Numerical impacts on tracer transport: Diagnosing the influence of dynamical core formulation and resolution on stratospheric transport

Aman Gupta*† and Edwin P. Gerber

Center for Atmosphere-Ocean Science, Courant Institute of Mathematical Sciences, New York, New York, USA

R. Alan Plumb

Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Peter H. Lauritzen

National Center for Atmospheric Research, Boulder, CO, USA

**Corresponding author address:* Aman Gupta, Theresienstrasse 37, Ludwig-Maximilian University, Munich 80333

E-mail: ag4680@nyu.edu

†Meteorological Institute Munich, Ludwig-Maximilian University, Munich, BY, Germany

ABSTRACT

Accurate representation of stratospheric trace gas transport in climate models is important for accurate ozone recovery modeling and future climate projections. Intermodel transport differences can arise due to differences in the slowly evolving diabatic circulation, rapid adiabatic mixing by planetary waves and numerical diffusion. This study investigates the impact of these processes on transport using an idealised tracer — age of air — and using the age-budget theory proposed in Linz et al. (2016) and Linz et al. (in prep.). Transport is assessed in two state-of-the-art dynamical cores using totally different numerical formulations : finite volume and spectral elements. Integrating the models in both free-running and nudged tropical winds configurations shows the strong impact of tropical dynamics on stratospheric transport and reveals key intermodel differences in transport. Using the age-budget theory, vertical and horizontal gradients of age are used to estimate the diabatic circulation strength, the adiabatic mixing flux and the numerical diffusive flux among models. Further, fractional contribution of each process is obtained by connecting the full 3-d model transport to the tropical leaky pipe (TLP) framework of Neu and Plumb (1999), but with vertically varying parameters. Up to three-fourths of the age difference among the free running integrations is due to differences in isentropic mixing among models. For runs nudged to identical tropical winds, 50% of the persisting intermodel age difference is still attributed to intermodel circulation differences. The analysis also establishes the importance of sufficiently high vertical resolution to constrain diffusive transport fluxes. It is concluded that the dynamical core formulation can still have a strong influence on model transport.

1. Introduction

Trace gases and aerosols play a critical role in air quality (Friedl 2007), the radiative balance of our planet, and, in the case of ozone, harmful UV radiation (Molina and Rowland 1974; Eyring et al. 2007). For species with lifetimes shorter than characteristic mixing time scale of the atmosphere, on the order of months in the troposphere to years in the stratosphere, an accurate simulation requires both the representation of sinks and source (emissions and atmospheric chemistry; e.g., Collins et al. 2017) and transport (the advection and mixing of trace gases by the atmospheric circulation, e.g., Plumb 2002). This study focuses on this second factor, building on recent work by Gupta et al. (2020) to establish an intercomparison test to assess trace gas transport by the primitive equation solvers, or dynamical cores, of atmospheric models.

Gupta et al. (2020), hereafter referred to as G20, explored the impact of model numerics and resolution on transport by the stratosphere-troposphere system, with particular focus on the middle atmosphere. The slower transport time scales of the Brewer-Dobson Circulation (BDC, Butchart (2014)), from months in the upper troposphere and lower stratosphere (UTLS) to years in the so-called deep branch of the BDC (Birner and Bönisch 2011), expose the impact of small errors in numerical transport. Trace gas transport by the BDC depends on both a slow overturning of mass from the tropics to the mid and high latitudes, and a fast, quasi-horizontal mixing of tracers along isentropic (constant potential temperature) surfaces. We refer to the slow overturning as the diabatic circulation, since air must be radiatively warmed in the tropics and cooled in the higher latitudes to ascend or descend through the highly stratified stratosphere, and the faster mixing processes as isentropic mixing, since it is associated with adiabatic transport and mixing. Both components are primarily driven by planetary wave breaking. The westward torque induced by breaking planetary waves forces the mass in the tropics to cross angular momentum surfaces and enter the higher latitudes; the "Rossby wave pump" of Holton et al. (1995). Stratospheric transport by a dynamical core depends on its ability to accurately capture both the overturning circulation and isentropic mixing.

G20 extended the dynamical core setup of Held and Suarez (1994) to account for a more active stratosphere (as in Polvani and Kushner 2002) and introduced an idealized tracer, proposing two benchmark tests to assess stratosphere-troposphere coupling and stratospheric transport. The idealized setup and tests, which are reviewed in Section 2, strip away all parameterizations of subgrid scale processes (including transport and chemistry) to isolate the role of numerics on model climatology, transport, and stratosphere-troposphere coupling. They found differences in transport between four different dynamical cores, including two state-of-the art cores. They concluded that the choice of numerics can significantly impact transport in the stratosphere.

Trace gas transport in G20 was quantified by the Age of Air (Hall and Plumb 1994; Waugh and Hall 2002), a measure of the transport timescales in the stratosphere discussed in more detail in Section 3. The models show large quantitative disagreement in the steady state age of air profile, which was found to vary by as much as 40% in the winter midlatitude stratosphere in the first "free running" benchmark test, as shown here in Figure 1. These extreme differences in transport were, however, mostly rooted in differences in the tropical climatology, as opposed to the accuracy of the employed numerical schemes.

The subtle momentum balance in the tropical stratosphere permits the development of the Quasi-Biennial Oscillation (QBO), a 28-month swing between easterly and westerly jets (Baldwin et al. 2001). Representation of the QBO in comprehensive models depends on both high vertical resolution and the parameterization of unresolved gravity waves (Butchart et al. 2018). The idealized setup of G20 lacks a parameterization of gravity waves, and, as a result, the dynamical cores produced steady climatological wind profiles which differed greatly between models with spectral and finite volume based numerical cores. G20 partially resolved this divergence by proposing a second benchmark test in which the tropical wind climatology is constrained (See Section 2), but differences between models with finite volume and spectral numerics persist.

In this study, we assess the inter-model differences in stratospheric transport found in the two G20 benchmark tests in more detail. We develop a framework to quantify the fractional contribution in transport differences to the differences in diabatic circulation, isentropic mixing, and numerical and unresolved diabatic fluxes among models. To do this, we apply the theory of age established by Linz et al. (2016) and Linz et al. (in prep.) to first estimate the strength of the diabatic circulation and the mixing fluxes in each model. The key is to study the flux of age in isentropic coordinates. Partitioning a given isentrope into regions of diabatic upwelling and downwelling allows a convenient separation of the diabatic and adiabatic components of transport. We discuss this in more detail in Section 3, where we revisit the theory of age and connect the meridional differences in age over these regions to the diabatic circulation strength and the vertical gradients of age in these regions to the midlatitude mixing fluxes.

In Section 4, we first estimate the fractional contribution of the diabatic diffusive fluxes to the overall tracer flux, highlighting the importance of high vertical resolution when studying stratospheric transport. In Section 5, we establish a framework to quantify the importance of different transport processes on the distribution of mean age using the Tropical Leaky Pipe (hereafter TLP) transport model of Neu and Plumb (1999). The framework clearly identifies the key role of isentropic mixing generated in response to the different tropical wind profiles in the free running experiments, explaining why the age differed by up to 40% between models. Once the tropical winds (and mixing) are constrained, however, fundamental differences in the diabatic circulation between models are most responsible for maintaining differences in large-scale stratospheric transport. Section 6 summarises our findings.

2. Model integrations and the benchmark tests

G20 proposed two benchmark tests to assess climatological transport in atmospheric general circulation models: a Free Running (FR) test and the SPecified tropical wind (SP) test. The tests were applied to four different dynamical cores with different underlying numerics, developed at two modeling centers, the Geophysical Fluid Dynamics Laboratory (GFDL) and the National Center for Atmospheric Research (NCAR). Here we focus only on two of the four models, the more modern cores developed at the centers, as they capture the key differences observed across all cores. We consider the exact same experiments as in G20, but review the key details here for completeness.

a. Model Details

We compare the behavior of two dynamical cores. The first is a finite volume model based on a cubed sphere grid. It was developed at GFDL and we refer to it as GFDL-FV3, or FV3 for short. The core employs finite volume schemes in both the vertical and the horizontal to solve the primitive equations. FV3 was built as the core of GFDL's Atmospheric Model, Version 3, AM3 (Donner et al. 2011), and a related non-hydrostatic version was recently adopted as the core of the National Center for Environmental Prediction Global Forecasting System.

The second model is a spectral element core developed at NCAR as part of the Community Atmospheric Model (CAM) framework. It is also based on a cubed sphere grid, but uses a spectral finite element method to discretize the primitive equations in the horizontal, and is coupled to finite volume Lagrangian advection in the vertical (Lauritzen et al. 2018). We refer to it as CAM-SE, or SE for short. It was designed to become the dynamical core of NCAR's Community Earth System Model, version 2 (CESM2) and used for many of the CESM2 integrations in the Coupled Model

Intercomparison Project, Phase 6 (CMIP6). Further details on both CAM-SE and GFDL-FV3 are provided in G20 (Section 3 and Appendix A).

G20 considered two additional cores, a pseudospectral model developed by GFDL and a finite-volume core based on a latitude-longitude grid, used in the CAM framework. They found, however, that the two finite volume cores and the two "spectral" cores (i.e., pseudospectral and spectral element) behaved very similarly to each other in terms of their climatological circulation, transport, and sensitivity to resolution. For brevity, in this study we only focus on the two recently developed dynamical cores.

b. Experiment Details

The cores are forced with identical thermal forcings (Held and Suarez 1994; Polvani and Kushner 2002), detailed in Section 4.1 of G20. Newtonian relaxation to an analytic temperature state (which can be interpreted as a state of raditative-convective equilibrium) generates a perpetual northern-winter climatology. The SP test differs from the FR test only in that a Rayleigh damping is applied to the zonal winds of the equatorial stratosphere. As detailed in G20 (Section 4.2 and Appendix B), the damping relaxes the winds above 200 hPa and within 15° of the equator to a steady easterly profile with a timescale of 40 days in the lower stratosphere, rising to 10 days above 3 hPa.

To quantify transport, a linearly increasing (in time) clock tracer is introduced near the surface ($p \geq 700$ hPa) as detailed in G20, Section 4.1. The clock tracer is used to compute the age of air, which provides a measure of transport timescales of trace gases in the stratosphere. The age of air itself can be interpreted as an idealized tracer, Γ, that "ages" linearly with time, $d\Gamma/dt = 1$, but is reset to zero when it comes in contact with the surface. We discuss age in greater detail in Section 3.

As in G20, the models were integrated for 10,000 days and the last 3,300 days were used for analysis. In order to test robustness of transport to model grid resolutions, each core was integrated with at least four different resolutions, as tabulated in Table 1, allowing us to assess the impact of doubling both the horizontal and vertical resolution. For CAM-SE, additional higher resolution runs were considered. For context, the models were integrated with horizontal resolution comparable to that used for comprehensive models in CMIP experiments (approximately 1°), but with vertical resolution that is generally higher than that used for climate integrations, even in the 40 level configurations. Since the results are less sensitive to the horizontal resolution, we focus on the higher horizontal resolution runs (i.e., $1^\circ \times 1^\circ$) with 40 and 80 vertical levels, unless otherwise specified.

c. Key Results from the Gupta et al. (2020) Benchmark Tests

The dynamical cores essentially fail the Held and Suarez (1994) test in the free running configuration: numerics and resolution fundamentally affect the climatological circulation. As the vertical resolution is increased from 40 to 80 levels, the spectral-based models and the finite volume-based models develop differences in the climatological state of the tropical stratosphere. With 40 vertical levels, all the models generate tropical easterlies (irrespective of the horizontal resolution) but with 80 levels, the two spectral models generate tropical westerlies (up to 20 m/s) while the two finite volume models maintain tropical easterlies (Figure 1a,b and Figure 2 of G20).

The impact of the tropical wind difference on the overall age profile is significant. The age of air in model integrations with tropical westerlies (up to 20 m/s) is up to 40% higher (6 years vs 8.5 years) than the models with tropical easterlies (Figure 1c). The difference in the age of air indicates significant differences in transport. G20 suggested that tropical westerlies induce enhanced tropical-extratropical wave mixing, as the westerlies provide a duct for extratropical

Rossby waves to break deep into the tropical atmosphere. The enhanced mixing results in an increase in the older age transported from the midlatitudes into the tropics. Recirculation of the air results in an older age throughout the stratosphere. This hypothesis was explored with the Specified Tropics (SP) test.

