Chapter 5: Brewer-Dobson Circulation

Chapter lead authors

Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	United Kingdom
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany

Co-authors

Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Mohamadou Diallo	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Florian Haenel	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Paul Konopka	Institute of Energy and Climate Research, Stratosphere,Forschungzentrum Jülich	Germany
Bernard Legras	Laboratoire de Météorologie Dynamique, IPSL, UPMC/ENS/CNRS/Ecole Polytechnique	France
Felix Ploeger	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Thomas Reddmann	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Gabriele Stiller	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
Marta Abalos	Department of Earth Physics and Astrophysics, Universidad Complutense de Madrid	Spain
Harald Boenisch	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration (NOAA)	USA
Hella Garny	(1) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) (2) Meteorological Institute, Ludwig-Maximilians University Munich	Germany
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Kazuyuki Miyazaki	Tropospheric Composition Group, Jet Propulsion Laboratory, California Institute of Technology *previously at: Japan Agency for Marine-Earth Science and Technology, Japan*	USA
Howard K. Roscoe	British Antarctic Survey	UK
Kaoru Sato	Department of Earth and Planetary Science, The University of Tokyo	Japan
Mengchu Tao	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Darryn Waugh	Department of Earth and Planetary Sciences, Johns Hopkins University	USA

Abstract. This chapter focuses on the evaluation and comparison of the stratospheric circulation, using diagnostics based on the residual mean meridional circulation (*e.g.*, tropical upwelling), and on stratospheric transport tracers such as the age-of-air (AoA). Off-line chemistry-transport models in Eulerian and Lagrangian frameworks are used to compute tracer diagnostics for major recent reanalyses. Results are compared to those from observation-based datasets derived from satellite, ground-based, balloon, and aircraft observations of long-lived tracers such as SF_6, CO_2, and N_2O. Particular attention is given to comparing past trends in AoA from the different reanalyses with different offline chemistry-transport models (CTMs) driven by the reanalyses.

Dynamics diagnostics show that in recent reanalysis products the Brewer-Dobson circulation (BDC) is consistent in terms of climatological-mean structures with overall coherent interannual variability in metrics such as tropical upwelling at 70 hPa. However, estimates of long-term trends in tropical upwelling are inconsistent among different products, showing either strengthening, weakening, or no trend. Residual circulation transit times (RCTTs), a measure of the integrated circulation strength throughout the stratosphere, show large variability across different products, although long-term trend structures in RCTTs indicate a strengthening of the BDC, especially within its shallow branch.

Our comparison of AoA tracer results has shown that recent reanalyses produce mean AoA in much better agreement with observations than the previous generation of reanalysis, showing the improvement achieved by the reanalysis systems in the representation of the BDC. However significant discrepancies in AoA and tracers distribution among reanalyses still remain. For the overall period (1989 - 2010) our offline results show large spread in values and sign of mean AoA trends, depending on the reanalysis and on the region of the stratosphere. For the MIPAS period (2002 - 2012) only ERA-Interim is in good agreement with the observed trends, independently of the offline model used. We point to possible causes of these discrepancies and provide recommendations for users and for reanalyses centres. Much investigation is still needed on BDC trends, and factors affecting them, including natural variability and changes in the observation system of assimilated data.

Contents

5.1 Introduction

5.1.1 General description of the BDC and approach

The Brewer-Dobson circulation (BDC) describes the major transport pattern in the stratosphere. The BDC was first postulated by *Brewer* (1949) and *Dobson* (1956) to explain measurements of water vapour and ozone in the stratosphere. The circulation is fundamentally driven by dissipating waves of tropospheric origin and broadly consists of large-scale tropical ascent and winter pole descent. The BDC is much weaker during boreal summer due to the different distribution of land masses and the associated differences in the generation of planetary and gravity waves between both hemispheres.

Modern reanalysis (reanalysis) products[1] include a spatially well-resolved stratosphere, which motivates an assessment of their BDC and its associated trends. Characteristics of the BDC may be obtained based on general circulation metrics, such as the residual mean meridional overturning streamfunction, from variables directly available from the reanalyses, and also from offline model simulations driven by the reanalyses meteorological fields (see *Section 5.2*). Using the residual mean streamfunction, *Iwasaki et al.* (2009) previously compared a general circulation model (GCM) with that from five reanalysis products: JRA-25, ERA-40, ERA-Interim, NCEP/NCAR, and NCEP/DOE. They found consistency in the extratropical winter circulation across the reanalyses, but also large discrepancies in low latitudes, and they found that trends were not reliable. Here we provide new results and discussions on mean streamfunction comparisons including additional modern reanalyses.

For multiannual stratospheric studies, an accurate representation of the BDC is essential for chemistry-transport models (CTMs) to achieve realistic tracer distributions. Offline CTMs take winds and temperatures from general circulation models or from meteorological analyses. The advantage of using analyses is that the CTM simulations are then linked to real meteorology and results are directly comparable to observations. Reanalyses extend this advantage into the past, enabling long-term simulations that provide valuable information on the temporal evolution of the atmospheric composition, helping to understand the present and predict the future. Therefore, CTMs rely on the quality of the reanalyses to obtain accurate tracers distributions. And, in turn, this reliance makes CTMs a powerful tool for evaluating the reanalyses themselves. This use of CTMs was proposed by *Monge-Sanz et al.* (2006) and applied during the preparatory phase of ERA-Interim (*Monge-Sanz et al.*, 2012; 2007; *Dee et al.*, 2011). Here, we have applied and extended such an approach using several complementary CTMs to evaluate modern reanalyses.

Recently, the use of reanalyses to nudge climate models is also becoming an emerging practice to constrain dynamics in climate simulations (*e.g.*, *Orbe et al.*, 2020; *Chrysanthou et al.*, 2019), which increases the need for accurate representation of circulation processes in the reanalyses. In some places, this Chapter includes as a point of comparison results based on chemistry-climate model (CCM) experiments from the Chemistry Climate Model Initiative (CCMI, *Morgenstern et al.*, 2017). Models included are (*cf.*, *Dietmüller et al.*, 2018): CMAM, EMAC, GEOSCCM, MRI-ESM1r1, NIWA-UKCA, SOCOL3, ULAQ-CCM, WACCM. Free-running climate models have the advantage that they provide more physically consistent estimates of metrics. Note that JRA-55AMIP effectively represents a climate model (with prescribed SSTs). ERA-20C and 20CR represent products that can be considered as half-way between free-running models and reanalyses.

5.1.2 Chapter objectives

In this Chapter we evaluate how well existing major reanalyses reproduce the BDC, and we provide an intercomparison among these reanalyses and against existing independent observations.

We have aimed at identifying potential causes for the differences we have found among reanalyses, as well as identifying key elements for a realistic representation of the BDC in the reanalysis systems, with a particular focus on model developments.

The final part of the Chapter provides a set of recommendations for reanalyses users and producers; for users to be aware of potential limitations in the datasets, and for producers to achieve further improvements in future reanalyses.

Beyond the intercomparison of the existing major reanalyses, this Chapter also contributes to increasing our scientific knowledge on stratospheric transport processes and provides an updated overview of studies looking into the BDC pattern using reanalyses. We have devoted a significant part of the Chapter to assess trends and variability in BDC diagnostics using the different reanalyses, aiming at shedding light onto the major research question of the apparent discrepancy between models and observations regarding the evolution of the BDC (*e.g.*, *Waugh*, 2009; *Butchart et al.*, 2006). Our analyses have contributed to identifying processes that affect the representation of the BDC and its evolution, and therefore processes that require further attention in future model development.

[1] To ease discussion of the results, we will distinguish "older" from "more recent" products frequently along the chapter, We generally consider ERA-Interim, MERRA-2, JRA-55, CFSR as "more recent" with ERA5 being the newest product. Full details on production dates for each reanalysis can be found in *Chapter 1* and *Fujiwara et al.* (2017).

The BDC governs the entry and distribution of air masses and constituents from the troposphere into and within the stratosphere. It also plays a major role in the exchange of key constituents, such as ozone, back into the troposphere. Therefore, changes in the BDC will affect the stratospheric concentrations of longer-lived trace gases whose sources are in the troposphere (*e.g.*, CFCs, CH_4, N_2O), as well as of their stratospheric products (*e.g.*, reactive chlorine gases, H_2O, reactive nitrogen gases). BDC changes will also affect the tropospheric concentrations of trace gases with large sources in the stratosphere (*e.g.*, ozone and water vapour). Since these gases have key impacts on atmospheric climate and chemistry, it is essential to understand what changes have occurred to the BDC in the past to be in a better position to predict those that will occur in future.

5.2 Diagnostics description

5.2.1 Dynamical variables

We use standard pressure level output and compute diagnostics consistently across all products (*Martineau et al.* 2018). Notably these pressure level data lack resolution in the shallow branch of the BDC (no level provided between 100 hPa and 70 hPa). Detailed tests were performed initially using ERA-Interim to study the impact of vertical resolution (model versus standard pressure levels), details of the numerical computation methods, and upper boundary conditions. We chose ERA-Interim because model-level diagnostics were available from previous work with slightly different numerical computation methods. These tests included comparing residual velocities computed independently by different groups.

The residual circulation mass streamfunction is defined based on the Transformed Eulerian Mean (TEM) framework as (*e.g.*, *Andrews et al.*, 1987):

$$\psi^* = \psi - ag^{-1}\cos\varphi\frac{\overline{v'\theta'}}{\partial_p\bar{\theta}} \quad (5.1),$$

with the Eulerian mean streamfunction given by:

$$\psi = ag^{-1}\cos\varphi\int_{TOA}^{p}\bar{v}\,dp' \quad (5.2).$$

Here, a is Earth's radius, g is acceleration due to gravity, φ is latitude, p is pressure, v is meridional velocity, and θ is potential temperature. Overbars denote zonal averages and TOA stands for top-of-atmosphere. The upwelling mass flux through a given level (*e.g.*, 70 hPa) is then defined as $\psi^*_{max} - \psi^*_{min}$, *i.e.*, as the difference between the residual streamfunction's maximum and minimum value on that level, which by definition corresponds to the net upward mass flux between the so-called turnaround latitudes. The turnaround latitudes mark those latitudes where residual mean flow is upward equatorward of them and downward poleward of them (Rosenlof, 1995).

Residual (TEM) circulation velocities are calculated directly based on:

$$\bar{v}^* = \bar{v} - \partial_p\left(\frac{\overline{v'\theta'}}{\partial_p\bar{\theta}}\right) \quad (5.3)$$

and

$$\bar{w}^* = \bar{w} + \frac{1}{a\cos\varphi}\partial_\varphi\left(\frac{\overline{v'\theta'}\cos\varphi}{\partial_p\bar{\theta}}\right) \quad (5.4).$$

We use finite centered differences for the numerical derivatives, where the pressure-derivative is computed as $\partial_p = (\partial_p\sigma)\partial_\sigma$, with $\sigma = (p_0/p)^\kappa$, where p_0=1000 hPa, $\kappa = R/c_p \approx 2/7$, and the analytical expression for $\partial_p\sigma = -\kappa\sigma/p$ is used. Issues arise primarily at the upper boundary due to: a) implementation of vertical derivative, and b) missing levels in the output (*e.g.*, ERA-Interim pressure output levels only extend to 1 hPa, but the actual underlying model levels extend to 0.1 hPa). We deal with issue a) by introducing a notional top-of-atmosphere (TOA) layer between the model top (or highest output level) and p_{TOA}=0 hPa. We set the average meridional velocity for this layer to half the velocity at the model/output top, which corresponds to setting the flow to zero at p_{TOA}. To deal with issue b) the following steps for the $\bar{v}^*$ upper boundary condition have been determined to be "optimal" empirically (by comparison to full model levels in the case of ERA-Interim). First, note that $\partial_p\bar{\theta}$ appears inside another vertical derivative, but generally represents a very smooth field. We therefore apply a simple extrapolation beyond the second to last level using a power law:

$$\partial_p\bar{\theta}|_{top} = \partial_p\bar{\theta}|_{top-1}\left(\frac{\sigma_{top}}{\sigma_{top-1}}\right)^\alpha \quad (5.5),$$

where

$$\alpha = \frac{\ln\left(\frac{\partial_p\bar{\theta}|_{top-1}}{\partial_p\bar{\theta}|_{top-2}}\right)}{\ln\left(\frac{\sigma_{top-1}}{\sigma_{top-2}}\right)} \quad (5.6),$$

where the indices "top", "top–1", "top–2" refer to the top output level, the next level below that, and the next further level below that. Furthermore, we assume that the heat flux contribution vanishes at p_{TOA}:

$$\bar{v}^*_{top} = \bar{v}_{top} - \frac{\left(\frac{\overline{v'\theta'}}{\partial_p\bar{\theta}}\right)_{top-1}}{p_{top-1}} \quad (5.7).$$

For $\bar{w}^*$ the horizontal derivatives are also taken as centred differences, using linear extrapolation to obtain values at the poles.

Finally, residual circulation transit times (RCTTs) are obtained as in *Birner and Bönisch*, 2011: residual circulation trajectories are run backward from specified arrival latitudes, pressures, and twice per month. RCTTs provide an integrated measure of the residual velocities (including the effects of transient changes in $\bar{v}^*$ and $\bar{w}^*$ over the transport pathway). They help to diagnostically distinguish different branches of the BDC.

Trajectories are terminated when they intersect the local (time-dependent) tropopause. The RCTT is the transit time along these trajectories (*cf.*, *Rosenlof,* 1995). For the tropopause we use the thermal tropopause level obtained from zonal-monthly-mean temperatures on the provided pressure levels. Furthermore, we set $\bar{w}^*_{top}=0$ and also $\bar{v}^*$ to zero at the poles, to avoid trajectories leaving the domain. Note that boundary conditions at the bottom of the domain do not influence the (stratospheric) RCTT calculations as long as the bottom level is well below the tropopause, which is the case for all data sets used here.

Note that our analyses do not include a temperature-based metric, such as used by *Fu et al.* (2015), even though such a metric has the advantage of being quite well constrained by observations (*e.g.*, satellite data). The fact that temperatures are well constrained by observations results in very close agreement across reanalysis products, and so an intercomparison of these products is in this case less insightful. A related intercomparison for tropical tropopause temperatures is presented in *Chapter 8* of this report.

5.2.2 Transport tracers from offline simulations

5.2.2.1 Introduction to offline modelling tools

Reanalyses are used by a wide range of models to drive offline simulations to study atmospheric composition and transport processes. Since these offline simulations rely on the quality of the meteorological fields used to drive them, offline models can be used as a very valuable tool to assess how realistic meteorological fields from reanalysis are.

In this Chapter we use several offline models with recognised worldwide experience in stratospheric scientific studies and applications. A description of the models we use here (BASCOE, CLaMS, KASIMA, TOMCAT, TRACZILLA) and key references can be found in *Section 5.3*. All of them have taken part in numerous intercomparison projects and international activities (*e.g.*, WMO Ozone Assessments, CCMVal model intercomparisons, StratoClim EU project).

By using several offline models we obtain a certain spread in the performance of the different reanalysis, which helps to overcome the sensitivity that a particular reanalysis may have to a particular offline model configuration. To the extent possible, we will also aim to explain differences in the performance of reanalyses due to differences among the offline models, but this type of research is out of the scope of this S-RIP Report. It is however being pursued as a follow-up project by several co-authors of this Chapter (*Monge-Sanz et al.*, in prep).

5.2.2.2 Diabatic heating rates

The diabatic heating rate field, Q/c_p as described by the equation below, gives information on the model temperature tendencies and is a fundamental component of the temperature budget; this field is used by some offline models in this study to compute vertical velocities.

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \text{grad}\, T - \omega \left(\frac{\kappa T}{p} - \frac{\partial T}{\partial p} \right) = \frac{Q}{c_p} \qquad (5.8)$$

In the atmosphere, diabatic heating includes effects of radiative heating, latent heat fluxes and turbulent heat transport, however reanalyses archive total diabatic temperature tendency, and temperature tendency from radiation. This leaves the latent heat flux and the turbulent heat mix as one same contribution to the tendency from the reanalyses data:

$$\frac{Q}{c_p} = \frac{Q_{\text{rad}}}{c_p} + \left(\frac{Q_{\text{lat}}}{c_p} + \frac{Q_{\text{mix}}}{c_p} \right) \qquad (5.9)$$

Diabatic heating rates (K/day) generated by the reanalysis forecast models are in general provided at 6-hourly time resolution. This field is based on average temperature tendencies over the length of the assimilation window, therefore, *e.g.,* for a 6-hour window, diabatic heating rates data would be centred at 03:00, 09:00, 15:00, and 21:00 rather than at the standard synoptic times 00:00, 06:00, 12:00, and 18:00.

