Changes in Stratospheric Climate and Age-of-Air in Recent GEOS Systems since MERRA-2

Clara Orbe[1,2], Lawrence L. Takacs[3,4], Amal El Akkraoui[3], Krzysztof Wargan[3,4], Andrea Molod[3], Steven Pawson[3]

[1]NASA Goddard Institute for Space Studies, New York, NY
[2]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY
[3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD
[4]Science Systems and Applications, Inc., Lanham, MD, USA

Key Points:

- The stratospheric mean age-of-air simulated in GEOS is sensitive to the remapping scheme used within the finite-volume dynamical core.
- This sensitivity in the age-of-air is large (∼30%) and imprints on the simulated distributions of several long-lived trace gases (e.g., N_2O, CH_4).
- The age-of-air sensitivities primarily reflect changes in resolved wave convergence over the NH extratropical stratosphere.

Corresponding author: Clara Orbe, clara.orbe@nasa.gov

Abstract

Accurately modeling the large-scale transport of trace gases and aerosols is critical for interpreting past (and projecting future) changes in atmospheric composition. Simulations of the stratospheric mean age-of-air continue to show persistent biases in chemistry climate models, although the drivers of these biases are not well understood. Here we identify one driver of simulated stratospheric transport differences among various NASA Global Earth Observing System (GEOS) candidate model versions under consideration for the upcoming GEOS Retrospective analysis for the 21st Century (GEOS-R21C). In particular, we show that the simulated age-of-air values are sensitive to the so-called "remapping" algorithm used within the finite-volume dynamical core, which controls how individual material surfaces are vertically interpolated back to standard pressure levels after each horizontal advection time step. Differences in the age-of-air resulting from changes within the remapping algorithm approach $\sim$ 1 year over the high latitude middle stratosphere – or about 30% climatological mean values – and imprint on several trace gases, including methane (CH_4) and nitrous oxide (N_2O). These transport sensitivities reflect, to first order, changes in the strength of tropical upwelling in the lower stratosphere (70-100 hPa) which are driven by changes in resolved wave convergence over northern midlatitudes as (critical lines of) wave propagation shift in latitude. Our results strongly support continued examination of the role of numerics in contributing to transport biases in composition modeling.

Plain Language Summary

Large-scale transport plays a crucial role in distributing climatically important trace constituents in the atmosphere, especially in the stratosphere where transport largely determines the chemical lifetimes of trace gases. One summary of transport in the stratosphere is the "mean age" or the mean transit time since air at a point in the stratosphere was last in the troposphere. Current models used for simulating stratospheric composition produce a range of simulated ages, although these differences are poorly understood. Among other factors, model numerics play a critical role in transport, but few studies have explored the sensitivity of the mean age to the choice of numerical scheme employed within different dynamical cores. Here we use one model to show that the mean age is sensitive to the so-called "remapping" algorithm used within the finite-volume dynamical core that controls how individual material surfaces are vertically interpolated back to standard pressure levels after each horizontal advection time step. This reflects sensitivities in the representation of how waves propagate from the troposphere into the stratosphere. This work suggests that model numerics can be an important factor in contributing to differences in simulated transport among models.

1 Introduction

The chemical and radiative properties of the troposphere and lower stratosphere are strongly influenced by the stratosphere-troposphere exchange of mass and tracers (e.g., Morgenstern and Carver (2001); Hegglin et al. (2006); Pan et al. (2007)). Properly simulating the stratospheric circulation and its influence on atmospheric composition in earth system models is important for capturing past decadal trends in surface climate, particularly in response to changes in Southern Hemisphere ozone depletion (e.g., Son et al. (2009); Polvani et al. (2011)). In the Northern Hemisphere (NH), the stratospheric circulation's coupling to ozone could represent an important feedback on the climate's response to future increases in greenhouse gases (GHGs), especially over the North Atlantic (e.g., Chiodo and Polvani (2019)). On shorter subseasonal timescales, stratospheric ozone changes associated with strong polar vortex states may also modulate Arctic sea level pressure and surface temperatures (e.g., Ivy et al. (2017); Oehrlein et al. (2020)), so much so that seasonal forecast systems employing prognostic ozone show suggestions of increased signal-to-noise ratio in predictions of the North Atlantic Oscillation (B. M. Monge-Sanz et al. (2022)).

Key to accurately simulating a consistent representation of coupling between stratospheric dynamics and chemical trace gases is ensuring that a model's underlying transport circulation is properly represented. To this end, much effort has been paid to developing so-called "tracer-independent" metrics of transport (Holzer and Hall (2000)) such as the mean age-of-air (Hall and Plumb (1994)) and to applying these measures to rigorously evaluate model transport characteristics in chemistry climate models (CCMs) (e.g., Hall et al. (1999); Orbe et al. (2018); Dietmüller et al. (2018); Abalos et al. (2020)).

While the assessment of CCMs participating in the SPARC Chemistry Climate Model Validation (SPARC CCMVal) effort showed a marked improvement in simulated transport characteristics relative to previous intercomparisons (J. Neu et al. (2010)), more recent analysis of models participating in the SPARC Chemistry Climate Modeling Initiative (CCMI) (Eyring et al. (2013)) do not demonstrate any improvement (Dietmüller et al. (2018), see their Figure 3). In particular, although some models produce mean age values that agree well with observational estimates, the CCMI intermodel spread is $\sim$ 50%, with models generally simulating transport that is too vigorous relative to observations. While documenting these transport differences among models is straightforward, understanding the drivers of this spread remains a challenge and there is still no consensus on what is causing the large spread in simulated ages among the current generation of CCMs.

A key challenge in identifying the drivers of age-of-air – and other stratospheric transport – biases is that they reflect the time-integrated effects of advection by the residual mean circulation and eddy diffusive mixing, or the quasi-random transport due to the breaking of Rossby waves (e.g., Holton et al. (1995); Plumb (2002)). Given that the influences of mixing and advection are not easily separable, studies have come to different conclusions about sources of age biases in models. In particular, the analysis of the CCMVal models showed a strong correlation between the intermodel spread in the age-of-air and lower stratospheric tropical upwelling, whereas Dietmüller et al. (2018) showed that the age spread among the CCMI models was driven by differences in mixing. While future attempts to further distinguish between sources of age biases using either simplified "leaky pipe" models (Plumb (1996); J. L. Neu and Plumb (1999)) or more complete measures of the transport circulation such as the "age spectrum" (e.g., Hall and Plumb (1994); Waugh and Hall (2002))) may prove enlightening, at present there is no consensus on what is causing large simulated age-of-air biases in models.

One potential limitation of previous work based on multi-model intercomparisons is that many aspects of model formulation can influence both stratospheric upwelling and mixing. Thus, while intercomparisons are useful for identifying common model biases, understanding the drivers of these biases is difficult absent single model-based process studies. Among these, several aspects of model formulation have been identified as influencing simulated mean age distributions. As the mean age is sensitive to vertical motion in the lowermost stratosphere, these include large sensitivities to vertical resolution (Orbe et al. (2020)) and to spurious vertical mixing either introduced in vertical coordinate transformations in offline chemical transport models (B. Monge-Sanz et al. (2007)) or through use of assimilated winds performed either in offline (e.g., Legras et al. (2004)) or online data assimilation and "nudged" configurations (e.g., Pawson et al. (2007); Orbe et al. (2017); Davis et al. (2022)). These age sensitivities can be still further amplified, depending on whether or not parameterized gravity waves are included (Eichinger et al. (2020)).

By comparison, sensitivities of the mean age to underlying tracer numerics have been less well examined, although Eluszkiewicz et al. (2000) documented a large sensitivity in simulated age-of-air values to the choice of advection scheme. More recently, Gupta et al. (2020) showed differences of $\sim 25\%$ in the age-of-air across identical experiments performed using four different dynamical cores, especially between those using spectral versus finite-volume schemes. The experiments employed in that study, however, were highly idealized and it is not clear if the strong influence of tracer numerics that they identified is also realized in more comprehensive model simulations with moist physics, especially in the context of model development as carried out in operational modeling centers.

To better elucidate this influence of tracer numerics on the transport properties simulated in a comprehensive global model context, here we document the sensitivity of the stratospheric mean age in several recent versions of the NASA Global Earth Observing System (GEOS) general circulation model (Molod et al., 2015). The versions examined here represent different stages in model development since the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2; Gelaro et al. (2017)).

Our focus on transport evaluation is partly in wake of the upcoming release of the GEOS Retrospective analysis for the early 21st Century (GEOS-R21C), which is an atmospheric reanalysis that includes many advances over MERRA-2, and serves as a step towards MERRA-3, a planned coupled Earth system reanalysis. Along with having implications for GEOS-R21C, our analysis also includes the GEOS-IT Version 5.29.4 dating from October 19, 2022 (hereafter simply "GEOS-IT"), which will be used to drive an off-line chemistry reanalysis with a full chemistry model (GEOS-Chem) and an advanced Constituent Data Assimilation component to update the chemistry fields. Since this chemical reanalysis will be produced in replay-mode (one-way coupling) whereby the meteorology fields are used to define the background atmospheric flow (Orbe et al., 2017), it is imperative that recent tags of GEOS produce a credible representation of transport processes.

In particular, here we document how in the process of evaluating candidate systems for GEOS-R21C we found that the mean age was ~ 1 year younger (or $\sim 30\%$ smaller) than the values simulated in the model version used to produce MERRA-2 (Figure 1). The model versions shown in Figure 1 reflect more than 10 years' worth of accumulated changes in model development, most notably changes in radiation, parameterized convection and, as we focus on here, changes in the model's description of vertical advection. More precisely, we show that slight modifications to the so-called "remapping" algorithm, which is used to transform advected fields from Lagrangian levels to the new pressure levels after each horizontal advection time step, are the primary driver of the age-of-air changes exhibited in recent GEOS-R21C candidate model versions. Thus, whereas Gupta et al. (2020) highlighted large differences between dynamical cores employing spectral versus finite-volume (FV) numerics, our results show that large transport differences can occur even within a given FV dynamical core, a result which may have broader implications for other general circulation models employing FV numerics. We begin by discussing methods in Section 2 and present key results and conclusions in Sections 3 and 4, respectively.

2 Methods

2.1 Model Configurations

Here we present results from several versions of GEOS spanning MERRA-2 to more recent candidates for GEOS-R21C. Among these model versions, a subset are more "official" as they have been documented and/or employed in recent model intercomparisons and are highlighted in Figure 1. In particular, these include a model version that was used in Phase 1 of CCMI and documented in Orbe et al. (2017) (Fig. 1, red line). A more recent model version that was used in the CCMI Phase 2 simulations (correspondence with Michael Manyin) is also shown (Fig. 1, green line).

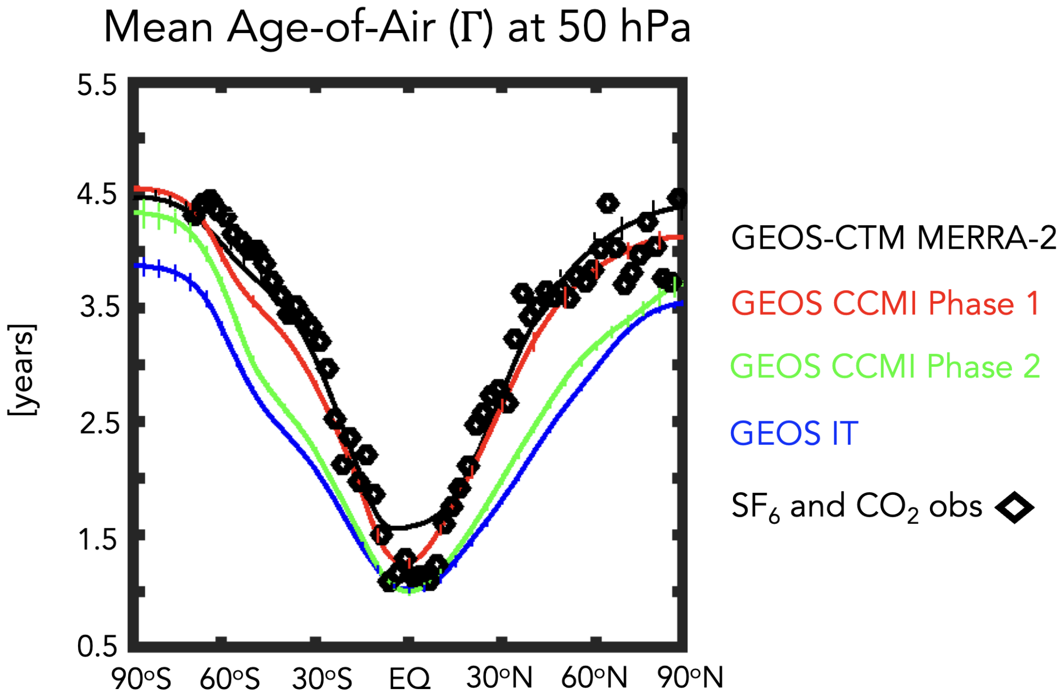

Figure 1. The 2000-2010 climatological annual mean meridional profile of the stratospheric mean age-of-air (Γ), evaluated at 50 hPa. Results from a GEOS-CTM integration constrained with MERRA-2 meteorological fields (black line) as well as free-running GEOS simulations using a model configuration for CCMI Phase 1 (red line), CCMI Phase 2 (green line) and the GEOS-IT Version 5.29.4 dating from October 19, 2022 (hereafter simply "GEOS-IT", blue line) are shown. All simulations are constrained with the same (observed) historical sea surface temperatures. Diamonds correspond to SF_6 and CO_2 in situ based estimates of Γ from Boering et al. (1996) and Engel et al. (2009). Vertical dashed lines denote $\pm\sigma$, the standard deviation of Γ over 2000-2010, for each model simulation.