The SP test differs only in one key aspect: the tropical winds are constrained to a specified analytical easterly profile. As a result of this nudging, a closer agreement in tropical winds – and in turn in age of air – was achieved, as seen in Figure 1d-f. The age among the two state-of-the-art cores, however, still exhibits fundamental differences. As detailed in G20, Section 7, the climatological circulation and transport behavior of both cores appears to be stable to further increase in the horizontal and vertical resolution: in a rough sense, they converge to a well defined state under the SP test. This state, however, differs between the two models: the age of air in particular varies by as much as 15% in the winter midlatitudes. The goal of this study is to explore and understand the differences in age between the GFDL-FV3 and CAM-SE dynamical cores in the FR and SP tests.

3. Stratospheric age budget analysis in isentropic coordinates

A schematic of the large-scale stratospheric circulation and transport is shown in Figure 6. Following Neu and Plumb (1999), we partition the stratosphere into an upwelling (diabatic warming) region in the tropics and downwelling (diabatic cooling) regions in the extratropics. The net overturning, or diabatic circulation, is completed by a flux of mass from the tropics to extratropics, illustrated with the streamlines in the left panel. There is also substantial mixing of air between the upwelling and downwelling regions, illustrated with curly arrows. While this has no impact on the net overturning of mass, it is associated with an adiabatic transport of tracers along isentropes. In the tropical leaky pipe model (Neu and Plumb 1999), the geometry is simplified by lumping the two downwelling regions together (right panel of Figure 6). The net overturning flow in the meridional, μ_{net}, is viewed as difference between the total entrainment of mass from the tropics to the extratropics, μ_{ent}, and the the total mixing flux of air that recirculates back into the tropical pipe, μ_{mix}. The model also allows for the vertical diffusion of age, as represented by vertical, bi-directional curly arrows in the right panel.

The goal of this section is to review how the TLP framework can be applied to the three-dimensional circulation of the atmosphere in diabatic coordinates using the age of air, as established by Linz et al. (2016, in prep.). This allows us to infer how differences in the net diabatic circulation, mixing, and vertical diffusion influence transport in the dynamical cores.

a. Meridional age difference and vertical diffusion

The age of air quantifies the total time elapsed since an air mass was last in contact with the surface (or tropopause, depending on the definition). It represents the mean of a distribution of transit times for air parcels in the air mass. In the stratosphere, the age can be related to the residence times of ozone and ozone destroying substances (most of which have near-surface emissions). It has been extensively used by both observation and model based studies as a diagnostic for atmospheric transport (Waugh et al. 1997; Hall et al. 1999; Engel et al. 2009; Orbe et al. 2013; Linz et al. 2017). Linz et al. (2016) reformulate the TLP in isentropic coordinates to develop a theory to estimate the strength of the diabatic circulation solely from the age-of-air. Working in potential temperature (or equivalently, entropy) coordinates in the vertical is advantageous primarily because it helps decouple two key transport processes, the slow diabatic advection across isentropes vs. fast adiabatic mixing along isentropes (McIntyre and Palmer 1983; Sparling et al. 1997; Haynes and Shuckburgh 2000).

As detailed in Section 2 of Linz et al. (2016), for a steady-state circulation, and neglecting vertical diffusion (or diabatic dispersion), the age difference between the midlatitude and the tropics is solely determined by the diabatic circulation strength. The key is to properly define the gross age difference, $\Delta\Gamma = \Gamma_d - \Gamma_u$, where Γ_u and Γ_d are *mass-flux weighted* representative ages over the upwelling and downwelling partitions of an isentrope, respectively,

$$\Gamma_u(\theta) = \frac{\int_u \rho_\theta \dot{\theta} \Gamma \, dA}{\int_u \rho_\theta \dot{\theta} \, dA} \ , \ \Gamma_d(\theta) = \frac{\int_d \rho_\theta \dot{\theta} \Gamma \, dA}{\int_d \rho_\theta \dot{\theta} \, dA}. \tag{1}$$

Here $\dot{\theta}$ is the diabatic velocity, dA is the infinitesimal area element and $\int_u$ ($\int_d$) represents special integration only over the upwelling (downwelling) partition of the θ-isentrope, i.e., where $\dot{\theta}$ is positive (negative). The gross meridional age difference can be directly related to the globally integrated mass above the θ-isentrope, $M(\theta)$, and the mass throughput across the isentrope per unit time, $\mathcal{M}(\theta)$:

$$\Delta\Gamma(\theta) = \Gamma_d(\theta) - \Gamma_u(\theta) = \frac{M(\theta)}{\mathcal{M}(\theta)} \tag{2}$$

The diabatic mass flux $\mathcal{M}$ is the total upward mass flux, $\mathcal{M}_u$, or equivalently, the total downward mass flux, $\mathcal{M}_d$, assuming steady state conditions. It is obtained by integrating over the upwelling (downwelling) partition of the stratosphere: $\mathcal{M} = \int_u \rho_\theta \dot{\theta} \, dA$, the denominator in definition of Γ_u (and likewise for $\mathcal{M}$).

The quantity $\frac{M(\theta)}{\mathcal{M}(\theta)}$ (which has units of time) represents the Mean Residence Time (MRT) of air in the atmosphere above the θ-isentrope. It presents as an upper bound on the meridional age difference $\Delta\Gamma$ in the limiting case where the age is only transported by mean diabatic advection and isentropic mixing. The diffusive fluxes, otherwise, act towards reducing $\Delta\Gamma$. In this respect, Equation (2) can be used to estimate the contribution of unresolved diabatic fluxes + numerical diffusive fluxes towards the total age difference $\Delta\Gamma$. We use this rationale in Section 4 to estimate the contribution of unrepresented diffusive tracer fluxes for the two climate models.

b. Vertical age gradient and the tropics-midlatitudes mixing flux

While the meridional age difference is directly connected to the diabatic circulation strength (Linz et al. 2016), the vertical gradients of the net upwelling age can be used to quantify the efficiency with which planetary wave fluxes transport midlatitude air into the tropics (Linz et al. in prep.). In the upper winter stratosphere, isentropic mixing is effected by breaking of upward propagating planetary waves. In the lower stratosphere, this mixing is primarily driven by baroclinic eddies and synoptic scale wave breaking in both hemispheres (Plumb 2002).

Linz et al. (in prep.) show that, for a steady-state circulation in the limit of no vertical diffusion, the aging of air in the tropical pipe i.e. $d\Gamma_u/d\theta$ can be expressed as a sum of aging by vertical advection and aging by mixing between the two regions. Mathematically, this is expressed as

$$\frac{d\Gamma_u}{d\theta} = \frac{\sigma_u}{\mathcal{M}} + \mu_{mix}\frac{\Delta\Gamma}{\mathcal{M}}, \tag{3}$$

where σ_u is the isentropic density $\rho_\theta = \frac{1}{g}|dp/d\theta|$ horizontally integrated over the upwelling partition and μ_{mix} is the mass flux per unit entropy that mixes the midlatitude air with the tropical air. The first term on the right hand side represents the net aging of air were it purely advected by diabatic fluxes through the tropical pipe. The second term represents the total aging by mixing of air between the two regions by the mixing flux μ_{mix}. In the special case of vanishing $\Delta\Gamma$ or vanishing μ_{mix}, the air ages solely due to the slow vertical advection of air (Plumb 1996).

Having an *apriori* knowledge of the age structure and diabatic mass fluxes (i.e., the results from our model integrations), we compute the mixing flux μ_{mix} as a residual from Equation 3. It is often transformed to a "mixing efficiency" ε, defined by Neu and Plumb (1999) as the ratio of equatorward mass transport to *net* poleward mass transport. Here, μ_{net} is the net poleward mass flux, represented by the black poleward arrows in Figure 6(left). It can be calculated directly from model output, and equals the difference of the poleward directed mass flux and the equatorward directed mixing flux, i.e., $\mu_{net} = \mu_{ent} - \mu_{mix}$ (Neu and Plumb 1999).

c. Applying the theory: An Illustration with the GFDL-FV3 Core

To provide a concrete example of the theory, we illustrate the transport metrics from Equations 2 and 3 for the 1° L80 GFDL-FV3 core integration in Figure 2. The metrics will be compared for the two models in the following sections; here we focus on how they are computed and their overall climatological structure. In order to compute transport metrics in isentropic coordinates using model output on pressure-levels, we employ a mass-preserving binning scheme (Yamada and Pauluis 2015) to interpolate from pressure to isentropic coordinates. The details of the scheme are provided as supplementary material.

Figure 2(a) shows the time mean, zonal mean age simulated by the finite volume core. The solid black lines demarcate the mean regions of diabatic upwelling (the tropical region with diabatic velocity $\dot{\theta} > 0$) and diabatic downwelling (higher latitudes with $\dot{\theta} < 0$). Air is youngest at the tropical tropopause, the primary entry point for air into the stratosphere, and older in the extratropics, where mass, on average, descends. The age always increases with potential temperature (i.e., the stratosphere is stratified with respect to age), however, because of mixing.

In the tropics, this stratification reflects the vertical transport of air upward into the stratosphere; the air ages as it is slowly advected up into the tropical atmosphere by the mean overturning circulation. In the extratropics, stratification is maintained by isentropic mixing, which exchanges younger air from the ascending branch of the circulation in the tropics with older descending air in the extratropics. It is for this reason that the weakest vertical gradient is observed in the polar vortex (Northern Hemisphere above 900 K), as strong gradients in potential vorticity associated with the vortex inhibit mixing. The perpetual boreal winter forcing leads to an artificially weak circulation in the austral hemisphere, and the tropical pipe extends all the way to the pole above 1200 K.

We compute the mass-flux weighted average age over regions of upwelling and downwelling to obtain Γ_u and Γ_d respectively, plotted in Figure 2(b). The gross ages increase with height, and Γ_d is consistently older than Γ_u throughout the vertical, as expected. At about 500K, there is an abrupt increase in the rate of aging in the vertical, more readily seen in Γ_d compared to the tropical age Γ_u. The increase can be interpreted as a diffusive boundary layer between the older air in the stratosphere and the much younger tropospheric air below (Neu and Plumb 1999). The air in the troposphere is very well mixed across latitudes. As most of the tropospheric air enters the stratosphere through the tropical pipe, the transition is much less abrupt in the tropics than it is in the midlatitudes.

The meridional age difference $\Delta\Gamma$ increases from 380K up to 500K (where it is the highest), subsequently decreasing above. This is shown more clearly in Figure 2(c) which compares $\Delta\Gamma$ (dashed) to the mean residence time $M/\mathcal{M}$ (solid). Both quantities have very similar structure throughout the vertical but the age difference is constantly less than the upper bound $M/\mathcal{M}$ at all levels. The difference between the solid and dashed curves can be attributed to the diffusive diabatic flux of age, an additional vertical transport unaccounted for in our theory. The presence of diabatic diffusion (Sparling et al. 1997) and numerical diffusion of age reduces the age difference between the two regions, as they enhance the net vertical transport of age by diabatic circulation (Linz et al. 2016).

Figure 2(d) shows the inferred mixing flux μ_{mix} (solid) and the net poleward flux μ_{net} for the finite volume core, derived from the vertical gradients of Γ_u. Both the fluxes rapidly decay in the vertical, reflecting reduction in the diabatic flux associated with decreasing density. The mixing flux is weaker than the net poleward flux at all isentropic levels, though the two are of the same order of magnitude in the lower stratosphere. The mixing flux all but dies away above 600 K. The un-physical negative values reflect a breakdown of the model associated with the fact that the tropical pipe (and extratropics) are not uniformly mixed, i.e., the age is not uniform along isentropes within the tropics/extratropics, as seen in Figure 2(a). The model assumes that air entrained from the tropics leaves with the mean age of tropical air; that air on the boundary of the tropical pipe is somewhat older than the mean causes us to underestimate the transport of age out of the pipe by the diabatic circulation. The mixing flux is computed as a residual, and so can become negative to balance the error. These negative (albeit near zero) values should be interpreted as very weak mixing, the amplitude of "negative" flux crudely quantifying uncertainty in our framework.