Offline models operating on isentropic vertical coordinates use these temperature tendencies (heating rates) to calculate the cross-isentropic velocity (*Section 5.3*). It is therefore important to understand the differences that this field exhibits in the different reanalysis datasets we have used, as differences in this field will result in differences in transport and mixing, as well as in thermodynamic differences that impact tropical entry and ascent of atmospheric constituents.

5.2.2.3 Mean age-of-air

The mean age-of-air (AoA) is a standard diagnostic for stratospheric circulation widely used by models in the stratosphere. It gives information on the time spent by air parcels in the stratosphere after entering through the tropical tropopause from lower levels (*e.g.*, *Waugh and Hall*, 2002).

The main advantage of this diagnostic is that it can be computed from measurements of certain atmospheric tracers, *e.g.*, CO_2 and SF_6 tracers. These two long-lived constituents approximately fulfil the linearly conserved conditions, they have linearly increasing concentrations in the troposphere and no stratospheric sources or sinks, and can therefore be used to derive the stratospheric mean AoA. By measuring their concentrations in the stratosphere we can trace back how long air parcels have been residing in the stratosphere.

Doing this at different stratospheric locations provides a picture of the strength of the circulation in this atmospheric region. The annual mean cross-section of the mean AoA obtained in this way should look similar to the one inferred by *Waugh and Hall* (2002) (**Figure 1**). These two gases, CO_2 and SF_6, are complementary as the tropospheric annual cycle of CO_2 can affect age values in the lower stratosphere, while SF6 concentrations in the upper stratosphere are affected by mesospheric loss (*Reddman et al.*, 2001). The effect of the mesospheric SF_6 sink contributes to make mean AoA values older at higher stratospheric altitudes; we will discuss this effect in *Section 5.5.2.3.*

Full-chemistry models can compute the simulated AoA from the corresponding model CO_2 and SF_6 tracers concentrations, or they can use an idealised linear tracer. By using an idealised tracer, results are chemistry independent, then results from different models can be compared only in terms of transport differences, not chemistry differences among models. This use of an idealised AoA tracer has been employed in large stratospheric model intercomparisons like CCMI, CCMVal and CCMVal-2 (*Eyring et al.*, 2006; *Dietmüller et al.*, 2018), where CTMs and CCMs were compared in terms of their performance in the stratosphere.

To compare different reanalyses we can run several simulations with an offline CTM driven by the different datasets, keeping the CTM configuration unchanged so that the transport differences will be due to the different meteorological datasets used to drive the offline model. This approach was applied for instance to advise ECMWF during the preparatory phase of the ERA-Interim production (*Monge-Sanz et al.*, 2007). We are using a similar approach in this part of the Chapter to evaluate the S-RIP reanalyses datasets. The CTMs used for the offline simulations are described in *Section 5.3.*

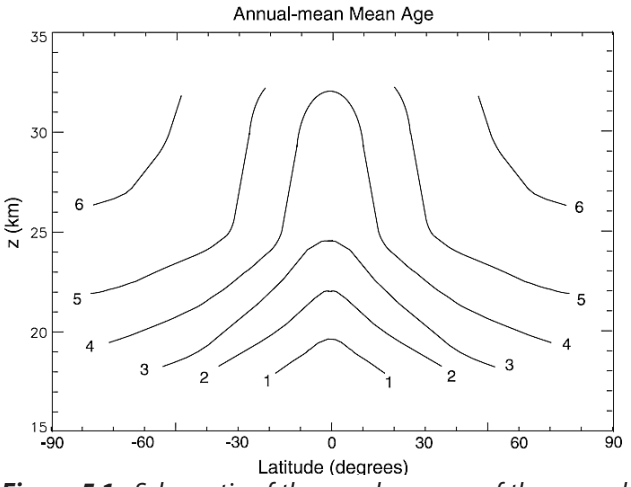

Figure 5.1: *Schematic of the zonal average of the annual mean of the mean AoA distribution (years), as inferred from observations as described in Waugh and Hall (2002). Figure from Waugh and Hall (2002). ©American Geophysical Union. Used with permission.*

Mean AoA trends

A significant part of this Chapter deals with the open and active scientific question that concerns potential trends in the mean AoA. It was initially motivated by the apparent disagreement between most climate models and existing long records of mean AoA observation based datasets (*e.g.*, *Engel et al.*, 2009; *Waugh et al.*, 2009) that was a matter of active debate when the S-RIP project started in 2012. This disagreement was also suggested by global observation datasets of AoA and by CTM simulations driven by ERA-Interim (*Stiller et al.*, 2012; *Monge-Sanz et al.*, 2012; *Diallo et al.*, 2012).

Whether observation datasets are showing long-term changes in the BDC or natural variability, and why most models are not capturing the same trend behaviour are among the scientific questions we address, to the extent possible, with the reanalyses evaluations and intercomparisons performed for this Chapter. Therefore, here we also evaluate the different reanalysis in terms of their ability to reproduce observed time evolution and trends in the mean AoA for the reanalysis period.

5.2.2.4 Age spectrum

The age spectrum is the statistical distribution of transit times for an air parcel from a source location, typically the Earth surface or the tropical tropopause, to a given location in the stratosphere (*Kida*, 1983). This concept was mathematically developed by *Hall and Plumb* (1994), who defined the age spectrum as a Green's function $G(x; x_0; t)$ that, for a tracer mixing ratio x, propagates in time a boundary condition from a source region x_0 (typically the tropical tropopause) into the stratosphere.

The mean age $\Gamma(x; x_0)$ at a certain stratospheric location is then the average over the age spectrum at that location:

$$\Gamma(x, x_0) = \int_0^\infty tG(x, x_0, t)dt \qquad (5.10).$$

The first moment of the spectrum is the mean age-of-air, as described in *Section 5.2.2.3.* Although the age-spectrum gives a more complete view of the stratospheric circulation than the mean age-of-air, it cannot be actually measured. It is the mean age value obtained from the spectra that we can compare against observation based AoA values. Nevertheless, an intercomparison of age-spectra derived with different reanalyses can yield valuable information on the different representation of the stratospheric circulation in each dataset.

5.2.2.5 Stratospheric Water Vapour tracer

A complete overview of stratospheric water vapour (SWV) in the different reanalyses is provided in *Chapter 4* of this Report and in *Davis et al.* (2017).

In the current Chapter we focus on the SWV distributions obtained from offline CTMs driven by the different reanalyses. This gives additional information, for each reanalysis dataset, on their ability to transport real constituents into and within the stratosphere, as well as additional information on the usefulness of each dataset for offline model applications. It also needs to be taken into account that SWV depends on several variables, including tropopause temperature in the different reanalysis, and it will not be only a diagnostic for stratospheric transport.

Entry values of water vapour through the tropical tropopause exhibit a seasonally varying signal, imposed by the seasonality in the tropopause temperatures. This makes water vapour concentration values in the tropical lower stratosphere appear as if they had been marked by a tape-recorder (*Mote et al.,* 1996; 1998). Over the tropical LS region, this so-called "tape recorder" diagnostic (timeseries, amplitude and phase of water vapour concentrations), provides information on the propagation of air masses into the stratosphere. This diagnostic is also one of the standard tests applied to stratospheric models to evaluate the representation of the subtropical mixing barrier. The vertical propagation of the tape recorder signal allows the estimation of the vertical ascent over the tropics. When deriving the tape-recorder with an idealised sinusoidal tracer, we can avoid its dependency on exact tropopause temperatures and the corresponding tape-recorder signal reflects only transport aspects. The phase lag of the tape-recorder signal is a good way to quantify the vertical velocity, while the amplitude decay mainly characterizes the strength of the tropical mixing barrier. Therefore the tape-recorder can be used as an additional way to assess the BDC over tropical latitudes. The SWV tape-recorder is also one of the stratospheric transport diagnostics that we can test against satellite observations such as from the HALOE and MLS instruments, or the merged SWOOSH dataset.

5.3 Offline models description

5.3.1 Description of the BASCOE model

BASCOE is a kinematic transport model (*Skachko et al.*, 2014). Its advection module is the Flux-Form Semi-Lagrangian (FFSL) scheme (*Lin and Rood*, 1996) configured to follow closely the recommendations of *Rotman et al.* (2001). The FFSL advection scheme is run on the native vertical grid of each reanalysis and a common low-resolution latitude-longitude grid with 2° x 2.5° increments. It requires to input the surface pressure and horizontal velocity on a so-called Arakawa C-grid, *i.e.,* the zonal wind u must be staggered in longitude and the meridional wind v must be staggered in latitude. The FFSL algorithm evaluates internally the corresponding mass fluxes and derives the vertical winds (w) from mass conservation. Hence the reanalysis datasets must be pre-processed from spectral or high-resolution gridded fields to the low-resolution C-grid.

Special attention was paid to the preprocessing of the reanalyses to make sure that the different types of wind fields were expressed in a consistent manner for the BASCOE transport algorithm. For the five reanalysis datasets used by BASCOE (ERA-Interim, JRA-55, MERRA, MERRA-2 and CFSR) a preprocessing algorithm based on *Segers et al.* (2002) is used, with additional preliminary derivation of the spherical harmonics coefficients of vorticity, divergence and surface pressure for reanalyses other than ERA-Interim. In all cases, these spectral coefficients are truncated at wavelength 47 to avoid aliasing on the 2° x 2.5° target BASCOE grid.

In the BASCOE simulations the AoA is derived from an idealized tracer with a concentration that increases linearly with time at the surface. To allow quick propagation of this boundary condition to the free troposphere, eddy vertical diffusion is modeled in the lower half of the troposphere with a vertical diffusion coefficient K_{zz} decreasing from $10\,m^2\,s^{-1}$ at the surface to zero at the pressure level halfway between the surface and the tropopause. There is no other representation of convection in the BASCOE model nor any explicit mechanism for horizontal diffusion.

5.3.2 Description of the CLaMS model

The Chemical Lagrangian Model of the Stratosphere (CLaMS) is a Lagrangian transport model with trace gas transport based on the motion of 3-D forward trajectories and an additional parameterization of subgrid scale atmospheric mixing, which relates mixing to deformations in the large-scale flow (*Konopka et al.*, 2004; *McKenna et al.*, 2002). The model uses an isentropic vertical coordinate, with vertical transport driven by the total diabatic heating rate (*Ploeger et al.*, 2010). Meteorological fields from the reanalyses are read in 3-hourly timesteps (horizontal winds and diabatic heating rates).

For this study, CLaMS simulations use fields from ERA-Interim (*Dee et al.*, 2011), JRA-55 (*Kobayashi et al.*, 2015) and MERRA-2 (*Gelaro et al.*, 2017) reanalyses. For driving the CLaMS model simulations, horizontal winds and diabatic heating rates from the reanalysis forecast are used on native model levels and with a horizontal resolution of 1° x 1° in latitude and longitude. The AoA results from the different simulations have been interpolated to potential temperature levels (same for all reanalyses) and monthly zonal mean climatologies have been created.

CLaMS provides evaluation of the representation of SWV, mean age of air (AoA) and age spectrum. The evaluation is based on the modelled quantities as monthly and zonal means from 1979 to 2015. In the stratosphere and the UTLS, potential temperature is employed as the vertical coordinate of CLaMS, and the cross-isentropic velocities are derived from the total diabatic heating rates provided by the reanalysis products, including effects of radiative heating, turbulent heating and heating release.

The model configuration follows the model setup described in *Pommrich et al.* (2014) with 100 km horizontal and 250 m vertical resolutions around the tropical tropopause. The age spectrum diagnostic computation used by CLaMS is described in *Ploeger and Birner* (2016); accordingly AoA spectrum is calculated for each reanalysis from multiple tracer pulses and the mean age value is obtained from the spectrum (*Ploeger et al.*, 2019). It is worth noting that in the CLaMS simulations shown here, an upper boundary condition is imposed for the mean AoA values, by prescribing top of the model values with MIPAS derived AoA.

5.3.3 Description of the KASIMA model

The Karlsruhe Simulation of the Middle Atmosphere (KASIMA) model is a three-dimensional mechanistic model of the middle atmosphere solving the primitive equations including middle atmosphere chemistry (*Kouker et al.*, 1999). For the simulations used here, the model was run on isobaric surfaces from 7 km to 120 km with a vertical resolution of 750 m in the stratosphere, gradually increasing to 3.8 km at the upper boundary. The horizontal resolution in the simulation is 5.4 ° × 5.4° (T21). The model is coupled to the specific meteorology by using the analyzed geopotential field at the lower boundary (7 km) and applying analyzed vorticity, divergence and temperature fields from ECMWF ERA-Interim below 1 hPa. Above 1 hPa radiative heating rates were calculated using a 2D climatology for ozone and H_2O. The parameterization of the gravity-wave drag is based on the formulation of *Holton* (1982). The parameterization has been modified compared to the version described in *Kouker et al.* (1999) in order to better describe the cross-mesopause transport often observed after sudden stratospheric warmings (SSWs). The spectral distribution of the vertical momentum flux is now described with a Gaussian function of a centroid of 7 m s^{-1} and a standard deviation of 50 m s^{-1} with phase speeds of 0, 20, 40, 60 and 80 m s^{-1}. The filter condition for critical phase speeds has been extended to be applied when the absolute difference between the speeds is less than 10 m s^{-1}. The latter condition effectively prevents gravity waves of low phase speed from propagating and breaking in the lower mesosphere. Only gravity waves of higher phase speed then break at higher altitudes, causing an elevated stratopause to build. In addition, the numerical implementation of the vertical diffusion has been re-formulated for better mass conservation according to *Schlutow et al.*, 2014.

KASIMA has used the following artificial tracers to derive the mean age-of-air: T1 is an idealized tracer exhibiting a linear trend. For T1 the mean age Γ is just the lag time $\Lambda 1$. T2 is a tracer initialized with a time series of mixing ratio data of SF_6 complemented by the data taken from the NOAA HATS (Halocarbons and other Atmospheric Trace Species) data set data set (**https://www.esrl.noaa.gov/gmd/hats/**). No chemical loss is assumed for that tracer and a lag time $\Lambda 2$ is deduced. With this tracer we study non-linearity effects in the trend curve. The tracer T3 is defined as T2, but includes a chemical loss as described by *Reddmann et al.* (2001), using the version including all relevant reactions. Chemical loss of a mean age tracer with a positive trend results in an apparent higher age as the tracer shows a lower mixing ratio than expected. SF_6 exhibits a significant mesospheric loss by electron attachment and subsequent reactions as described by *Reddmann et al.* (2001). Tracer T3 includes these loss mechanisms and the (apparent) lag time is calculated as for the inert tracer case. Whereas tracer T3 should be the most realistic tracer to be compared with SF_6 observations there are caveats as the loss mechanism of SF_6 is subject to significant uncertainties. With the tracer T3 we test the influence of the mesospheric loss on the derivation of stratospheric mean age.

As SF_6 shows a pronounced inter-hemispheric difference in the mixing ratio in the troposphere, the inter-hemispheric difference was imprinted to the mixing ratio of SF_6 at the lower boundary inside the troposphere in the form $A \cdot \tanh(1.5 \sin(\phi))$ with ϕ the geographic latitude and $A = 0.55$ years for the ideal linear tracer, and a mixing ratio difference for tracers T1 and T2 corresponding to an amplitude of the hemispheric difference of approximately 1 year. The tracers were formally initialized for 1965, and the first two years of the ERA-Interim reanalyses were used repeatedly till 1979 to bring the tracers to an approximately steady state. Two years were used for spin-up to include an approximate full QBO period.

5.3.4 Description of the TOMCAT/SLIMCAT model

TOMCAT/SLIMCAT is a 3D offline CTM (*Chipperfield*, 2006). The CTM is flexible in terms of the winds datasets it can use, however, the ECMWF datasets have been the only ones that this model has extensively used since 1999, when they were extended into the stratosphere up to 0.1 hPa, and even more so after the completion of ERA-40 for multiannual runs for long-term chemical investigations (*e.g., Feng et al.*, 2007; *Chipperfield et al.*, 2005). The reanalysis fields are read in typically every 6 h, but this can be adapted to other available frequencies (*e.g., Monge-Sanz et al.*, 2012). In TOMCAT/SLIMCAT, read-in fields are interpolated in time to intermediate time steps (of 60 minutes in the case of the runs considered here).

The horizontal grid of the CTM is completely variable in resolution and in latitudinal regularity. The ECMWF (re) analyses are read in as spectral coefficients, which are then converted to grid-point fields by a spectral transform on to the CTM prescribed latitudinal grid using pre-tabulated integrals of the associated Legendre functions (*Chipperfield*, 2006). In this way the CTM is not restricted only to the usual Gaussian latitudes. Also, the number of vertical levels is flexible and the vertical coordinate can be either $\sigma - p$ (TOMCAT mode) or $\sigma - \theta$ (SLIMCAT mode). Vertical motion is calculated from the divergence of the horizontal winds.