We begin by comparing 10-year (2000-2010) climatological mean zonally averaged age-of-air profiles at 50 hPa across this subset of model versions, derived from 30-year long atmosphere-only (AMIP) integrations constrained with observed sea surface temperatures (Figure 1). First, we note that the profiles for the CCMI Phase 1 version of the model are very close to observations (black stars), consistent with the "GEOSCCM" documented age characteristics reported in Dietmüller et al. (2018) (see their Figure 3). In addition, while passive tracers were not integrated within MERRA-2, results using the GEOS chemistry transport model (GEOS-CTM, Kouatchou et al. (2015)) constrained with MERRA-2 meteorological fields (black line) also exhibits good agreement with observed values. This good agreement between the CTM-generated age-of-air and the observations is consistent with results from a previous GEOS-CTM simulation (constrained with MERRA) as documented in Orbe et al. (2017).

Moving to more recent development versions of the model (green and blue lines), however, reveals a reduction in the mean age by $\sim$ 1 year over both southern and northern high extratropical latitudes, or a decrease of $\sim$ 20-30% relative to the MERRA-2 constrained simulation and the observations. As discussed earlier, the green line refers to the CCMI Phase 2 model version, whereas the blue line refers to GEOS-IT. Note that this decrease in the climatological age in both model versions far exceeds the (internal) variations in mean age that occur interannually (vertical bars on solid lines).

Since MERRA-2, numerous updates have been introduced in the GEOS model. This includes replacing the Chou Suarez (Chou & Suarez, 1994) and Chou (Chou, 1990, 1992) radiation codes with the Rapid Radiative Transfer Model for GCMs (RRTMG; Iacono et al. (2008)), which impacts the stratosphere; the introduction of the Grell-Freitas deep-convection code (Grell and Freitas (2014); Freitas et al. (2018)), which has a minimal impact on this study; and upgrades to the vertical remapping option from that used in MERRA-2 (hereafter denoted as GMAO FV3 Cubic) to the current updated GDFL remapping option (hereafter denoted as GFDL FV3 Cubic). Note that both MERRA-2 and current systems use the GFDL FV3 dynamical core (Lin et al., 2017), but differ in terms of this remapping option.

As we show in Section 3, the latter transition from the remapping scheme used in MERRA-2 (GMAO FV3) to the scheme used in the current GFDL FV3 core has the largest impact on the simulated age-of-air values. This degradation in simulated transport within the current GFDL FV3 core primarily reflects differences in the implementation of the remapping algorithm, which is used to vertically interpolate individual material surfaces after each horizontal advection step back to the model's reference Eulerian coordinate. In its implementation in MERRA-2 within the GMAO FV3 core (Table 1), this remapping involves 1) fitting piecewise parabolic (hereafter PPM) functions to input layer-mean values of U, V, Q and tracers; 2) calculating PPM functions to output layer edges; 3) integrating PPM functions between output layer edges to produce new layer-mean values of U, V, Q and tracers; 4) calculating total energy (TE) at input mid-layer pressures; 5) calculating TE at output mid-layer pressures using cubic interpolation and applying an a-posteriori integral conservation; and, finally, 6) remapping temperatures from total energy via T = (TE - K - Φ)/C_p. Here T, U, V, Q, C_p, K and Φ correspond to temperature, zonal wind, meridional wind, specific humidity, specific heat capacity and kinetic and potential energy, respectively.

Differences in the implementation of this remapping algorithm within the current GFDL FV3 core results in degraded simulated stratospheric transport. To demonstrate this in as clean a fashion as possible, we use targeted experiments with the GMAO FV3 system to show that the degradations in simulated transport when using the remapping option from the GFDL FV3 core resemble changes that occur when using a lower order (quadratic vs. cubic) interpolation scheme in Step 5 of the remapping algorithm (Table 1). This otherwise innocuous change in interpolation order in turn highlights the large sensitivities in transport that can occur even *within* a given finite-volume numerical scheme,

Table 1. GMAO FV3 Core Finite-Volume Remapping Algorithm: The remapping algorithm examined in this study controls how individual material surfaces are vertically interpolated back to standard pressure levels. Employing linear – and to a lesser extent – quadratic interpolation in Step 5 produces stratospheric transport characteristics that are more consistent with the most recent GEOS model configurations (green and blue lines, Figure 1), whereas a cubic interpolation is more consistent with older configurations (red and black lines, Figure 1) and with MERRA-2. Here T, U, V, Q, C_p, K and Φ correspond to temperature, zonal wind, meridional wind, specific humidity, specific heat of air at constant pressure and kinetic and potential energy, respectively.

Step	REMAP Procedure (MERRA-2, GMAO FV3 Core)
1	Fit PPM functions to input layer-mean U, V, Q and tracers
2	Calculate PPM to output layer edges
3	Integrate PPM functions between output layer edges to produce new layer-mean U, V, Q and tracers
4	Calculate $TE = C_pT + K + \Phi$ at input mid-layer pressures
5	Calculate TE at output mid-layer pressures using cubic interpolation and a-posteriori integral conservation
6	Construct "remapped" T via $T = (TE - K - \Phi)/C_p$

in our case resulting in large differences in lower stratospheric upwelling and a $\sim 30\%$ reduction in the simulated mean age of air.

2.2 Model Experiments

2.2.1 AMIP vs. EMIP

We begin our analysis by interpreting the results shown in Figure 1, which are all based on historical AMIPs that were performed at the same cubed sphere C180 (approximately half-degree) horizontal resolution. As they represent more official model versions they serve as an important motivation for the experiments that follow. However, there are numerous (potentially compensating) development changes between these model versions which renders it nearly impossible to cleanly identify drivers of differences in their simulated transport.

To this end, in order to investigate the drivers of the differences in Figure 1 we perform targeted modeling experiments aimed at disentangling the influence of individual model development changes on stratospheric transport properties (Table 2). In order to evaluate impacts on transport climate statistics, we consider both a set of climatological AMIP (rows 1-4) as well as so-called "EMIP" (rows 5-7) experiments.

In particular, we carry out 30-year-long AMIP simulations at C180 resolution which we use to infer the climate characteristics of the different model configurations. The "EMIP" experiments – ensembles of 3-month-long integrations initialized on approximately November 15 of each year between 1985 and 2015 – are also used to infer impacts on simulated transport climate. As they are more computationally efficient than AMIPs since all 30

Table 2. GEOS Model Experiments: Targeted GEOS model experiments based off a control experiment (row 1) were carried out to identify the influence of radiation (row 2) and changes in the remapping algorithm used since MERRA-2 (row 3), as well as their combined influence (row 4). The influence of the remapping algorithm changes is then interpreted using a simpler set of sensitivity experiments, performed using the GMAO FV3 core, in which only the order of the interpolation scheme used to calculate TE at output mid-layer pressure levels is altered (rows 5-7). Experiments in rows 1-4 are 30-year-long AMIPs run at C180 resolution, whereas rows 5-7 refer to 30-member 3-month-long (DJF) EMIP experiments. Both AMIPs and EMIPs are used for climate statistic evaluation (see Appendix B for more on the correspondence between the two). EMIP experiments are run at both C180 and C360 horizontal resolutions.

Experiment Name	Configuration	Experiment Type	Hor. Resolution
CTRL	Control, GFDL FV3 Core	AMIP (30 yrs.)	C180
CSRAD	Chou-Suarez (1994) Radiation (RAD)	AMIP (30 yrs.)	C180
M2REMAP	GMAO FV3 Core (cubic)	AMIP (30 yrs.)	C180
CSRAD+M2REMAP	Chou-Suarez (1994) RAD GMAO FV3 Core (cubic)	AMIP (30 yrs.)	C180
LINEAR	GMAO FV3 Core (linear)	EMIP (30 mem.)	C180, C360
QUADRATIC	GMAO FV3 Core (quadratic)	EMIP (30 mem)	C180, C360
CUBIC	GMAO FV3 Core (cubic)	EMIP (30 mem)	C180, C360

3-month-long integrations may be run in parallel, they are performed at both C180 and C360 (approximately quarter-degree) resolutions in order to examine the sensitivity of our results to changes in horizontal resolution.

As shown in Appendix B, comparisons of the December-January-February (DJF) vertical profile of tropical upwelling show excellent agreement between EMIP and AMIP integrations carried out using the same model configuration (Appendix Figure B1). This somewhat incidental result represents, to the best of our knowledge, the first time that EMIP-based statistics have been shown to converge well to those from AMIPs for the stratospheric metrics considered in this study. This suggests that EMIPs, relative to AMIPs, may be used to provide a computationally more efficient initial assessment of the impacts of model changes on the stratospheric circulation.

2.2.2 Model Development Changes

Moving next to the precise model development changes examined, we begin by defining a control experiment (CTRL; Table 2, row 1), which best corresponds to the blue line shown in Figure 1. Then we define three new AMIP experiments, which aim to distinguish between the age-of-air changes resulting from changes in radiation versus changes in the handling of the REMAP algorithm that occurred in the transition from MERRA-2 to the current FV3 cores (Section 3.2.1).

Specifically, these include experiments in which we a) replace RRTMG with the radiation from Chou and Suarez (1994) (CSRAD; Table 2, row 2), b) replace the current FV3 REMAP approach with the settings used in the GMAO FV3 core when running MERRA-2 (i.e., M2REMAP; Table 2, row 3) and c) combine these two changes (CSRAD+M2REMAP; Table 2, row 4).

As we show in Section 3.2.1, the M2REMAP experiment produces larger changes in age-of-air, compared to the experiment in which only the radiation is altered (CSRAD). Interpreting this result, however, is not straightforward since there are several differences in the implementation of the remapping algorithm between the GMAO and GFDL FV3 cores that are interdependent and, thus, difficult to isolate cleanly. To this end, in order to simplify the problem we focus the remainder of our investigation (Section 3.2.2) on examining a clean set of EMIP experiments that are all performed using the GMAO FV3 core remapping settings and that differ from each other only in terms of the order of the interpolation that is used to calculate TE at the mid-layer pressure levels (Step 5, Table 1). More precisely, we compare configurations using a linear (LINEAR; Table 2, row 5), quadratic (QUADRATIC; Table 2, row 6) and cubic interpolation (CUBIC; Table 2, row 7) scheme, with the latter corresponding to the approach that was used in MERRA-2. To assess the robustness of our findings to changes in horizontal resolution, all three sensitivity experiments are run at both C180 and C360 resolutions.

These three numerical schemes are derived from the generic Lagrangian interpolation equation:

$$\mathcal{Q}(\mathcal{P}) = \sum_{k'} a_{k+k'}\,\mathcal{Q}_{k+k'} \qquad k' = 0, \pm 1, \pm 2, \ldots \tag{1}$$

where $\mathcal{P}$ represents the target output location in $\ln(p)$ and $\mathcal{Q}_{k+k'}$ denotes the surrounding grid-point values at input locations. The coefficients $a_{k+k'}$ are derived through Taylor Series expansions using non-uniform grid spacing given by:

$$a_{k+k'} = \frac{\prod\limits_{m} (\mathcal{P}_{k+m} - \mathcal{P})}{\prod\limits_{m} (\mathcal{P}_{k+m} - \mathcal{P}_{k+k'})} \qquad m = 0, \pm 1, \pm 2, \ldots \quad m \neq k' \tag{2}$$

For the three schemes, the grid points used are: LINEAR $(k, k-1)$, QUADRATIC $(k+1, k, k-1)$, and CUBIC $(k+1, k, k-1, k-2)$. In all cases, the grid points are chosen such that the target location resides between layers k and $k-1$.

Note that, while the LINEAR and QUADRATIC experiments do not actually correspond to any of the model versions shown in Figure 1, they highlight the large sensitivity of the mean age to changes in the interpolation scheme that may otherwise seem innocuous. They also provide further evidence of the strong influence of changes in tropical lower stratospheric upwelling strength on the stratospheric mean age in GEOS.

