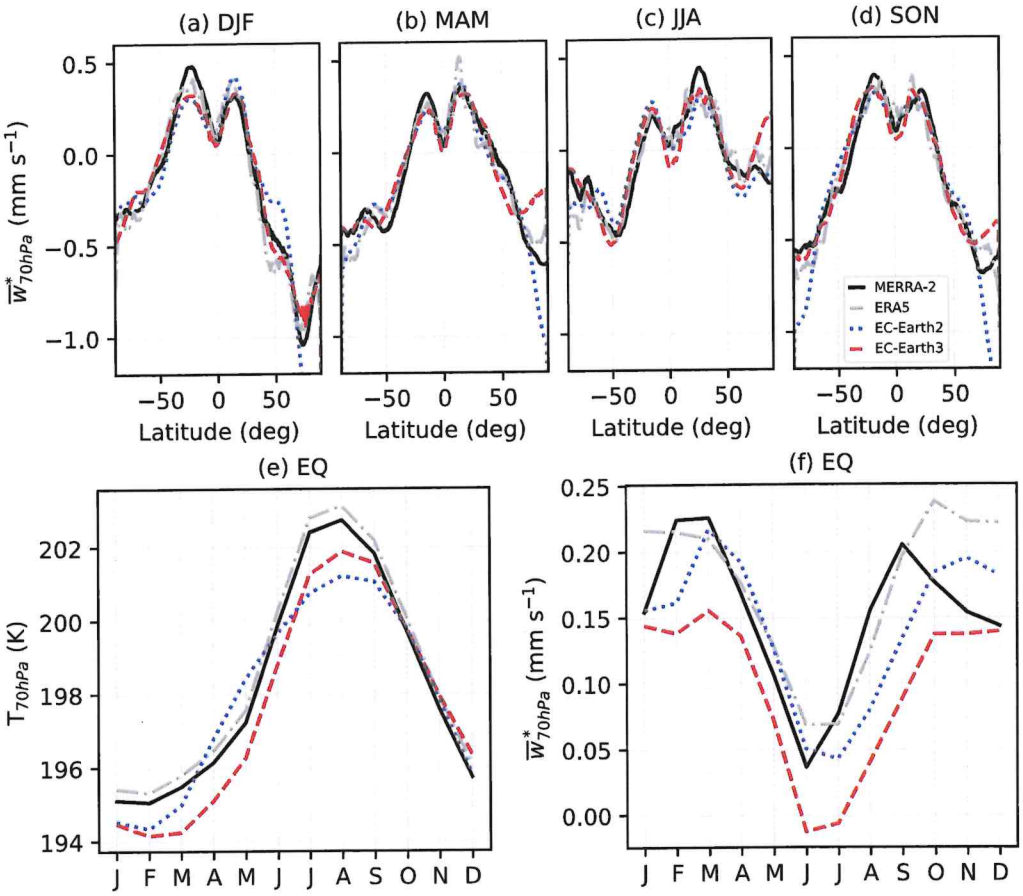

Figure 3. Residual vertical velocity at 70 hPa as a function of latitude for EC-Earth, MERRA-2 and ERA5 for the different seasons (a–d). Annual cycles of (e) temperature and (f) residual vertical velocity in the equatorial region (mean over 10°S–10°N) and at the same level.

et al., 2019; Richter et al., 2014), circulation changes can be attributed to the introduction of non-orographic GWD in EC-Earth3. As discussed by Polichtchouk et al. (2017) for IFS, changes between these EC-Earth versions are expected as adjusting the NOGWD scheme and further increasing resolution are beneficial to reduce stratospheric circulation biases.

An important diagnostic for stratosphere-troposphere exchange is the mean residual vertical velocity at 70 hPa (Holton et al., 1995), which is frequently used for model evaluation, as it is related with the intensity of the shallow branch of the Brewer-Dobson circulation (BDC). In Figures 3a–3d we report its seasonal averages for EC-Earth and the two reanalyses. We can observe that the seasonality of the model is realistic, but also that EC-Earth underestimates the upwelling strength in the summer hemisphere. The maximum occurs in boreal late fall and early winter, as found for other CMIP6 models (Abalos et al., 2021). Note that in JJA there is actually downwelling between 0° and 10°N for EC-Earth3, something which is not visible when averaging over tropical latitudes (Yoshida et al., 2018). This feature was already identified in chemistry-climate models (SPARC, 2010), and may be attributed to effects from GWD forcing (Diallo et al., 2021) or be related to resolved extratropical wave pumping (Grise & Thompson, 2013). Meridional profiles for ERA5 agree with those for MERRA-2 but show larger variability, likely due to its higher spatial and vertical resolution, highlighting how the representation of dynamical processes in the stratosphere differs across reanalyses. EC-Earth2 simulates a very strong downwelling at polar latitudes in the equinoctial seasons. Focusing on the equatorial region (Figures 3e and 3f), models appear to be colder than the reanalyses throughout the whole year apart from boreal autumn, even if residual upwelling is typically weaker than in the reanalyses. Upwelling goes to zero in EC-Earth3 in June and July. Similar plots for the extratropical latitudes are shown in Figure S2 in Supporting Information S1.

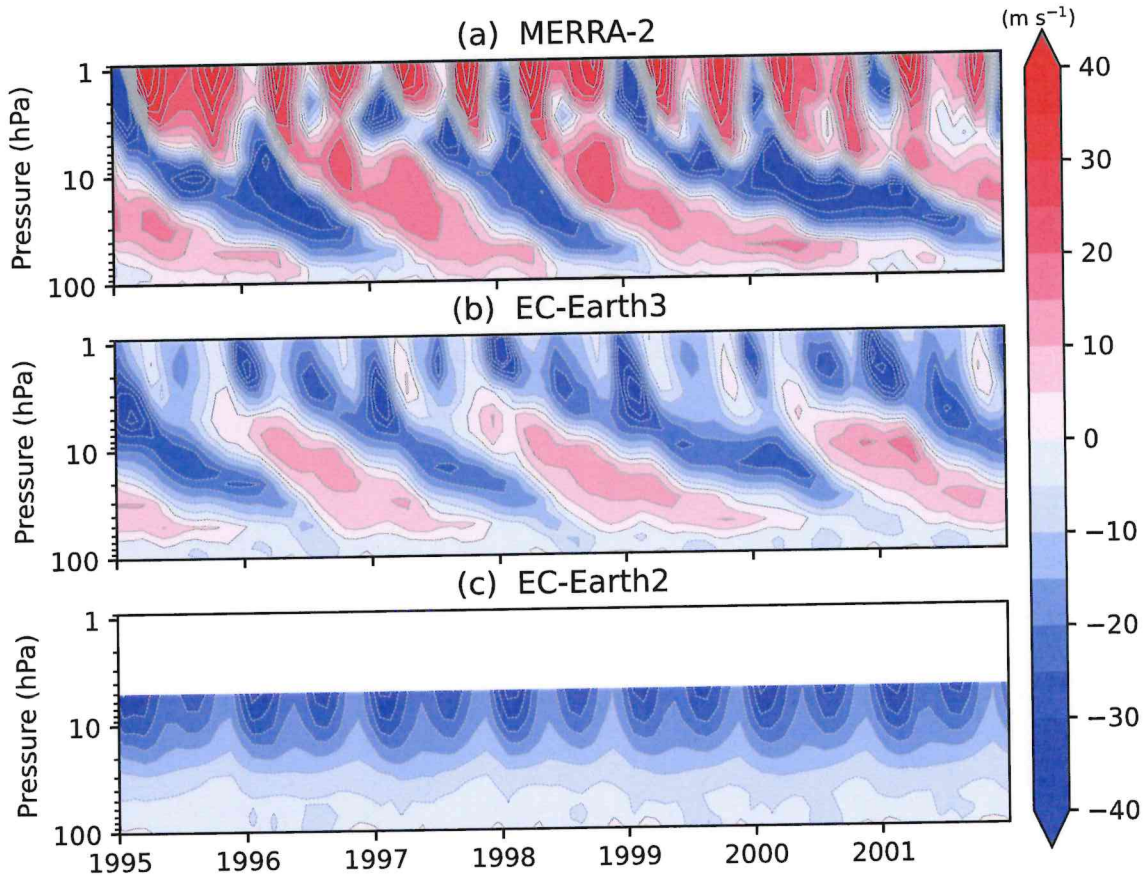

Figure 4. Zonal mean zonal wind in the equatorial stratosphere as a function of time and pressure for (a) MERRA-2, (b) EC-Earth3 and (c) EC-Earth2.