In the case of ECMWF datasets, the divergence field is directly taken from the reanalyses or operational analyses. The conservation of second-order moments non-diffusive advection scheme by *Prather* (1986) is used in the CTM runs.

TOMCAT/SLIMCAT also includes a module for the calculation of particle trajectories, which allows for a Lagrangian as well as the default Eulerian approach. The trajectory position is computed from the same meteorological data used to force the Eulerian simulations; horizontal and vertical motion are calculated at the centre of the Eulerian grid and then interpolated to the trajectory position in that particular grid cell. An explicit fourth-order Runge–Kutta method (*Fisher et al.*, 1993) is used to advance the trajectory position forward (or backward) in time. The same general configuration options (vertical coordinate, vertical motion) are also available for the Lagrangian runs.

5.3.5 Description of the TRACZILLA model

TRACZILLA is a Lagrangian transport model derived from FLEXPART (*Pisso and Legras*, 2008). The simulations used here are performed by launching parcels from a 3-D grid on 32 isentropic levels in the stratosphere (from 300 K to 1420 K) on a 2 ° by 3 ° latitude-longitude grid, every ten days over the 22-year period 1989 - 2010. The trajectories are integrated until they cross the tropopause, determining the age of the parcel at that time, or until they reach a maximum age of ten years. In the simulations shown here, the motion of air parcels is governed by horizontal velocity fields and radiative heating rates from three reanalyses: ERA-Interim, MERRA and JRA-55.

For ERA-Interim, TRACZILLA uses meteorological data on model levels up to 0.1 hPa at 1 ° resolution in latitude and longitude and 3-hourly temporal resolution. For JRA-55, the model uses data on model levels up to 0.1 hPa at 0.56 ° resolution in latitude and longitude and 6-hourly temporal resolution. In MERRA, we use data on pressure levels up to 0.1 hPa (heating rates not available above) at 0.62 ° horizontal resolution and 3-hourly temporal resolution. 200 million trajectories have been used in this TRACZILLA S-RIP study.

TRACZILLA calculates the mean age-of-air by averaging over all parcels that cross the tropopause; the contribution of parcels that have not crossed it is calculated based on the well-established approximation of an exponential tail based on *Scheele et al.* (2005). TRACZILLA applies correction techniques to the trajectories calculation: first a uniform horizontal heating is applied on pressure levels to correct the lack of mass conservation when using radiative heatings in the stratosphere. Second, the trajectories which go above 0.5 hPa (*i.e.,* 2300 K) are discarded (clipped). This is a common correction technique in Lagrangian studies, *e.g.*, in *Schoeberl and Dessler* (2011) trajectories were clipped above 1800 K. The clipping level at 0.5 hPa was chosen in order for ERA-Interim to provide the best fit of the reconstructed AoA values to the aircraft and balloon observations derived from CO_2, N_2O and CH_4 during the SOLVE campaign (*Andrews et al.*, 2001).

5.4 Description of tracers observations

This Section provides a brief overview of the independent observation-based datasets we have used to validate the AoA and tracer distributions from the offline model simulations. We also include key references for more detailed descriptions of these measurement datasets.

5.4.1 "Standard" mean AoA observations for model intercomparisons

Mean AoA can be calculated from measurements of concentrations of long-lived tracers with an approximately linear increasing trend at the surface, such as CO_2 or SF_6. Between 1992 and 1998 NASA ER-2 aircraft and high-altitude balloons measured concentrations of CO_2 and SF_6. The ER-2 measurements were part of the campaigns Stratospheric Photochemistry Aerosol and Dynamics Experiment (SPADE), Airborne Southern Hemisphere Ozone Experiment/Measurements for Assessing the Effects of Stratospheric Aircraft (ASHOE/MAESA), Stratospheric Tracers of Atmospheric Transport (STRAT) and Photochemistry of Ozone Loss in the Arctic Regions in Summer (POLARIS). Balloon flights were part of the Observations of the Middle Stratosphere (OMS) experiments.

Multidecadal datasets were compiled from these balloon soundings and aircraft flights using both CO_2 and SF_6 measurements (*e.g.*, *Andrews et al.*, 2001; *Ray et al.*, 2014; *Ray et al.*, 1999) that have been widely used in model intercomparison studies (*e.g., Dietmüller et al.*, 2018; *Eyring et al.*, 2006), and have become a standard reference to monitor model development and circulation processes in the stratosphere (*e.g., Ploeger et al.*, 2019; *Chabrillat et al.*, 2018; *Butchart*, 2014; *Monge-Sanz et al.*, 2012).

5.4.2 Long timeseries of mean AoA in the Northern Hemisphere

For the NH mid-latitudes, a long time series of mean AOA derived from balloon-borne measurements of CO_2 and SF_6 exists that dates back to the mid 1970's *(Engel et al.*, 2009; 2017). The balloon-borne observations used in *Engel et al.* (2009) were taken in the region between 24 km and 35 km over NH midlatitudes, where the vertical gradient in mean AoA was found to be very small. The balloon data were limited to a total of 28 flights over a 30 year period, from 1975 to 2006, and showed a positive trend of 0.24 years per decade for this region.

Although this trend was estimated to be within the observational uncertainty, it pointed to an important potential disagreement between observations and most models (*e.g.*, *Waugh*, 2009). The dataset used in *Engel et al.* (2009) has been more recently extended using the new AirCore in-situ measurements (*Engel et al.*, 2017), which have also helped to narrow the trend uncertainty from the previous dataset. These NH data have become widely used by offline model studies concerned with the active debate of BDC trends in reanalyses (*e.g.*, *Ploeger et al.*, 2019; *Chabrillat et al.*, 2018; *Mahieu et al.*, 2014; *Monge-Sanz et al.*, 2012; *Diallo et al.*, 2012).

Balloon measurements described described here reached maximum altitudes of 31 km, while the aircraft missions reached up to 21 km, which limits the altitude range covered by these datasets to the LS and middle stratosphere. Also the latitude range is limited as the OMS flights covered only three latitude values (65 °N, 35 °N and 7 °S) and the measurements used in *Engel et al.* (2009; 2017) are limited to the NH midlatitudes LS region. We therefore need additional observations that provide mean AoA values derived from global coverage measurements, based on the MIPAS satellite observations.

5.4.3 MIPAS AoA dataset based on tracer observations

Global coverage time series have been derived from satellite observations of SF_6 retrievals from the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS; *Fischer et al.,* 2008) satellite instrument, which provided an updated global dataset for the period 2002 - 2012. MIPAS was an instrument on board the Envisat satellite, measuring the mid infrared emission of the atmosphere against the space background. The measurements were done in limb scanning mode covering an altitude range of cloud top (or about 6 km in cloud-free cases) to about 72 km. The emission signatures of molecules in the atmosphere were used to retrieve the spatial distribution of up to 30 different trace gases and temperature with good global coverage from pole to pole, also during (polar) night. The mission extended from July 2002 to April 2012.

Information on the stratospheric mean AoA is obtained from the spatio-temporal distribution of the SF_6 tracer, measured by MIPAS with a vertical resolution of 4 km to 6 km and a single profile precision of about 10 - 20 %. Although the single profile precision is rather low, the huge number of profiles measured (more than 2 million profiles over the MIPAS mission lifetime) provided very valuable information on AoA from zonal mean distributions. The SF_6 distributions were retrieved from the upper troposphere up to about 50 km. Above 35 km, the systematic errors become larger and the vertical resolution deteriorates; for this reason quantitative analysis of SF_6 and AoA above 35 km is not recommended. Very detailed descriptions of this global dataset can be found in *Stiller et al.* (2012) and *Haenel et al.* (2015).

5.4.4 BAS Polar tracer observations

Stratospheric measurements of polar summer NO_2 were used by *Cook and Roscoe* (2009; 2012) to derive trends in the BDC. NO_2 data were measured from a zenith-sky spectrometer set up at Faraday in the Antarctic (65.25 °S, 64.27 °W) between 1990 and 1995, and then from the nearby site of Rothera (67.57 °S, 68.13 °W) since 1996, providing almost continuous measurements of Antarctic NO_2 since 1990 (*Roscoe*, 2004; *Roscoe et al.*, 2001).

Stratospheric column of NO_y over these Antarctic stations were obtained from measurements of NO_2 taken during the period 1990 - 2007. A photochemical box model and observed ozone and temperature profiles were used to determine column values. The years 1991 and 1992 were excluded because of the large amounts of volcanic aerosols from the Pinatubo eruption still present in the stratosphere. Full details and discussions related to this dataset of measurements by the British Antarctic Survey (BAS) can be found in *Cook and Roscoe* (2009; 2012).

5.4.5 Stratospheric water vapour tape-recorder observations

The seasonally varying signal of the water vapour in the tropical stratosphere, the so-called "tape recorder" signal (*Mote et al.,* 1996; 1998), reflects how rapidly air masses are transported upwards from the tropical tropopause into the stratosphere. The tape recorder is thus a good measure for the strength of the BDC over the tropics.

Observationally based values of the time series of $2CH_4 + H_2O$ measured by HALOE from 1992 - 1997 were analysed by *Mote et al.* (1998) with an empirical orthogonal function method. The amplitude and phase of the tape recorder signal were derived from this method, together with estimations from in-situ CO_2 observations (*Boering et al.*, 1996). These data have been extensively used for model validations and intercomparisons.

More recently, the Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database provided by the NOAA Chemical Sciences Laboratory (CSL) extends the coverage period merging vertically resolved water vapor data from the SAGE-II/III, UARS HALOE, UARS MLS, and Aura MLS satellite instruments starting from 1984 to present (*Davis et al.*, 2016). The homogenization process described by *Davis et al.* (2016) is designed to minimize artificial jumps in time and account for inter-satellite biases. The merged SWOOSH data thus provide a long-term SWV time series with reliable representations of interannual to decadal variability. We use the SWOOSH zonal-mean monthly mean time series of merged water vapor mixing ratios to assess offline model simulations of SWV tracer distributions and variability.

5.5 Comparison results

5.5.1 Results from dynamical variables

5.5.1.1 Climatological description

Figure 5.2 summarizes the climatological structure of the BDC for the multi-reanalysis-mean (REM) during December-January-February (DJF) and June-July-August (JJA) in terms of $\overline{w}^*$ (converted to mm s^{-1}) and ψ^*. Here and in the following we define climatologies based on the period 1980 2010 and include MERRA-2, ERA-Interim, JRA-55, CFSR in the REM. Turnaround latitudes are shown based on the extrema of ψ^*. Note that because ψ^* is calculated from $\overline{v}^*$ it is not everywhere consistent with $\overline{w}^*$ (see *Eqs. 5.3, 5.4*), so that the turnaround latitudes do not everywhere match $\overline{w}^*=0$. Overall, these fields are consistent with previous studies showing the climatological BDC structure from individual reanalyses (*e.g.*, *Miyazaki et al.*, 2016; *Abalos et al.*, 2015; *Iwasaki et al.*, 2009).

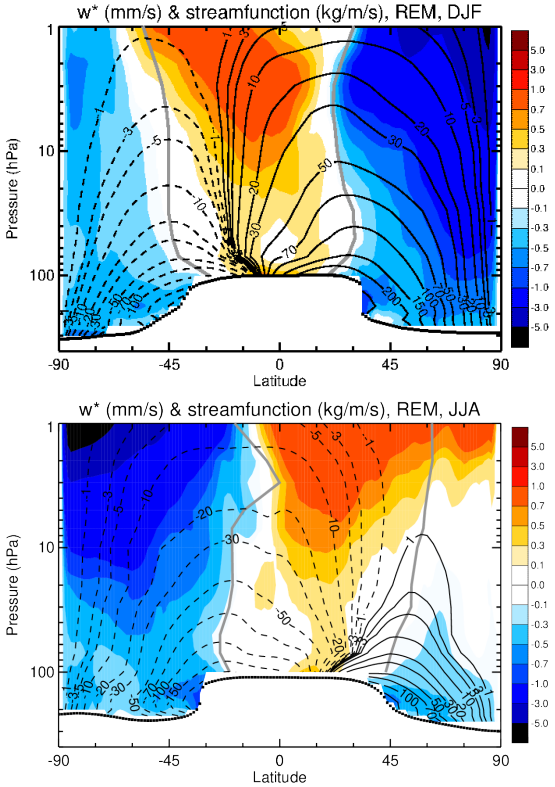

Figure 5.2: *Climatological (1980-2010) REM structures of $-p\overline{w}^*/H$ (with scale height $H=7$ km, color shading) and ψ^* (divided by Earth's radius to ease comparison to previous literature, contours) for December-January-February (top) and June-July-August (bottom), respectively. Full gray lines mark the turnaround latitudes based on the ψ^* fields. Thick gray dots mark the tropopause location based on climatological temperatures. Fields are only shown above the tropopause.*

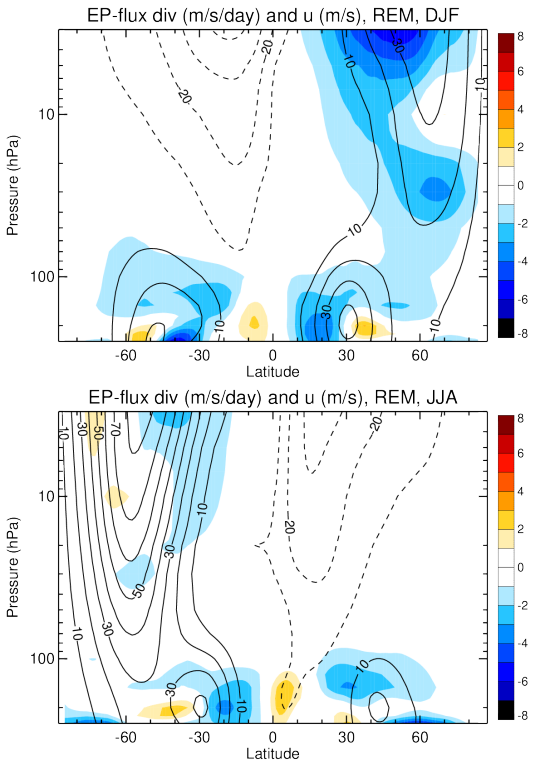

Figure 5.3: *Climatological (1980-2010) REM structures of EP-flux divergence (color shading) and zonal mean zonal wind (contours) for December-January-February (top) and June-July-August (bottom), respectively. Note that the lowest shown level is 250 hPa, which is the tropopause in the extratropics.*

The BDC is driven by wave forcing, which, for the resolved waves, can be quantified based on the Eliassen-Palm flux (EP-flux) divergence shown in **Figure 5.3** for the REM for DJF and JJA together with the zonal mean zonal winds. The structure of EP-flux divergence roughly indicates separate wave forcing for the shallow versus deep circulation branch (*cf.*, *Plumb* 2002; *Konopka et al.*, 2015): the lowermost stratospheric forcing of the shallow branch is present in both seasons and hemispheres, whereas the mid to upper stratospheric forcing of the deep branch is only present during each hemisphere's winter season.

We quantify the wave forcing of the shallow versus deep branch and their seasonality by creating respective hemispheric and vertical averages: 70-100 hPa for the shallow branch and 3-50 hPa for the deep branch (note that 3 hPa is the highest diagnosed level). Note, that these refer to the total resolved wave forcing[2]; the individual contributions due to Rossby and gravity waves are studied in detail in *Sato and Hirano* (2019). Their seasonal climatological evolutions are shown in **Figure 5.4** for all diagnosed reanalyses. Overall, wave forcing is quite consistent between different reanalysis products. The largest spread is found for the shallow branch forcing in the NH throughout the year, as well as for the deep branch forcing in the NH winter and for the shallow branch forcing in the Southern Hemisphere (SH) spring and summer.

[2] These may include a gravity wave contribution insofar as these waves are resolved.