Finally, in all experiments using the MERRA-2 remapping approach (i.e., M2REMAP, CSRAD+M2REMAP, LINEAR, QUADRATIC, CUBIC) additional modifications to the divergence damping coefficients were used so as to best ensure consistency with what was used in MERRA-2. Specifically, these include changes to the number of layers for vertical subgrid mixing, the coefficient for barotropic mode damping, the use of 2^{nd} vs. 6^{th} order divergence damping and the strength of the divergence damping coefficients. Compared to the impact of moving from the CTRL to M2REMAP experiments, these additional changes have a much smaller impact on lower stratospheric upwelling in the model (not shown) and hereafter will not be discussed further.

2.3 Analysis Approach

2.3.1 Transport Diagnostics

To diagnose the transport circulation we focus primarily on the age-of-air (Hall and Plumb (1994)). This is inferred from an idealized global "clock" or ideal age tracer (Γ) (Thiele and Sarmiento (1990)) that is defined with respect to the bottom model level as follows: initially, the ideal age tracer is set to zero throughout the troposphere and thereafter held to zero over the entire Earth's surface, subject to a constant aging of 1 year/year throughout the atmosphere. We present here the statistically stationary (equilibrated) value of $\Gamma(\mathrm{r})$, which is equal to the average time since the air at a location r in the stratosphere last contacted the Earth's surface. In addition to the mean age, we also show results from an idealized e90 tracer that is uniformly emitted over the entire surface layer and decays exponentially at a rate of 90 days^{-1} such that concentrations greater than 125 ppb and less than 50 ppb tend to reside in the lower troposphere and stratosphere, respectively (Prather et al. (2011)). As this tracer features strong near-tropopause gradients and takes significantly less time to equilibrate, compared to the mean age, it is useful for evaluating stratosphere-troposphere-exchange and transport within the upper troposphere/lower stratosphere (Abalos et al. (2017, 2020); Orbe et al. (2020)).

Both the mean age and e90 tracers were integrated in all of the AMIP experiments shown in Figure 1 and listed in Table 2 (rows 1-4), which were run using the same idealized passive tracer package described in Orbe et al. (2017). Note that the mean age tracer was not integrated in the EMIP experiments given its much longer characteristic timescale in the stratosphere ($\sim$ 3-5 years). As such, the EMIP simulations, which do not exceed one year, are not appropriate for evaluating the time-integrated transport characteristics reflected in the age-of-air.

In addition to carrying the idealized tracers, two of the experiments shown in Figure 1 were also run with full interactive chemistry and correspond to the two CCMI (Phase 1 and Phase 2) integrations (red and green lines, Figure 1). Both simulations employ the same Global Modeling Initiative (GMI) chemical mechanism (Strahan et al. (2013)) and are therefore useful in evaluating the impact of age differences on real trace gas distributions. In particular, as shown in Section 3.1 results from these experiments show significant imprints of the age-of-air changes on nitrous oxide (N_2O) and methane (CH_4).

2.3.2 Circulation Diagnostics

As we show in Section 3, the changes in age-of-air across the different model versions are strongly tethered to changes in the advective component of the circulation, which we quantify using the Transformed Eulerian Mean (TEM) estimate of the Lagrangian transport of mass by the circulation. Thus, in addition to more standard Eulerian metrics of the circulation (e.g., zonal winds and temperatures), we focus on the vertical component of the TEM residual velocity, defined as $\overline{w}^* = \overline{w} + \frac{\partial(\psi cos\phi)}{acos\phi\partial\phi}$, where $\psi = \overline{v'\theta'}/\frac{\partial\overline{\theta}}{\partial p}$ is the eddy stream function, θ refers to potential temperature, a is the Earth's radius and overbars and primes denote zonal means and deviations therefrom, respectively (Andrews et al. (1987)). In addition, we interpret the behavior in w^* using the Eliassen-Palm flux divergence ($\nabla\cdot$F), whose horizontal (F(ϕ))and vertical (F(p)) components are respectively defined as F(ϕ) = $acos\phi[\frac{\partial u}{\partial p}\psi - \overline{u'v'}]$ and F(p) = $acos\phi([f - \frac{\partial\overline{ucos\phi}}{acos\phi\partial\phi}]\psi - \overline{u'\omega'})$.

2.4 Observations and Reanalyses

While our focus is on interpreting and understanding the different model configurations, we incorporate observations to provide context when possible, although we do not present an exhaustive evaluation of the model's transport characteristics (for that see earlier studies including Orbe et al. (2017, 2018)). However, as the tracers are not directly integrated in MERRA-2 (with the exception of ozone), we compare against independent observational estimates. For the mean age we first compare simulated meridional age profiles at 50 hPa with values derived from in situ aircraft measurements of carbon dioxide (CO_2), averaged in 2.5 degree latitude bins over the altitude range 19.5 to 21.5 km (Boering et al. (1996), see also Figure 5 in Hall et al. (1999)).

We also briefly evaluate impacts of transport biases on the simulated trace gas distributions for the CCMI Phase 1 and 2 experiments. The simulated fields of methane (CH_4) are compared with the climatologies derived for 1991–2002 from the Halogen Occultation Experiment (HALOE) on board the Upper Atmosphere Research Satellite (UARS) (Grooß and Russell III (2005)). Comparisons of simulated nitrous oxide (N_2O) are made

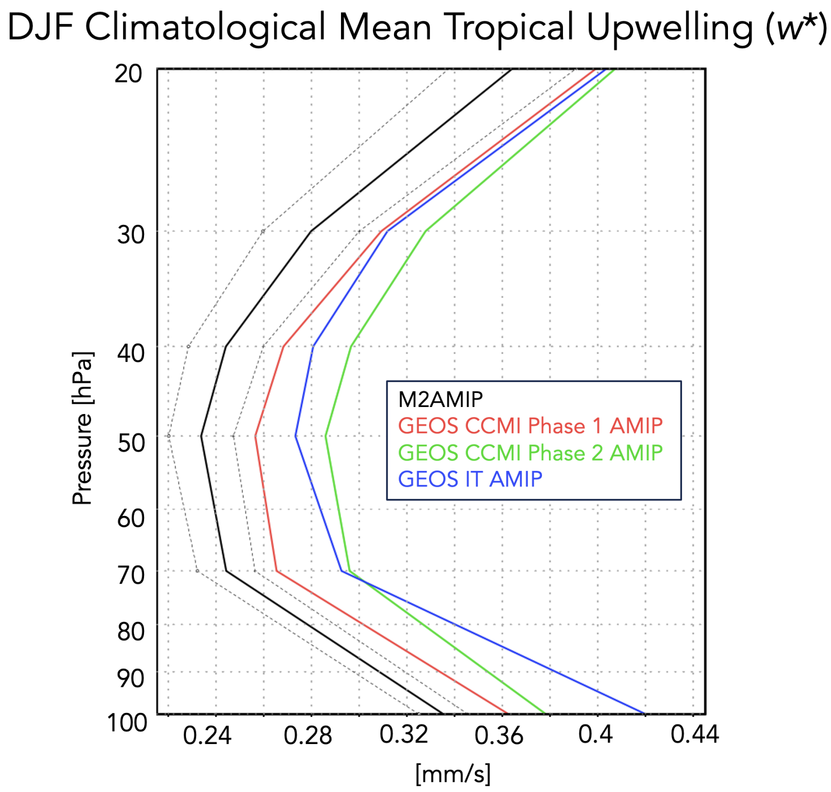

Figure 2. The DJF 1985-1994 climatological mean vertical residual mean velocity, w^*, averaged at each level between the turnaround latitudes for GEOS free-running AMIP simulations using the model configurations corresponding to the CCMI Phase 1 (red) and Phase 2 (green) submissions and to GEOS-IT (blue). M2AMIP is shown in black, with black dashed lines denoting ±1 standard deviation.

against climatologies derived from the Microwave Limb Sounder (MLS) on the Earth Observing System (EOS) Aura satellite. Climatologies over the same period (2005–2015) are used to evaluate both the model and the observations. We use the 190-GHz retrieval from Version 4.2 because the 640-GHz data set ends in summer 2013 due to the failure of the N_2O primary band.

For the circulation diagnostics nearly all comparisons are made relative to the MERRA-2 data assimilation (DAS) reanalysis product, noting that comparisons against ERA-5 (not shown) reveal a similar picture. One exception, however, is the vertical component of the TEM circulation (w^*), which shows some differences in vertical structure between the MERRA-2 DAS and a 30-member ensemble of (free-running) AMIP integrations produced using the MERRA-2 model, hereafter referred to as M2AMIP (Collow et al., 2017)(Appendix Figure A1, right). As the free-running model results shown in Figure 1 show more consistency with the vertical profile of M2AMIP, not MERRA-2, we compare w^* in all free-runnning GEOS experiments with M2AMIP, noting that for non-derived measures (i.e., winds, temperatures), the raw MERRA-2 output is used.

The differences in w^* between M2AMIP and the MERRA-2 DAS may reflect the influence of temperature increments in the DAS (MERRA-2) which can drive spurious vertical transport in assimilated products (Weaver et al., 1993; Orbe et al., 2017). In particular, Weaver et al. (1993) showed that the imbalance between the thermal and velocity fields at the time an observation is ingested during the assimilation cycle can excite unwanted inertial-gravity wave modes that manifest strongly in the residual vertical winds. While this impact of the increments may explain the differences in w^*, we reserve a more-in depth examination for future work as our focus in on the transport characteristics of the free-running GEOS system.

3 Results

3.1 Reduction of Stratospheric Mean Age in GEOS Models Since MERRA-2

We begin by interpreting the reduction in mean age exhibited in more recent model versions in terms of changes in the strength of upwelling in the tropical lower stratosphere. In particular, the reductions in Γ (Figure 1) are consistent with increases in the strength of lower stratospheric tropical upwelling, with w^* becoming progressively stronger in more recent model versions, relative to MERRA-2 (Figure 2). Note that these increases in w^* across model versions exceed those due to internal variability alone (black dotted lines, Fig. 2). Furthermore, while the increases in w^* occur throughout the stratosphere, we focus on the changes occurring between 70 and 100 hPa as these are most relevant to determining the tropical upward mass flux and associated strength of the mean overturning circulation.

Though perhaps naive, the relationship between lower stratospheric upwelling and the mean age suggested by comparing Figure 1 and Figure 2 is consistent with the long-term behavior of Γ inferred from both historical and projected future climate simulations (Butchart et al. (2010); Abalos et al. (2021)). A strong relationship between the strength of lower stratospheric ascent and the mean age was also shown to hold in the CCMVal models (see Fig. 5.20 in J. Neu et al. (2010)). Nevertheless, it is important to note that a clear relationship between w^* and Γ is not a priori expected, as the age-of-air is also known to be very sensitive to mixing, which may be important in interpreting differences among the CCMI Phase 1 models (Dietmüller et al. (2018)).

The differences in w^* highlighted in Figure 2 are associated with enhanced Eliassen-Palm flux convergence over NH midlatitudes (Figure 3). Increased wave convergence is evident not only within the subtropical lower stratosphere ($< 30°$N, 50-100 hPa) but also over higher latitudes and altitudes ($\sim 40°$-70°N, 20-50 hPa). The fact that differences

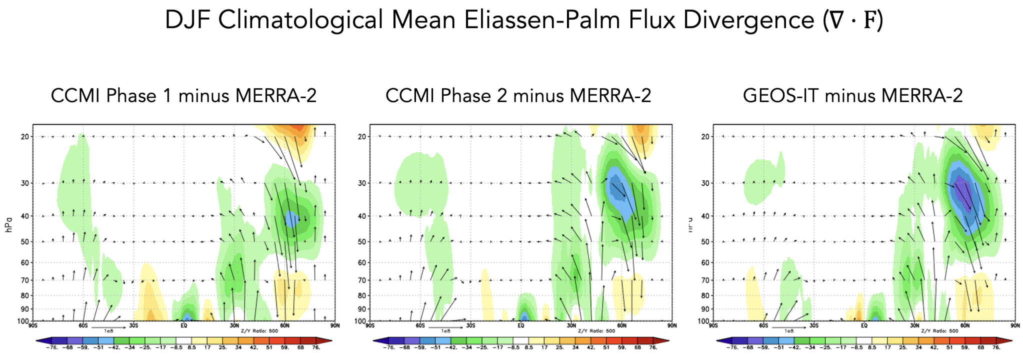

Figure 3. Colors show anomalies in the DJF climatological mean Eliassen-Palm (EP) flux divergence between the CCMI Phase 1 (left), CCMI Phase 2 (middle) and GEOS-IT AMIP (right) model versions, relative to MERRA-2. Arrows denote anomalies in the vertical and meridional EP flux vectors (relative to MERRA-2).

in extratropical wave convergence imprint on tropical upwelling is consistent with our understanding of the so-called "downward control" principle (Haynes et al. (1991)).