4. Mean age gradients: Quantifying vertical diffusion of age

We show the gross ages Γ_u (solid) and Γ_d (dashed) for the GFDL-FV3 (orange) and CAM-SE (green) in Figure 3(a). Since the FV3 core exhibits very similar age profiles for both the FR and SP test, we only show the free running case. For CAM-SE, however, we show results from both the FR and SP integrations. At 380 K, Γ_u and Γ_d in both cores (and for both the FR and SP tests) are nearly identical, indicating that the age of air entering the tropical stratosphere is similar for all integrations (i.e., the models simulation of transport in the troposphere are fairly equivalent). The age profiles, however, diverge quickly with height. In CAM-SE, the FR test (dark green curves) exhibits substantially older air throughout the stratosphere in both the tropics (solid) and the extratropics (dashed). This difference from the FV3 core is reduced when the tropical winds in the CAM-SE core are prescribed in the SP test, albeit not entirely (light green curves with markers). We discuss these differences in detail in Section 5, but first focus on the age difference, $\Delta\Gamma$, pictured in Figure 3(b). This varies substantially less between the integrations, despite the differences in the mean age.

The vertical transport of age in the stratosphere is effected primarily by the tracer advective tendencies due to the diabatic upwelling of mass. In addition, turbulent diffusive fluxes and diabatic dispersion (Sparling et al. 1997) can also introduce additional cross-isentropic diffusive fluxes in the region. We expect this diffusion to be negligible in the tropics (where isentropes line up with pressure), though possibly larger in the extratropics where gradients along pressure and temperature are not always identical. In climate models, depending on the vertical resolution, such physical diffusive fluxes might be partially or completely unresolved. Instead, unphysical (numerical) diffusion in models potentially induces considerable cross-gradient diffusive flux of tracers. These effects can be quantified in the models by comparing the observed age difference, $\Delta\Gamma$, with that implied by advection alone, the mean residence time $M/\mathcal{M}$, as shown in Figure 3(b). The diffusive flux of age, proportional to the negative of vertical age gradient, $-\partial_\theta \Gamma$, is always negative (acting downwards) because the age monotonically increases with height. So, diffusion acts towards reducing the meridional age difference relative to the mean residence time computed from advection alone.

The age difference $\Delta\Gamma$ (solid curve) steadily increases from approximately 1 to 1.25 years at 380 K to 2-2.25 years at 500 K. Above 500 K, it steadily decreases in the vertical. This reflects the variation of the net mass flux $\mathcal{M}$ relative to the density of air (or equivalently, mass above the isentrope, M), i.e., variations in the upward velocity w, albeit inversely. There is a decrease in the upward velocity relative to density through the lower stratosphere, associated with the fall off in synoptic scale wave forcing, and then a relative increase in the middle to upper stratosphere, associated with wave forcing from planetary waves above. In the lower-to-middle stratosphere, the age difference (and residence times) are different between the FV3 and SE integrations, with SE exhibiting a slightly larger $\Delta\Gamma$ and $M/\mathcal{M}$ than FV3. This reflects a weaker diabatic circulation ($\mathcal{M}$) in SE, not any material differences in M.

The dashed curves, which show the mean residence time of air above a given isentrope, have a very similar vertical structure to $\Delta\Gamma$, but with slightly higher magnitude. The separation between the solid and corresponding dashed curves, i.e., the residual in Equation 2, represents the contribution to the age difference due to both diabatic dispersion (a physical phenomena) and any numerical diffusive fluxes in the models.

We estimate the vertical diffusive flux of age as the residual in Equation 2. To obtain the fractional contribution of diffusion, we compute the ratio $\mathcal{D}$ of the net vertical diffusive flux of age to the net resolved vertical advective flux of age across a given isentrope. We define the non-dimensional ratio $\mathcal{D}$ as :

$$\mathcal{D} = \frac{(\Gamma_u + \Gamma_d)(\frac{M}{\Delta\Gamma} - \mathcal{M})}{2(\mathcal{M}_u\Gamma_u + \mathcal{M}_d\Gamma_d)} \tag{4}$$

where the numerator (including the factor of 2 in the denominator) is the net diffusive flux inferred from Equation 2 and the denominator (excluding the factor of 2) is the net advective age flux up and down through an isentrope.

Figure 4(a) shows the ratio $\mathcal{D}$ for 1° FV3 (orange) and SE (green) cores at three different vertical resolutions. Our analysis shows that in the lower stratosphere, the diffusive tracer flux of age accounts for up to 10% of the resolved tracer flux. Moreover, this fraction rapidly decreases with height, more so for the 40 level runs (dotted curves). This finding is consistent with the diffusive boundary layer identified in Figure 3 which indicates that the shapes for Γ_u and Γ_d in the lower stratosphere are largely determined by the diffusive stratospheric-tropospheric exchange of age.

As the models' vertical resolution is doubled from 40 to 80 (and subsequently from 80 to 160), a rapid decrease in the contribution from the numerical vertical fluxes is noticed. For integrations with 160 levels in the vertical, the contribution from vertical diffusion is practically insignificant enough to not affect vertical transport. Figure 4(a) highlights the importance of sufficiently high vertical resolution in idealized and comprehensive stratosphere-resolving climate models used to study UTLS- and stratospheric-transport. Moreover, it indicates that the upper tropospheric and lower stratospheric region is relatively more sensitive to the effects and consequences of model diffusion, as compared to the middle and upper stratosphere.

Figure 4(b) shows the same diffusive-to-advective age flux ratio, but for 80 vertical level FV3 and SE cores integrated at three different horizontal resolutions. We observe the seemingly opposite dependence on resolution: refining the *horizontal* grid increases *vertical* diffusion. As the horizontal resolution is increased from 2° to 1° (and subsequently from 1° to 0.5° for CAM-SE), the diffusive tracer flux increases in both the FV3 and SE core. This dependence of the vertical diffusion on aspect ratio suggests that the net vertical diffusion may be effected by quasi-horizontal diffusion across isentropic surfaces near the grid scale. Allowing additional horizontal motions increases effective diffusion if the vertical resolution is not increased at the same time to properly resolve the finest allowed scales. This perverse effect highlights the importance of considering the aspect ratio when refining the grid of an atmospheric simulation, as highlighted by Lindzen and Fox-Rabinovitz (1989).

While vertical diffusion varies with resolution in both cores, the effect is qualitatively the same, and quantitatively similar, particularly in the lower stratosphere. Moreover, vertical diffusion does not explain the difference in tracer transport between the two cores. We thus turn to differences in the diabatic circulation and adiabatic mixing.

5. Quantifying the impact of mixing and the diabatic circulation

Following Linz et al. (2016) and Linz et al. (in prep.), we connect the full three-dimensional transport in the dynamical cores to the one-dimensional Tropical Leaky Pipe (TLP) model of Neu and Plumb (1999). This allows us to use the simplicity of the TLP to analyze and categorize the transport differences between the GFDL-FV3 and CAM-SE models.

a. A Leaky Pipe Emulator

Our goal is to construct an analogue to the TLP model from three dimensional dynamical core integrations. We thus allow for vertical variations in all the model parameters, which are listed in Table 2. We first compute the optimal value for each parameter based on the dynamical core output. Allowing the coefficients to vary in the vertical precludes closed form, analytic solutions, so we numerically solve the equations using a second-order implicit Crank-Nicolson temporal integrator to obtain the mean tropical and midlatitude ages, Γ_T and Γ_M. After verifying these solutions against the corresponding ages in dynamical cores, Γ_u and Γ_d, we can characterize the importance of differences in mixing, circulation, and other factors by systematically perturbing the parameters.

The vertical coordinate of the TLP is switched from height to entropy (potential temperature θ) and all the isentropic averages are obtained using the binning technique of Yamada and Pauluis (2015); see the supplementary material for details. The average vertical velocity $\dot{\theta}$ is computed over the upwelling and downwelling regions, weighted by mass. Similarly, the isentropic densities σ_u and σ_d are computed by horizontally integrating the isentropic density ρ_θ over the upwelling and downwelling partitions respectively. The mass distribution α between the tropical pipe and the extratropics is estimated as the ratio $\alpha = \sigma_u/(\sigma_u + \sigma_d)$. The scale height in our model stratosphere does not significantly deviate from 6.5 km and so H=6.5 km is chosen.

The mass-flux weighted ages Γ_u and Γ_d, analogous to the tropical and midlatitude ages Γ_T and Γ_M of the original Neu and Plumb model, were already described. The mixing flux from the midlatitudes into the tropics, μ_{mix}, is estimated using Equation 3 (and illustrated in Figure 5), while the net mass flux μ_{net} is computed as the vertical derivative of diabatic mass flux $\mathcal{M}(\theta)$. The two quantities are used to compute the TLP parameter ε, the mixing efficiency. Finally, the vertical diffusion parameters K_T and K_M are tuned to ensure that the model data estimates of Γ_u and Γ_d are in agreement with Γ_T and Γ_M obtained by numerically integrating the TLP equations. We find that for the midlatitudes, a fixed diffusion coefficient of $K_M = 1.5 \times 10^{-2}$ m^2 s^{-1} provides the best fit. The optimal tropical diffusion coefficient K_T, however, slightly varies among the different dynamical cores (but the choice does not materially affect the results).

As discussed in Section 3c, the three dimensional circulation cannot be perfectly mapped onto the 1-d framework. A key assumption of Neu and Plumb (1999) is that fast mixing within the tropical pipe and the midlatitudes completely homogenizes the age in each region; all latitudes in the tropics (midlatitudes) are assumed to have the same age Γ_T (Γ_M) and it does not matter from where air is entrained or mixed between the layers. This is not the case for the simulated stratosphere in our idealized climate models. Figure 2(a) illustrates meridional variations of age in the two regions. Air entrained out of the the tropics (and out of the midlatitudes) will be older than the mean tropical age Γ_u (and younger than the mean midlatitude age Γ_d) leading to a systematic positive biased flux of age between the layers. As a result, the mixing flux μ_{mix} will be systematically negatively biased, to the point that it can become unphysically negative. While this reflects a weakness in the model, it does not affect our ability to reconstruct the age in the two integrations and assess the impact of varying all our parameters.

Several past studies have explored the TLP in the context of models and observations. Ray et al. (2010), for example, applied the TLP to the observed ozone profiles and inferred mean age profiles to connect the multidecadal ozone variability to changes in mean circulation and isentropic mixing. Similarly, Garny et al. (2014) used a linear fit to infer a (vertically invariant) value for the mixing efficiency ε in climate model simulations. They did not account for vertical variations in transport which can prevent a direct comparison between the model age and the age in TLP.

b. Mixing Fluxes vs. Diabatic Fluxes

The mixing flux μ_{mix} is plotted alongside the net poleward flux, μ_{net} for the free running GFDL-FV3 and CAM-SE dynamical cores at 1° L80 resolution in Fig 5(a). The two models exhibit very similar net meridional fluxes μ_{net} (solid curves). μ_{net} is equal to the vertical convergence of the diabatic mass flux $\mathcal{M}$, by mass continuity; the two free running integrations thus exhibit a fairly comparable diabatic circulation, at least in terms of the gross overturning. Minor differences in μ_{net} do affect the transport, but are swamped in the difference in the mixing flux between the two free running integrations.

The mixing fluxes μ_{mix} (dashed lines) differ tremendously between 400 K and 600 K. At 500 K, for instance, $\mu_{mix} = 0.5 \times 10^7$ kg/s-K for the FV3 core, compared to 2.5×10^7 kg/s-K for the SE core: at this level, 5 times the mass is being exchanged between the tropics and midlatitudes in the SE core, even though the net flux from tropics to the extratropics is equivalent. The mixing flux quickly decays to near zero in the mid-stratosphere, but this occurs c. 550 K in the FV3 core vs. 600 K in the SE core. Given the exponential fall off in the mass flux (associated with the exponential fall off in density), it is more illustrative to plot the mixing efficiency $\varepsilon = \mu_{mix}/\mu_{net}$, shown in Figure 5(b). The mixing efficiency measures the "leakiness" of the tropical pipe. In the limit $\varepsilon \to 0$, the tropics and midlatitudes are entirely isolated, apart from the net poleward flow of mass associated with the diabatic circulation, as in the model of Plumb (1996). In the $\varepsilon \to \infty$ limit, explored by Plumb and Ko (1992), the mixing timescale entirely dominates the diabatic flux.