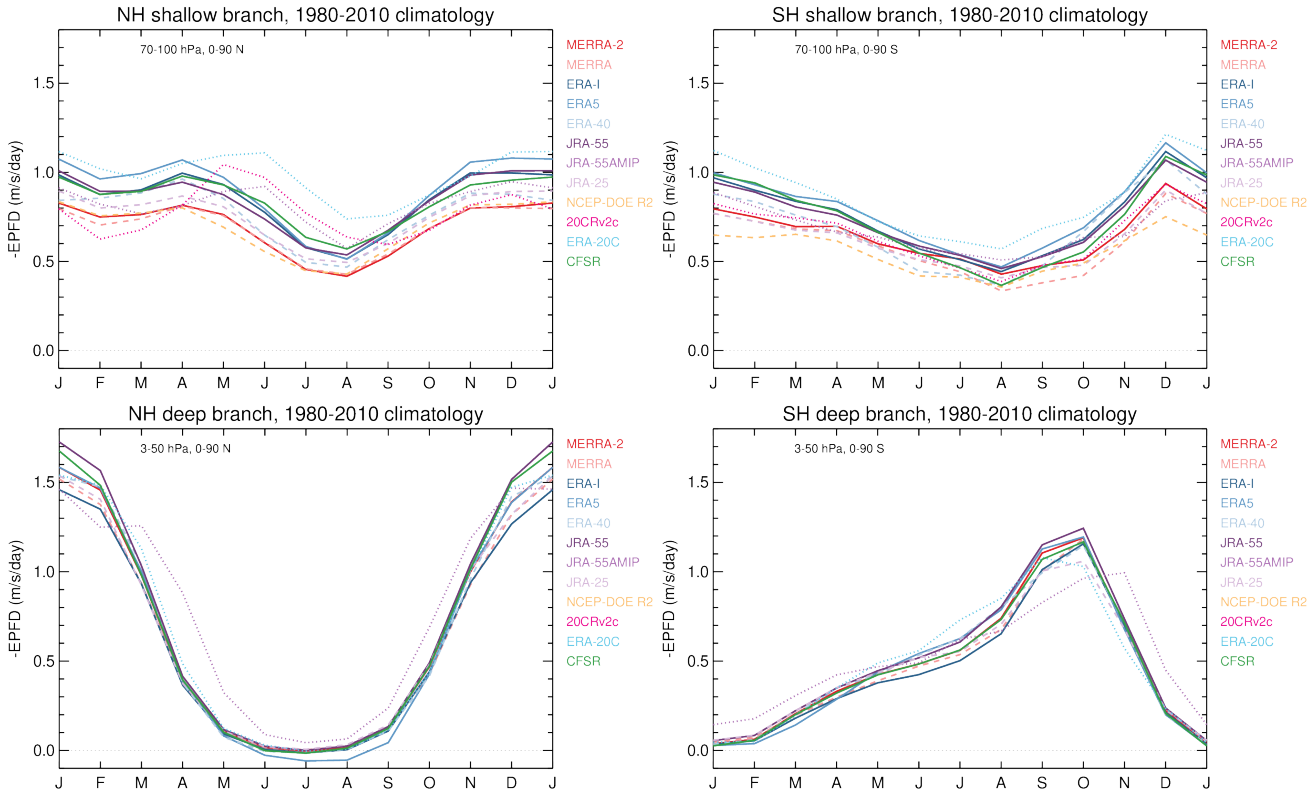

Figure 5.4: *Climatological (1980 - 2010) seasonal evolutions for each reanalysis of the EP-flux divergence for the shallow and deep BDC branches and for each hemisphere separately. Full lines refer to more recent reanalysis products, dashed lines to older reanalysis products, and dotted lines to other products.*

ERA-Interim, JRA-55, and CFSR agree very well except for the deep branch during NH winter, whereas MERRA-2 shows overall less wave forcing. Interestingly, ERA5 shows consistently stronger wave driving of the shallow branch throughout the year in both hemispheres, strongest in NH winter. This could perhaps be due to contributions from partially resolved gravity waves in this much higher resolution product. JRA-55AMIP (the freely evolving atmosphere model version of JRA-55) shows a persistent bias in its seasonality, with a delayed drop-off in wave forcing during spring for the NH deep branch and a delayed peak in wave forcing during spring for the SH deep branch.

The mass overturning at 70 hPa shows a considerable degree of uncertainty between the different products: although the qualitative structure with extratropical downwelling and tropical upwelling is consistent, structural aspects of the upwelling vary strongly in some cases (**Figure 5.5**). For example, the local minimum in tropical upwelling near the equator (*Ming et al.*, 2016) is very pronounced for MERRA-2 and CFSR (minimum roughly zero), but is only weakly present in JRA-55. Of the older products JRA-25 does not exhibit a local minimum, ERA-40 shows a noisy and too narrow upwelling structure, and MERRA shows local downwelling over the equator. The peak in SH tropical upwelling strength during JJA is about one order of magnitude smaller in JRA-55 compared to CFSR (not shown).

Tropical upwelling at 70 hPa is known to exhibit strong seasonality (*e.g.*, *Rosenlof* 1995). However, this arises primarily due to its contribution in the SH where upwelling is much stronger during the NH cold season (**Figure 5.6**). Here, SH upwelling is simply calculated via the streamfunction difference between the equator and its SH minimum. Likewise, NH upwelling is based on the streamfunction difference between its NH maximum and the equator. NH tropical upwelling shows a weak seasonal cycle with stronger upwelling during the SH cold season in some products (most pronounced in JRA-55 and MERRA-2, similar to the climate models), but seasonality is generally inconsistent between products.

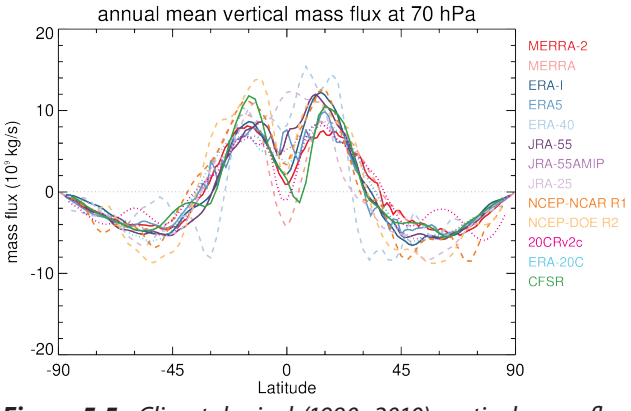

Figure 5.5: *Climatological (1980 - 2010) vertical mass flux at 70 hPa as a function of latitude for the annual mean. Line styles as in* ***Figure 5.4****.*

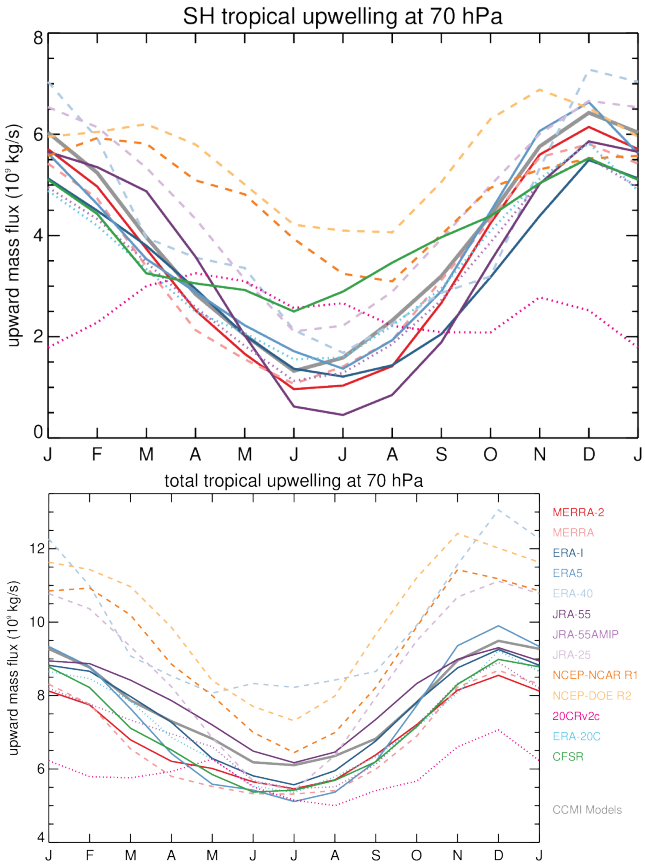

Figure 5.6: *Climatological (1980-2010) tropical upwelling characteristics for each individual product at 70 hPa. Top: individual hemispheric contributions. Bottom: total tropical upwelling between turnaround latitudes (i.e., sum of hemispheric contributions shown in the top panels). Gray lines show the multi-model-mean (MMM) of the CCMI models.*

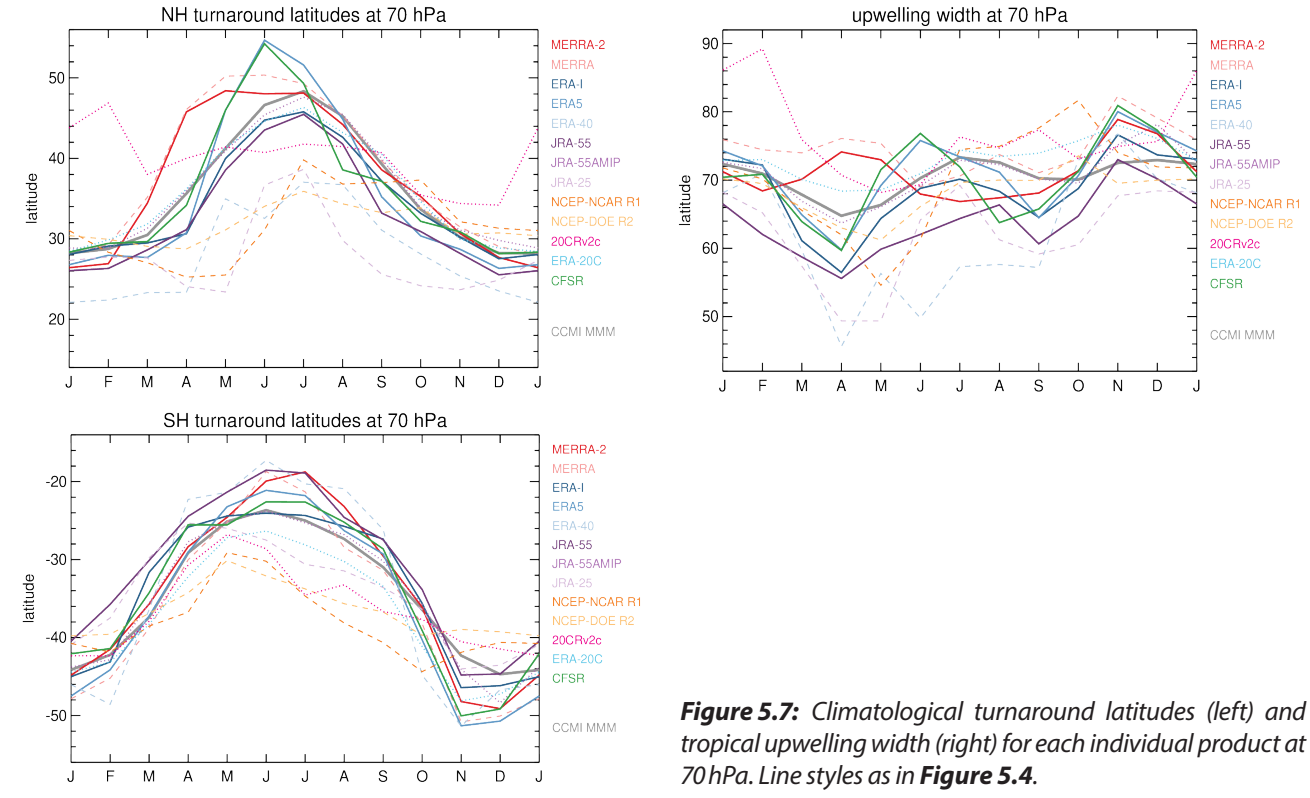

Figure 5.7: *Climatological turnaround latitudes (left) and tropical upwelling width (right) for each individual product at 70 hPa. Line styles as in* ***Figure 5.4****.*

When combined, total tropical upwelling has a consistent upwelling seasonal cycle. Its amplitude is significantly larger in the older products (dashed lines in **Figure 5.6**, bottom), although the newest product (ERA5) also shows a larger amplitude than its predecessor (ERA-Interim), MERRA-2, JRA-55, CFSR, and the chemistry climate models (CCMs). The more recent products show a very similar seasonal cycle to the CCMs. 20CR stands out as essentially completely missing upwelling seasonality and generally showing much too weak upwelling. CFSR shows a consistent seasonality for total tropical upwelling, however, this arises due to compensating biases between the hemispheres during May-September. ERA-20C and JRA-55AMIP agree overall quite well with the newer reanalysis products. Note that seasonality in extratropical downwelling is generally very consistent across products (not shown), including its hemispheric differences.

The annual cycle in total tropical upwelling at 70 hPa is primarily determined by the annual cycle in local upwelling strength, as opposed to the annual cycle in the upwelling width (see **Figure 5.7**). Even though seasonal variations in turnaround latitudes are large in each hemisphere, with turnaround latitudes farthest poleward during summer and closest to the equator in winter, the upwelling width (distance between turnaround latitudes) generally shows much weaker seasonal variations. Furthermore, seasonality of the width is inconsistent between different products and the range between different products is of similar magnitude as seasonal variations. Large disagreement occurs especially during northern spring where NH turnaround latitudes exhibit very different seasonal transitions between different products: *e.g.*, MERRA and MERRA-2 are already close to their maximum poleward position during April, whereas most other products, including the CCMI MMM, only reach these positions during June-July. 20CR represents an outlier in that it lacks the correct NH seasonality.

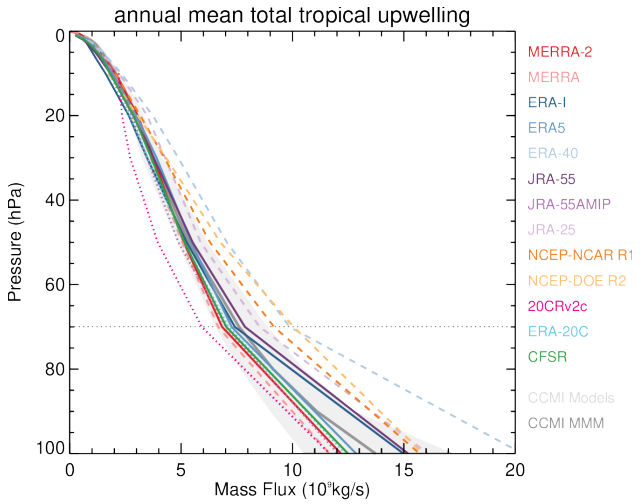

Figure 5.8: *Climatological annual mean total tropical upwelling as a function of pressure (between turnaround latitudes at each level). Line styles as* ***Figure 5.4****. Gray shading shows range of CCMI models with the thick gray line marking the MMM.*

Figure 5.6 showed that the total tropical upwelling at 70 hPa is spread over a fairly wide range between different products, with the older products showing much larger upwelling than the newer products and the climate-like runs (JRA-55AMIP, ERA-20C). This spread is even larger at 100 hPa but tends to decrease at higher altitudes (**Figure 5.8**). The vertical gradient of total tropical upwelling gives the net poleward mass flux from the upwelling region to extratropical latitudes. This gradient is generally much stronger in the older products (dashed lines) above 70 hPa. This means that leakage out of the tropical pipe is much stronger in these products. Between the newer products, JRA-55, ERA-Interim, and CFSR have similar leakage, while MERRA-2 shows a somewhat smaller leakage. The difference in upwelling between 100 - 70 hPa could be interpreted as an estimate for the net shallow branch divergence. However, it is important to note that because of the large gradient in turnaround latitudes between 100 - 70 hPa such an estimate includes both, poleward and downward mass fluxes. A large degree of variability arises due to the downward component, which also partly explains why there is generally a large spread in the mass flux gradient between 100 - 70 hPa in **Figure 5.8**. There is no clear change in this mass flux gradient from the older to the newer products. Part of the discrepancy also results from discrepancies in the turnaround latitudes (upwelling width) and their difference between 100 - 70 hPa. The upwelling width is much smaller at 100 hPa than at 70 hPa for most products. However, some products show only a small difference (*e.g.*, CFSR), whereas others show a very large difference (*e.g.*, ERA-Interim).

The upwelling strength at 100 hPa also shows a wide spread across the CCMs (gray shading in **Figure 5.8**). Presumably, differences in Hadley cell strength and vertical extent also play into the 100 hPa upwelling spread as the spread decreases markedly between 100 hPa and 70 hPa. The model diagnostics also contain the 90 hPa and 80 hPa levels, which demonstrates that the shallow branch divergence is likely weaker than diagnosed based on the 100 hPa and 70 hPa levels. ERA-Interim and JRA-55 are both close to the MMM throughout the lower stratosphere, although these two products show stronger upwelling than other recent reanalysis products (including ERA5) and the MMM at 100 hPa. For ERA-Interim and JRA-55 we had model level output available, which confirms that the shallow branch divergence based on the upper half of the 100 - 70 hPa layer is significantly weaker (about half) than that based on its lower half. Furthermore, interannual variability, similar to the spread across products, is much larger near 100 hPa. MERRA-2 and CFSR are both near the low end of model upwelling strengths. Higher up in the stratosphere (above ~ 10 hPa) the upwelling strength in the models is significantly larger than in the reanalyses, indicating a more isolated tropical pipe in the models.