In particular, the strength of the residual mean streamfunction (Ψ^*) is, via downward control, directly related to the vertically integrated eddy-induced total zonal force above that level and has contributions both from the (resolved wave) Eliassen-Palm flux divergence (Figure 3) as well as the gravity wave drag scheme's parameterized waves (not shown). The tropical upward mass flux – defined as Ψ^*_{max}-Ψ^*_{min} evaluated at the turnaround latitudes (e.g. Rosenlof (1995)) – is therefore directly dependent on the wave forcing aloft.

One subtlety to note is that the wave convergence changes shown in Figure 3 occur at high latitudes and are directly associated with downwelling over the polar region. It is then via mass balance that anomalously strong downwelling associated with enhanced flux convergences must be accompanied by enhanced upwelling in the tropics. This indirect impact of higher latitude wave drag reflects an "extratropical pumping" mechanism (Holton et al., 1995), which is illustrated more clearly in Section 3.2.2 in the context of the LINEAR, QUADRATIC and CUBIC experiments.

While the reduction in Γ (Figure 1) of $\sim 30\%$ at 50 hPa is significant, it is neither clear if this change is representative of other altitudes within the stratosphere nor how this age bias imprints on real chemical species. To this end, we begin by comparing the full latitude-pressure distribution of changes in Γ and another passive tracer (e90) (Figure 4) between the CCMI Phase 1 and Phase 2 model configurations (red and green lines, Figure 1). In particular, we find that the changes in both passive tracers – large reductions in Γ within both hemispheres (Fig. 4, top right) and increased values of e90 within the lower stratosphere (Fig. 4, bottom right) – are reflective of an overall increase in the strength of the transport circulation. This is highlighted in the CCMI Phase 2 – 1 model differences for the passive tracer distributions (Fig. 4, right panels) which are shown in the absence of robust observational constraints of Γ at higher altitudes (or any observational constraints for e90, for that matter). The reduced/increased stratospheric burdens of the age and e90 tracers are consistent with stronger upwelling in the CCMI Phase 2 model configuration (Figure 2).

While the observational constraints on Γ presented in Figure 1 and the departure of w^* away from MERRA-2 suggest that transport properties of the newer model configurations are moving in the wrong direction, it is relevant to ask whether or not the

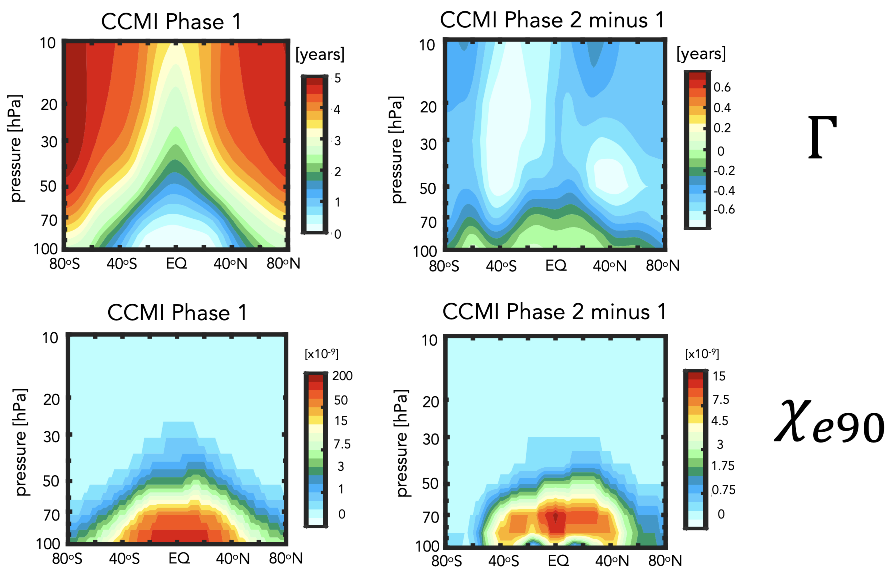

Figure 4. The climatological mean (2000-2010) distribution of the mean age-of-air (Γ) (left, top) and e90 idealized tracers (left, bottom) for the CCMI Phase 1 model configuration. Climatological differences between the CCMI Phase 2 and Phase 1 model configurations are shown in the right panels. Note that a nonlinear colorbar has been used in the e90 subplots.

trace gas satellite measurements also support this conclusion. Indeed, comparisons with observations show larger biases in N_2O (Fig. 5, top panels) and CH_4 (Fig. 5, bottom panels), increasing from 10% to 30% in the CCMI Phase 2 model configuration, depending on the species. Recall that the same chemistry mechanism is used in both CCMI Phase 1 and 2 simulations.

The patterns of the trace gases biases are generally consistent with the biases in the mean age (Fig. 4). This comports with well-known correlations between the mean age and stratospheric trace gases, reinforcing the fact that model transport inaccuracies can significantly affect simulations of important long-lived chemical species in the stratosphere (Hall et al. (1999)).

The fact that the mean age changes have a significant imprint on the simulated trace gases is consequential for the GEOS-R21C system. However, the configurations shown in Fig. 1-5 differ in many respects (physics, resolution, radiation, FV remapping algorithm) and it is difficult to meaningfully interpret what is driving the changes in w^* (and the tracers). We therefore move next to the targeted model experiments (Table 2) in order to interpret the model development steps that resulted in these transport circulation changes.

3.2 Identifying Drivers of Upwelling and Tracer Changes Since MERRA-2

3.2.1 Radiation versus REMAP Algorithm

As discussed in Section 2, among the model changes that were made since MERRA-2, the changes in radiation and the FV remapping algorithm are most likely to directly have impacted the stratospheric circulation. We therefore begin by assessing which of these changes dominates the decreases in Γ shown in Figure 1.

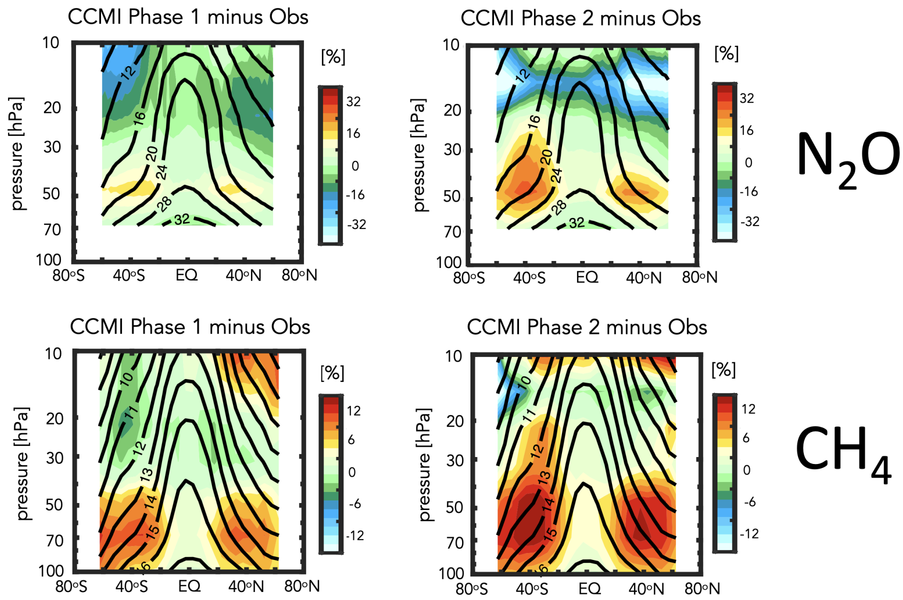

Figure 5. Colors shown anomalies in the simulated distributions of nitrous oxide (N_2O) (top) and methane (CH_4) (bottom), relative to the MLS and HALOE observed values, respectively, for the CCMI Phase 1 (left) and Phase 2 (right) GEOS model configurations. Climatological mean observed values are shown in the black contours.

Figure 6 shows the distribution of Γ for experiments in which the longwave, shortwave, and REMAP updates since MERRA-2 have successively been undone. Relative to the control experiment (CTRL; Table 2, row 1), the transition back to Chou and Suarez (1994) in the shortwave and Chou (1990, 1992) in the longwave results in an increase in the mean age of $\sim$ 0.5 years throughout the stratosphere (CSRAD; Table 2, row 2). Though significant, this change in Γ is smaller than the change that results from applying the remapping approach used in MERRA-2 (M2REMAP; Table 2; row 3), in which the mean age increases by $\sim$ 1 year. The combined impacts of both changes (CSRAD+M2REMAP; Table 2 row 4) is roughly linear, with age values of $\sim$ 5.5 years over high latitudes at 50 hPa, consistent with the values simulated by the GEOS-CTM MERRA-2 integration (black line, Figure 1) and with the CCMI Phase-1 version of the model (red line, Figure 1).

Next we ask if the behavior of Γ exhibited in Figure 6 can be interpreted in terms of changes in the strength of lower stratospheric tropical upwelling and extratropical wave convergence, as our previous analysis of the CCMI experiments suggested. Indeed, Figure 7 shows that values of upwelling decrease in the CSRAD and M2REMAP experiments, relative to the CTRL integration. The increase in upwelling resulting from both changes (CSRAD+M2REMAP) is still larger, consistent with the larger age decreases in that experiment. This change in the behavior of w^* within the tropical stratosphere can be interpreted in terms of changes in the Eliassen Palm flux convergence over NH midlatitudes (not shown), which features smaller values in the CSRAD, M2REMAP (and CSRAD+MSREMAP) experiments. Note that our examination of the changes in w^* are derived from EMIP integrations, which we showed previously converge (for DJF) to the statistics derived from corresponding AMIP experiments.

It is important to note that, while the reduced values of w^* in the M2REMAP experiment (Fig. 7), represent an improvement, relative to the CTRL integration, these

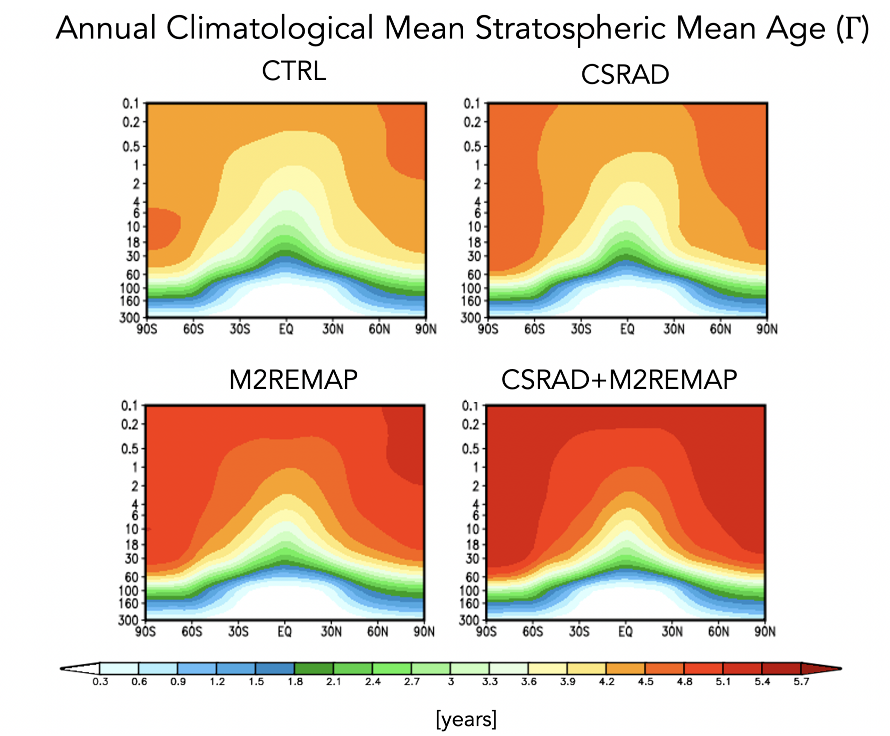

Figure 6. Colors show the simulated 2000-2010 climatological annual mean distributions of the mean age-of-air (Γ) for the CTRL (top left; Table 2, row 1), CSRAD (top right; Table 2, row 2), M2REMAP (bottom left; Table 2, row 3) and combined CSRAD+M2REMAP (bottom right; Table 2, row 4) experiments.

changes are, in isolation, accompanied by a cold bias (∼4-6 K) in the tropical lower stratosphere (Fig. 8a). Unlike the upwelling changes, this represents a degradation in model skill, increasing the temperature bias in that region, relative to MERRA-2 (Fig. 8 b,c). While this temperature bias is concerning in isolation, it is decoupled from the changes in residual mean upwelling which, rather, are more sensitive to the changes in *extratropical* wave convergences (described further in the next section). Furthermore, the increased temperature bias in the M2REMAP experiment appears to be tethered to other updates that were made in the CTRL model (specifically the radiation progression from Chou-Suarez to RRTMG). In particular, the M2AMIP ensemble, which also employs the GMAO cubic remapping option, features a much smaller temperature bias in the tropical lower stratosphere, relative to MERRA-2 (∼ 0-2K) (Fig. 8d). This suggests that the amplified temperature bias moving from the CTRL to M2REMAP experiments needs to be interpreted in the context of other model development changes that were introduced, particularly in the radiation scheme and its coupling to convection.