3.2. Forcing of the Tropical Zonal Winds

The simulation of a realistic quasi-biennial oscillation (QBO) is an important testbed for high-top models (Bushell et al., 2022). A general evaluation of the realism of the QBO in EC-Earth3 was presented in Döscher et al. (2022), and that in a slightly early version, EC-Earth v3.1, was considered in Christiansen et al. (2016) and Serva et al. (2020) but in this section the DynVarMIP diagnostics are analyzed for additional insights. The zonal mean zonal wind at the equator (averaged between 3°S and 3°N) is shown in Figure 4 for about three QBO cycles (i.e., from 1995 to 2002). Results for EC-Earth3 show clear improvements over EC-Earth2, with a QBO-like oscillation between 70 and 5 hPa, thanks to the introduction of a NOGWD parameterization (Orr et al., 2010), further tuned for EC-Earth3 (Davini et al., 2017), but with GW characteristics not linked to their sources (such as convection). As discussed by Döscher et al. (2022) the QBO descent in EC-Earth3 is faster than in observations. On top on that, we can note that the amplitude of the oscillation in EC-Earth3 is smaller than in MERRA-2, especially in the lower stratosphere. This amplitude bias is common to other QBO-resolving models (Bushell et al., 2022) and it has implications for the influence that the QBO can exert on the tropical tropopause layer and the surface (Gray et al., 2018; Serva et al., 2022) in this model. As shown by Döscher et al. (2022), above 5 hPa the zonal wind is mostly easterly in EC-Earth3, with no clear semiannual oscillation (SAO) in the upper stratosphere.

To better visualize the differences in variability between the models and reanalysis, time series for the 1 hPa (in the SAO region), 30 hPa (in the QBO region) and 70 hPa (in the lower stratosphere) are reported in Figure 5. At 1 hPa there is some seasonality in the easterlies of EC-Earth3 (with a semi-annual modulation of their intensity), while MERRA-2 shows prevailing westerlies and alternating easterlies with variable intensity. At 30 hPa, improvements are clear (Richter et al., 2020) and results for EC-Earth3 are similar to MERRA-2, despite a smaller amplitude, while an easterly regime is seen in EC-Earth2. In the lower stratosphere at 70 hPa EC-Earth3 clearly underestimates the amplitude of the QBO.

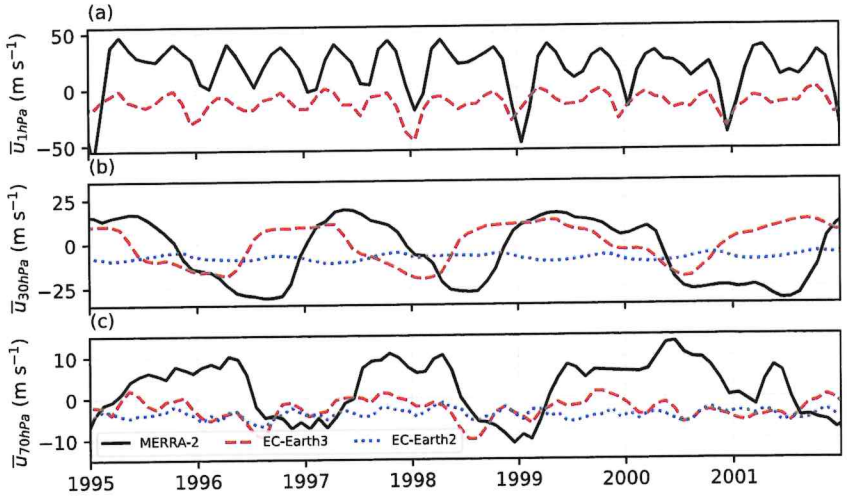

Figure 5. Time series of the equatorial zonal mean zonal wind at pressure levels (a) 1, (b) 30 and (c) 70 hPa.

The availability of momentum budget diagnostics allows us to better understand the source of equatorial wind biases (Kawatani et al., 2011). Time series of momentum budget terms at SAO and QBO heights for reanalyses and EC-Earth3 are shown in Figures 6 and 7, respectively. For the 1 hPa pressure level, we find that GWD in EC-Earth3 (essentially from the non-orographic scheme at the equator) is much smaller than in MERRA-2, with most easterly forcing provided by the advective term (mostly by meridional wind, not shown), and weaker westerly acceleration provided by the Eliassen-Palm (EP) flux divergence (Smith et al., 2022). For example, in 1998 and

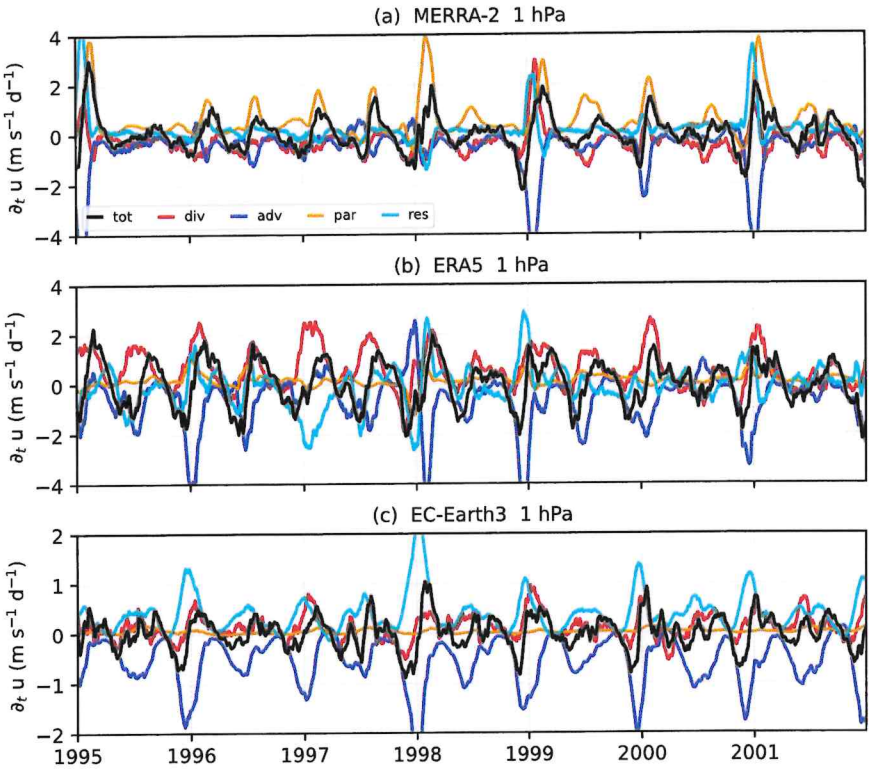

Figure 6. Time series of the equatorial zonal mean zonal wind total tendency (*tot*, black), acceleration by EP flux divergence (*div*, red), advective term (*adv*, blue), parameterized forcing (*par*, orange) and the residual term (*res*, light blue) needed to close the budget at 1 hPa. All terms are smoothed in time with a 30 day moving average. Note the different *y*-axis extent between the plots.

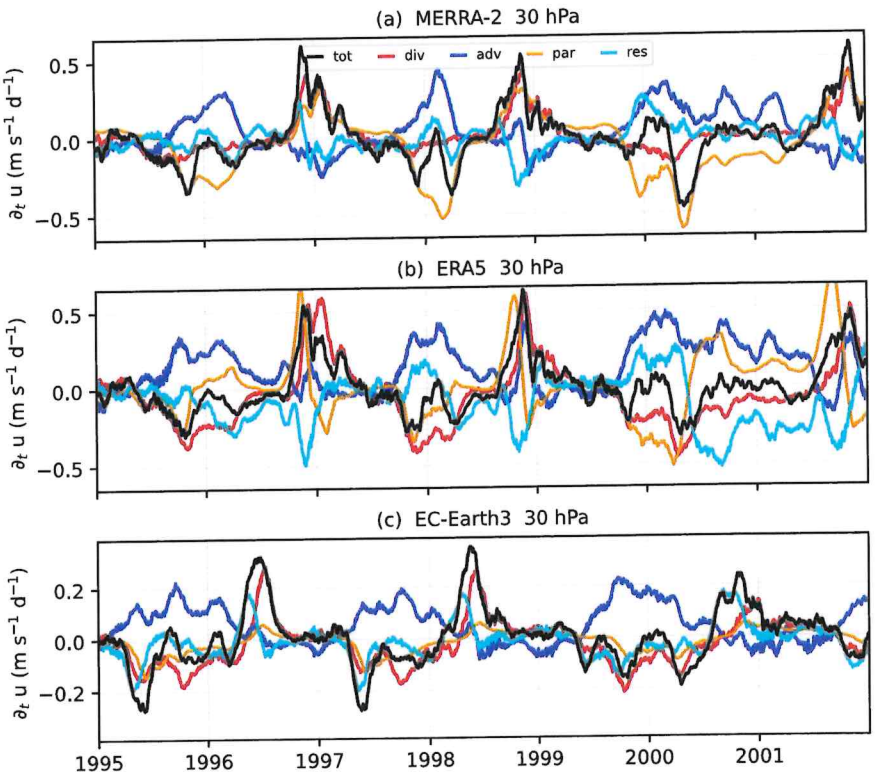

Figure 7. As in Figure 6, but for the 30 hPa pressure level. Note the different *y*-axis extent.