Focusing first on free running integrations (dashed curves in Figure 5b), we see that the mixing flux peaks relative to the diabatic flux in the lowermost stratosphere ($\approx$ 390K) in both models. The key difference between them is in how the mixing falls off above. The peak in ε is associated with enhanced mixing due to synoptic scale wave breaking at the top of the subtropical jets. Above, the mixing efficiency steadily decreases up to 550-600 K, where it ultimately vanishes. The reduction in mixing in the middle and upper stratosphere is consistent with the findings of Haynes and Shuckburgh (2000), who quantified mixing using the effective diffusivity proposed in Nakamura (1996). Their computations also indicated a very low eddy diffusivity in the tropics above 600 K.

The solid curves in Figure 5(b) show the mixing efficiency in the two integrations where the tropical winds were constrained to an easterly profile. There is little difference in mixing between the SP and FR in FV3, consistent with the fact that both models exhibit similar winds in the tropics. (The wind profile was chosen to match the winds observed in all integrations with low vertical resolution; this profile was relatively insensitive to resolution in the FV based cores. As detailed in G20, this choice is not meant to imply that easterlies are the "correct" profile.) The difference between the FR and SP integrations of the SE core indicate that the tropical wind structure strongly controls the rate of mixing. The transition from a climatological state with westerly to easterly winds centered about 600 K (Figure 1b vs. d) drops the ceiling of the high mixing rates, bringing the spectral element core with specified winds into a state similar to the finite volume based core. Westerly winds shift the critical lines deeper into the tropics, allowing deeper penetration of Rossby waves into the tropics, leading to enhanced mixing. Enhanced mixing is associated with the westerly phase of the QBO (Plumb and Bell 1982); here the locked westerlies lead to a climatological increase in mixing.

As discussed in Section 3c, the unphysically negative values of the mixing flux, most pronounced in the free running integration with the spectral element core, reflect the limitation of our method. High mixing between 400 K and 600 K in the SE integration leads to a situation where the age is not well mixed across the tropical pipe. The age distribution is more sharply peaked in the tropics in Figure 1(b) relative to the other integrations, exacerbating this error. We believe that the correct interpretation is that mixing is essentially zero above 600 K, at least until the upper stratosphere above 900 K.

c. The global impact of enhanced mixing in the lower stratosphere in the FR integrations

The Leaky Pipe emulator allows us to diagnose the impact of the difference in mixing between 400 and 600 K on tracer transport throughout the stratosphere. Using the parameters computed from the 3D model integrations (tabulated in Table 2), we first establish that one can reconstruct the age in the tropics and extratropics, Γ_u and Γ_d, with our numerical solver of the variable coefficient TLP model. As illustrated in Figure 6, there is a good match with Γ_T and Γ_M, respectively; here we have kept the notation of Neu and Plumb (1999) to distinguish the ages from the dynamical core integrations and the TLP emulator. The fit is so good for the tropical pipe in the SP integrations that Γ_u and Γ_T cannot be distinguished. This is not a surprising result – the parameters were fit with knowledge of the ages Γ_u and Γ_d – but establishes that the emulator works well with fixed diffusivity coefficients in the vertical.

To assess the contribution of each transport process to the spread in age distributions in the model integrations, we take the following approach. Starting with the parameters obtained for the FV3 core, we perturb the parameters, one at a time (or in related groups), computing the impact of each process on the mean age of the tropics and extratropics. Figure 7 shows the results for the free running integrations. For example, the orange curve in panel (a) shows Γ_T in the TLP emulator, our fit to the mass weighted age Γ_u of the FR integrations of the FV3 core. The blue dashed curves shows the TLP emulator tropical age when the mixing efficiency ε is changed from its value in the FV3 fit to the SE fit, all other parameters left the same. The difference in mixing alone explains the bulk of the difference between the two integrations, as shown explicitly in Figure 7(b).

Before commenting on each process, we first establish that the results are sufficiently linear to separate them. The purple dotted curve in Figure 7(b) shows the sum of differences associated with each process, compared to total difference between Γ_T fit for the SE and FV3 models in the green curve. For the FR integrations, the sum of the parts is always very good ($\approx$ 95% of the total) up to 1000 K. Above this altitude, nonlinearity starts to matter (i.e., the change in age associated with mixing is exacerbated by differences in other processes). For the SP integrations, where differences are smaller, the result is extremely linear and the sum is omitted for clarity in Figure 8.

Three parameter groups characterize the key processes that govern stratospheric transport:

i. the diabatic circulation: μ_{net} and α (which characterizes the geometry of the flow)

ii. isentropic mixing : ε, and

iii. the input of age at the tropopause: the age at θ=380 K

The impact of switching these parameters from the optimal FV3 value to those based on SE are shown in Figure 7. The differences associated with the diabatic circulation (in red) and boundary conditions (in yellow) do not explain the tropical and midlatitude age differences between the FV3 and SE free running integrations. A striking jump in age, however, occurs when the SE mixing profile is imposed onto FV3. The difference in mixing (blue curves in Figure 7) accounts for almost 75%-80% of the net age difference between FV3 and SE free running integrations (green curves).

The contribution due to difference in diabatic circulation strength is small, not greater than 10% in the middle stratosphere, though weaker overturning in the lower stratosphere in the SE model does amplify the aging associated with mixing. The contribution from differences in boundary conditions (i.e., age transport in the troposphere) is relatively small as well, not accounting for more than 10% of the age difference. The minor difference left unaccounted by i-iii is the difference in model diffusion between the CAM-SE and GFDL-FV3 cores. Nonlinear effects matter in the upper troposphere, as differences in age due to mixing and circulation lead to an overall greater difference between the cores.

d. The importance of circulation differences in the SP integrations

We repeated the TLP analysis for the SP runs in Figure 8. For the integrations with similar tropical wind climatologies, the net difference in age is much smaller as compared to the FR runs. Figures 8 suggests that most of the age differences among the SP runs are caused by differences in the diabatic circulation (dashed red) and boundary conditions associated with transport through the troposphere (dashed yellow) between the cores. Differences in boundary conditions provides a straight offset of 0.2 years (up to 50%) in both the tropics and the midlatitudes. Further, differences in diabatic circulation accounts for almost 50% of the total age differences in the middle stratosphere. Differences due to mixing are relatively small up to 900 K. Enhanced mixing in the SE core leads to an aging of air in the lower stratosphere, younger age in the upper stratosphere (dashed blue curve in Figure 8).

Even though the relative contribution of diabatic velocity differences is lower for the FR runs than for the SP runs, the absolute contribution remains quite similar. More simply put, for both the FR and SP runs, differences in diabatic circulation accounts for as much as 0.25 yrs (0.4 yrs) of the net difference in Γ_T (Γ_M). This suggest that differences in the diabatic circulation are unrelated to differences in the tropical wind climatology (and unaffected by the mild damping used to constrain the winds). The diabatic circulation, $\mathcal{M}$, for the two models is compared in Figure 9. The FV3 core consistently exhibits a stronger circulation, particularly in the middle stratosphere (450-700K).

To compare the meridional structure of the diabatic circulation among models following G20, we compute the diabatic streamfunction (Townsend and Johnson 1985; Pauluis et al. 2009), shown in Figure 9(b) and (c). For the free running integrations, the key difference between the models is in the tropics. The tropical westerlies induce a secondary mean meridional circulation in the tropics (Plumb and Bell 1982). The presence of this mean meridional circulation can be seen in Figure 9(c) as a *shoulder* in the green dashed curve at 500 K between the equator and 30°N, which is not noticed at 600K (Figure 9(b)). When the winds in the tropics are constrained, the shoulder in the circulation for CAM-SE at 500 K disappears and a strengthened circulation for the core is noticed at both 500 and 600 K. For the FV3 core (which has similar tropical winds for both FR and SP runs) the circulation strength is quite similar. The two state-of-the-art models with different numerics still maintain noticeable differences in the circulation strength in the middle stratosphere, despite being identically forced. While for the FR runs, these differences are overshadowed by major differences in mixing, for the SP integrations, these differences account for up to half of the overall age difference between the models.

6. Conclusion

We have assessed how the formulation of an atmospheric model's dynamical core impacts stratospheric transport using the theory of age transport in isentropic coordinates developed by Linz et al. (2016) and Linz et al. (in prep.). We examined two modern dynamical cores, GFDL's Finite Volume 3 (FV3) core and the Community Atmosphere Model's Spectral Element (CAM-SE) core, which employ very different underlying numerical methods to solve the primitive equations (Section 2). Both models were run at comparable resolutions and driven with identical diabatic and tracer forcings, as prescribed in the transport benchmark tests established by Gupta et al. (2020). Despite the carefully prescribed test environment, the two cores diverge substantially in their representation of stratospheric transport, particularly in a "free running" configuration where the models produce a very different representation of the zonal winds in the tropical stratosphere, as highlighted in Figure 1 and detailed by Gupta et al. (2020). Variations in transport can arise from differences in the simulation of the explicitly resolved circulation (both the slow overturning circulation, or diabatic circulation, and rapid mixing of air along isentropes), and implicit differences in transport by the numerical schemes (trace gas representation, grid scale diffusion, and other errors). A careful analysis of the age budget in isentropic coordinates and a vertically varying formulation of the tropical leaky pipe model (TLP; Neu and Plumb 1999) allowed us to diagnose the individual impact of each factor on the circulation.

As the transport is sensitive to vertical resolution, particularly in CAM-SE, we first quantified the contribution of numerical diffusion in Section 4. While the contribution from diffusion is tied to vertical resolution (Figure 4), it does not explain the gross differences between the models. At moderate vertical resolution (40 vertical levels), the diffusive tracer flux is on the order of 10% of the resolved diabatic fluxes in both models. Its contribution rapidly decreases with increasing vertical resolution, becoming negligible in 160 level runs. Since most comprehensive climate models still employ less than 40 levels in the vertical, our findings highlight the importance of vertical resolution for studying stratospheric transport. It is differences in the explicitly resolved tracer transport, however, that dominate differences between GFDL-FV3 and CAM-SE in the benchmark experiments.

As established by Linz et al. (2016) and Linz et al. (in prep.) and reviewed in Section 3, the vertical gradient of the gross upwelling age, Γ_u, allows us to quantify the adiabatic mixing flux, μ_{mix}, the rapid transport of tracer between the tropics and extratropics along isentropic surfaces by breaking waves. The key to relating age in a three-dimensional model to the TLP is to weight the flow by the mass flux when computing the average age Γ_u. Equation (3) relates the vertical age gradient to the aging purely by diabatic advection versus "aging by mixing", allowing us to estimate the mixing flux in the two models as a residual. With the vertically varying TLP emulator, we can then quantify the impact of differences in the mean diabatic circulation vs. differences in mixing.

For the free running (FR) integrations, differences in transport are dominated by mixing. The mixing flux varied immensely in the middle stratosphere (450-650 K): CAM-SE exchanges up to five-times more air between the midlatitudes and the tropics relative to GFDL-FV3 (Figure 5). This enhanced mixing accounts for up to 75% of the total observed age differences among the FR model runs (Figure 7). The specified tropical wind test integrations (SP test) establish that it is the zonal wind profile in the tropics that controls this difference in mixing. The formation of westerly jets in CAM-SE allows deeper penetration of Rossby waves into the upwelling branch of the diabatic circulation (i.e., the tropical pipe), enhancing mixing. Once the tropical winds are constrained to easterlies in the two models, the models simulate nearly identical mixing.

Even in the constrained SP tests, however, the simulation of mean stratospheric age differs by up to 15% between the two models, a difference that remains fairly constant as the resolution is increased in both the horizontal and vertical. The TLP analysis suggests that differences in net diabatic circulation account for the bulk of the age difference (50%, Figure 8). Differences in mixing still account for a non-negligible fraction of the difference, but the second largest contribution are differences in the boundary conditions at the tropopause (i.e., differences in transport through the troposphere, which are also dominated by the mean overturning circulation in the tropics).