3.2.2 FV REMAP Algorithm: Sensitivity of Climate Statistics

Having shown in the previous section that the largest changes in the mean age were realized through the differences in implementation of the remapping algorithm between the GMAO FV3 core used in MERRA-2 and in current FV3 core configurations, we now investigate further the sensitivity of the transport circulation to the choice of remapping interpolation scheme. In particular, we compare simulations run with the GMAO FV3 remapping settings in which total energy is calculated at new mid-layer pressures using cubic, quadratic and linear interpolation prior to the aposterior integral conservation (Table 2, rows 5-7). In addition, in this section we seek to understand how the changes in the Eliassen-Palm flux convergence over NH midlatitudes arise via analysis of the large-scale wind structure.

Figure 9 (left panel) shows a clear sensitivity in tropical upwelling to the choice of interpolation scheme, with w^* progressively increasing in strength moving from the CUBIC to QUADRATIC to LINEAR schemes. This sensitivity is robust across horizontal resolutions as the same suite of experiments performed at C360 exhibit the same sensitivity (Fig. 9, right panel). While no current model version actually employs a linear scheme, this suite of experiments highlights the strong sensitivity to choice of interpolation scheme within the remapping algorithm; to the best of our knowledge, this result has not been reported in the literature.

Furthermore, as we show next, this clean set of experiments allow us to inquire mechanistically into the processes that are driving the changes in wave convergence over midlatitudes, unencumbered by differences in horizontal resolution, physics, etc.

Interestingly, the increases in w^* moving from the CUBIC to QUADRATIC schemes not only manifests in free-running AMIP simulations, but also in AMIP simulations in which GEOS is constrained (or replayed, following Orbe et al. (2017)) to MERRA-2 meteorological fields, using the MERRA-2 GMAO cubic interpolation (Fig. 10, blue line) and the GFDL FV3 core remapping approach (Fig. 10, red line). While there is a general increase in w^* in the former, however, both simulations, lie within the range of MERRA2-DAS, suggesting that replay does act to ameliorate some of the upwelling biases manifest in the underlying unconstrained models. While not the focus of this study, this impact of the remapping approach on simulations run in replay mode will be examined further in future work.

Consistent with our expectations based on the analysis of the previous experiments, the drivers of the changes in w^* are related to increased wave convergence moving from the CUBIC to QUADRATIC to LINEAR schemes (Figure 11). Over extratropical latitudes, the zonal force associated with this enhanced wave convergence is associated with enhanced downwelling at high latitudes that, through mass balance, is accompanied by

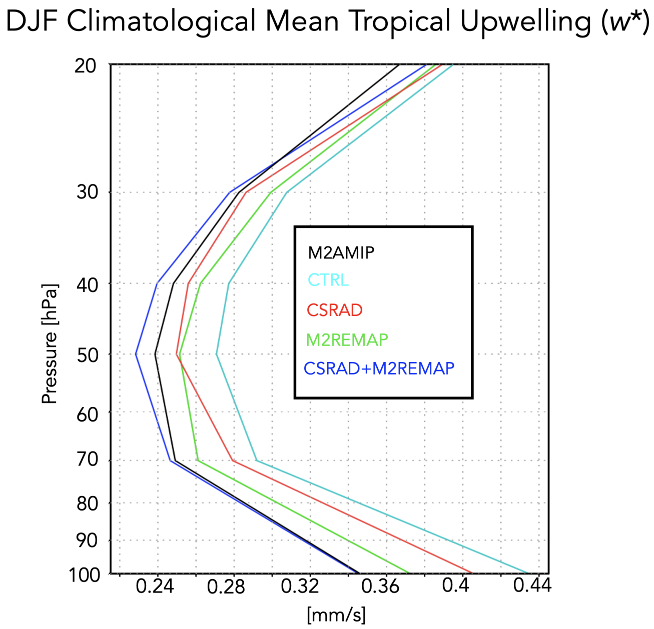

Figure 7. The DJF 1985-2015 climatological mean vertical residual mean velocity, w^*, averaged at each level between the turnaround latitudes for the CTRL (cyan line; Table 2, row 1), CSRAD (red line; Table 2, row 2), M2REMAP (green line; Table 2, row 3) and combined CSRAD+M2REMAP (blue line; Table 2, row 4) experiments. M2AMIP is shown in black.

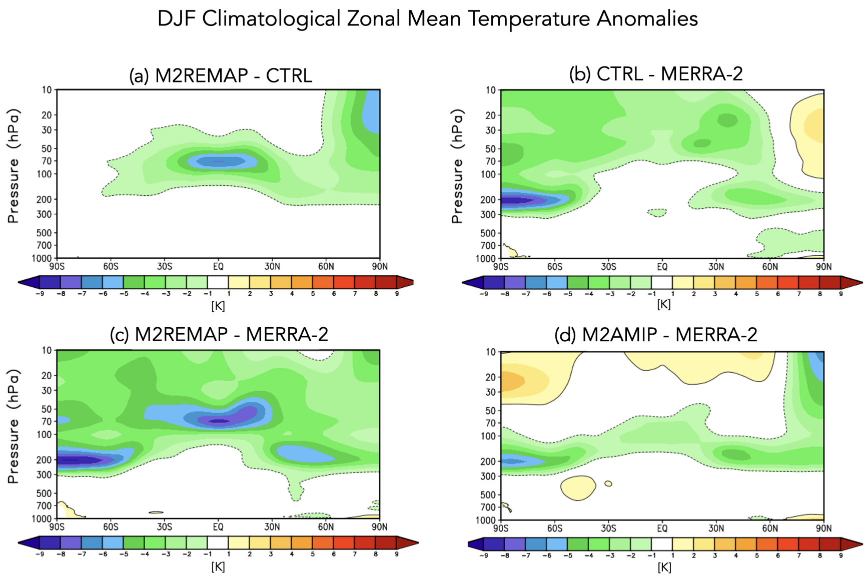

Figure 8. Anomalies in DJF 1985-2015 climatological mean zonal mean temperatures: M2REMAP - CTRL (a), CTRL - MERRA-2 (b), M2REMAP - MERRA-2 (c) and M2AMIP - MERRA-2 (d).

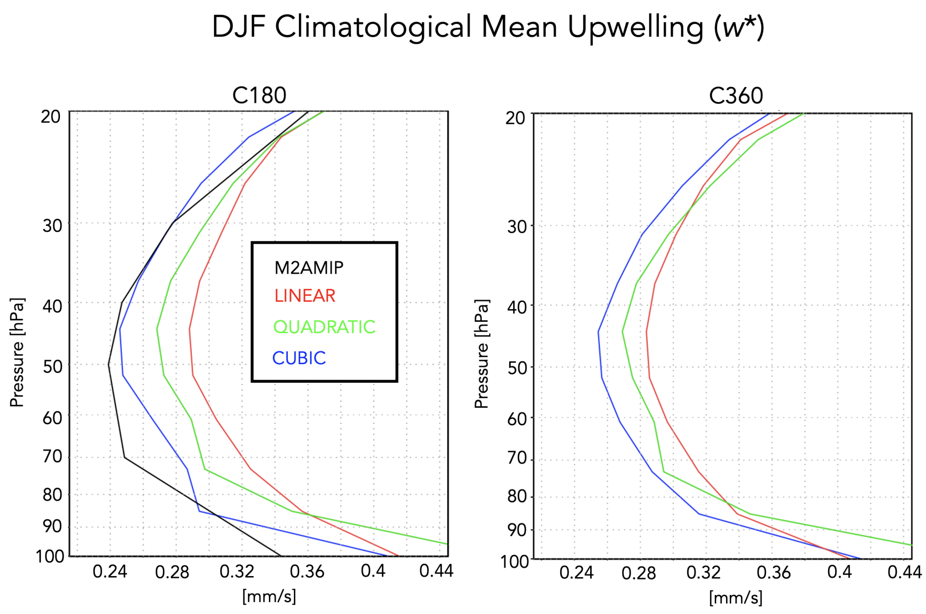

Figure 9. The DJF 1985-2015 climatological mean vertical residual mean velocity, w^*, averaged at each level between the turnaround latitudes for the LINEAR (red line; Table 2, row 5), QUADRATIC (green line; Table 2, row 6) and CUBIC (blue line; Table 2, row 7) experiments. M2AMIP is shown in black. Results from C180 and C360 EMIP experiments are shown in the left and right panels, respectively.

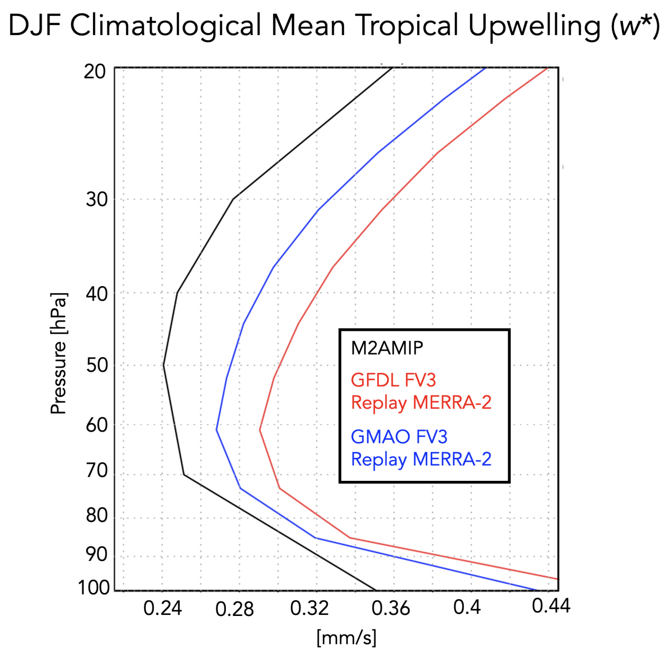

Figure 10. The DJF 2005-2015 climatological mean vertical residual mean velocity, w^*, averaged at each level between the turnaround latitudes for two GEOS replay AMIP simulations constrained with MERRA-2 meteorological fields using remapping approaches from the MERRA-2 GMAO FV3 cubic core (blue line) and the GFDL FV3 core (red line). MERRA-2 DAS is shown in black.

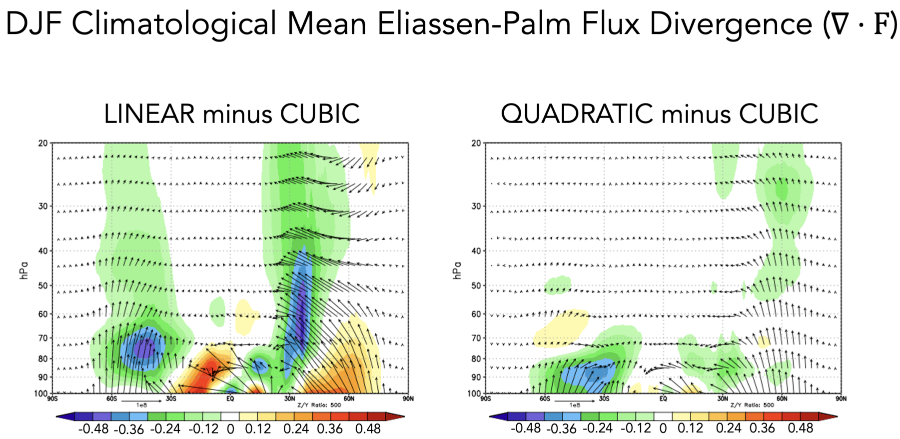

Figure 11. Colors shown anomalies in the DJF 1985-2015 climatological mean Eliassen-Palm (EP) flux divergence in the LINEAR (left) and QUADRATIC (right) experiments, relative to the CUBIC model experiment. Arrows denote anomalies in the vertical and meridional EP flux vectors.

enhanced upwelling in the tropics. This indirect impact of higher latitude wave drag is evident in Appendix Figure C1, which shows stronger upwelling/downwelling in the LINEAR and QUADRATIC experiments over the tropics/polar region.

Next we exploit the fact that these experiments only differ with respect to the interpolation scheme in order to inquire further into the drivers of the wave convergence changes. To this end, Figure 12 compares profiles of the zonal mean zonal wind between the CUBIC, QUADRATIC and LINEAR experiments, averaged over the region of enhanced wave convergence (i.e. 20°N-60°N). The experiments featuring stronger wave convergence (LINEAR and QUADRATIC) are also simulations with stronger zonal winds, relative to MERRA-2, especially above 70 hPa. This change in winds occurs at both C180 (Fig. 12, left panel) and C360 (Fig. 12, right panel) resolutions.