To further quantify the mass flux within the shallow branch and to avoid the large sensitivity near the 100 hPa level, we consider the poleward residual flow at 70 hPa evaluated at the turnaround latitudes (**Figure 5.9**).

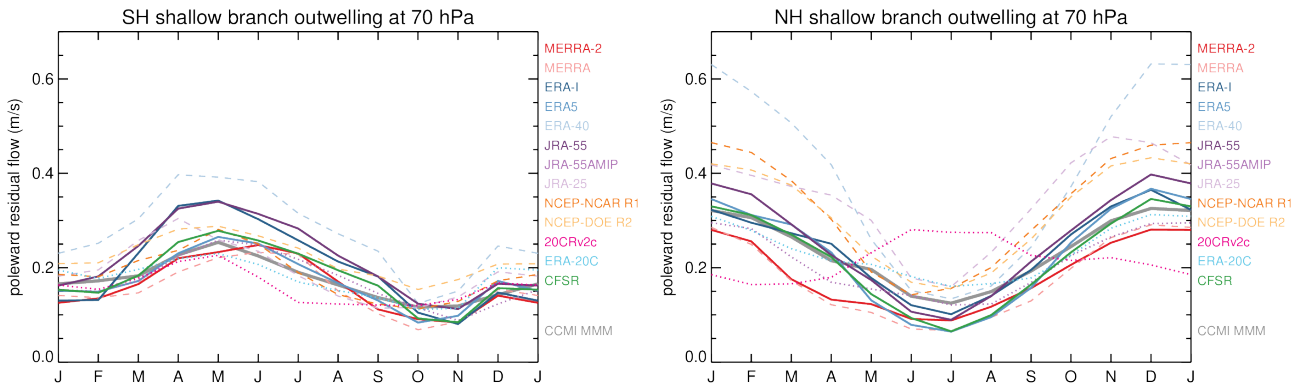

Figure 5.9: *Climatological seasonal cycles of shallow branch tropical "outwelling" (poleward residual flow through the turnaround latitudes at 70 hPa) for the SH (left) and the NH (right). Line styles as* ***Figure 5.4****.*

The SH shallow branch poleward mass flux shows a maximum between fall and early winter (strongest in ERA-Interim, JRA-55, ERA-40) and a minimum in late spring. In the NH, outwelling is generally strongest during winter and weakest during summer, except for 20CR, which is the only product to show the opposite seasonality. The older products tend to show much stronger NH outwelling throughout the year compared to the newer products. Both MERRA products show consistently weaker NH outwelling during the cold season compared to other recent products. Taking both SH and NH together the outwelling diagnostics confirms that MERRA and MERRA-2 are at the low end of both lowermost stratospheric upwelling and shallow branch outwelling. Overall, shallow branch wave driving as quantified by EPFD (**Figure 5.4**) only explains part of the spread and variation in outwelling. This is likely because at certain latitudes gravity wave drag becomes important (not diagnosed here), while closer to the equator adjustments in relative vorticity become important (which modifies the relation between $\bar{v}^*$ and EPFD).

So far we have concentrated on diagnostics that directly quantify the (local) strength of the BDC. The RCTT diagnostic provides estimates of the integrated circulation strength. **Figure 5.10** shows the annual mean structure of RCTTs for the climatological REM. Since RCTTs result from backward trajectories with transit times up to several years near the poles, we discard the first few years of the time series. In order to still obtain a 30-year climatology we use the period 1986 - 2016 for the REM. The overall structure of the RCTTs agrees well with that from CCMs (*cf.*, *Birner and Bönisch*, 2011; *Dietmüller et al.*, 2018): a strong vertical gradient in the tropical pipe, which is similar to that of AoA, and a strong meridional gradient with strongly increasing RCTTs toward the poles in both hemispheres. Interestingly, both hemispheres have about equal RCTT structures and if anything the NH shows larger RCTT values near the pole in the lowermost stratosphere, perhaps due to the fact that the NH circulation reaches deeper into the upper stratosphere and the mesosphere (see **Figure 5.2**). The double peak in tropical upwelling shows up as a double peak in RCTTs with smallest values near 20 °N/S and a local maximum near the equator (at a given level between the tropopause and ~ 10 hPa).

Figure 5.11 shows climatological annual mean RCTT structures and differences from the REM for individual reanalyses, comparing the more recent products to the older products. Overall, the more recent products tend to be much more consistent compared to the older products. ERA-40 and JRA-25 show much smaller RCTTs compared to the REM, indicating that their BDC is too strong. MERRA's tropical upwelling is biased, primarily because of local downwelling over the equator (see **Figure 5.5**), and this leads to a large positive bias in RCTTs over the equator. Of the more recent products, JRA-55 shows the smallest RCTTs, consistent with strongest tropical upwelling (*cf.*, **Figure 5.8**). JRA-55 is also the only product that does not exhibit the double peak in tropical upwelling, and likewise in RCTTs in the tropics, with a local maximum near the equator. ERA-Interim tends to show the largest RCTTs, except for the NH mid-latitude lower stratosphere (NH shallow branch). ERA-Interim's RCTTs show a pronounced hemispheric asymmetry consistent with a stronger shallow circulation branch in the NH compared to the SH.

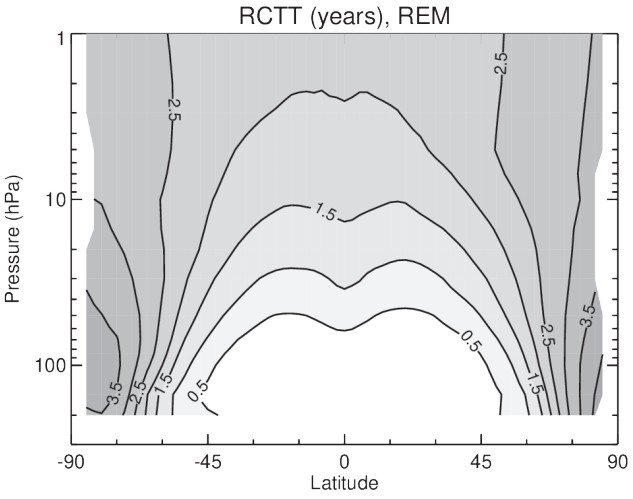

Figure 5.10: *Climatological (1986 - 2016) annual mean residual circulation transit time (RCTT, in years) distribution for the multi-reanalysis mean (REM).*

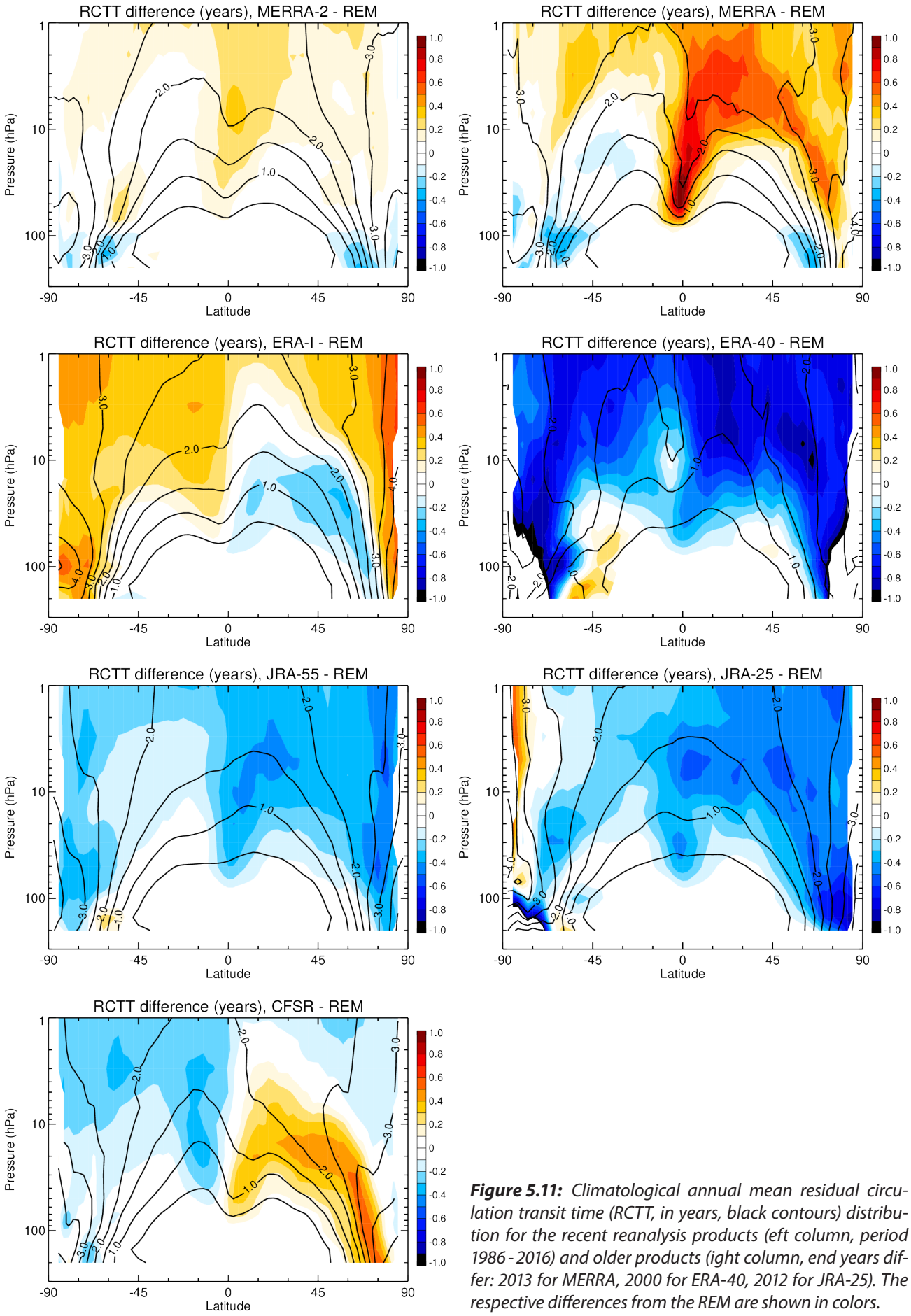

Figure 5.11: *Climatological annual mean residual circulation transit time (RCTT, in years, black contours) distribution for the recent reanalysis products (eft column, period 1986 - 2016) and older products (ight column, end years differ: 2013 for MERRA, 2000 for ERA-40, 2012 for JRA-25). The respective differences from the REM are shown in colors.*

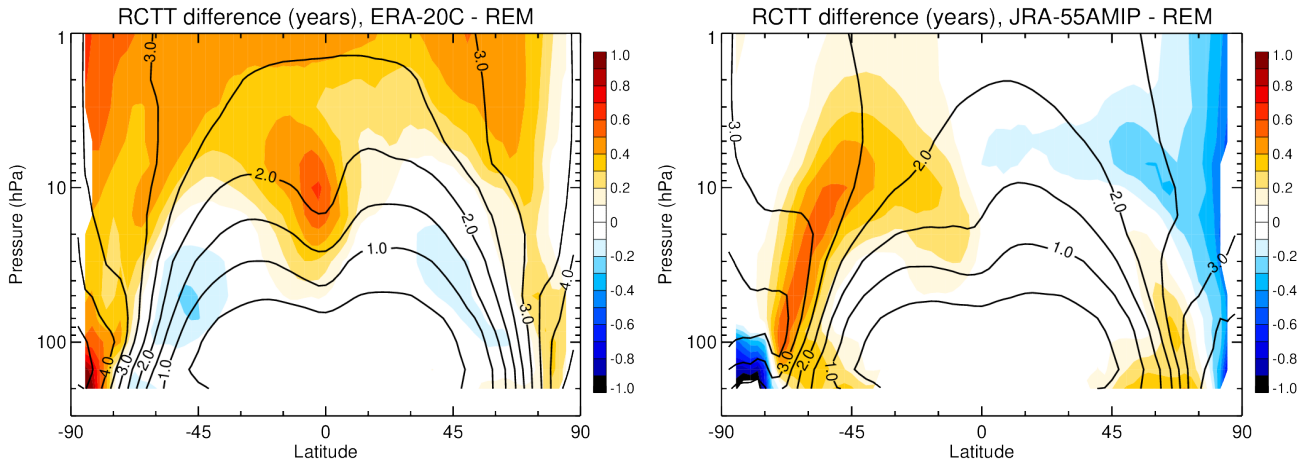

Figure 5.12: *As* ***Figure 5.11*** *but for ERA-20C (left, end year is 2010) and JRA-55AMIP (right, end year is 2012).*

This asymmetry disappears in ERA5, which agrees well with the REM overall, although RCTTs are still higher in the upper stratosphere (not shown, but climatological contours are included below in **Figure 5.17**). MERRA-2 shows a similar structure compared to its older version, but with smaller RCTTs, especially over the equator. CFSR shows smaller RCTTs throughout much of the SH, but larger RCTTs in the NH, especially along the deep circulation branch. CFSR exhibits a hemispheric asymmetry that is opposite of that in ERA-Interim.

RCTTs are overall similar between ERA-20C and ERA-Interim (**Figure 5.12**), which are based on a similar underlying model, although the strong hemispheric asymmetry in ERA-I is not present in ERA-20C. The free-running model version of JRA-55 does show an asymmetry with smaller RCTTs (stronger circulation) in the NH. Differences to the REM exist in the SH mid-latitudes and NH high latitudes.

5.5.1.2 Tropical upwelling trends

Even though some observational evidence for a strengthening of the BDC exists, modern reanalyses do not consistently show such strengthening. Specifically, ERA-Interim shows inconsistent trends compared to other reanalysis products, depending on the upwelling measure used (*Abalos et al.*, 2015).

Figure 5.13 shows that interannual variability in tropical upwelling at 70 hPa is large and likely spurious in some of the older reanalysis products, such as ERA-40 (perhaps due to older data assimilation systems). Corresponding trends are therefore not trustworthy. This variability is reduced and more consistent (see below) among the more recent reanalysis products. Furthermore, these more recent products lie within the range of CCMs with JRA-55 closely following the MMM, whereas MERRA-2 and CFSR consistently lie near the lower edge of model time series. ERA-Interim, which is the only recent product that shows a negative trend, initially closely follows the MMM and JRA-55, but from the late 1990's onward more closely follows the other three recent products. ERA5 shows overall similar variability to ERA-Interim but with consistently smaller upwelling values between ~1980 - 2005 transitioning to larger upwelling values from 2006 onward. ERA-20C and JRA -55AMIP also show a similar time series to those models with weaker overall upwelling. 20CR is generally biased low compared to all other products.

Visually, JRA-55, MERRA-2, and CFSR all show positive trends in tropical upwelling indicating a strengthening of the BDC. These trends are quantified in **Table 5.1** and for these three products are in the 2 - 3 %/decade range. ERA-Interim, on the other hand, shows a negative upwelling trend of the same order of magnitude (*cf.*, *Abalos et al.*, 2015), indicating a weakening of the BDC in this product. ERA5 shows a weak negative trend that is, however, not statistically significant.

To compare the reanalysis trends to those from CCMs we also calculated associated upwelling trends (see **Table 5.2**), but in this case for the longer period of 1960 - 2009 (the common period of 1980 - 2009 between the CCMs and the recent reanalyses is marginally short to obtain robust trends, *cf.*, *Hardiman et al.*, 2017).

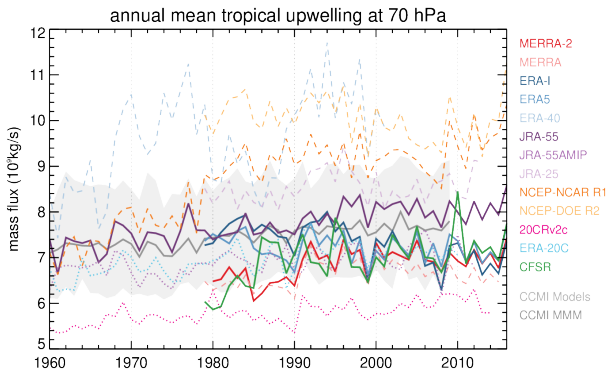

Figure 5.13: *Time series of annual mean tropical upwelling mass flux at 70 hPa (between turnaround latitudes). Line styles as* ***Figure 5.4****. The gray shading denotes the range of CCMI models with the multi-model mean shown as thick gray line.*

Table 5.1: *1980 - 2016 trends (in %/decade) of total tropical upwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude.*

MERRA-2	+ **2.5** ± 1.3
ERA-I	− **3.4** ± 1.4
ERA5	− 0.7 ± 1.3
JRA-55	+ **2.3** ± 0.9
CFSR	+ **3.4** ± 2.0

The MMM trend is significantly smaller than those based on reanalyses, but some individual models (*e.g.*, WACCM) reach a similar upwelling trend of ~ 2 %/decade. ERA-20C also exhibits a trend of ~2 %/decade over this time period, with JRA-55 (the only product, which has both examined time periods available) producing a similar trend than over the 1980-2016 period (see above). The free-running version of JRA-55 exhibits a trend near the lower end of, but consistent with, CCM trends.