2001, during westerly SAO transitions (Figure 5a), GWD acceleration and resolved forcing are dominant in MERRA-2 and ERA5, respectively, while the residual term is large in EC-Earth3 (but winds do not turn to westerly). The residual forcing, calculated as the total tendency ($\partial_t u$) minus parameterized, advection and divergence terms, is always westerly in EC-Earth3, due to the Rayleigh drag applied in the upper model levels. For MERRA-2 at 1 hPa GWD provides most westerly acceleration, while advection and flux divergence terms are typically opposite. In the budget of ERA5 the resolved flux divergence term is more important, probably due to the finer spatial resolution of the model. Note that in reanalyses the residual term also includes contribution from the data assimilation procedure.

Near the QBO core (i.e., at 30 hPa, Figure 7), again tendency terms are smaller in the model than in MERRA-2 and ERA5. While the GWD is important (exceeding 0.5 m s^{-1} d^{-1}) in both easterly and westerly phases in reanalyses (Coy et al., 2016), it is comparatively weaker (below 0.1 m s^{-1} d^{-1}) in EC-Earth3. A westerly transition occurs in 1998/99 (Figure 5b), and it is evident that the contribution from GWD is smaller than both reanalyses (Figures 7a and 7b). In both reanalysis and EC-Earth3 the advection term (due to both vertical and meridional components, not shown) is opposed by resolved waves flux divergence, with residual terms systematically getting larger before the westerly peak of the total tendency.

To summarize, tendencies are generally larger throughout the stratosphere in MERRA-2 and ERA5 than in EC-Earth, and in the model the contribution of GWD is particularly weak. The residual term also provides a significant contribution in the tropical wind budget, as also discussed by Pahlavan et al. (2021) and Giorgetta et al. (2022), but its nature is not well understood. Note that the non-orographic GW schemes in the IFS and the MERRA-2 model are different: while in the latter the GW launch momentum flux is enhanced in the tropical band (Molod et al., 2015), in IFS it is gradually reduced at equator with increasing spatial resolution (Palmeiro et al., 2022). While this has shown to improve the realism of the extratropical stratosphere in IFS (Johnson et al., 2019), the insufficient wave forcing negatively affects the tropical zonal wind variability in EC-Earth3. The relative importance of the different forcing terms in EC-Earth3, with the residual term being larger than the tendency from NOGWD, remains the same throughout the depth of the QBO (not shown).

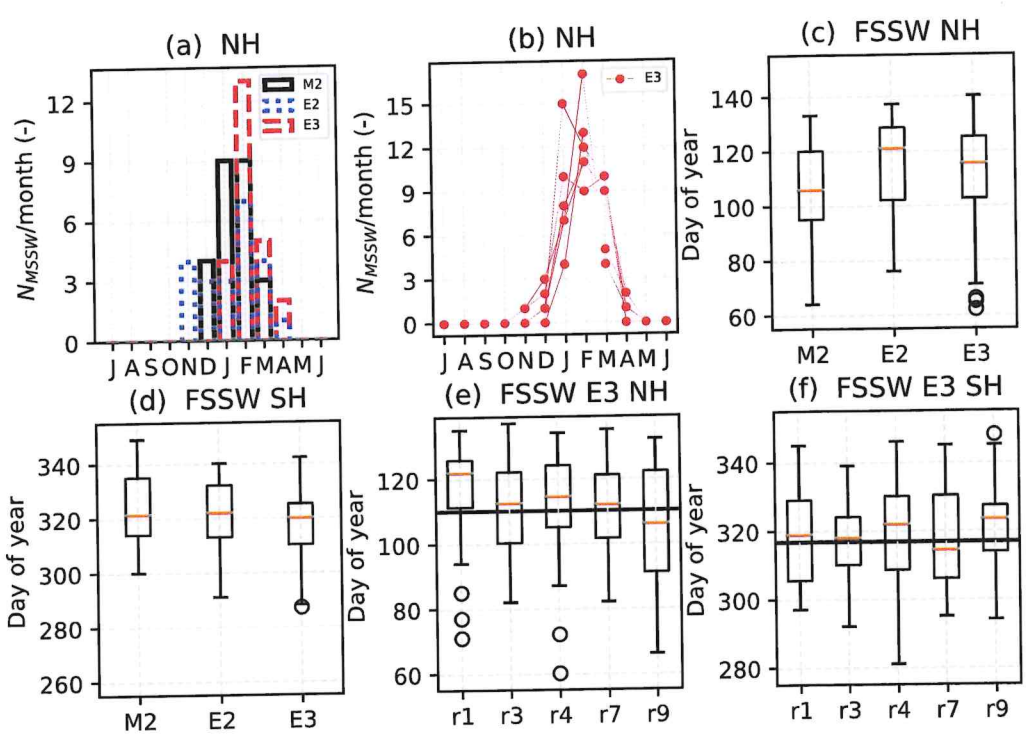

Figure 8. Number of major SSW (MSSW) as a function of the calendar month in the northern hemisphere (NH) for (a) for MERRA-2 (M2), EC-Earth2 (E2) and EC-Earth3 (E3), and (b) number of events in the EC-Earth3 five-member ensemble (note that dots are overlapping for some months). Boxplots of the final sudden stratospheric warming (FSSW) date in MERRA-2, EC-Earth2 and EC-Earth3 in the (c) NH and (d) southern hemisphere (SH). Outliers are indicated with circles. Boxplots for FSSW dates for all EC-Earth3 AMIP realizations in (e) the NH and (f) the SH, with dates for r10i1p1f1 shown with the horizontal line. Note that the EC-Earth2 record is ~10 years shorter than other data sets.

3.3. Wintertime Variability in the Extratropics

As discussed in the previous sections, model developments helped to reduce the biases of EC-Earth, even if the wave-driven circulations are weaker than the reanalysis. The availability of diagnostic outputs makes it possible to better study the intraseasonal variability associated with polar vortex anomalies (Baldwin et al., 2021), manifesting as sudden stratospheric warmings (SSWs). These rapid changes of the extratropical stratospheric flow can alter tropospheric weather (Domeisen & Butler, 2020) and their effect can influence the global climate (Yoshida & Mizuta, 2021). SSWs are generally classified as major, minor and final events. While final warmings occur every year marking the passage to the springtime easterly circulation, significant perturbations to the westerly flow can occur in all winter months. In the NH and at 10 hPa, major SSWs are conventionally defined by the reversal of the zonal mean zonal wind at 60°N, while in minor SSWs winds remain westerly but the thermal gradient between 60° and 90°N changes sign and gets larger than 25 K (Butler et al., 2015). These definitions were based on studies of the NH extratropics, and since no agreed definition is available for the SH (Jucker et al., 2021; Rao et al., 2019), we apply the same criterion but considering 60° and 90°S for the austral hemisphere. More complex methods for evaluating SSW in models were also proposed (Kim et al., 2017) and could be used to extend our algorithm, even if their results were not much different from standard methods. Note that we only consider full winter seasons, for example, NH winter 1979/80 for MERRA-2 is discarded in our method as incomplete. The annual cycle of extratropical zonal mean temperature and zonal wind in the stratosphere at 10 hPa are shown in Figure S3 in Supporting Information S1, and similar to Palmeiro et al. (2020) we find that in the NH the vortex intensity is slightly larger in early winter in the high-top model compared to reanalysis, while the SH vortex is generally stronger in the low-top model. The NH vortex is more persistent in the low-top configuration, as also seen in Figure 8c. We note that the experimental design of Palmeiro et al. (2020) differs from ours, as they considered centennial simulations with climatological forcings and v3.1 of EC-Earth.