Structurally, the increase in zonal wind strength over northern extratropical midlatitudes is reflective of a poleward shift in the zonal winds as the critical latitude, i.e. where the zonal wind is zero, shifts northward in the QUADRATIC and, especially, LINEAR integrations, relative to the CUBIC experiment (Figure 13). Since stationary waves only propagate in westerly zonal flow, the latitude where zonal flow is zero acts a boundary for wave propagation (Hardiman et al. (2014)). As a result, this shift in critical latitude results in enhanced wave propagation and convergence over middle and high latitudes.

Figures 12 and 13 highlight how the changes in zonal winds in the LINEAR and QUADRATIC experiments reflect a degradation in model skill, relative to MERRA-2, throughout the entire stratosphere. The changes in upwelling, mean age, chemical trace gases and zonal winds thus provide a coherent and self-consistent picture suggestive of a degradation in the representation of the stratospheric circulation since MERRA-2. That is, an increased bias in the stratospheric northern zonal winds are, via their influence on wave convergence, compromising changes in the strength of the mean meridional overturning circulation and its impact on composition. It is interesting to note that the wind biases also extend into the troposphere and show degraded skill relative to MERRA-2 in the LINEAR and QUADRATIC experiments (Figure 13). Examination of other fields

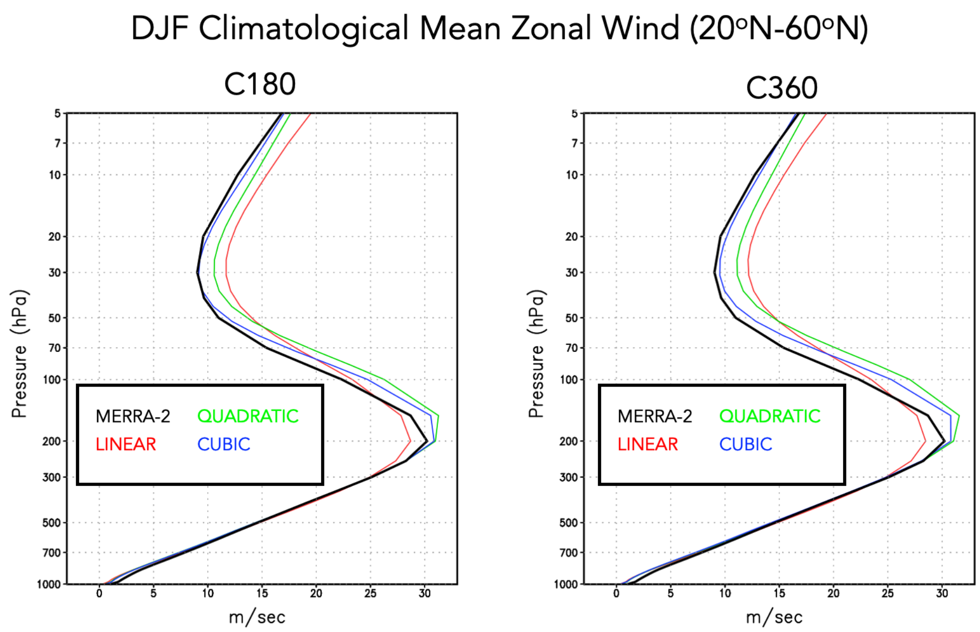

Figure 12. Vertical profiles of the DJF 1985-2015 climatological mean zonal mean zonal winds in the LINEAR (red), QUADRATIC (green) and CUBIC (blue) experiments, averaged between 20°N and 60°N. MERRA-2 is shown in the black line. Results for both C180 (left) and C360 (right) experiments are shown.

(i.e. tropopause biases, Appendix Figure D1) present somewhat more of a nuanced story that depends on other changes that made to the model since MERRA-2, as shown in Figure 8 and discussed in Section 3.2.1. The improvements in the zonal winds, however, are most relevant for setting the upwelling characteristics within the tropical lower stratosphere via their influence on wave propagation into that region.

Finally, to better understand why these impacts on the winds have such a consequence for the wave convergence properties within the stratosphere, next we examine the zonal structure of these biases in the middle stratosphere (Figure 14. This reveals that the enhanced winds in the LINEAR (and, to a lesser extent, QUADRATIC) integrations are concentrated over the North Pacific at both C180 (Fig. 14, top) and C360 (Fig. 14, bottom) resolutions (a similar picture emerges within the troposphere, not shown). As this region is the primary region dominating the stationary component of the upward flux of vertical wave activity (Plumb (1985), see their Figure 4) it is perhaps not surprising that this region is having a profound impact on the mean overturning circulation. Again, as with the zonal mean wind changes, the increases in wind strength over the North Pacific represent degraded model skill relative to MERRA-2. Note that comparisons with ERA-5 reveal a similar bias (not shown).

4 Conclusions

Here we have performed an analysis aimed at understanding differences in the representation of the stratospheric circulation in recent candidate systems for GEOS-R21C, relative to older versions of GEOS similar to the model used to produce MERRA-2. Using targeted experiments oriented at disentangling various model development updates, we have identified a key role played by changes in the implementation of the remapping

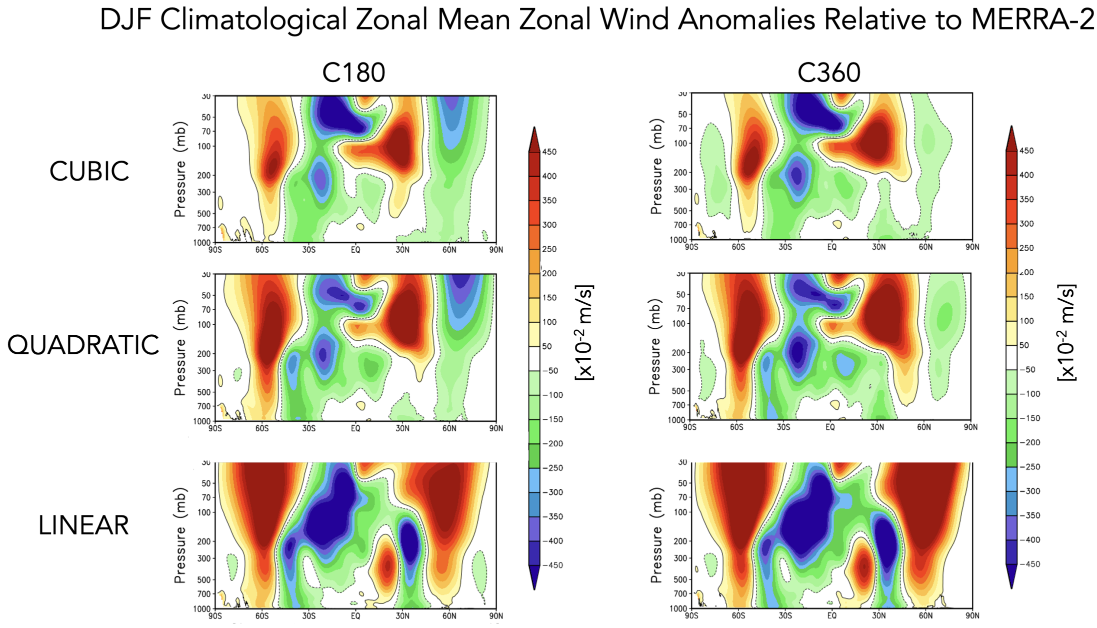

Figure 13. Colors shown anomalies in the DJF 1985-2015 climatological mean zonal mean zonal winds in the CUBIC (top), QUADRATIC (middle) and LINEAR (bottom) experiments, relative to MERRA-2. Results for both C180 (left) and C360 (right) experiments are shown.

algorithm within the model's finite-volume dynamical core. Our key results are as follows:

#1. The stratospheric mean age-of-air in GEOS is sensitive to the degree of the interpolation scheme that is used to calculate layer-mean values of total energy, U, V and tracers. Different treatment of the vertical remapping algorithm result in mid-stratospheric (50 hPa) age-of-air differences of $\sim$ 1 year over high latitudes, or about 30% climatological mean values.

#2. The increased age-of-air biases in more recent GEOS configurations are reflected in the increased biases in simulated trace gases, including CH_4 and N_2O.

#3. The age-of-air sensitivities reflect, to first order, changes in the strength of tropical upwelling associated with the Brewer-Dobson circulation which are in turn are driven by changes in EP flux convergence over northern midlatitudes. Changes in wave convergence reflect shifts in (critical lines of) wave propagation that originate in the troposphere over the Pacific Ocean, a region of strong upward wave activity.

#4. The degradation of upwelling statistics manifest in AMIPs, also translate to degradations in configurations of GEOS in which the meteorological fields are constrained or "replayed" to MERRA-2.

An important caveat should be mentioned in relation to Conclusion 3 listed above, which highlights the leading role played by *extratropical* wave convergences in the behavior of tropical upwelling in the model. While this is consistent with our understanding that the Brewer-Dobson Circulation is, to first order, determined by extratropical wave flux convergences, in practice this means that other undesirable biases may emerge in the tropics (i.e., tropical temperatures, Figure 8). Our analysis of the M2AMIP ensemble, however, suggests that this tropical temperature bias is not a necessary conse-

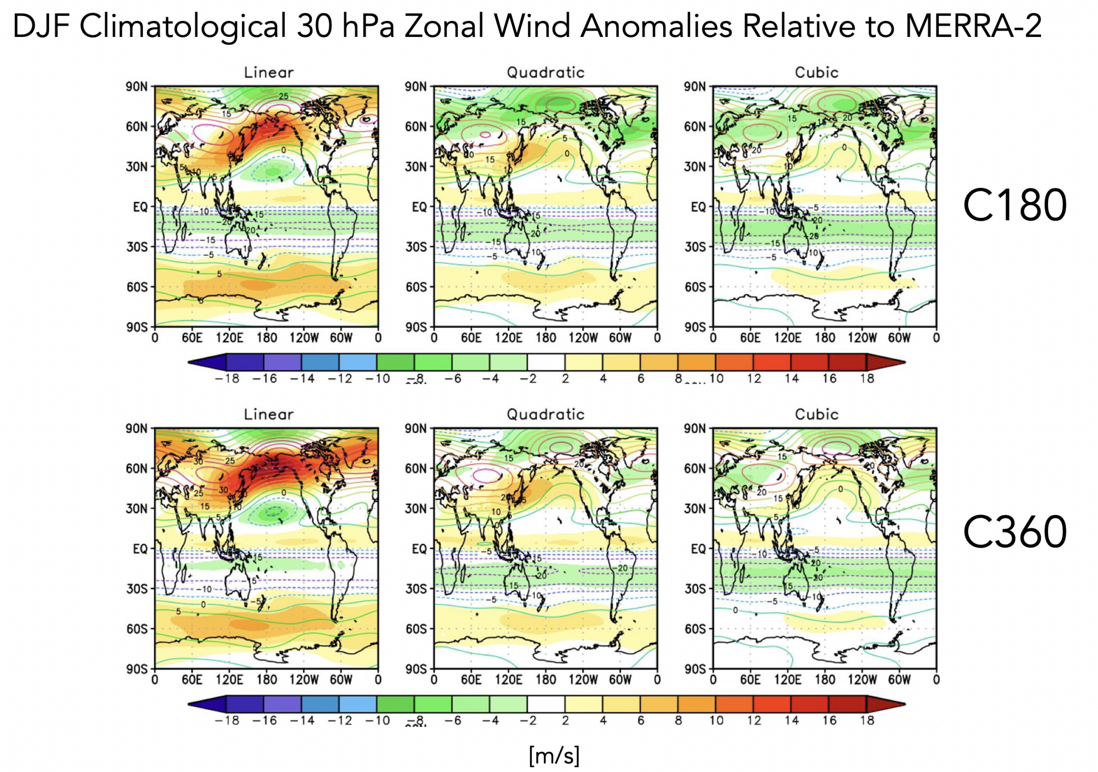

Figure 14. Colors shown anomalies in the DJF 1985-2015 climatological mean zonal winds at 30 hPa in the CUBIC (right), QUADRATIC (middle) and LINEAR (left) experiments, relative to MERRA-2. Results for both C180 (top) and C360 (bottom) experiments are shown.

quence of the GMAO FV3 remapping scheme but, rather, emerges in tandem with other changes that were made in the model. Nonetheless, Figure 8 suggests that improvements in transport associated with the GMAO remapping scheme must be carefully weighed in the broader context of other model development choices.

Interestingly, preliminary analysis suggests that our findings may also translate to replay configurations of GEOS (Figure 10), although the effect is muted, relative to free-running configurations. As a rigorous evaluation of the stratospheric circulation in replay and DAS configurations is beyond the scope of the current study, future work will therefore focus on assessing the extent to which the free-running model biases reported here are expressed when the model is run in data assimilation mode. It also bears emphasizing that our findings do have immediate implications for the (free-running) subseasonal forecast and coupled chemistry climate applications of the GEOS model currently in operation.