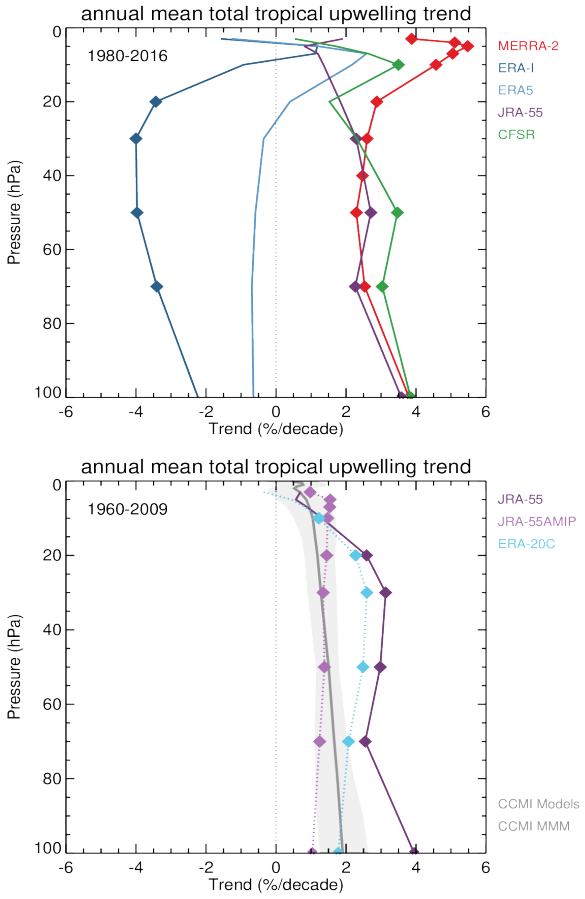

Figure 5.14: *Trends in annual mean total tropical upwelling as a function of pressure (between turnaround latitudes at each level). Top: for the more recent reanalysis products and the period 1980 - 2016. Bottom: comparing climate models with reanalyses and other products for the period 1960 - 2009 (gray shading shows range of CCMI model trends with the thick gray line marking the MMM). Symbols indicate trends that are statistically significantly different from zero (based on 2σ uncertainty). For the MMM line all levels have statistically significant trends.*

Table 5.2: *1960 - 2009 trends (in %/decade) of total tropical upwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude. Note: individual model trends range from + 1.1 ± 0.5 (CMAM) to + 2.1 ± 0.7 (WACCM).*

CCMI MMM	+ **1.7** ± 0.4
JRA-55AMIP	+ **1.2** ± 0.8
JRA-55	+ **2.5** ± 0.6
ERA-20C	+ **2.1** ± 0.8

The trends in 70 hPa tropical upwelling are overall consistent with those at other stratospheric levels as shown in **Figure 5.14**. Specifically, MERRA-2, JRA-55, and CFSR all show mostly consistent, statistically significant positive upwelling trends for the period 1980 - 2016 over much of the stratosphere. These trends are somewhat stronger at 100 hPa (between + 3.5 - 4 %/decade), vertically coherent in the + 2 - 3 %/decade range up to 20 hPa, above which the different reanalysis products disagree about the trend. In contrast, ERA-Interim shows negative upwelling trends that are statistically significant between 70 - 20 hPa. These results are consistent with, and serve as an update of, the results presented in (*Abalos et al.*, 2015) based on a slightly shorter time period. The new ERA5 product shows small negative trends between 100 - 30 hPa, although none of them are statistically significant in the analysed pressure range. We again also consider the period 1960 - 2009 to compare to CCMI results (**Figure 5.14**, bottom). The MMM shows upwelling trends between + 1 - 2 %/decade throughout the stratosphere, which slightly decrease with height. JRA-55AMIP's trend, on the other hand, is generally within the range of CCMI trends, except for at 100 hPa. JRA-55's trends are much larger in the lower stratosphere, well outside the range of CCMI trends. ERA-20C lies between the two JRA products.

The disagreement in overall magnitudes and trends between even the recent reanalysis products, raises questions about their ability to capture long-term climate variations. A perhaps less stringent test is to examine the interannual variability of the different products. **Table 5.3** reveals that interannual variability is reasonably well correlated between ERA-Interim, ERA5, JRA-55, and MERRA-2, but not so much between CFSR and these products.

Table 5.3: *Correlation coefficients for interannual variability of total tropical upwelling at 70 hPa between recent reanalysis products. Time series have been detrended before calculating correlations. Bolded values indicate statistical significance at the 95 % confidence interval.*

	ERA-I	**MERRA-2**	**JRA-55**	**CFSR**
ERA-5	**0.75**	**0.66**	**0.65**	0.21
ERA-I	-	**0.67**	**0.70**	**0.50**
MERRA-2	-	-	**0.82**	0.14
JRA-55	-	-	-	0.25

Table 5.4: *1980-2016 trends (in %/decade) of total shallow branch outwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude.*

MERRA-2	+**7.2**± 2.3
ERA-I	+ 2.2 ± 2.3
ERA5	–1.3 ± 1.9
JRA-55	+**3.2**± 1.4
CFSR	+ 1.1± 2.3

5.5.1.3 Tropical outwelling and RCTT trends

Tropical upwelling at 70 hPa and above primarily measures the deep branch of the BDC. As before we use the poleward residual flow through the turnaround latitudes at 70 hPa as a measure of the shallow branch outwelling; its time series of the combined NH+SH outwelling are shown in **Figure 5.15**. Similar to lower stratospheric upwelling strength, net tropical outwelling is significantly weaker in the more recent reanalysis products compared to their predecessors (except for MERRA). The more recent products agree in their overall strength with the CCMI models, as do the other reanalysis-related products (JRA-55AMIP, ERA-20C, 20CR). Visually, MERRA-2 exhibits a strong increasing trend between 1980 - 2000, and JRA-55 exhibits a long-term trend over the entire depicted record. With the exception of ERA5, all recent reanalysis products show positive trends for the period 1980 - 2016 (see **Table 5.4**), although this trend is not statistically significant in ERA-Interim and CFSR. The weak negative trend in ERA5 is likewise not statistically significant.

Over the longer period from 1960 - 2009 JRA-55 exhibits a consistent trend in net shallow branch outwelling with the shorter period (both between 3 - 4 %/decade, *cf.*, **Table 5.5**). While some CCMI models almost reach this strong accelerating trend (WACCM), the MMM trend is somewhat weaker and JRA-55AMIP's trend is at lower end of CCMI model trends. ERA-20C exhibits a positive outwelling trend within the range of CCMI model trends.

Similar to the 70 hPa upwelling time series, we have also analyzed interannual variability in net shallow branch outwelling (see correlation coefficients listed in **Table 5.6**). Co-variability in this case is weak across many recent reanalysis's. The highest correlation coefficient is 0.68 between

Table 5.5: *1960 - 2009 trends (in %/decade) of total shallow branch outwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude. Note: individual model trends range from + 0.6 ± 0.6 (CMAM) to +2.7±1.2 (WACCM).*

CCMI MMM	+**1.9** ± 0.5
JRA-55AMIP	+ 0.8 ± 0.8
JRA-55	+**3.7** ± 0.9
ERA-20C	+**1.2** ± 0.7

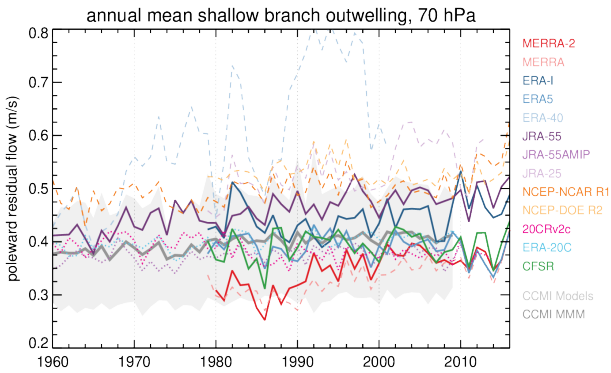

Figure 5.15: *Time series of annual mean tropical outwelling at 70 hPa (total poleward residual flow through the turnaround latitudes). Line styles as* ***Figure 5.4****. The gray shading denotes the range of CCMI models with the multi-model mean shown as thick gray line.*

MERRA-2 and ERA5. The low correlations in this case exist despite coherent variability in EP-flux divergence (not shown), indicating that unresolved processes and/or model biases are primarily responsible for the lack of co-variability in shallow branch outwelling.

We next examine time series of RCTTs. The 50 hPa (~ 20 km altitude) level is often used to compare AoA estimates (see *Section 5.5.2*). **Figure 5.16** shows the time series of annual global mean RCTTs from different products. Consistent with strongest upwelling JRA-55 shows smallest RCTTs that are steadily decreasing over time, consistent with a strengthening of the BDC. CFSR and MERRA-2 both show much larger interannual variations and a large negative trend in the 1980's and 1990's. ERA-Interim is closer to JRA-55 in the beginning of the record, but approaches CFSR and MERRA-2 toward the end of it, in the latter period showing similarly strong interannual variations. ERA5 tends to be more consistent with MERRA-2 and CFSR than with ERA-Interim in this metric, especially from the mid-1990's forward. CCMI models show a wide range of global mean RCTTs at 50 hPa, encompassing essentially all reanalysis products.

RCTT trends are examined for the period 1982 - 2016 (the first few years of the records need to be discarded because of the backward trajectory setup of the calculations), listed in **Table 5.7**.

Table 5.6: *Correlation coefficients for interannual variability of total shallow branch outwelling at 70 hPa between recent reanalysis products. Time series have been detrended before calculating correlations. Bolded values indicate statistical significance at the 95 % confidence interval.*

	ERA-I	**MERRA-2**	**JRA-55**	**CFSR**
ERA-5	0.23	**0.68**	**0.34**	**0.53**
ERA-I	-	**0.29**	0.26	**0.30**
MERRA-2	-	-	**0.43**	**0.55**
JRA-55	-	-	-	**0.52**

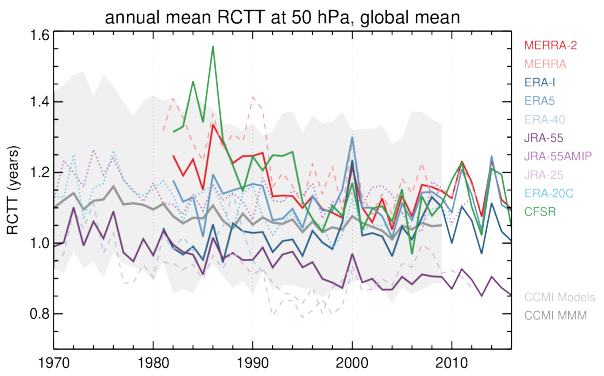

Figure 5.16: *Time series of annual gobal mean RCTTs at 50 hPa. Line styles as* ***Figure 5.4****. The gray shading denotes the range of CCMI models with the multi-model mean shown as thick gray line.*

Qualitatively, these RCTT trends are consistent with the respective upwelling trends at 70 hPa (*cf.*, **Table 5.1**): we find negative transit time trends indicating a strengthening of the BDC in MERRA-2, JRA-55, and CFSR, a positive trend indicating a weakening of the BDC in ERA-Interim, and a non-significant trend in ERA5. CFSR's trend (not shown), although formally statistically significant, has a large uncertainty due to the questionable interannual and decadal variability in the beginning of the record (see **Figure 5.16**); its trend values are therefore not included here or in other RCTT trend estimates. All recent reanalysis products indicate much weaker trends since the year 2000 (*cf.*, **Figure 5.16**), and all of them show a pronounced maximum in that year, reflecting the weaker upwelling values the year before (presumably due to the strong La Niña event in 1999). The difference in magnitude of BDC trends pre and post 2000 is consistent with recent arguments regarding the role of ozone depletion for BDC trends (*e.g.*, *Abalos et al.*, 2019; *Polvani et al.*, 2019; *Garfinkel et al.*, 2017).

Over the longer period 1970 - 2009 JRA-55 shows an even stronger negative RCTT trend at 50 hPa (**Table 5.8**). This strong BDC acceleration is not found in the free-running version JRA-55AMIP, although this data set also shows a negative RCTT trend. Moreover, JRA-55AMIP is consistent with the MMM of the CCMI models. ERA-20C's corresponding trend falls somewhere in the middle of those trends.

MERRA-2, ERA-Interim, ERA5, and JRA-55 show reasonably strong interannual co-variability in global mean RCTTs at 50 hPa with correlation coefficients ranging between 0.53 (between MERRA- and ERA-Interim) and 0.85 (between

Table 5.7: *1982 - 2016 trends (in %/decade) of 50 hPa global mean RCTTs with their 2σ uncertainties based on recent reanalysis products (see text for details). Bolded values indicate trends exceeding their 2σ uncertainty in magnitude. Note, CFSR is not included here because it shows questionable decadal variability.*

MERRA-2	**−2.9** ± 1.9
ERA-I	**+ 2.1** ± 1.8
ERA5	−0.2 ± 1.8
JRA-55	**−3.6** ± 1.0

Table 5.8: *1970 - 2009 trends (in %/decade) of global mean RCTTs at 50 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude.*

CCMI MMM	**−2.2**±0.5
ERA-20C	**− 3.0** ±1.2
JRA-55	**−4.2** ± 1.0
JRA-55AMIP	**− 1.7**± 1.1

ERA5 and ERA-Interim, *cf.*, **Table 5.9**). CFSR (not shown) variability agrees well with the other products in the latter part of the record (*cf.*, **Figure 5.16**).

The latitude-pressure structure of individual products' RCTT trends are shown in **Figure 5.17**. Overall, MERRA-2 and JRA-55 show mainly negative trends, in some cases reaching - 20 %/decade (*e.g.*, MERRA-2 in the NH subtropical lower stratosphere). JRA-55 tends to show smallest climatological RCTTs in both hemispheres (black contours), whereas ERA-Interim shows largest RCTTs for the SH deep branch with MERRA-2 showing largest RCTTs for the NH deep branch. MERRA-2 shows a large negative trend in the first half of the record for the NH deep branch, which does not continue over the latter half of the record (not shown). ERA-Interim shows primarily weakly positive trends, except for in the lowermost mid-latitude stratosphere, consistent with a weakening of its deep branch but a strengthening of its shallow branch. A similar picture emerges with ECMWF's new product, ERA5, with perhaps a wider area of negative trends in the shallow branch. The shallow branch strengthening is fairly consistent across products, except for the free-running models (as above). The strengthening of the shallow branch seen in the RCTT trends appears to be the only robust trend that is consistent across all recent reanalysis products.

This robust strengthening trend of the shallow branch is confirmed for the longer period (1970 - 2009) in JRA-55 (**Figure 5.17**). In fact, these longer-term trends are generally larger in magnitude for JRA-55. However, a consistent trend across the available products for this longer period only exists for the SH deep branch. Interestingly, the MMM of the CCMI models and JRA-55AMIP show a positive trend in RCTTs for parts of both hemisphere's shallow branches (indicative of weakening), suggesting a robust mismatch between the CCMs and reanalyses.