Tracing the difference in the diabatic circulation back to the numerical formulation of the dynamical core is beyond the scope of our analysis. The strength of the diabatic circulation is set by the efficiency of large scale Rossby waves in mixing potential vorticity in the stratosphere, driving the extratropical pump which lifts mass up into the tropics and back down in the extratropics

(Haynes et al. 1991; Holton et al. 1995). The stronger diabatic circulation in GFDL-FV3 is consistent with its slightly weaker climatological polar vortex, in that both are responses to stronger drag on the mean flow by eddies. These differences must relate to differences in the efficiency of wave breaking. The potential for a positive feedback between a weaker vortex, which allows further Rossby wave propagation, which in turn further weakens the vortex, could exacerbate subtle differences in the numerics. The enhanced wave driving implies more mixing of potential vorticity by wave breaking; assuming a fixed mixing efficiency, this would lead to enhanced mixing of tracer as well, which will partially offset the reduction of aging by the stronger diabatic circulation.

In the lower stratosphere (near 390 K), however, the SE core exhibits greater mixing efficiency (Figure 5(b)), leading to additional aging of the flow up to about 600 K (Figure 8). Above this height, the mixing efficiency is comparable in both cores (implying more mixing in the FV3 core), which does partially offset the age bias associated with the weaker diabatic circulation of SE.

Our application of the TLP forms a natural extension to the work of Ray et al. (2010) which used the TLP to study the changes in mean circulation and isentropic mixing to study multidecadal ozone variability, even explaining the effects of extreme events like volcanic explosions on the stratospheric ozone variability. Our framework, which considers full vertical variations of the TLP parameters, provides an effective way to assess the individual impact on age due to the key dynamical processes in the stratosphere. Conversely, the framework developed in this study can also be used to more accurately diagnose multidecadal changes in the vertical structure of stratospheric circulation and mixing, using satellite-based observations of stratospheric tracers or comprehensive climate models.

Our results establish the importance of the tropical stratosphere on stratospheric transport (Punge et al. 2009) and the sensitivity of tropical stratospheric climatology to model numerics (Yao and Jablonowski 2015). The momentum balance of the tropical stratosphere is a stiff test for model numerics and is vital for both dynamics (e.g., the Quasi-biennial Oscillation) and transport, particularly in the relative isolation of the tropical upwelling region. The tropical winds impact stratospheric transport primarily by modulating the mixing fluxes and the meridional extent of their propagation. Minor differences in the tropical winds can have consequences for global stratospheric transport primarily because most of the air enters the stratosphere in the tropics, and subtle changes in mixing can have a large impact on the stratosphere above. Most stratospheric ozone is produced in the tropical stratosphere, and its transport by the mean overturning circulation and isentropic mixing sets the global climatology.

The formulation of a dynamical core impacts the stratospheric transport both through grid scale diffusion and truncation errors (e.g., conservation), and through the influence of the model numerics on the global circulation. With state-of-the-art cores developed by the Geophysical Fluid Dynamics Laboratory (FV3) and the National Center for Atmospheric Research (CAM-SE), differences in transport appear to be dominated by latter: even with identical forcing, subtle differences in the climatological circulation and mixing lead to significant differences in transport. It is tempting to hope that, given reasonably high resolution, tuning a model to the correct climatological circulation is sufficient to accurately model transport. Chrysanthou et al. (2019), however, suggest that the manner it which you correct model biases matters a lot. They found significant differences in diabatic circulation strength among an ensemble of models nudged towards the same reanalysis. Furthermore, Marianna Linz (personal communication) found a degradation of transport in Specified Dynamics (SD-) WACCM relative to free running WACCM. In both of these cases, there is a danger that nudging a model to the correct climate state leads to an imbalance between model's transport and dynamics. The key is to get the right climatological state *self-consistent* with the underlying dynamics of the model.

Acknowledgments. We acknowledge support of the US National Science Foundation through grant AGS-1852727 to New York University. We also thank Marianna Linz and Olivier Pauluis for insightful discussion and constructive criticism of the research.

References

Baldwin, M. P., and Coauthors, 2001: The quasi-biennial oscillation. *Reviews of Geophysics*, **39 (2)**, 179–229, doi:10.1029/1999RG000073.

Birner, T., and H. Bönisch, 2011: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere. *Atmospheric Chemistry and Physics*, **11 (2)**, 817–827, doi: 10.5194/acp-11-817-2011.

Butchart, N., 2014: The Brewer-Dobson circulation. *Reviews of Geophysics*, **52 (2)**, 157–184, doi:10.1002/2013RG000448.

Butchart, N., and Coauthors, 2018: Overview of experiment design and comparison of models participating in phase 1 of the SPARC Quasi-Biennial Oscillation initiative (QBOi). *Geoscientific Model Development*, **11 (3)**, 1009–1032, doi:10.5194/gmd-11-1009-2018.

Chrysanthou, A., and Coauthors, 2019: The effect of atmospheric nudging on the stratospheric residual circulation in chemistry–climate models. *Atmospheric Chemistry and Physics*, **19 (17)**, 11 559–11 586, doi:10.5194/acp-19-11559-2019.

Collins, W. J., and Coauthors, 2017: AerChemMIP: Quantifying the effects of chemistry and aerosols in CMIP6. *Geoscientific Model Development*, **10 (2)**, 585–607, doi:10.5194/gmd-10-585-2017.

Donner, L. J., and Coauthors, 2011: The Dynamical Core, Physical Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL Global Coupled Model CM3. *Journal of Climate*, **24 (13)**, 3484–3519, doi:10.1175/2011JCLI3955.1.

Engel, A., and Coauthors, 2009: Age of stratospheric air unchanged within uncertainties over the past 30 years. *Nature Geoscience*, **2 (1)**, 28–31, doi:10.1038/ngeo388.

Eyring, V., and Coauthors, 2007: Multimodel projections of stratospheric ozone in the 21st century. *Journal of Geophysical Research: Atmospheres*, **112 (D16)**, doi:10.1029/2006JD008332.

Friedl, L., 2007: Aerosols, air quality, and international policy. *International Seminar on Nuclear War and Planetary Emergencies ? 36th Session*, The Science and Culture Series ? Nuclear Strategy and Peace Technology, WORLD SCIENTIFIC, 151–162, doi:10.1142/9789812709233_0023.

Garny, H., T. Birner, H. Bönisch, and F. Bunzel, 2014: The effects of mixing on age of air. *Journal of Geophysical Research: Atmospheres*, **119 (12)**, 7015–7034, doi:10.1002/2013JD021417.

Gupta, A., E. P. Gerber, and P. H. Lauritzen, 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, doi:10.1002/qj.3881.

Hall, T. M., and R. A. Plumb, 1994: Age as a diagnostic of stratospheric transport. *Journal of Geophysical Research*, **99 (D1)**, 1059, doi:10.1029/93JD03192.

Hall, T. M., D. W. Waugh, K. A. Boering, and R. A. Plumb, 1999: Evaluation of transport in stratospheric models. *Journal of Geophysical Research: Atmospheres*, **104 (D15)**, 18 815–18 839, doi:10.1029/1999JD900226.

Haynes, P., and E. Shuckburgh, 2000: Effective diffusivity as a diagnostic of atmospheric transport: 1. Stratosphere. *Journal of Geophysical Research: Atmospheres*, **105 (D18)**, 22 777–22 794, doi:10.1029/2000JD900093.

Haynes, P. H., M. E. McIntyre, T. G. Shepherd, C. J. Marks, and K. P. Shine, 1991: On the "Downward Control" of Extratropical Diabatic Circulations by Eddy-Induced Mean Zonal Forces. *Journal of the Atmospheric Sciences*, **48 (4)**, 651–678, doi:10.1175/1520-0469(1991)048⟨0651:OTCOED⟩2.0.CO;2.

Held, I. M., and M. J. Suarez, 1994: A Proposal for the Intercomparison of the Dynamical Cores of Atmospheric General Circulation Models. *Bulletin of the American Meteorological Society*, **75 (10)**, 1825–1830, doi:10.1175/1520-0477(1994)075⟨1825:APFTIO⟩2.0.CO;2.

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

Lauritzen, P. H., and Coauthors, 2018: NCAR Release of CAM-SE in CESM2.0: A Reformulation of the Spectral Element Dynamical Core in Dry-Mass Vertical Coordinates With Comprehensive Treatment of Condensates and Energy. *Journal of Advances in Modeling Earth Systems*, **10 (7)**, 1537–1570, doi:10.1029/2017MS001257.

Lindzen, R. S., and M. Fox-Rabinovitz, 1989: Consistent Vertical and Horizontal Resolution. *Monthly Weather Review*, **117 (11)**, 2575–2583, doi:10.1175/1520-0493(1989)117⟨2575:CVAHR⟩2.0.CO;2.

Linz, M., R. A. Plumb, E. P. Gerber, F. J. Haenel, G. Stiller, D. E. Kinnison, A. Ming, and J. L. Neu, 2017: The strength of the meridional overturning circulation of the stratosphere. *Nature*

Geoscience, **10 (9)**, 663–667, doi:10.1038/ngeo3013.

Linz, M., R. A. Plumb, E. P. Gerber, and A. Sheshadri, 2016: The Relationship between Age of Air and the Diabatic Circulation of the Stratosphere. *Journal of the Atmospheric Sciences*, **73 (11)**, 4507–4518, doi:10.1175/JAS-D-16-0125.1.

Linz, M., R. A. Plumb, A. Gupta, E. P. Gerber, Edwin P., and D. E. Kinnison, in prep.: Stratospheric adiabatic mixing rate derived from the vertical age gradient. *Journal of Geophysical Research: Atmospheres*.

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

Molina, M. J., and F. S. Rowland, 1974: Stratospheric sink for chlorofluoromethanes: Chlorine atom-catalysed destruction of ozone. *Nature*, **249 (5460)**, 810–812, doi:10.1038/249810a0.

Nakamura, N., 1996: Two-Dimensional Mixing, Edge Formation, and Permeability Diagnosed in an Area Coordinate. *Journal of the Atmospheric Sciences*, **53 (11)**, 1524–1537, doi:10.1175/1520-0469(1996)053⟨1524:TDMEFA⟩2.0.CO;2.

Neu, J. L., and R. A. Plumb, 1999: Age of air in a "leaky pipe" model of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, **104 (D16)**, 19 243–19 255, doi:10.1029/1999JD900251.

Orbe, C., M. Holzer, L. M. Polvani, and D. Waugh, 2013: Air-mass origin as a diagnostic of tropospheric transport. *Journal of Geophysical Research: Atmospheres*, **118 (3)**, 1459–1470, doi:10.1002/jgrd.50133.

Pauluis, O., A. Czaja, and R. Korty, 2009: The Global Atmospheric Circulation in Moist Isentropic Coordinates. *Journal of Climate*, **23 (11)**, 3077–3093, doi:10.1175/2009JCLI2789.1.

Plumb, R. A., 1996: A "tropical pipe" model of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, **101 (D2)**, 3957–3972, doi:10.1029/95JD03002.

Plumb, R. A., 2002: Stratospheric Transport. *Journal of the Meteorological Society of Japan. Ser. II*, **80 (4B)**, 793–809, doi:10.2151/jmsj.80.793.

Plumb, R. A., and R. C. Bell, 1982: A model of the quasi-biennial oscillation on an equatorial beta-plane. *Quarterly Journal of the Royal Meteorological Society*, **108 (456)**, 335–352, doi: 10.1002/qj.49710845604.

Plumb, R. A., and M. K. W. Ko, 1992: Interrelationships between mixing ratios of long-lived stratospheric constituents. *Journal of Geophysical Research: Atmospheres*, **97 (D9)**, 10 145–10 156, doi:10.1029/92JD00450.

Polvani, L. M., and P. J. Kushner, 2002: Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. *Geophysical Research Letters*, **29 (7)**, 18–1–18–4, doi:10.1029/2001GL014284.

Punge, H. J., P. Konopka, M. A. Giorgetta, and R. Müller, 2009: Effects of the quasi-biennial oscillation on low-latitude transport in the stratosphere derived from trajectory calculations. *Journal of Geophysical Research: Atmospheres*, **114 (D3)**, doi:10.1029/2008JD010518.