For each data set, we identified major, minor and final warmings in both hemispheres. The calculation basically follows Charlton and Polvani (2007), and results are validated by comparing reanalysis results with previous

works (Butler & Domeisen, 2021; Butler et al., 2017, 2019; Hirano et al., 2021). Dates for both major and final warmings in the NH for MERRA-2 are generally in good agreement (within 1–2 days) with previous reanalysis studies, including more recent events (Rao et al., 2018, 2019). Results in this section are shown for the full period available from reanalysis (1980–2021) and simulations (1977–2008 and 1979–2017 for EC-Earth2 and EC-Earth3, respectively).

Figure 8a shows the number of major SSWs (MSSWs) as a function of the month for the northern hemisphere. The monthly distribution in the NH (Figure 8a) indicates most events occur in February in both MERRA-2 and EC-Earth3. Differently from Palmeiro et al. (2020), we find that EC-Earth2 overestimates their frequency in November and both versions in March, while previous works found a peak in December for the high-top version. The monthly SSW distribution of the models is not statistically different from that of MERRA-2 according to both a χ^2 and a bootstrapping test. The SSW seasonality may however be realization-dependent, as for the EC-Earth3 ensemble (Figure 8b) the number of events is maximized between January and March, more similar to Palmeiro et al. (2020). Such ensemble spread can arise from internal variability or sampling uncertainty, for example, associated with annular modes (Deser et al., 2012). In the SH (not shown) only one major warming is observed in the period considered, while EC-Earth2 produces more events later in winter.

The date of the final SSW (FSSW) marks the time of passage to the easterly circulation, and it is related with important processes such as the magnitude of the seasonal ozone depletion (Butler & Domeisen, 2021). Figures 8c–8f shows FSSW dates for the model and reanalysis, as well as for the EC-Earth ensemble. In the NH (Figure 8c) the vortex demise occurs between day of year 100–120 (i.e., mid-late April) for the three data sets, while for the SH (Figure 8d) the easterlies set up around day 320 (mid-November). Compared to EC-Earth2, the high-top model shows a tendency for an earlier transition to easterlies in both hemispheres. Results for r10i1p1f1 are similar to those for the other CMIP6 realizations in both hemispheres (Figures 8e and 8f). Interestingly, in these AMIP runs EC-Earth does not show systematic FSSW delays in the SH as present in other models (de la Cámara et al., 2016).

The occurrence of MSSWs is followed by a vertical descent of stratospheric anomalies towards the surface, influencing weather and climate. Composites of zonal wind anomalies (with respect to the annual cycle) during MSSW events are shown in Figures 9a–9c. The vertical descent of easterly anomalies is relatively fast (within 1 month from the central date) in both model versions, and in EC-Earth3 they dissipate quicker than in the reanalysis and EC-Earth2 in the stratosphere. In both versions some positive anomalies are seen near surface levels after the central date, indicating a less robust influence on the surface climate than the reanalysis. Interestingly, the temporal evolution of the anomalies in EC-Earth2 is qualitatively similar to MERRA-2. The overly weak surface response in both EC-Earth versions is deduced from the sea level pressure composites (Figures 9d–9f), as the positive anomalies in the Arctic region are underestimated and the north Pacific sector response is inconsistent with that of MERRA-2. The high latitude positive anomalies are slightly more robust in EC-Earth3, but the Atlantic and Eurasian response is more pronounced in the low-top configuration. The spatial response in the stratosphere to MSSWs is reported in Figure S4 in Supporting Information S1, with both configurations missing the extratropical response and larger anomalies over the Arctic in the low-top configuration. Analysis of the NH minor and SH major warming events (not shown) reveals that easterly anomalies during minor SSWs are generally insignificant, and that for MSSWs in the SH the anomalies in EC-Earth2 disappear below 100 hPa, similar to the NH case (Figure 9c).

To gain understanding on the processes leading to SSWs (Martineau et al., 2018), we consider in Figure 10 the momentum budget at 30 hPa and 60°N. This pressure level, rather than 10 hPa, is chosen to avoid artefacts occurring near the model top in the low-top model configuration. Previous works such as Christiansen (2001) have shown that momentum deposition from EP flux divergence is responsible for deceleration of the winter westerlies leading to SSW events. Forcing from flux divergence is balanced by the advection of momentum from meridional residual velocity, which accelerates the vortex westerlies. These dynamics can be observed for both the MERRA-2 and the ERA5 reanalyses, for which results are very similar. The budget partitioning in EC-Earth models are qualitatively similar to reanalyses but the magnitude of individual terms is smaller, with somewhat less variability in the low-top configuration. Note that both the residual and parameterized forcings do not seem relevant in the development of MSSWs. Figure 10d shows that the EP flux divergence and residual advection terms evolve quite slowly in EC-Earth2, as for example, they do not go to zero ~5 days after the central date as in the other cases. Analysis of EP fluxes (not shown) does not indicate that wave reflection (Kodera et al., 2008) is

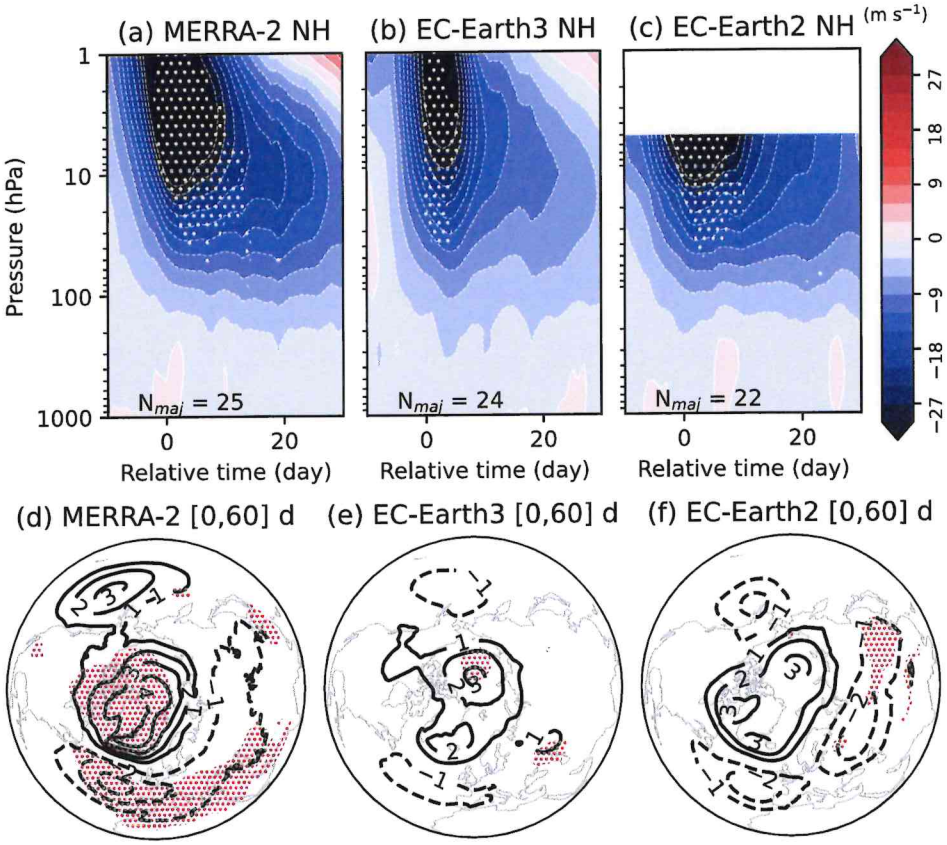

Figure 9. Time-pressure composites of the zonal mean zonal wind anomalies during northern hemisphere (NH) MSSWs in (a) MERRA-2, (b) EC-Earth3 and (c) EC-Earth2. The number of events for each data set is shown in the bottom left of each panel. Plots (d–f) report the corresponding NH sea level pressure (units hPa) composites between days 0–60 after MSSW central dates. Day zero marks the central date of the MSSW event, that is, the date of zonal wind reversal. In all plots, hatchings indicate significant differences at the 5% level based on bootstrapping of random winter dates.

leading the EP flux divergence, hinting at forcing from the lower stratosphere (Palmeiro et al., 2020). Figure S5 in Supporting Information S1 reports anomalies of the budget terms as a function of time and height in EC-Earth2 for the same location, and it shows that they change substantially in a layer below 10 hPa. The evolution of EP flux divergence and advection terms near the model top is likely spurious and may be due to issues from numerical differentiation.