Table 5.9: *Correlation coefficients for interannual variability between 1982 - 2016 of global mean RCTTs at 50 hPa between recent reanalysis products. Time series have been detrended before calculating correlations. CFSR is not included (see text for details). Bolded values indicate statistical significance at the 95 % confidence interval.*

	ERA-I	**MERRA-2**	**JRA-55**
ERA-5	**0.85**	**0.77**	**0.70**
ERA-I	-	**0.53**	**0.64**
MERRA-2	-	-	**0.71**

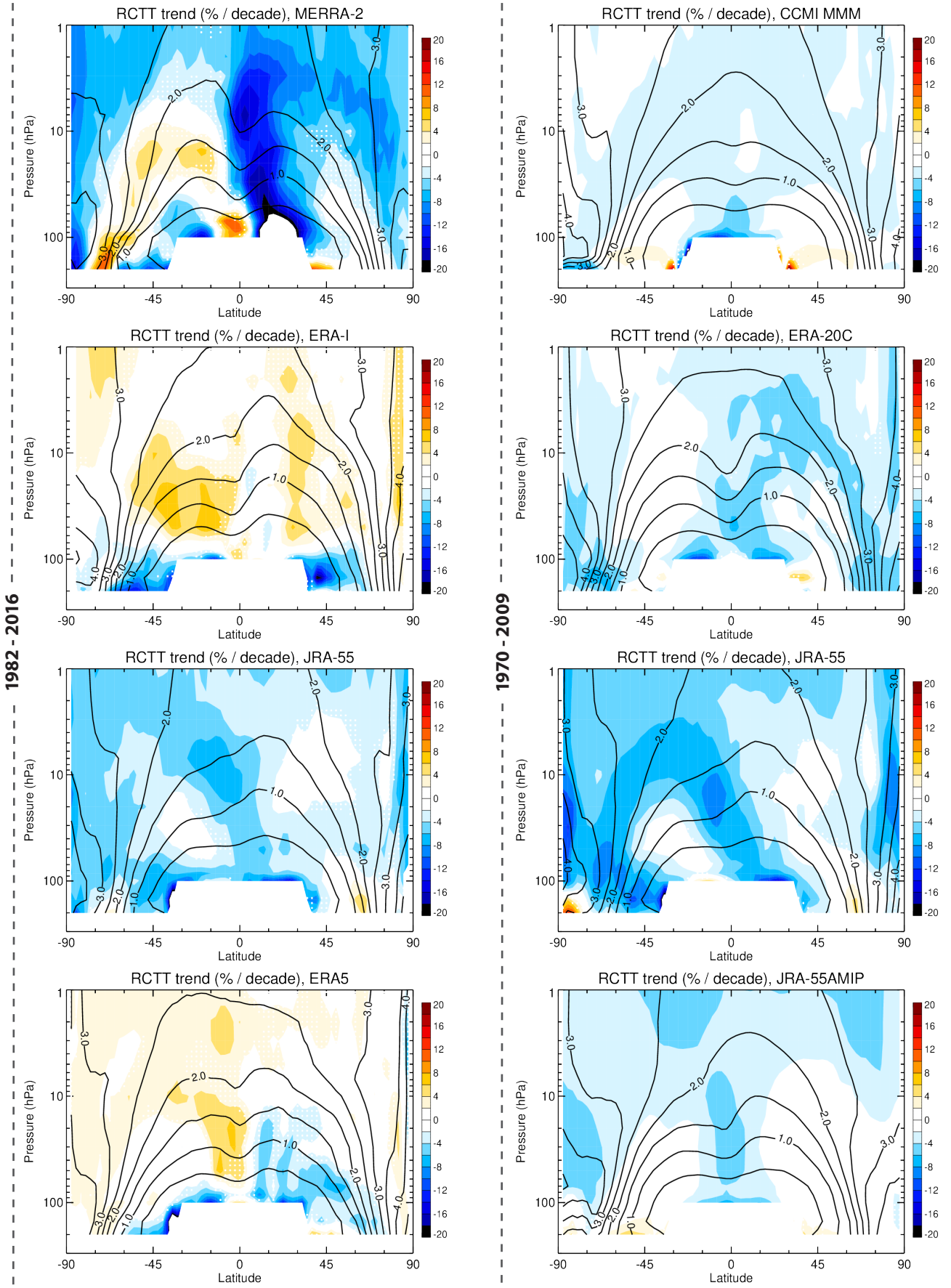

Figure 5.17: *Trends in annual mean RCTT as a function of latitude and pressure (color shading, in %/decade) for the period 1982 - 2016 in the recent reanalysis products (eft column, note that ERA5 is included instead of CFSR, see text), as well as the period 1970 - 2009 in the products shown (ight column). Each product's climatology over the respective period is depicted as black contours.*

For the deep branches the MMM of the CCMI models shows robust negative trends in RCTTs, indicating opposing trends between parts of the shallow and deep branches in those models (at least by this metric of the BDC).

5.5.2 Results from transport tracers simulations

5.5.2.1 Heating rates

Heating rates[3] from reanalysis are not only a stratospheric circulation diagnostic in itself, but they are also one of the fields used to drive some of the offline models employed for our tracer transport simulations. The CLaMS and TRACZILLA offline models use heating rates for their advection schemes. TOMCAT/SLIMCAT also uses diabatic heating rates when run in "SLIMCAT" mode. Therefore, a comparison of diabatic heating rates in the different reanalyses datasets contributes to identify differences in stratospheric transport in the considered simulations.

Figure 5.18 shows the annual cycle of the diabatic heating rate, $\dot{\theta}$ (K/day), in isentropic coordinates, at the tropical UTLS based on daily data covering 1980 - 2010; data come from the ERA-Interim, JRA-55, MERRA-2 and CFSR reanalyses.

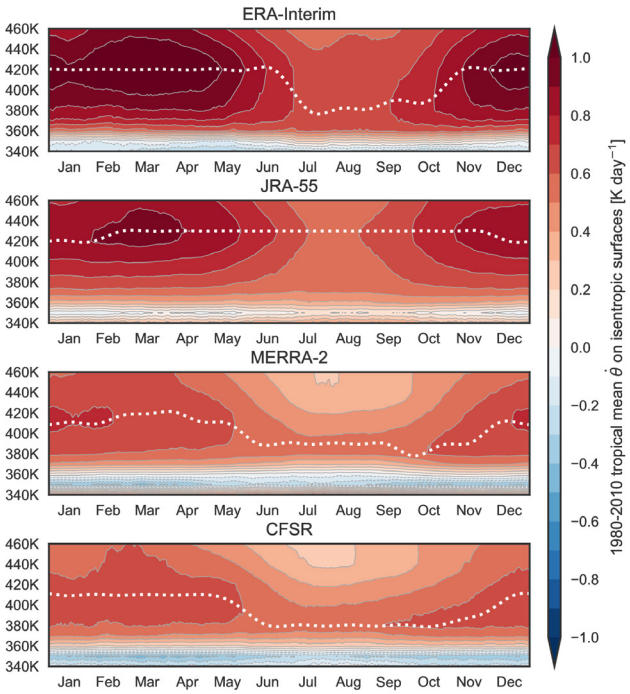

Figure 5.18: *Annual cycle of the tropical mean (30 °S - 30 °N) of the diabatic heating rate field (K/day) on isentropic surfaces, from 340 - 460 K. The field has been averaged over the period 1980 - 2010 for the ERA-Interim (top row), JRA-55 (second row), MERRA-2 (third row) and CFSR (bottom panel) reanalyses. The white dotted line in each panel shows the annual cycle of the local maximum within the lower stratosphere for the corresponding reanalysis.*

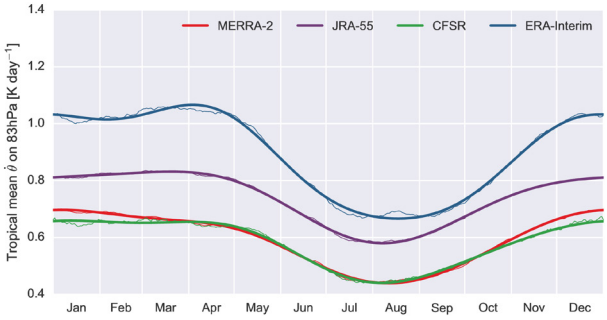

Figure 5.19: *Annual cycle of the tropical mean (30° S-30° N) of the diabatic heating rate field (K/day) at the 83 hPa level from ERA-Interim (blue), JRA-55 (purple), MERRA-2 (red) and CFSR (green). Annual cycles are based on day-of-year means (thin lines) and smoothed using FFT-based low pass filters (thick lines).*

The field has been averaged over the broader tropical region (30 °S-30 °N). The white dotted line in each panel shows the annual cycle of the local maximum within the lower stratosphere for the corresponding reanalysis. Details on the way these fields have been calculated can be found in *Wright and Fueglistaler* (2013) and *Dessler et al.* (2014), and a detailed discussion on this fields in *Martineau et al.* (2018). The figure shows that the strongest annual cycle and the largest values for this field correspond to the ERA-Interim reanalysis, the structure of the cycle is similar for JRA-55 although with weaker and smaller values, especially over the months with maximum values. MERRA-2 and CFSR, show smaller $\dot{\theta}$ values and a weaker annual cycle than the other two reanalyses for the whole vertical profile. The corresponding annual cycle for the 83 hPa level is shown in **Figure 5.19**, showing both the annual cycles based on day-of-year means (thin lines) and those smoothed using FFT-based low pass filters. As in **Figure 5.18**, there are significant differences among reanalyses, with ERA-Interim showing the highest values and the most pronounced seasonal cycle, and CFSR showing the lowest values and least pronounced seasonal cycle. MERRA-2 is surprisingly very similar to CFSR and JRA-55 values are in between the CFSR and the ERA-Interim ones, although the amplitude of the seasonal cycle in JRA-55 is as low as for CFSR and MERRA2.

The way the $\dot{\theta}$ field has evolved with time for the different reanalyses is shown in **Figure 5.20** as the time series of the tropical mean (30 °S-30 °N) of the diabatic heating rate field (K/day) at the 83 hPa level for the diagnosed field (darker solid lines) and for the forecast field (lighter solid lines). The corresponding linear trends (K/day per decade) are also shown in this figure. All reanalyses except ERA-Interim show considerable differences between the diagnosed and the forecast fields, and even for ERA-Interim the corresponding linear trends are different for both sets. Figures shown here evidence the large differences that exist between reanalyses for the annual cycle of the diabatic heating rates. There are also large differences between reanalyses regarding the diurnal cycle of diabatic heating rates in the UTLS in convective regions (*Tegtmeier et al.*, 2020).

[3] See the footnote on diabatic heating rates in reanalyses in *Chapter 12, Section 12.1.3.*

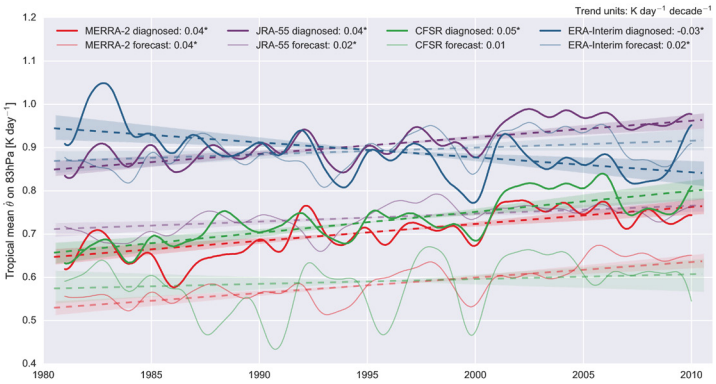

Figure 5.20: *Time series of the tropical mean (30° S-30° N) of the diabatic heating rate field (K/day) at the 83 hPa level for the diagnosed field (darker solid lines) and for the forecast field (lighter solid lines). The corresponding linear trends (K/day·decade) are also shown (dashed lines). Time series have been low-pass filtered via a 24-month rolling mean using a Hamming window. Trends are calculated from annual mean values via the Theil-Sen estimator (95 % confidence intervals estimated via bootstrapping). The best-estimate trend values are shown in the legend. Reanalyses shown and colour scale as in* ***Figure 5.19****.*

5.5.2.2 Mean age-of-air from observations

In this section we discuss mean AoA results obtained from recent observation based studies and datasets described in *Section 5.4*. We use these in later sections to compare our results from the offline models driven by the different reanalyses.

"Standard" observations for model intercomparisons

Observation-based mean AoA is derived from concentration measurements of long-lived tracers with an approximately linear increase at the surface, such as CO_2 or SF_6. For CO_2 one needs to take into account the surface seasonal cycle, which can still affect derived AoA values in the lower stratosphere, while SF_6 is affected by the mesospheric sink which makes derived values in the upper stratosphere biased towards older values. Multidecadal datasets were compiled from balloon soundings or aircraft flights using both CO_2 and SF_6 measurements (*e.g., Ray et al.*, 2014; *Andrews et al.*, 2001; *Ray et al.*, 1999; *Boering et al.*, 1996; *Elkins et al.*, 1996; *Harnisch et al.*, 1996). These observational datasets have been used for model validation in numerous studies and SPARC model intercomparison activities (*e.g., Ploeger et al.*, 2019; *Chabrillat et al.*, 2018; *Dietmüller et al.*, 2018; *Monge-Sanz et al.*, 2012; 2007; *Eyring et al.*, 2006; *Waugh and Hall*, 2002).

British Antarctic Survey (BAS) polar measurements

Cook and Roscoe (2009, 2012) used stratospheric measurements of polar summer NO_2 to derive trends in the BDC. Stratospheric column of NO_y over the Antarctic station of Rothera (67° S) were derived from measurements of NO_2 taken during 1990 - 2007; a photochemical model and observed ozone and temperature profiles were used to determine column values (**Figure 9** in *Cook and Roscoe*, 2009). Years 1991 and 1992 were excluded from their calculations because of the large amounts of volcanic aerosols from the Pinatubo eruption still present in the stratosphere.

A reconstruction from a multiple regression of these NO_y values, in which the solar cycle, the QBO, the ENSO, and a linear term are considered, is shown in **Figure 5.21**. The ratio of NO_y column to the BDC strength can be calculated following the methods in *Cook and Roscoe* (2009); they found a trend value in NO_y of – 1.1 ± 3.5 %/decade which translated into an increase in BDC of 1.4 ± 3.5 %/decade. Therefore, from the studies of *Cook and Roscoe* (2009, 2012), the conclusion was that the BDC exhibited no significant trend over the summer Antarctic for the period considered. However, they also pointed out the existence of an unexplained cycle with an amplitude of at least 15 % and a period longer than 17 years, with a minimum in BDC strength (maximum in NO_y values) around year 2000.

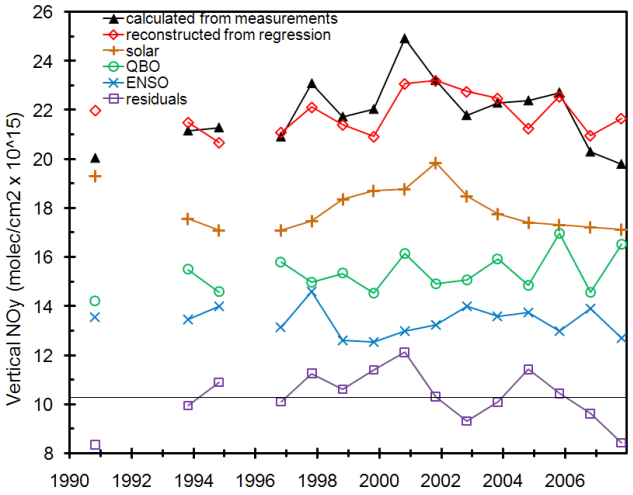

Figure 5.21: *Reconstructions from a multiple regression of the NO_y values from* ***Fig. 9*** *in Cook and Roscoe (2009) against solar cycle, QBO, ENSO, and a linear term (as shown in the key legend). The measurements and total reconstruction have the scale shown, but the separate terms in the reconstruction have been arbitrarily displaced for clarity. The linear trend term is the thin black line with the residuals centred on it; note that the residuals have been offset upwards and the trend is negligible. The conclusion was that the trend in speed of BDC was – 1.1 ± 3.5 %/decade. The speed also had a large unexplained cycle of amplitude > 15 % and period > 17 years. See Cook and Roscoe (2009, 2012) and main text for further details.*

Balloon observations over NH midlatitudes

The balloon-borne observations used in *Engel et al.* (2009) were taken in a region between 24 km and 35 km where the vertical gradient in mean age at NH midlatitudes was found to be very small, leading to little variability in this region. The balloon data were limited to a total of 28 flights and showed a positive trend of 0.24 years per decade for this region, which was, however, estimated to be non-significant.

These results have been recently updated by *Engel et al.* 2017), **Figure 5.22**, adding balloon-borne AirCore observations, to extend the previous data series so that it now covers more than 40 years. The corresponding updated trend is smaller than that from *Engel et al.* (2009) with a small positive value of 0.15 ± 0.18 years per decade. Although the trend is statistically non-significant, these observations are still in contrast to the strong negative trends in mean AoA derived from most climate model calculations (*e.g., Waugh*, 2009). The potential of the relatively cheap measurements of the AirCore instrument in *Engel et al.* (2017) makes them a promising way to keep monitoring AoA in the LS regions of interest.

An acknowledged caveat in the observations in the studies by *Engel et al.* (2009; 2017) and *Cook and Roscoe* (2009, 2012) is that the corresponding datasets covered only limited regions (midlatitudes or the Antarctic) and were also sparse in time. Global measurements from satellites are needed to provide a more complete picture of the BDC strength. For the period 2002 - 2012 AoA values derived from the SF_6 measurements taken by the MIPAS instrument on board Envisat are available (see next subsection).