Ray, E. A., and Coauthors, 2010: Evidence for changes in stratospheric transport and mixing over the past three decades based on multiple data sets and tropical leaky pipe analysis. *Journal of Geophysical Research: Atmospheres*, **115 (D21)**, doi:10.1029/2010JD014206.

Sparling, L. C., J. A. Kettleborough, P. H. Haynes, M. E. McIntyre, J. E. Rosenfield, M. R. Schoeberl, and P. A. Newman, 1997: Diabatic cross-isentropic dispersion in the lower stratosphere. *Journal of Geophysical Research: Atmospheres*, **102 (D22)**, 25 817–25 829, doi:10.1029/97JD01968.

Townsend, R. D., and D. R. Johnson, 1985: A Diagnostic Study of the Isentropic Zonally Averaged Mass Circulation during the First GARP Global Experiment. *Journal of the Atmospheric Sciences*, **42 (15)**, 1565–1579, doi:10.1175/1520-0469(1985)042⟨1565:ADSOTI⟩2.0.CO;2.

Waugh, D., and T. Hall, 2002: Age of Stratospheric Air: Theory, Observations, and Models. *Reviews of Geophysics*, **40 (4)**, 1–1–1–26, doi:10.1029/2000RG000101.

Waugh, D. W., and Coauthors, 1997: Mixing of polar vortex air into middle latitudes as revealed by tracer-tracer scatterplots. *Journal of Geophysical Research: Atmospheres*, **102 (D11)**, 13 119–13 134, doi:10.1029/96JD03715.

Yamada, R., and O. Pauluis, 2015: Momentum Balance and Eliassen–Palm Flux on Moist Isentropic Surfaces. *Journal of the Atmospheric Sciences*, **73 (3)**, 1293–1314, doi:10.1175/JAS-D-15-0229.1.

Yao, W., and C. Jablonowski, 2015: Idealized Quasi-Biennial Oscillations in an Ensemble of Dry GCM Dynamical Cores. *Journal of the Atmospheric Sciences*, **72 (6)**, 2201–2226, doi:10.1175/JAS-D-14-0236.1.

LIST OF TABLES

Numerics	**Resolution**	**Grid Size**	**Resolution**
GFDL - Cubed Sphere Finite Volume	C48L40	$192 \times 96 \times 40$	$2^\circ \times 2^\circ$
	C48L80	$192 \times 96 \times 80$	
	C90L40	$360 \times 180 \times 40$	$1^\circ \times 1^\circ$
	C90L80	$360 \times 180 \times 80$	
CAM - Spectral Element	NE16L40	$256 \times 129 \times 40$	$2^\circ \times 2^\circ$
	NE16L80	$256 \times 129 \times 80$	
	NE30L40	$512 \times 257 \times 40$	$1^\circ \times 1^\circ$
	NE30L80	$512 \times 257 \times 80$	
	NE30L160	$512 \times 257 \times 160$	
	NE60L80	$1024 \times 513 \times 80$	$0.5^\circ \times 0.5^\circ$

TABLE 1. The following 10 model runs using the two different dynamical cores : GFDL-FV3 and CAM-SE were considered in the study. The models were integrated for 30 years at these resolutions, in both the free running (FR) and specified tropics (SP) configuration as described in Section 2.

Physical Quantity	**Neu and Plumb (1999)**	**Model Analogue**
Scale height (H)	7 km	6.5 km
Vertical velocity (W_T)	0.3×10^{-3}m s^{-1}	$\dot{\theta}\left\|\frac{d\theta}{dp}\right\| = \frac{\int_u \rho_\theta \dot{\theta} dA}{\int_u \rho_\theta dA}\left\|\frac{d\theta}{dp}\right\|$
Mass distribution (α)	0.5	$\frac{\sigma_u}{\sigma_u+\sigma_d}$
Entrainment ratio (λ)	$\frac{-\alpha W_T}{H}$	-$\partial_\theta \mathcal{M}_u/\sigma_d$
Mixing (ε)	$\in [0,1]$	μ_{mix}/μ_{net}
Upwelling Age (Γ_T)	0 (z=0)	Γ_u(θ=380K)
Downwelling Age (Γ_M)	0 (z=0)	Γ_d(θ=380K)
Vertical diffusion (K_T)	Model dependent	Model dependent
Vertical diffusion (K_M)	0.5 m^2 s^{-1}	1.5×10^{-2} m^2 s^{-1}

TABLE 2. (middle column) The original leaky pipe parameters and (right column) the corresponding climate model analogues used to compute the vertically varying parameters in the leaky pipe model. The diffusion coefficient in the tropics was tuned to match the age from the models and from the solvers and is model dependent. For GFDL-FV3, $K_T = K_M$ and for CAM-SE, $K_T = 0.5K_M$.

LIST OF FIGURES

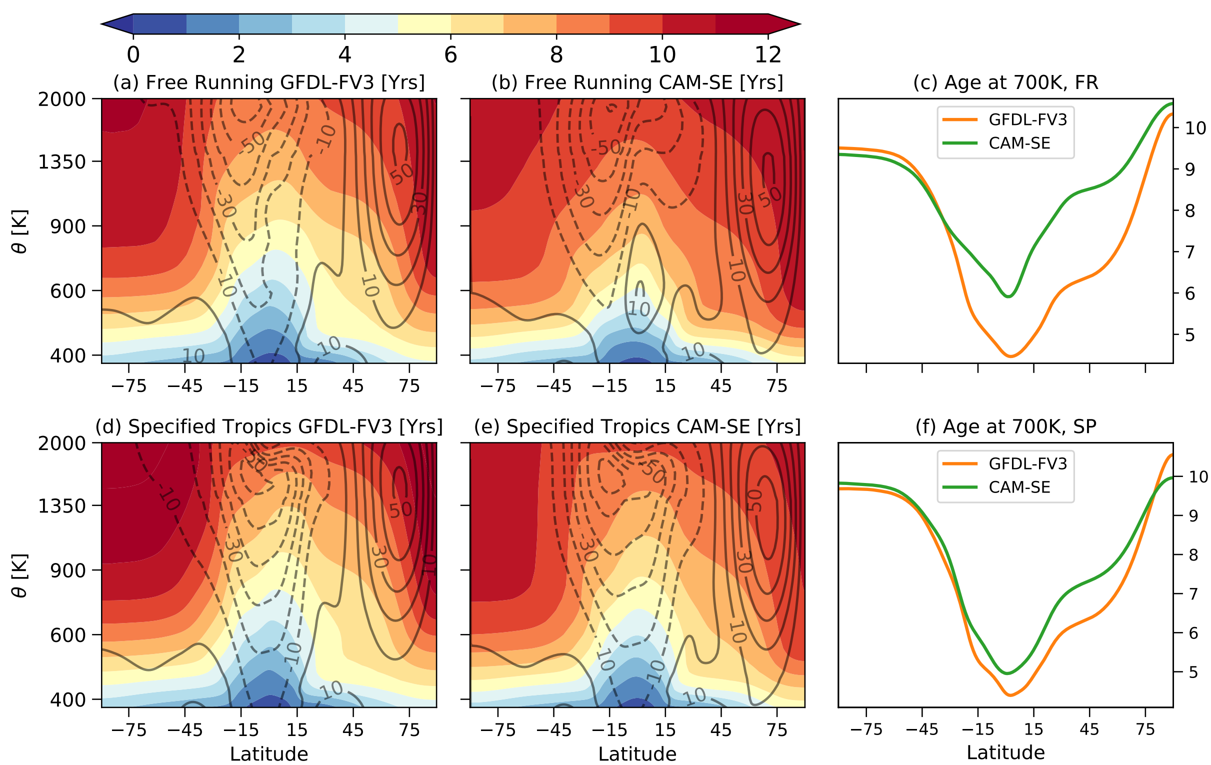
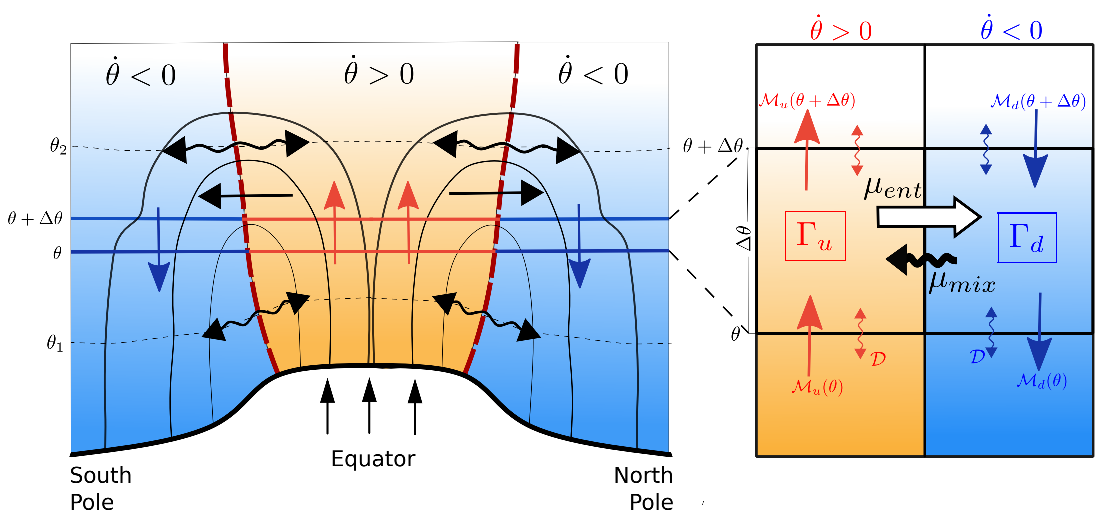
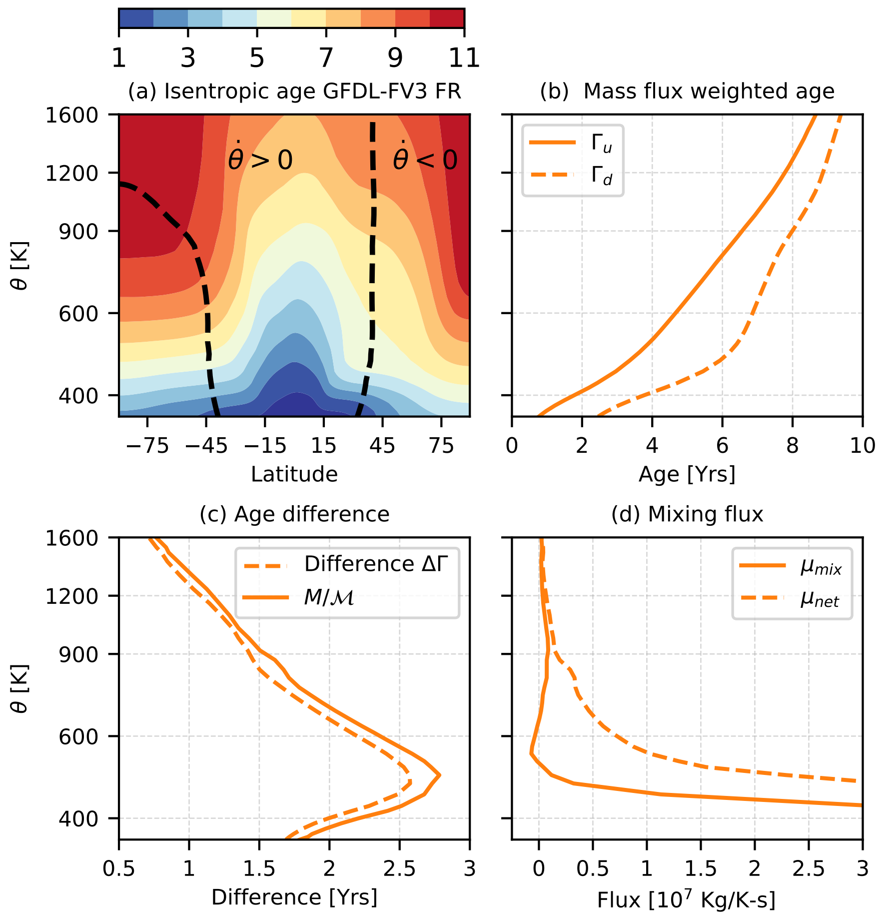


FIG. 1. Zonal mean age (in color) in years and zonal mean winds (in black) in m/s for the (a) free running GFDL-FV3, (b) free running CAM-SE core, (d) specified tropics GFDL-FV3 core and (e) specified tropics CAM-SE. Subplots (c) and (f) show the zonal mean age at 700K isentropic level for the FV3 (orange) and CAM-SE (green) core for the free running and specified tropics integrations respectively.