A summary of major and minor warming is reported in Table 1. For mid-winter events (Table 1), despite its westerly bias in the extratropics (Figure 1f, Figures S1f and S3b, S3d in Supporting Information S1), EC-Earth2 has a higher frequency of major warmings in both hemispheres, even larger than the reanalysis, while it underestimates the frequency of minor events. Together with the different momentum budget evolution, this may imply that dynamical regimes simulated by the low-top model are not realistic (Badin & Domeisen, 2016; Kim et al., 2017; Shepherd & Shaw, 2004), thus limiting the use of such configuration for reliably estimating future stratospheric circulation changes. Even if SSW dynamics in the low and high-top models are different, the final warming dates are fairly realistic as seen in Figure 8, suggesting the presence of compensating biases.

4. Conclusions

Atmosphere-only simulations of the EC-Earth model for CMIP5 (low-top configuration, EC-Earth2) and CMIP6 (high-top configuration, EC-Earth3) phases are compared to evaluate the impacts of model improvements. The analysis focuses on the stratosphere, it is facilitated by the availability of comprehensive diagnostic outputs prepared according to the DynVarMIP protocol (Gerber & Manzini, 2016) and it uses data from the recent MERRA-2 and ERA5 reanalyses as references.

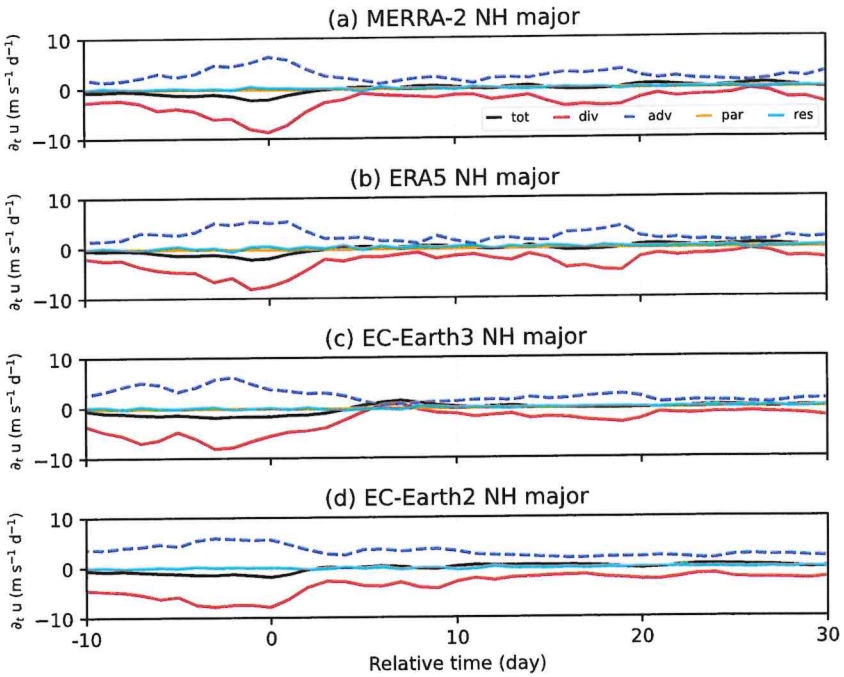

Figure 10. Time series of the momentum tendency terms at 30 hPa and 60°N during MSSWs in (a) MERRA-2, (b) ERA5, (c) EC-Earth3 and (d) EC-Earth2. Line colors for the various terms are as in Figure 6. Parameterized tendency is not present in EC-Earth2.

EC-Earth3 now includes a more realistic representation of stratospheric processes, thanks to the extension of the vertical domain up to the lower mesosphere, a finer vertical discretization and implementation of parameterization for processes such as non-orographic GWD that is essential for simulating the stratospheric circulation. As found for other climate models with an improved representation of the stratosphere (Kawatani et al., 2019; Richter et al., 2014), the simulated climatological conditions are much more realistic going from EC-Earth2 to EC-Earth3, with both zonal winds and temperature in better agreement with the reanalysis in the lower and middle stratosphere. Besides EC-Earth3 biases due to radiative processes identified in previous work (Palmeiro et al., 2022), residual circulation differences can also be associated to the temperature biases both in the tropics and extratropics. Uncertainties remain in model results for upper levels, but it should also be noted that reanalysis products are weakly constrained there due to the lack of observations.

A major improvement of EC-Earth3 is the presence of a realistic quasi-biennial oscillation (QBO) in the equatorial stratosphere, although it is too weak compared to the reanalysis at all levels, likely due to insufficient gravity wave forcing as diagnosed from DynVarMIP variables. As found for other high-top models (Serva et al., 2022), the amplitude of the QBO in the lower stratosphere is small compared to reanalysis, which is important for its influence on tropical climate. In the upper stratosphere, EC-Earth3 still presents an easterly bias which is associated with weak gravity wave drag and the influence of prescribed numerical damping (sponge layer), known to be relevant in the EC-Earth atmospheric model (Polichtchouk et al., 2017). In the tropics the magnitude of the residual term in the momentum budget of the QBO is relatively large, as found in other works (Giorgetta

Table 1
Number of Major and Minor SSWs in Both Northern (NH) and Southern (SH) Hemispheres

Hemisphere	MERRA-2 major	EC-Earth2 major	EC-Earth3 major	MERRA-2 minor	EC-Earth2 minor	EC-Earth3 minor
NH	25 (5.9)	22 (6.9)	24 (6.1)	18 (4.3)	7 (2.2)	19 (4.9)
SH	1 (0.2)	4 (1.2)	0 (0)	5 (1.2)	4 (1.2)	7 (1.8)

Note. Events per decade are shown in brackets for easier comparison, that is, after normalization by 42, 32 and 39 years for MERRA-2, EC-Earth2 and EC-Earth3, respectively.

et al., 2022), and its origin should be considered by future research, for example, with the analysis of the tendencies from the dynamical core.

Intraseasonal variability in the extratropics is more realistic in EC-Earth3 than in EC-Earth2, as the low-top model has too many major SSWs in both hemispheres, even if the simulation length is quite short and internal variability may be important (Palmeiro et al., 2020). Notably, the momentum budget partitioning in the low-top model differs from than of the high-top version and both reanalyses, likely due to spurious effects driven by the model top placed in the middle stratosphere. The different dynamical properties of the low-top could be related to the relative frequency of major (too many) and minor (too few) SSWs in this configuration. To our knowledge this circumstance has not been systematically studied with climate model data and in a multi-model context before, but we leave this topic of research to future work. Care should be taken when interpreting momentum budgets for configurations with low vertical resolution and near the model lid.

Our findings highlight that more comprehensive and extended model outputs are useful for understanding model biases and their origin, including evaluating the realism of stratospheric processes and to quantify benefits from model developments. However, given the complex interactions between waves and mean flow in the stratosphere, nudging experiments (Hitchcock et al., 2022; Morgenstern et al., 2017) would be useful to identify the origins of biases.

Appendix A: Integration of TEM Diagnostics Into the CMORization Process

The calculation of the transformed Eulerian mean (TEM) variables, required for estimating momentum and heat fluxes, is often performed offline after the simulation. In order to reduce the numerical error associated with meridional and vertical derivatives, it is advised (Gerber & Manzini, 2016) to retain the native horizontal resolution and pressure levels close to the original model levels. In EC-Earth (and many other models), this requires writing 3D prognostic variables in the spherical harmonics representation at six-hourly frequency. As a result, the model produces large GRIB files every simulated month which are too costly to store permanently and therefore require an efficient and robust post-processing workflow into the CMIP6-compliant format.

To our knowledge, there are no publicly available tools to perform the calculation on generic gridded data, and the processing scripts are not publicly available for other CMIP6 models, limiting the reproducibility of results. For the EC-Earth contribution to the CMIP6, a set of scripts (Serva, 2023) for the TEM diagnostics (*tem-diag*) have been developed and they have been integrated into the cmorization workflow provided by *ece2cmor3* (van den Oord & Bakhshi, 2017). This facilitates the production of the data across different platforms and for future efforts related with DynVarMIP.