In addition to its implications for GEOS, our results more generally highlight the key role played by model numerics in transport (e.g., Rood (1987)). The sensitivities in the age-of-air documented herein are also consistent in spirit with the findings in Gupta et al. (2020) who showed significant age differences occurring between spectral versus finite-volume numerics. Our results, however, suggest that there remain large sensitivities even within a given (FV) dynamical core.

Looking forward, our findings support and build on the recommendation proposed in Gupta et al. (2020) for the construction of dynamical core benchmark tests aimed at determining how underlying AGCM numerics impact climatological transport properties. In particular, in addition to the age-of-air, the authors propose a range of stratospheric circulation diagnostics that should be evaluated including the zonal mean zonal winds, eddy temperature variance and zonal spectra of eddy kinetic energy. Our analysis reveals an important role to be played by the climatological zonal mean wind structure as it impacts wave convergence over midlatitudes; we therefore also recommend explicit consideration of the Eliassen Palm flux convergence and tropical upwelling (w^*) fields as they may be crucial for interpreting age-of-air changes.

One somewhat incidental – but practical - result from our analysis is that the statistics of $\nabla \cdot \mathrm{F}$ and w^* are well approximated by ensembles of so-called EMIP integrations. As these are substantially easier to run than AMIPs these could provide a “first pass” when evaluating new proposed model development changes, without the immediate need to integrate AMIP-style experiments. We emphasize, however, that this statement should only apply to a first stage in model development as the age-of-air will reflect the time integrated impacts of both advection and mixing.

Finally, we conclude by noting that, while we have focused on sensitivities within the FV remapping algorithm, our results have highlighted important sensitivities to changes in radiation and, to a lesser extent, changes in parameterized convection. Though not the dominant drivers of the age-of-air changes identified here, the former could potentially influence the age both directly through changes in thermal structure and indirectly by modifying wave propagation and/or generation in the troposphere.

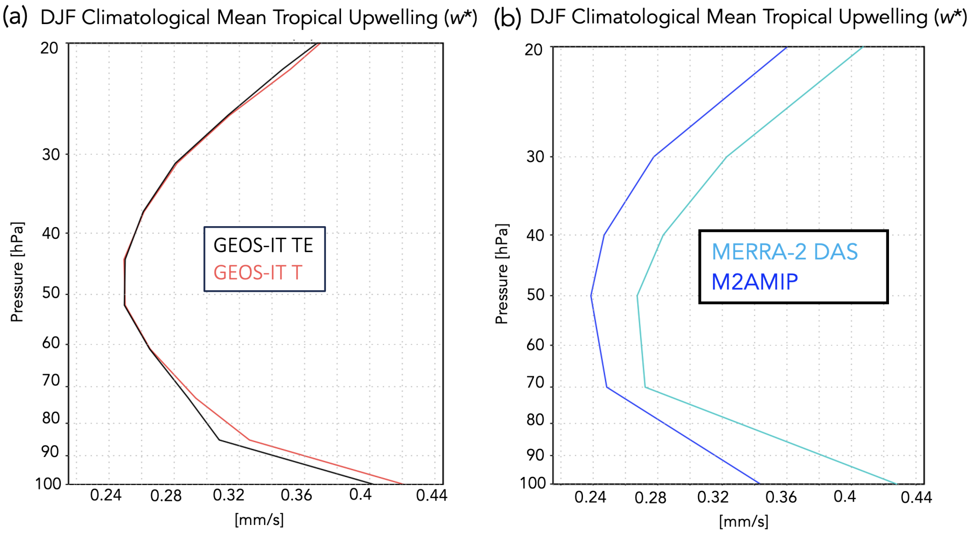

Figure A1. The DJF 1985-2015 climatological mean vertical residual mean velocity, w^*, averaged at each level between the turnaround latitudes compared between two experiments remapping to temperature (T) (red) versus total energy (TE) (black) (a) and between MERRA-2 DAS (cyan) and the M2AMIP ensemble (blue) (b). The underlying model code is consistent with the Version 5.29.4 GEOS-IT model.

Appendix A Sensitivities in Calculation of TEM Upwelling

There are various aspects of the calculation of the TEM circulation that warrant further comment. First, whereas the modeling experiments listed in Table 2 (rows 5-7) focus on the sensitivity of Step 5 within the GMAO FV3 core remapping algorithm to the choice of interpolation scheme, another difference between the GMAO and GFDL FV3 core remapping approaches is the use of TE versus T, respectively. To test the impact of this difference, we ran a new experiment which is identical to the CUBIC experiment (Table 2, row 7), except that T is remapped from input layer mean pressure locations to standard output layer mean locations directly using cubic interpolation (i.e., no computation of TE or a-posteriori energy conservation applied). Appendix Figure A1a shows that this has little impact on the strength of tropical upwelling, suggesting that the w^* differences associated with changes in the remapping algorithm are dominated by sensitivities to the choice of interpolation scheme, not the use of TE versus T.

Second, the vertical component of the TEM circulation (w^*) shows some differences in vertical structure between MERRA-2 and the 30-member M2AMIP ensemble (Appendix Figure A1b). This difference in vertical structure appears to reflect a difference between DAS and free-running configurations of the model, since other DAS configurations share a similar vertical structure (not shown). Given this difference, we ensure as apples-to-apples a comparison of simulated TEM velocities by comparing all AMIP results to other AMIP experiments.

Appendix B Correspondence between EMIP and AMIP Upwelling

Appendix Figure B1 shows the close correspondence in DJF climatological mean w^*, averaged at each level between the turnaround latitudes, from AMIP and EMIP ex-

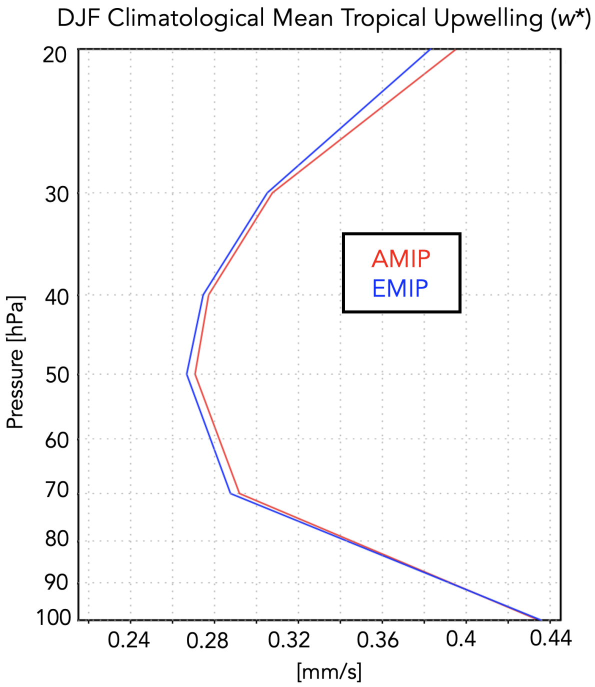

Figure B1. The DJF 1985-2015 climatological mean vertical residual mean velocity, w^*, averaged at each level between the turnaround latitudes for the CTRL experiment (Table 2, row 1). Results based on a 30-year-long AMIP experiment (red line) and a 30-member ensemble of three-month-long EMIP experiments (blue line) are shown.

periments using the same model configuration. This good agreement in upwelling is used to justify the analysis of the EMIP experiments listed in Table 2 (rows 5-7).

Appendix C Changes in Tropical and High Latitude Upwelling

Appendix Figure C1 compares the behavior in residual mean upwelling among the LINEAR, QUADRATIC and CUBIC experiments over the latitudes between the (tropical) turnaround latitudes (left) and poleward of the northern turnaround latitude (right). The ordering among experiments in both regions reflects how increases in downwelling at high latitudes are, through mass balance, accompanied by enhanced upwelling in the tropics.

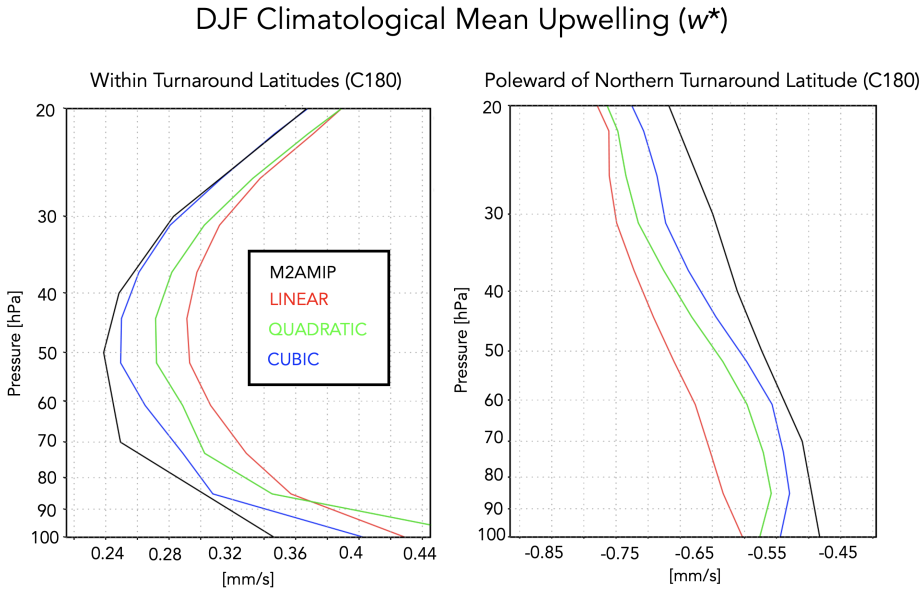

Figure C1. Left: The DJF 1985-2015 climatological mean vertical residual mean velocity, w^*, averaged at each level between the turnaround latitudes for the LINEAR (red line; Table 2, row 5), QUADRATIC (green line; Table 2, row 6) and CUBIC (blue line; Table 2, row 7) experiments. M2AMIP is shown in black. Right: As in left panel, except averaged over latitudes poleward of the northern turnaround latitude. Results in both panels are shown for C180 experiments.

Appendix D Tropopause Pressure

Appendix Figure D1 compares boreal winter tropopause pressure and temperature among the LINEAR, QUADRATIC and CUBIC experiments, relative to MERRA-2.

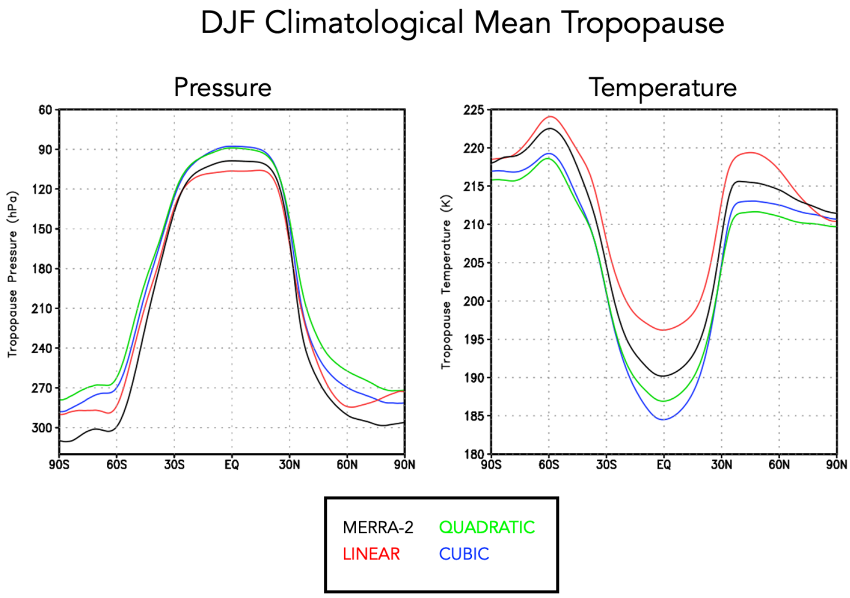

Figure D1. The DJF 1985-2015 climatological mean tropopause pressure (left) and temperature (right) in the CUBIC (blue), QUADRATIC (green) and LINEAR (red) experiments. MERRA-2 is shown in black. Results are shown for the C180 experiments.