$\dot{\theta} < 0$
$\dot{\theta} > 0$
$\dot{\theta} < 0$
θ_2
$\theta + \Delta\theta$
θ
θ_1
South Pole
Equator
North Pole
$\dot{\theta} > 0$
$\dot{\theta} < 0$
$\mathcal{M}_u(\theta + \Delta\theta)$
$\mathcal{M}_d(\theta + \Delta\theta)$
$\theta + \Delta\theta$
μ_{ent}
$\Delta\theta$
Γ_u
Γ_d
μ_{mix}
θ
$\mathcal{D}$
$\mathcal{D}$
$\mathcal{M}_u(\theta)$
$\mathcal{M}_d(\theta)$



870 FIG. 2. Demonstration of the transport metrics introduced in Section 3. (a) Zonal mean age (color), in years, in isentropic coordinates for the free running 1° L80 GFDL-FV3 and the curve demarcating the upwelling ($\dot{\theta} > 0$) and downwelling ($\dot{\theta} < 0$) regions. (b) Mass flux weighted ages Γ_u (solid) and Γ_d (dashed) for the FV3 run. (c) The age difference $\Delta\Gamma$ (dashed), in years, and the mean residence time $M/\mathcal{M}$ for the FV3 run. (d) The mixing flux μ_{mix} (solid) and the net poleward flux μ_{net} (dashed).

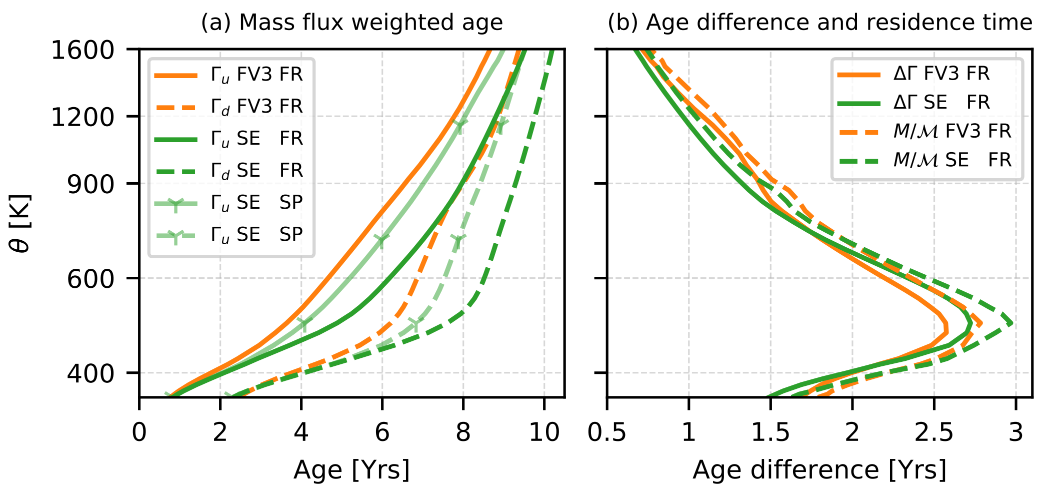


FIG. 3. (a) Gross upwelling and downwelling ages Γ_u (in solid) and Γ_d (in dashed) for GFDL-FV3 (in orange) and the CAM-SE (in green). The green curves with markers show the same quantities but for the SP runs in CAM-SE. (b) Age difference and residence time for the FR runs in the two models. The solid curves show the meridional age difference $\Delta\Gamma$ in the models and the dashed curves show the mean residence time $M/\mathcal{M}$ in the models.

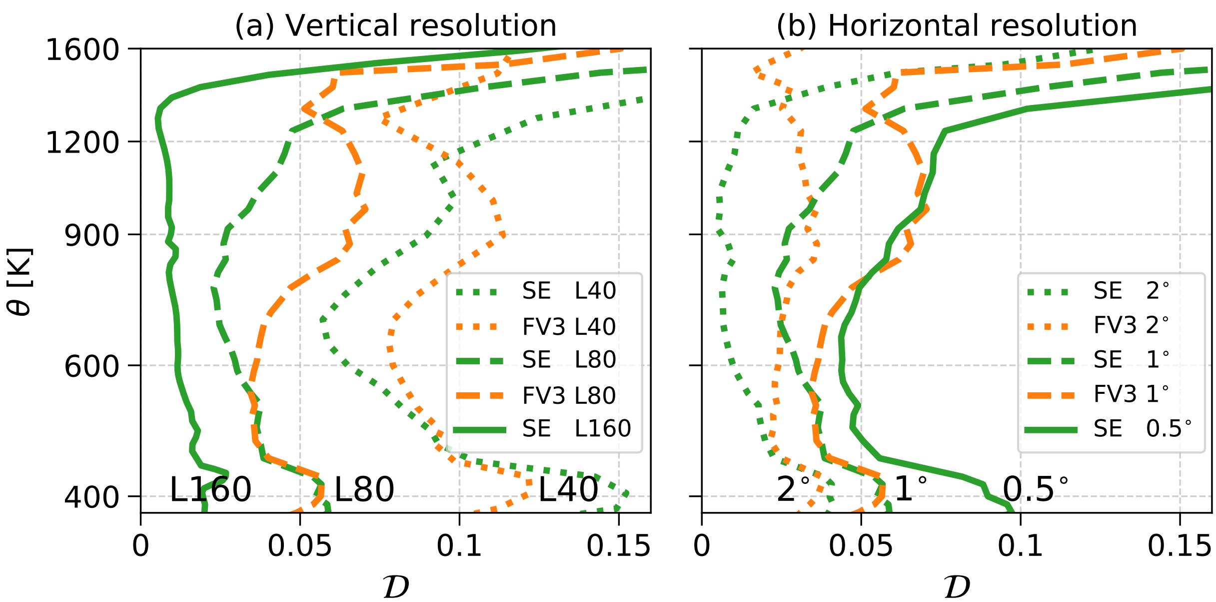


FIG. 4. Effect of changing (a) the vertical resolution and (b) the horizontal resolution on the diffusive + unresolved fluxes in the FV3 core (orange) and the SE core (green). The figure shows the non-dimensional ratio $\mathcal{D}$, defined in Equation 4. The dotted curves, dashed curves and solid curves in (a) represent SP runs with 40, 80 and 160 levels in the vertical at 1° horizontal resolution. The dotted curves, dashed curves and solid curves in (b) represent SP runs with 2°, 1° and 0.5° resolution in the horizontal resolution with 80 levels in the vertical. All the curves shown correspond to the specified tropics runs for the two models.

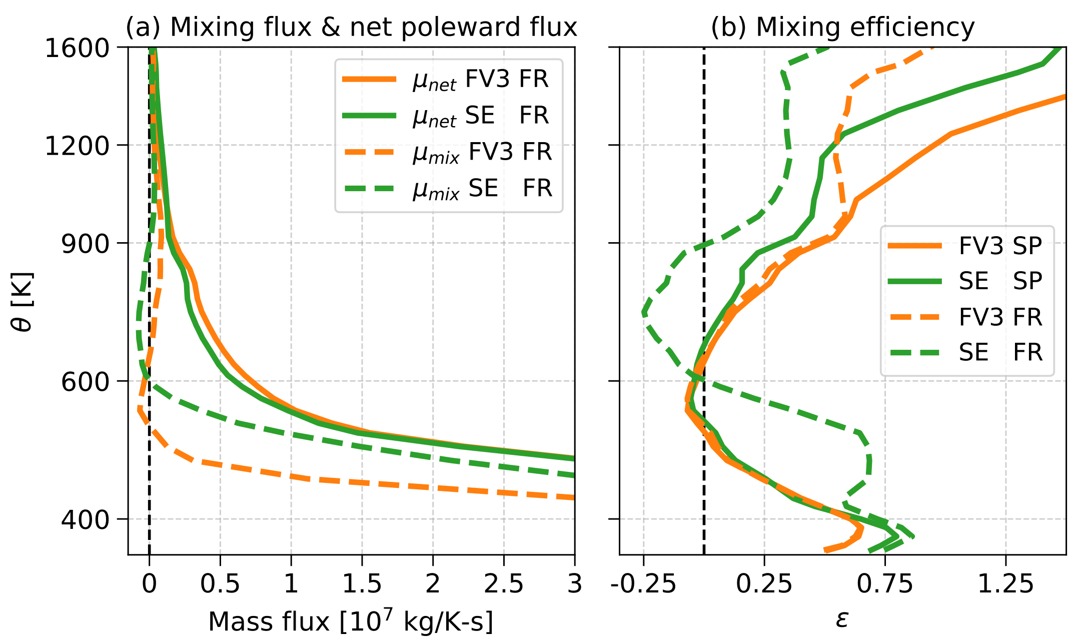


FIG. 5. (a) The mixing flux μ_{mix} and (b) the mixing efficiency ε for the GFDL-FV3 core (shown in orange) and CAM-SE core (shown in green). The net poleward flux μ_{net} in (a) is shown in solid curves and the mixing flux μ_{mix} is shown in dashed curves. The dashed (solid) curves in (b) show the mixing efficiency for the free running (specified) integrations.

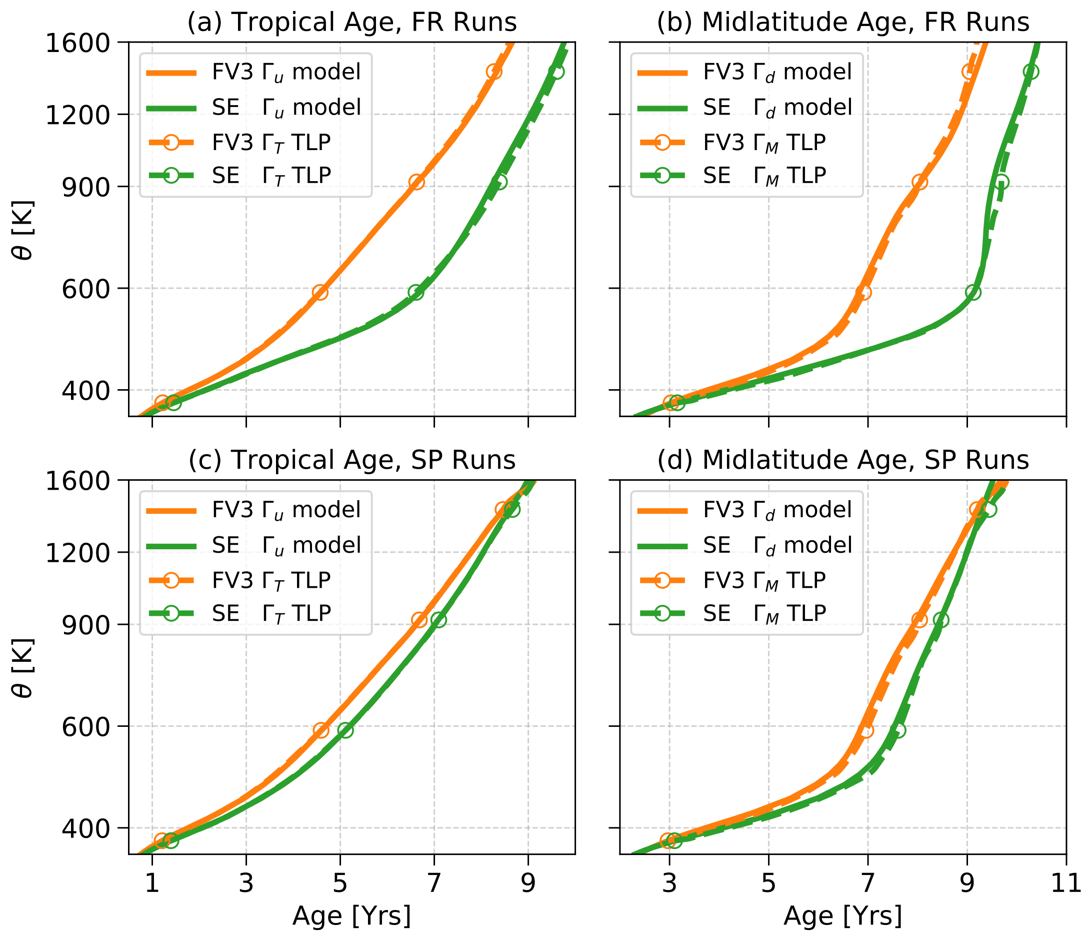


FIG. 6. Tropical and midlatitude age from numerically integrated leaky pipe emulator for the two models. The solid orange (green) curves show the gross ages Γ_u and Γ_d, obtained by mass-flux weighted averaging, for the GFDL-FV3 (CAM-SE cores) and the dashed orange (green) curves show the corresponding ages Γ_T and Γ_M, obtained from the TLP emulator when forced with the parameters obtained from model output.