Following Gerber and Manzini (2016), the eddy stream function is defined as $\psi = \overline{v'\theta'}/\partial_p\overline{\theta}$, and the TEM northward and vertical velocities are defined by $\overline{v}^* = \overline{v} - \partial_p\psi$ and $\overline{\omega}^* = \overline{\omega} + (a\cos\phi)^{-1}\partial_\phi(\psi\cos\phi)$. Here u, v, ω are the zonal, meridional and vertical velocity components, ϕ is latitude and p pressure, θ is potential temperature, and a is the Earth radius. The eastward wind tendencies due to advection are $\partial_t\overline{u}(\overline{v}^*) = \overline{v}^*\left[f - (a\cos\phi)^{-1}\,\partial_\phi\overline{u}\cos\phi\right]$ and $\partial_t\overline{u}(\overline{\omega}^*) = \overline{\omega}^*\partial_p\overline{u}$. Another term forcing the zonal mean zonal wind variations is the divergence of the EP fluxes, defined as $F_\phi = a\cos\phi\left(\partial_p\overline{u}\psi - \overline{u'v'}\right)$ and $F_p = a\cos\phi\left\{\left[f - (a\cos\phi)^{-1}\partial_\phi(\psi\cos\phi)\right]\psi - \overline{u'\omega'}\right\}$.

The scripts in *tem-diag* are written in *bash* and *python*. The implementation closely follows the indications provided by the DynVarMIP protocol, and while it is firstly aimed at EC-Earth outputs, could be applied to other gridded outputs as well. The main script (*tem_diag.py*) requires seven files in input: zonal means of temperature and zonal, meridional and vertical winds, and fluxes terms (meridional flux of zonal momentum, vertical flux of zonal momentum, meridional heat flux). The numerical (such as derivatives and integrations) and ancillary procedures (e.g., file I/O and other system operations) are collected in a python script (*tem_funcs.py*), and constants are defined in a third script (*tem_consts.py*). Note that the vertical dimension is pressure-based, therefore vertical velocity is a pressure tendency ($Pa\ s^{-1}$).

The calculation of flux terms and zonal mean dynamical variables from the large spectral model output files is the most data intensive part, and it is done through a wrapper and preprocessing bash script (*make_dynvars.sh*). In order to improve the TEM calculation itself, raw data (four dimensional temperature and wind components,

instantaneous and six-hourly) are preprocessed in order to obtain the required zonal mean files, such as eddy covariances. This allows us to substantially reduce the data used in the *python* scripts, a procedure which should be performed online (Yepes-Arbós et al., 2022), especially when running very high resolution simulations.

The wrapper script is called by *ece2cmor3*, which has been adapted to spawn other post-processing applications than its default, the Climate Data Operators (CDO (https://code.mpimet.mpg.de/projects/cdo/)) tool. The *tem-diag* scripts perform the full post-processing, including temporal aggregation into daily or monthly means, transforming fields into variables with the correct units, and all the required computations. After the scripts have been executed the *ece2cmor3* pipeline continues to equip the resulting *NetCDF4* files with the necessary metadata. Following its conventional approach, months are processed individually with *tem-diag* and then they are concatenated to form a yearly file for cmorization. Two libraries are required in the environment, namely *nco* (needed to set file attributes) and *scipy* (required for internal calculations). The processing of TEM variables requires adding to the list of IFS parameters (*ifspar.json*), each associated with a specific *grib* code and the external *bash* script. All the variables are instantaneous and no further adjustments, except for metadata correction, is done by *ece2cmor3*. In fact the TEM variables cmorization differs from that of other variables, as *tem-diag* produces files which do not need further processing (such as computation of rates). In conclusion, the integration of *tem-diag* with *ece2cmor3* provides a clear separation of concerns: the former is responsible for post-processing DynVarMIP variables from raw EC-Earth model output, and the latter is tasked with organizing the pipeline and attaching the appropriate metadata to those variables. For the end user, this interplay is (mostly) hidden and the DynVarMIP variables can be seamlessly combined with those from other MIPs.

The environment used for the processing is inherited from that activated to run *ece2cmor3*. A specific script *switch-on-off-DynVarMIP-mode.sh* is used to set up the environment of *ece2cmor3* for the DynVarMIP processing, by considering all model levels and the ancillary files with the mandatory variables.

Data Availability Statement

EC-Earth3 AMIP simulations are available online from: https://esgf-node.llnl.gov/projects/cmip6/. The MERRA-2 reanalysis data can be retrieved from GES DISC ((GMAO, 2015b), https://disc.gsfc.nasa.gov/datasets/M2I3NVASM_5.12.4/summary for model-level meteorological fields, (GMAO, 2015c), https://disc.gsfc.nasa.gov/datasets/M2T3NPUDT_5.12.4/summary for parameterized tendencies, and (GMAO, 2015a), https://disc.gsfc.nasa.gov/datasets/M2I3NPASM_5.12.4/summary for pressure-level meteorological fields, last accessed August 2023) and ERA5 can be accessed via MARS or the CDS (https://confluence.ecmwf.int/display/CKB/How+to+download+ERA5, last accessed November 2022). Processed data are available from these links: MERRA-2 (https://zenodo.org/record/6959944, (Serva, 2022b)), ERA5 (https://zenodo.org/record/7081436, (Serva, 2022a)) and EC-Earth2 (https://zenodo.org/record/7260929, (Serva & Wyser, 2022)). The *tem-diag* software is available at https://zenodo.org/records/10180386 (Serva, 2023).

Acknowledgments
We thank the anonymous reviewers and the editors for many useful suggestions to improve the paper. We would like to thank the EC-Earth community for supporting code development for the extension of model output and for performing simulations, and the reanalysis centers for making their products available. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modeling, coordinated and promoted CMIP6, and thank the Earth System Grid Federation (ESGF) for archiving the data and providing access. BC and SY are partly supported by the Danish National Center for Climate Research (NCKF). Acknowledgment is made for the use of ECMWF's computing and archive facilities under Special Project SPITSERV for EC-Earth3 simulation. SMHI's computations and data processing were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at NSC.

References

Abalos, M., Calvo, N., Benito-Barca, S., Garny, H., Hardiman, S. C., Lin, P., et al. (2021). The Brewer–Dobson circulation in CMIP6. *Atmospheric Chemistry and Physics*, *21*(17), 13571–13591. https://doi.org/10.5194/acp-21-13571-2021

Andrews, D. G., Holton, J. R., & Leovy, C. B. (1987). *Middle atmosphere dynamics* (Vol. 40). Academic press.

Badin, G., & Domeisen, D. I. V. (2016). Nonlinear stratospheric variability: Multifractal de-trended fluctuation analysis and singularity spectra. *Proceedings of the Royal Society A*, *472*(2191), 20150864. https://doi.org/10.1098/rspa.2015.0864

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., et al. (2021). Sudden stratospheric warmings. *Reviews of Geophysics*, *59*(1). https://doi.org/10.1029/2020RG000708

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. https://doi.org/10.1126/science.1063315

Bushell, A. C., Anstey, J. A., Butchart, N., Kawatani, Y., Osprey, S. M., Richter, J. H., et al. (2022). Evaluation of the quasi-biennial oscillation in global climate models for the SPARC QBO-initiative. *The Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1459–1489. https://doi.org/10.1002/qj.3765

Butler, A. H., Charlton-Perez, A., Domeisen, D. I., Simpson, I. R., & Sjoberg, J. (2019). Predictability of northern hemisphere final stratospheric warmings and their surface impacts. *Geophysical Research Letters*, *46*(17–18), 10578–10588. https://doi.org/10.1029/2019GL083346

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

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., & Match, A. (2015). Defining sudden stratospheric warmings. *Bulletin America Meteorology Social*, *96*(11), 1913–1928. https://doi.org/10.1175/BAMS-D-13-00173.1

Butler, A. H., Sjoberg, J. P., Seidel, D. J., & Rosenlof, K. H. (2017). A sudden stratospheric warming compendium. *Earth System Science Data*, *9*(1), 63–76. https://doi.org/10.5194/essd-9-63-2017

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

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., et al. (2013). On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *118*(6), 2494–2505. https://doi.org/10.1002/jgrd.50125