Acronyms

AMIP Atmospheric Model Intercomparison Project
CH_4 methane
CCMs chemistry climate models
CCMI Chemistry Climate Modeling Initiative
CCMVal Chemistry Climate Model Validation
CO_2 carbon dioxide
CTRL control
CTM chemistry transport model
DAS Data assimilation
DJF December-January-February
EMIP ensemble AMIP
EOS Earth Observing System
EP Eliassen-Palm
FV finite-volume
FP Forward Processing
GEOS Global Earth Observing System
GEOS-R21C GEOS Retrospective analysis for the 21st Century
GMI Global Modeling Initiative
HALOE Halogen Occultation Experiment
MERRA-2 Modern-Era Retrospective Analysis for Research and Applications v2
MLS Microwave Limb Sounder
N_2O nitrous oxide
NH northern hemisphere
PPM piecewise parabolic
RRTMG Rapid Radiative Transfer Model for GCMS

SW shortwave
TE total energy
TEM Transformed Eulerian Mean
UARS Upper Atmosphere Research Satellite

Open Research Section

All output from the GEOS simulations presented has been reprocessed as GrADS control files and binary datasets that are located publicly at https://gmao.gsfc.nasa.gov/gmaoftp/corbe/StratAge/Data/. The MERRA-2 data is publicly available at https://disc.gsfc.nasa.gov/datasets?project=MERRA-2.

Acknowledgments
C.O. thanks Lawrence Coy and William Putman for their insight which helped in interpreting the results and guiding experimental design. The authors also thank the high-performance computing resources provided by NASA's Advanced Supercomputing (NAS) Division and the NASA Center for Climate Simulation (NCCS) as well as NASA's Modeling, Analysis and Prediction (MAP) program, which supports the Global Modeling Assimilation Office and core chemistry-climate and chemistry-modeling activities.

References

Abalos, M., Calvo, N., Benito-Barca, S., Garny, H., Hardiman, S. C., Lin, P., … others (2021). The Brewer–Dobson circulation in CMIP6. *Atmospheric Chemistry and Physics*, *21*(17), 13571–13591.

Abalos, M., Orbe, C., Kinnison, D. E., Plummer, D., Oman, L. D., Jöckel, P., … others (2020). Future trends in stratosphere-to-troposphere transport in CCMI models. *Atmospheric Chemistry and Physics*, *20*(11), 6883–6901.

Abalos, M., Randel, W. J., Kinnison, D. E., & Garcia, R. R. (2017). Using the artificial tracer e90 to examine present and future UTLS tracer transport in WACCM. *Journal of the Atmospheric Sciences*, *74*(10), 3383–3403.

Andrews, D., Holton, J., & Leovy, C. (1987). Middle Atmosphere Dynamics. *Academic Press*, *60*, 489. doi: 10.1175/1520-0469(2003)060⟨0103:CEOOAL⟩2.0.CO;2

Boering, K. A., Wofsy, S., Daube, B., Schneider, H., Loewenstein, M., Podolske, J., & Conway, T. (1996). Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide. *Science*, *274*(5291), 1340–1343.

Butchart, N., Cionni, I., Eyring, V., Shepherd, T., Waugh, D., Akiyoshi, H., … others (2010). Chemistry–climate model simulations of twenty-first century stratospheric climate and circulation changes. *Journal of Climate*, *23*(20), 5349–5374.

Chiodo, G., & Polvani, L. M. (2019). The response of the ozone layer to quadrupled CO2 concentrations: Implications for climate. *Journal of Climate*, *32*(22), 7629–7642.

Chou, M.-D. (1990). Parameterizations for the absorption of solar radiation by O2 and CO2 with application to climate studies. *Journal of Climate*, *3*(2), 209–217.

Chou, M.-D. (1992). A solar radiation model for use in climate studies. *Journal of Atmospheric Sciences*, *49*(9), 762–772.

Chou, M.-D., & Suarez, M. J. (1994). An efficient thermal infrared radiation parameterization for use in general circulation models.

Collow, A. B. M., Mahanama, S. P., Bosilovich, M. G., Koster, R. D., & Schubert, S. D. (2017). *An evaluation of teleconnections over the united states in an* *ensemble of AMIP simulations with the MERRA-2 configuration of the GEOS atmospheric model* (Tech. Rep.).

Davis, N. A., Callaghan, P., Simpson, I. R., & Tilmes, S. (2022). Specified dynamics scheme impacts on wave-mean flow dynamics, convection, and tracer transport in CESM2 (WACCM6). *Atmospheric Chemistry and Physics*, *22*(1), 197–214.

Dietmüller, S., Eichinger, R., Garny, H., Birner, T., Boenisch, H., Pitari, G., ... others (2018). Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models. *Atmospheric Chemistry and Physics*, *18*(9), 6699–6720.

Eichinger, R., Garny, H., Šácha, P., Danker, J., Dietmüller, S., & Oberländer-Hayn, S. (2020). Effects of missing gravity waves on stratospheric dynamics: Part 1, Climatology. *Climate Dynamics*, *54*(5), 3165–3183.

Eluszkiewicz, J., Hemler, R. S., Mahlman, J. D., Bruhwiler, L., & Takacs, L. L. (2000). Sensitivity of age-of-air calculations to the choice of advection scheme. *Journal of the Atmospheric Sciences*, *57*(19), 3185–3201.

Eyring, V., Lamarque, J.-F., Hess, P., Arfeuille, F., Bowman, K., Chipperfield, M. P., ... others (2013). Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) community simulations in support of upcoming ozone and climate assessments. *SPARC Newsletter*, *40*(January), 48–66.

Freitas, S. R., Grell, G. A., Molod, A., Thompson, M. A., Putman, W. M., Santos e Silva, C. M., & Souza, E. P. (2018). Assessing the Grell-Freitas convection parameterization in the NASA GEOS modeling system. *Journal of Advances in Modeling Earth Systems*, *10*(6), 1266–1289.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... others (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454.

Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, *14*(10), 5233–5250.

Grooß, J.-U., & Russell III, J. M. (2005). A stratospheric climatology for O 3, H2O, CH4, NOx, HCl and HF derived from HALOE measurements. *Atmospheric Chemistry and Physics*, *5*(10), 2797–2807.

Gupta, A., Gerber, E. P., & Lauritzen, P. H. (2020). Numerical impacts on tracer transport: A proposed intercomparison test of Atmospheric General Circulation Models. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 3937–3964.

Hall, T. M., & Plumb, R. A. (1994). Age as a diagnostic of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, *99*(D1), 1059–1070.

Hall, T. M., Waugh, D. W., Boering, K. A., & Plumb, R. A. (1999). Evaluation of transport in stratospheric models. *Journal of Geophysical Research: Atmospheres*, *104*(D15), 18815–18839.

Hardiman, S. C., Butchart, N., & Calvo, N. (2014). The morphology of the Brewer–Dobson circulation and its response to climate change in CMIP5 simulations. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1958–1965.

Haynes, P., McIntyre, M., Shepherd, T., Marks, C., & Shine, K. P. (1991). On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *Journal of the Atmospheric Sciences*, *48*(4), 651–678.

Hegglin, M. I., Brunner, D., Peter, T., Hoor, P., Fischer, H., Staehelin, J., ... Weers, U. (2006). Measurements of NO, NOy, N2O, and O3 during SPURT: implications for transport and chemistry in the lowermost stratosphere. *Atmospheric Chemistry and Physics*, *6*(5), 1331–1350.

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.

Holzer, M., & Hall, T. M. (2000). Transit-time and tracer-age distributions in geophysical flows. *Journal of the atmospheric sciences*, *57*(21), 3539–3558.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, *113*(D13).

Ivy, D. J., Solomon, S., Calvo, N., & Thompson, D. W. (2017). Observed connections of arctic stratospheric ozone extremes to Northern Hemisphere surface climate. *Environmental Research Letters*, *12*(2), 024004.

Kouatchou, J., Molod, A., Nielsen, J., Auer, B., Putman, W., & Clune, T. (2015). *GEOS-5 chemistry transport model user's guide* (Tech. Rep.).

Legras, B., Pisso, I., Berthet, G., & Lefèvre, F. (2004). Variability of the lagrangian turbulent diffusivity in the lower stratosphere. *Atmospheric Chemistry and Physics Discussions*, *4*(6), 8285–8325.

Lin, S.-J., Putman, W., & Harris, L. (2017). *The gfdl finite-volume cubed-sphere dynamical core.* NWS/NCEP/EMC.

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.

Monge-Sanz, B., Chipperfield, M., Simmons, A., & Uppala, S. (2007). Mean age of air and transport in a CTM: Comparison of different ECMWF analyses. *Geophysical Research Letters*, *34*(4).

Monge-Sanz, B. M., Bozzo, A., Byrne, N., Chipperfield, M. P., Diamantakis, M., Flemming, J., … others (2022). A stratospheric prognostic ozone for seamless Earth system models: performance, impacts and future. *Atmospheric Chemistry and Physics*, *22*(7), 4277–4302.

Morgenstern, O., & Carver, G. D. (2001). Comparison of cross-tropopause transport and ozone in the upper troposphere and lower stratosphere region. *Journal of Geophysical Research: Atmospheres*, *106*(D10), 10205–10221.

Neu, J., Strahan, S., Braesicke, P., Douglass, A., Huck, P., Oman, L., … Tegtmeier, S. (2010). SPARC CCMVal (2010), SPARC Report on the Evaluation of Chemistry-Climate Models: Chapter 5: Transport. SPARC.

Neu, J. L., & Plumb, R. A. (1999). Age of air in a "leaky pipe" model of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, *104*(D16), 19243–19255.

Oehrlein, J., Chiodo, G., & Polvani, L. M. (2020). The effect of interactive ozone chemistry on weak and strong stratospheric polar vortex events. *Atmospheric Chemistry and Physics*, *20*(17), 10531–10544.

Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., & Molod, A. M. (2017). Large-scale atmospheric transport in GEOS replay simulations. *Journal of Advances in Modeling Earth Systems*, *9*(7), 2545–2560.

Orbe, C., Rind, D., Jonas, J., Nazarenko, L., Faluvegi, G., Murray, L. T., … others (2020). GISS Model E2.2: A climate model optimized for the middle atmosphere—2. Validation of large-scale transport and evaluation of climate response. *Journal of Geophysical Research: Atmospheres*, *125*(24), e2020JD033151.

Orbe, C., Yang, H., Waugh, D. W., Zeng, G., Morgenstern, O., Kinnison, D. E., … others (2018). Large-scale tropospheric transport in the Chemistry–Climate Model Initiative (CCMI) simulations. *Atmospheric Chemistry and Physics*, *18*(10), 7217–7235.

Pan, L. L., Wei, J., Kinnison, D., Garcia, R., Wuebbles, D., & Brasseur, G. P. (2007). A set of diagnostics for evaluating chemistry-climate models in the extratropical tropopause region. *Journal of Geophysical Research: Atmospheres*, *112*(D9).

Pawson, S., Stajner, I., Kawa, S. R., Hayashi, H., Tan, W.-W., Nielsen, J. E., … Livesey, N. J. (2007). Stratospheric transport using 6-h-averaged winds from a data assimilation system. *Journal of Geophysical Research: Atmospheres*, *112*(D23).

Plumb, R. A. (1985). On the three-dimensional propagation of stationary waves. *Journal of Atmospheric Sciences*, *42*(3), 217–229.

Plumb, R. A. (1996). A "tropical pipe" model of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, *101*(D2), 3957–3972.

Plumb, R. A. (2002). Stratospheric transport. *Journal of the Meteorological Society of Japan. Ser. II*, *80*(4B), 793–809.

Polvani, L. M., Waugh, D. W., Correa, G. J., & Son, S.-W. (2011). Stratospheric ozone depletion: The main driver of twentieth-century atmospheric circulation changes in the southern hemisphere. *Journal of Climate*, *24*(3), 795–812.

Prather, M. J., Zhu, X., Tang, Q., Hsu, J., & Neu, J. L. (2011). An atmospheric chemist in search of the tropopause. *Journal of Geophysical Research: Atmospheres*, *116*(D4).

Rood, R. B. (1987). Numerical advection algorithms and their role in atmospheric transport and chemistry models. *Reviews of geophysics*, *25*(1), 71–100.

Rosenlof, K. H. (1995). Seasonal cycle of the residual mean meridional circulation in the stratosphere. *Journal of Geophysical Research: Atmospheres*, *100*(D3), 5173–5191.

Son, S.-W., Tandon, N. F., Polvani, L. M., & Waugh, D. W. (2009). Ozone hole and Southern Hemisphere climate change. *Geophysical Research Letters*, *36*(15).

Strahan, S., Douglass, A., & Newman, P. (2013). The contributions of chemistry and transport to low arctic ozone in March 2011 derived from aura MLS observations. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1563–1576.

Thiele, G., & Sarmiento, J. (1990). Tracer dating and ocean ventilation. *Journal of Geophysical Research: Oceans*, *95*(C6), 9377–9391.

Waugh, D., & Hall, T. (2002). Age of stratospheric air: Theory, observations, and models. *Reviews of Geophysics*, *40*(4), 1–1.

Weaver, C. J., Douglass, A. R., & Rood, R. B. (1993). Thermodynamic balance of three-dimensional stratospheric winds derived from a data assimilation procedure. *Journal of Atmospheric Sciences*, *50*(17), 2987–2993.