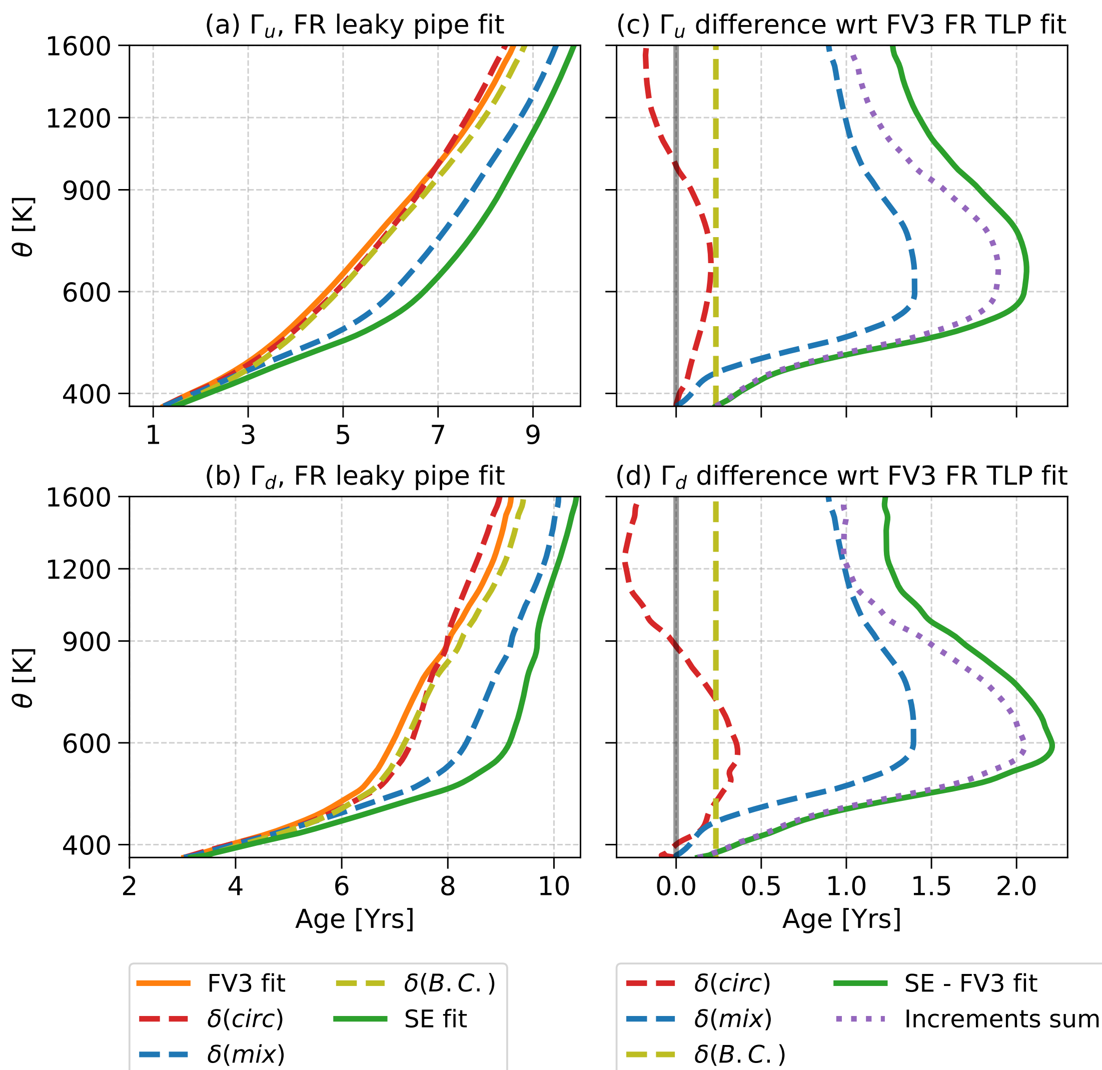


FIG. 7. Numerical solutions obtained from perturbation analysis using the TLP emulator. (a) and (b) show the Γ_T and Γ_M respectively obtained from the numerically integrated TLP for the set of individual increments i.e. orange and green curves show the ages obtained using GFDL-FV3 and CAM-SE obtained parameters and red, blue and yellow curves show the age obtained when the diabatic circulation strength, isentropic mixing and 380 K boundary conditions from CAM-SE are *individually* forced onto the GFDL-FV3 core. (c) and (d) respectively show the difference obtained (with respect to the FV3 fit) for each of the increments corresponding to the same color as in (a) and (b). For (c) and (d), the green curve shows the age difference between the green and orange curves in (a) and (b) i.e. the difference between the SE and FV3 age from the emulator. The dotted violet curve in (c) and (d) is the sum of individual age increments obtained by substituting SE core's diabatic circulation, mixing efficiency, boundary conditions and tuned diffusivity coefficients onto FV3 in the TLP emulator.

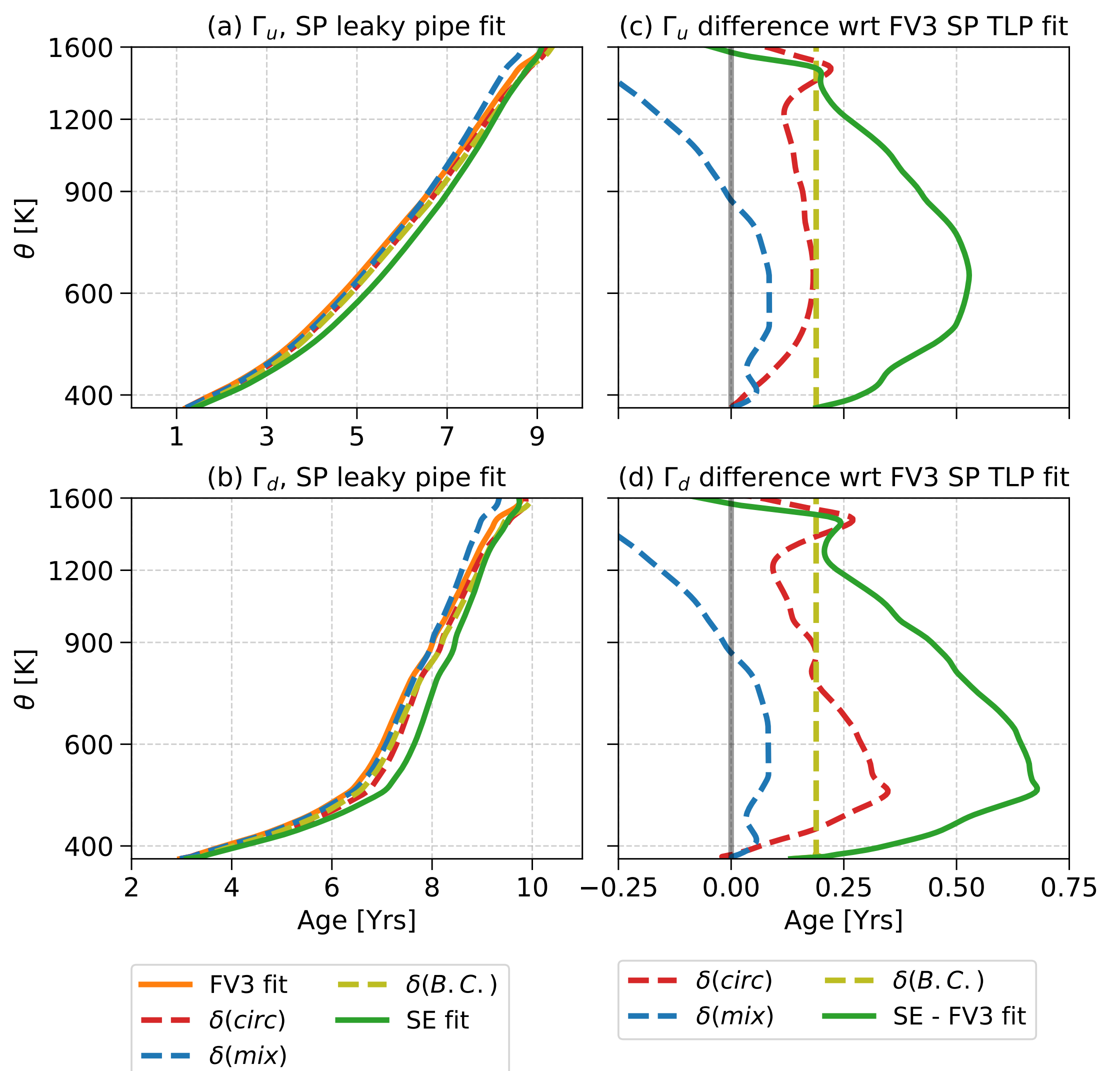


FIG. 8. Same as Figure 8 but for the specified tropics (SP) model runs.

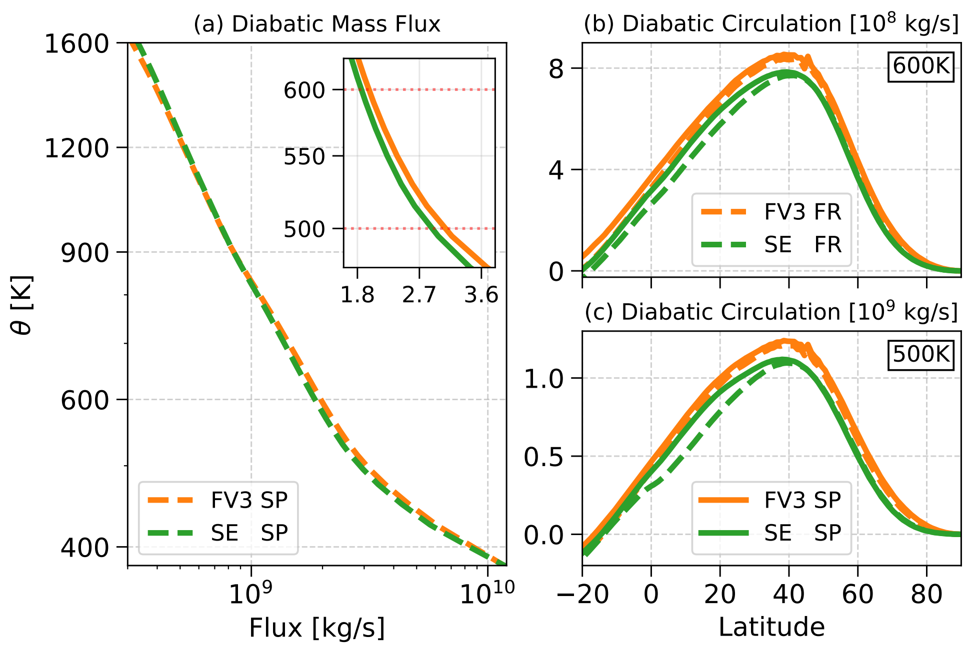


FIG. 9. (a) Horizontally integrated diabatic mass flux in the upwelling region, $\mathcal{M}_u$, for the GFDL-FV3 (orange) and the CAM-SE (green) cores. The inset shows a close-up of the fluxes between 450 K to 650 K. Subplots (b) and (c) compare the full diabatic streamfunction $\psi(\phi,\theta)$ for (b) θ=600 K and (c) θ=500 K for the FR runs (in dashed) and SP runs (in solid).