Christiansen, B. (2001). Downward propagation of zonal mean zonal wind anomalies from the stratosphere to the troposphere: Model and reanalysis. *Journal of Geophysical Research*, *106*(D21), 27307–27322. https://doi.org/10.1029/2000JD000214

Christiansen, B., Yang, S., & Madsen, M. S. (2016). Do strong warm ENSO events control the phase of the stratospheric QBO? *Geophysical Research Letters*, *43*(19), 10. https://doi.org/10.1002/2016GL070751

Coy, L., Wargan, K., Molod, A. M., McCarty, W. R., & Pawson, S. (2016). Structure and dynamics of the quasi-biennial oscillation in MERRA-2. *Journal of Climate*, *29*(14), 5339–5354. https://doi.org/10.1175/JCLI-D-15-0809.1

Davini, P., von Hardenberg, J., Corti, S., Christensen, H. M., Juricke, S., Subramanian, A., et al. (2017). Climate SPHINX: Evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model. *Geoscientific Model Development*, *10*(3), 1383–1402. https://doi.org/10.5194/gmd-10-1383-2017

de la Cámara, A., Lott, F., Jewtoukoff, V., Plougonven, R., & Hertzog, A. (2016). On the gravity wave forcing during the southern stratospheric final warming in LMDz. *Journal of the Atmospheric Sciences*, *73*(8), 3213–3226. https://doi.org/10.1175/JAS-D-15-0377.1

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Diallo, M., Ern, M., & Ploeger, F. (2021). The advective Brewer–Dobson circulation in the ERA5 reanalysis: Climatology, variability, and trends. *Atmospheric Chemistry and Physics*, *21*(10), 7515–7544. https://doi.org/10.5194/acp-21-7515-2021

Domeisen, D., & Butler, A. (2020). Stratospheric drivers of extreme events at the Earth's surface. *Communications Earth & Environment*, *59*(1), 59. https://doi.org/10.1038/s43247-020-00060-z

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., et al. (2022). The EC-Earth3 Earth system model for the coupled model intercomparison project 6. *Geoscientific Model Development*, *15*(7), 2973–3020. https://doi.org/10.5194/gmd-15-2973-2022

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Garcia, R. R. (2021). On the response of the middle atmosphere to anthropogenic forcing. *Annals of the New York Academy of Sciences*, *1504*(1), 25–43. https://doi.org/10.1111/nyas.14664

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

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., et al. (1999). An overview of the results of the atmospheric model intercomparison project (AMIP I). *Bulletin America Meteorology Social*, *81*(1), 29–55. https://doi.org/10.1175/1520-0477(1999)080\textless0029:AOOTRO\textgreater2.0.CO;2

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Geller, M. A., Alexander, M. J., Love, P. T., Bacmeister, J., Ern, M., Hertzog, A., et al. (2013). A comparison between gravity wave momentum fluxes in observations and climate models. *Journal of Climate*, *26*(17), 6383–6405. https://doi.org/10.1175/JCLI-D-12-00545.1

Gerber, E. P., & Manzini, E. (2016). The dynamics and variability model intercomparison project (DynVarMIP) for CMIP6: Assessing the stratosphere-troposphere system. *Geoscientific Model Development*, *9*(9), 3413–3425. https://doi.org/10.5194/gmd-9-3413-2016

Giorgetta, M. A., Sawyer, W., Lapillonne, X., Adamidis, P., Alexeev, D., Clément, V., et al. (2022). The ICON-A model for direct QBO simulations on GPUs (version icon-cscs:baf28a514). *Geoscientific Model Development*, *15*(18), 6985–7016. https://doi.org/10.5194/gmd-15-6985-2022

GMAO. (2015a). MERRA-2 inst3_3d_asm_Np: 3d,3-Hourly,Instantaneous,Pressure-Level, Assimilation,Assimilated meteorological fields [Dataset]. Goddard Earth Sciences Data and Information Services Center. https://doi.org/10.5067/QBZ6MG944HW0

GMAO. (2015b). MERRA-2 inst3_3d_asm_Nv: 3d,3-Hourly,Instantaneous,Model-Level, Assimilation,Assimilated meteorological fields [Dataset]. Goddard Earth Sciences Data and Information Services Center. https://doi.org/10.5067/WWQSXQ8IVFW8

GMAO. (2015c). MERRA-2 tavg3_3d_udt_Np: 3d,3-Hourly,Time-Averaged,Pressure-Level, Assimilation,Wind tendencies [Dataset]. Goddard Earth Sciences Data and Information Services Center. https://doi.org/10.5067/CWV0G3PPPWFW

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

Grise, K. M., & Thompson, D. W. J. (2013). On the signatures of equatorial and extratropical wave forcing in tropical tropopause layer temperatures. *Journal of the Atmospheric Sciences*, *70*(4), 1084–1102. https://doi.org/10.1175/JAS-D-12-0163.1

Haynes, P., Hitchcock, P., Hitchman, M., Yoden, S., Hendon, H., Kiladis, G., et al. (2021). The influence of the stratosphere on the tropical troposphere. *Journal of the Meteorological Society of Japan Series II*, *99*(4), 803–845. https://doi.org/10.2151/jmsj.2021-040

Hazeleger, W., Wang, X., Severijns, C., Stefanescu, S., Bintanja, R., Sterl, A., et al. (2012). EC-Earth V2.2: Description and validation of a new seamless earth system prediction model. *Climate Dynamics*, *39*(11), 2611–2629. https://doi.org/10.1007/s00382-011-1228-5

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *The Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hirano, S., Kohma, M., & Sato, K. (2021). Interannual variability of stratospheric final warming in the southern hemisphere and its tropospheric origin. *Journal of Climate*, *34*(15), 6115–6128. https://doi.org/10.1175/JCLI-D-20-0945.1

Hitchcock, P., Butler, A., Charlton-Perez, A., Garfinkel, C. I., Stockdale, T., Anstey, J., et al. (2022). Stratospheric nudging and predictable surface impacts (SNAPSI): A protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts. *Geoscientific Model Development*, *15*(13), 5073–5092. https://doi.org/10.5194/gmd-15-5073-2022

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., & Pfister, L. (1995). Stratosphere-troposphere exchange. *Reviews of Geophysics*, *33*(4), 403–439. https://doi.org/10.1029/95RG02097

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, *12*(3), 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Jucker, M., Reichler, T., & Waugh, D. W. (2021). How frequent are Antarctic sudden stratospheric warmings in present and future climate? *Geophysical Research Letters*, *48*(11), e2021GL093215. https://doi.org/10.1029/2021GL093215

Kawatani, Y., Hamilton, K., Gray, L. J., Osprey, S. M., Watanabe, S., & Yamashita, Y. (2019). The effects of a well-resolved stratosphere on the simulated boreal winter circulation in a climate model. *Journal of the Atmospheric Sciences*, *76*(5), 1203–1226. https://doi.org/10.1175/JAS-D-18-0206.1

Kawatani, Y., Hamilton, K., & Watanabe, S. (2011). The quasi-biennial oscillation in a double CO2 climate. *Journal of the Atmospheric Sciences*, *68*(2), 265–283. https://doi.org/10.1175/2010JAS3623.1

Kawatani, Y., Hirooka, T., Hamilton, K., Smith, A. K., & Fujiwara, M. (2020). Representation of the equatorial stratopause semiannual oscillation in global atmospheric reanalyses. *Atmospheric Chemistry and Physics*, *20*(14), 9115–9133. https://doi.org/10.5194/acp-20-9115-2020

Keeble, J., Hassler, B., Banerjee, A., Checa-Garcia, R., Chiodo, G., Davis, S., et al. (2021). Evaluating stratospheric ozone and water vapour changes in CMIP6 models from 1850 to 2100. *Atmospheric Chemistry and Physics*, *21*(6), 5015–5061. https://doi.org/10.5194/acp-21-5015-2021

Kim, J., Son, S.-W., Gerber, E. P., & Park, H.-S. (2017). Defining sudden stratospheric warming in climate models: Accounting for biases in model climatologies. *Journal of Climate*, *30*(14), 5529–5546. https://doi.org/10.1175/JCLI-D-16-0465.1

Kobayashi, C., & Iwasaki, T. (2016). Brewer–Dobson circulation diagnosed from JRA-55. *Journal of Geophysical Research: Atmospheres*, *121*(4), 1493–1510. https://doi.org/10.1002/2015JD023476

Kodera, K., Mukougawa, H., & Itoh, S. (2008). Tropospheric impact of reflected planetary waves from the stratosphere. *Geophysical Research Letters*, *35*(16). https://doi.org/10.1029/2008GL034575

Long, C. S., Fujiwara, M., Davis, S., Mitchell, D. M., & Wright, C. J. (2017). Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmospheric Chemistry and Physics*, *17*(23), 14593–14629. https://doi.org/10.5194/acp-17-14593-2017

Lu, Y., Wu, T., Jie, W., Scaife, A. A., Andrews, M. B., & Richter, J. H. (2020). Variability of the stratospheric quasi-biennial oscillation and its wave forcing simulated in the Beijing climate center atmospheric general circulation model. *Journal of the Atmospheric Sciences*, *77*(1), 149–165. https://doi.org/10.1175/JAS-D-19-0123.1

Manzini, E., Karpechko, A. Y., Anstey, J., Baldwin, M. P., Black, R. X., Cagnazzo, C., et al. (2014). Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *119*(13), 7979–7998. https://doi.org/10.1002/2013JD021403

Martin, J. E. (2006). *Mid-latitude atmospheric dynamics: A first course*. Wiley.

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

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geoscientific Model Development*, *8*(5), 1339–1356. https://doi.org/10.5194/gmd-8-1339-2015

Morcrette, J.-J., Barker, H. W., Cole, J. N. S., Iacono, M. J., & Pincus, R. (2008). Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Monthly Weather Review*, *136*(12), 4773–4798. https://doi.org/10.1175/2008MWR2363.1

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., et al. (2017). Review of the global models used within phase 1 of the chemistry–climate model initiative (CCMI). *Geoscientific Model Development*, *10*(2), 639–671. https://doi.org/10.5194/gmd-10-639-2017

O'Brien, J. J. (1970). Alternative solutions to the classical vertical velocity problem. *Journal of Applied Meteorology and Climatology*, *9*(2), 197–203. https://doi.org/10.1175/1520-0450(1970)009⟨0197:ASTTCV⟩2.0.CO;2

Orr, A., Bechtold, P., Scinocca, J., Ern, M., & Janiskova, M. (2010). Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, *23*(22), 5905–5926. https://doi.org/10.1175/2010JCLI3490.1

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

Palmeiro, F. M., García-Serrano, J., Bellprat, O., Bretonnière, P.-A., & Doblas-Reyes, F. J. (2020). Boreal winter stratospheric variability in EC-EARTH: High-Top versus Low-Top. *Climate Dynamics*, *54*(5), 3135–3150. https://doi.org/10.1007/s00382-020-05162-0

Palmeiro, F. M., García-Serrano, J., Rodrigo, M., Abalos, M., Christiansen, B., & Yang, S. (2022). Boreal winter stratospheric climatology in EC-EARTH: CMIP6 version. *Climate Dynamics*, *60*(3–4), 883–898. https://doi.org/10.1007/s00382-022-06368-0

Polichtchouk, I., Hogan, R., Shepherd, T., Bechtold, P., Stockdale, T., Malardel, S., et al. (2017). What influences the middle atmosphere circulation in the IFS? https://doi.org/10.21957/MFSNFV15O

Rao, J., Garfinkel, C. I., Chen, H., & White, I. P. (2019). The 2019 new year stratospheric sudden warming and its real-time predictions in multiple S2S models. *Journal of Geophysical Research: Atmospheres*, *124*(21), 11155–11174. https://doi.org/10.1029/2019JD030826

Rao, J., Ren, R., Chen, H., Yu, Y., & Zhou, Y. (2018). The stratospheric sudden warming event in February 2018 and its prediction by a climate system model. *Journal of Geophysical Research: Atmospheres*, *123*(23). https://doi.org/10.1029/2018JD028908

Reichler, T., Kim, J., Manzini, E., & Kröger, J. (2012). A stratospheric connection to Atlantic climate variability. *Nature Geoscience*, *5*(11), 783–787. https://doi.org/10.1038/ngeo1586

Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020). Progress in simulating the quasi-biennial oscillation in CMIP models. *Journal of Geophysical Research: Atmospheres*, *125*(8), e2019JD032362. https://doi.org/10.1029/2019JD032362

Richter, J. H., Solomon, A., & Bacmeister, J. T. (2014). Effects of vertical resolution and nonorographic gravity wave drag on the simulated climate in the Community Atmosphere Model, version 5. *Journal of Advances in Modeling Earth Systems*, *6*(2), 357–383. https://doi.org/10.1002/2013MS000303

Serva, F. (2022a). Transformed Eulerian mean data from the ERA5 reanalysis (daily means) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7081436

Serva, F. (2022b). Transformed Eulerian mean data from the MERRA-2 reanalysis (daily means) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.6959944

Serva, F. (2023). Transformed Eulerian mean diagnostics (tem-diag) [Software]. Zenodo. https://doi.org/10.5281/zenodo.10180386

Serva, F., Anstey, J. A., Bushell, A. C., Butchart, N., Cagnazzo, C., Gray, L. J., et al. (2022). The impact of the QBO on the region of the tropical tropopause in QBOi models: Present-day simulations. *The Quarterly Journal of the Royal Meteorological Society*, *148*(745), 1945–1964. https://doi.org/10.1002/qj.4287

Serva, F., Cagnazzo, C., Christiansen, B., & Yang, S. (2020). The influence of ENSO events on the stratospheric QBO in a multi-model ensemble. *Climate Dynamics*, *54*(54), 2561–2575. https://doi.org/10.1007/s00382-020-05131-7

Serva, F., & Wyser, K. (2022). Transformed Eulerian mean data from CMIP5 EC-Earth v2.3 AMIP experiment [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7260929

Seviour, W. J. M., Butchart, N., & Hardiman, S. C. (2012). The Brewer–Dobson circulation inferred from ERA-Interim. *The Quarterly Journal of the Royal Meteorological Society*, *138*(665), 878–888. https://doi.org/10.1002/qj.966

Shepherd, T. G., & Shaw, T. A. (2004). The angular momentum constraint on climate sensitivity and downward influence in the middle atmosphere. *Journal of the Atmospheric Sciences*, *61*(23), 2899–2908. https://doi.org/10.1175/JAS-3295.1

Smith, A. K., Holt, L. A., Garcia, R. R., Anstey, J. A., Serva, F., Butchart, N., et al. (2022). The equatorial stratospheric semiannual oscillation and time-mean winds in QBOi models. *The Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1593–1609. https://doi.org/10.1002/qj.3690

SPARC. (2010). SPARC CCMVal report on the evaluation of chemistry-climate models (Tech. Rep. Nos. 5, WCRP-30/2010, WMO/TD No. 40,). Retrieved from https://www.sparc-climate.org/publications/sparc-reports/sparc-report-no-5/

Thompson, D. W. J., Baldwin, M. P., & Solomon, S. (2005). Stratosphere–troposphere coupling in the southern hemisphere. *Journal of the Atmospheric Sciences*, *62*(3), 708–715. https://doi.org/10.1175/JAS-3321.1

van den Oord, G., & Bakhshi, R. (2017). Parallel post-processing of the Earth climate model output. *Procedia Computer Science*, *108*, 2473–2477. https://doi.org/10.1016/j.procs.2017.05.146

Wright, C. J., & Hindley, N. P. (2018). How well do stratospheric reanalyses reproduce high-resolution satellite temperature measurements? *Atmospheric Chemistry and Physics*, *18*(18), 13703–13731. https://doi.org/10.5194/acp-18-13703-2018

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., & Döscher, R. (2020). On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6. *Geoscientific Model Development*, *13*(8), 3465–3474. https://doi.org/10.5194/gmd-13-3465-2020

Yepes-Arbós, X., Van Den Oord, G., Acosta, M. C., & Carver, G. D. (2022). Evaluation and optimisation of the I/O scalability for the next generation of earth system models: IFS CY43R3 and XIOS 2.0 integration as a case study. *Geoscientific Model Development*, *15*(2), 379–394. https://doi.org/10.5194/gmd-15-379-2022

Yoshida, K., & Mizuta, R. (2021). Do sudden stratospheric warmings boost convective activity in the tropics? *Geophysical Research Letters*, *48*(16). https://doi.org/10.1029/2021GL093688

Yoshida, K., Mizuta, R., & Arakawa, O. (2018). Intermodel differences in upwelling in the tropical tropopause layer among CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *123*(24). https://doi.org/10.1029/2018JD029044