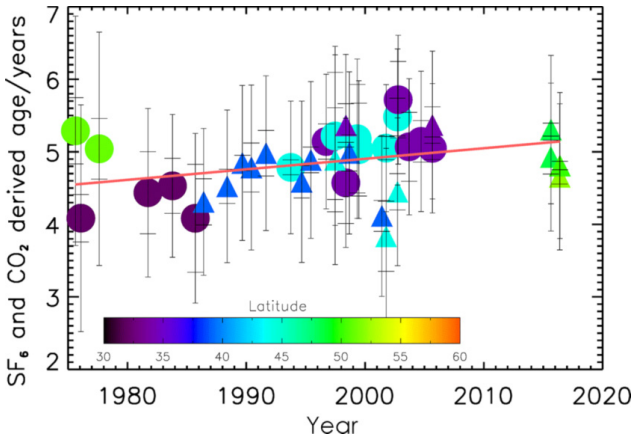

Figure 5.22: *Time series of mean age derived from balloon observations. Data before 2010 are from Engel et al. (2009), while data from 2015 and 2016 are derived from the Air-Core measurements in Engel et al. (2017). Each data point represents the average mean AoA between 5 - 30 hPa. Error bars represent the variability (inner error bars), and the uncertainty (outer error bars) as discussed in Engel et al. (2009). A non-significant trend of 0.15 (± 0.18) years per decade is derived from these observations. Figure from Engel et al. (2017).*

In this Chapter we have extensively used this MIPAS dataset to validate our offline model simulations obtained with the different reanalyses.

MIPAS mean AoA dataset

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) was an instrument on board of the Envisat satellite, measuring the mid infrared emission of the atmosphere against the space background. The measurements were done in limb scanning mode covering an altitude range of cloud top (or about 6 km in cloud-free cases) to about 72 km. The emission signatures of molecules in the atmosphere were used to retrieve the spatial distribution of up to 30 different trace gases and temperature with good global coverage from pole to pole, also during (polar) night. The mission extended from July 2002 to April 2012.

Information on the stratospheric mean AoA is obtained from the spatio-temporal distribution of the SF_6 tracer, measured by MIPAS with a vertical resolution of 4 km to 6 km and a single profile precision of about 10 - 20 %. Although the single profile precision is rather low, the huge number of profiles measured (more than 2 million profiles over the MIPAS mission lifetime) provided very valuable information on AoA from zonal mean distributions. The SF_6 distributions were retrieved from the upper troposphere up to about 50 km. Above 35 km, the systematic errors become rather large, and the vertical resolution deteriorates; for this reason quantitative analysis of SF_6 and AoA above 35 km is not recommended.

AoA has been derived from SF_6 zonal monthly averages using a surface SF_6 reference curve from the NOAA/GMD network. The combined global mean surface SF_6 from NOAA (**http://www.esrl.noaa.gov/gmd/hats/combined/SF6.html**) was used to calculate the time lag between the time of stratospheric measurement and the time when the same SF_6 amount had been observed on the surface. It was confirmed that the MIPAS-measured SF_6 mixing ratios of the free tropical troposphere and their trends agree well with the surface SF_6 abundances, and a small bias correction was applied before using the surface reference. For a strictly linearly growing tracer, this time lag is identical with the first moment of the AoA spectrum, the mean age of stratospheric air. Since SF_6 is not strictly linearly growing, the AoA calculation was corrected by convolving the SF_6 surface time series with ideal AoA spectra; more details are given in *Stiller et al.* (2012). In this way, monthly-mean zonal-means of mean AoA were provided for 10° latitude bands and 1 - to - 2 km altitude steps, for the period July 2002 to March 2012.

In **Figure 5.23** the mean AoA derived from MIPAS observations at 20 km altitude is compared to earlier airborne AoA observations taken during a large number of aircraft and balloon campaigns (*Waugh and Hall*, 2002; *Andrews et al.*, 2001; *Hall et al.*, 1999; *Ray et al.*, 1999).

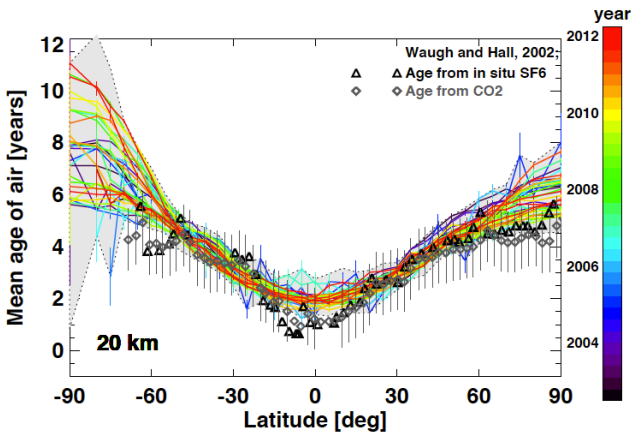

Figure 5.23: *Comparison of MIPAS-derived AoA (coloured lines) as a function of latitude at 20 km of altitude with airborne AoA measurements on basis of SF_6 (triangles) and CO_2 (diamonds). Shown for MIPAS is the monthly zonal mean for every third month between July 2002 and March 2012. The grey shaded area is the range of variability of all MIPAS zonal mean data. The vertical error bars are measurement uncertainties for CO_2-derived airborne AoA measurements. Airborne measurements are shown after Waugh and Hall (2002) and Hall et al. (1999). Figure from Haenel et al. (2015, Supplement).*

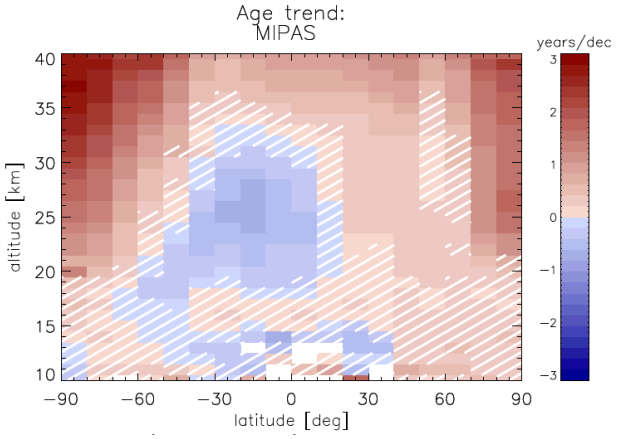

Figure 5.24: *Linear trend of AoA from a multivariate linear regression including seasonal variation and QBO effects for the period 2002 to 2012 from MIPAS data. Hatching indicates bins where the trend is not statistically significant in terms of its 2-sigma uncertainty. Figure from Haenel et al. (2015).*

The MIPAS values as a function of latitude are displayed for every second month, showing the considerable temporal variability of AoA at 20 km. In the midlatitudes, the MIPAS-derived AoA agrees very well with the SF_6-derived AoA from the aircraft campaigns, while it is higher compared to the CO_2-derived AoA (but in the same range as the difference between the SF_6- and CO_2-derived AoA from the aircraft campaigns). At southern polar latitudes, the variation of MIPAS-derived AoA is very large, however, these latitudes were not covered by the aircraft measurements. At northern polar latitudes, the MIPAS derived AoA are older than the aircraft data, although the SF_6-derived aircraft data still fall into the variability range of MIPAS. The most significant difference between MIPAS-derived and aircraft-derived AoA, however, is in the tropics. Here MIPAS-derived AoA is considerably older and this leads to a lower meridional gradient in AoA. Since there are more than ten years in between the measurements of MIPAS and most of the aircraft data (*e.g.*, see **Table 1** in *Waugh and Hall*, 2002), it cannot be determined whether this discrepancy is due to a change in atmospheric behaviour, *e.g.*, stronger inmixing of extratropical air into the tropical pipe, or if it is an artefact in one of the two datasets. The reference use for SF_6 to calculate the AoA values may also play a role: for the aircraft data, SF_6 measurements at the tropical tropopause were used as reference, while for MIPAS the global mean SF_6 time series at the surface provided by NOAA Global Monitoring Laboratory (GML) has been used.

The MIPAS monthly zonal means of AoA have been studied with respect to their short-term variability (seasonal cycle, QBO impact) and their decadal linear trend (*Haenel et al.*, 2015; *Stiller et al.*, 2012). The corresponding MIPAS-derived linear trends for the latitude/altitude bins are provided in **Figure 5.24**. There are wide areas where the AoA trend is significantly different from zero. Interestingly, a positive AoA trend is found all over the stratospheric Northern extratropics. These trends in the Northern midlatitudinal middle stratosphere agree well with the 30-year trends derived by *Engel et al.* (2009) from balloon-borne cryogenic air sampling data (see **Figure 7** in *Haenel et al.*, 2015) and are significantly positive. In the Southern Hemisphere, the Northern tropics, and the UTLS in both hemispheres, negative AoA trends are found, in agreement with most climate model predictions (*e.g.*, *Waugh*, 2009). Both mean AoA values and the trends derived from SF_6 can be affected by the mesospheric SF_6 loss (*e.g.*, *Reddmann et al.*, 2001; *Ravishankara et al.*, 1993). The SF_6-depleted air subsiding from the mesosphere in polar winters is thus misinterpreted as very old air; similarly, as the amount of SF_6 depletion scales with the absolute SF6 abundance that is increasing, the mesospheric loss leads to an apparent positive age trend. Both issues affect not only the polar winter air, but also the rest of the stratosphere, to the extent the previously mesospheric air is mixed into lower latitudes and altitudes after the polar vortex break-down. There have only been a few studies providing some estimation of the impact of the SF_6 mesospheric sink on absolute AoA and AoA trends in the stratosphere; see related discussion with the KASIMA results later in *Section 5.5.2.3*, and also in *Stiller et al.* (2012) and *Kovacs et al.* (2017).

5.5.2.3 Mean AoA from offline models

To assess the reanalyses' ability to reproduce atmospheric tracers distribution and evolution, we have used a set of different offline models (*Section 5.3*). By using several offline models we obtain a certain spread in the performance of the different reanalyses, which helps to overcome the sensitivity that a particular reanalysis may have to a particular offline model configuration.

The use of different types of offline models (*e.g.*, kinematic and diabatic models, in which the vertical motion is derived from the wind velocity or diabatic heating rate fields respectively) also allows us to narrow the source of the transport differences to particular fields in the reanalyses.

We have not included ERA-40 data (*Uppala et al.*, 2005) in our comparisons as this reanalysis was already shown to provide unrealistically fast stratospheric transport by numerous studies (*e.g.*, *Chipperfield,* 2006; *Scheele et al.,* 2005; *Meijer et al.,* 2004; *van Noije et al.,* 2004). The TOMCAT CTM was the first one to show the improvements obtained in the stratospheric transport representation with the newer generation ERA-Interim reanalysis compared to the ERA-40 reanalysis (*Monge-Sanz et al.*, 2007).

As an overall comparison of the different participating offline models, we examine how they reproduce the mean AoA from simulations driven with the ERA-Interim reanalysis. **Figure 5.25** shows the cross section of mean AoA values obtained from all the offline models driven by meteorological fields from this reanalysis, averaged over the period 1989 - 2010. All models reproduce similar distributions although some differences are also seen: Eulerian kinematic models (BASCOE, KASIMA and TOMCAT) simulate overall younger mean AoA values than the diabatic Lagrangian models (CLaMS and TRACZILLA). The shape of the tracer isolines is narrower over the tropics for the Lagrangian models, and the tropical peak shows a slight tilt towards South, while for the Eulerian models this tilt is less pronounced and goes northwards.

In **Figure 5.25**, the top right panel shows the annual-mean zonal-mean distribution of the mean AoA as simulated by the KASIMA CTM driven by ERA-Interim reanalysis fields for the period 1980 - 2010. It shows the typical bell form with maximum mean age values slightly older than 5 years; and the asymmetry with older mean age in southern polar latitudes, related to the persistent downwelling of old air from the mesosphere during the Antarctic winter polar vortex. This figure shows the AoA distribution obtained with the ideal tracer T1 (see *Section 5.3*) and it can therefore be compared to the analogous distribution obtained with the TOMCAT CTM.

The bottom left panel in **Figure 5.25** shows the mean AoA zonal mean distribution averaged over the period 1989 - 2010 from the ERA-Interim TOMCAT CTM simulation. Maximum values older than 5.0 years are reached above 58 km over the tropics, and above 30 km in the SH high latitudes; younger values are found over the NH high latitudes than over the SH. Thus TOMCAT mean AoA values are in good agreement with KASIMA, although KASIMA yields slightly older values over high latitudes in both hemispheres. TOMCAT AoA values are also similar to those obtained with the TRACZILLA "raw" simulation (without corrections), and with the CLaMS ERA-Interim simulation.

The mean AoA distribution obtained with TOMCAT for the period 2002 - 2007 is in good overall agreement with that from BASCOE (**Figure 5.26**), although BASCOE AoA maximum values are slightly younger (less than 0.5 years) than TOMCAT for all latitudes, except over the SH polar region where they are around 0.5 years older.

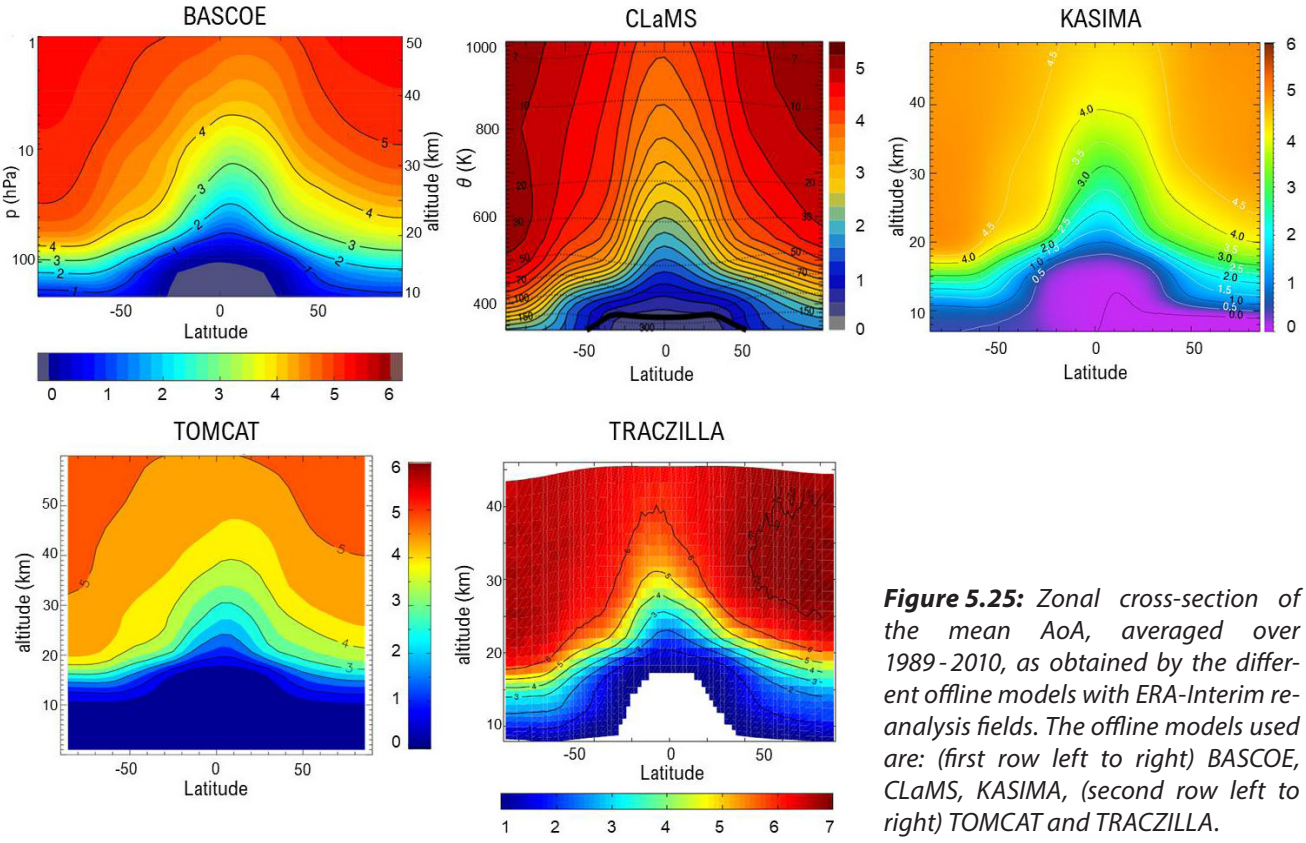

Figure 5.25: *Zonal cross-section of the mean AoA, averaged over 1989 - 2010, as obtained by the different offline models with ERA-Interim reanalysis fields. The offline models used are: (first row left to right) BASCOE, CLaMS, KASIMA, (second row left to right) TOMCAT and TRACZILLA.*

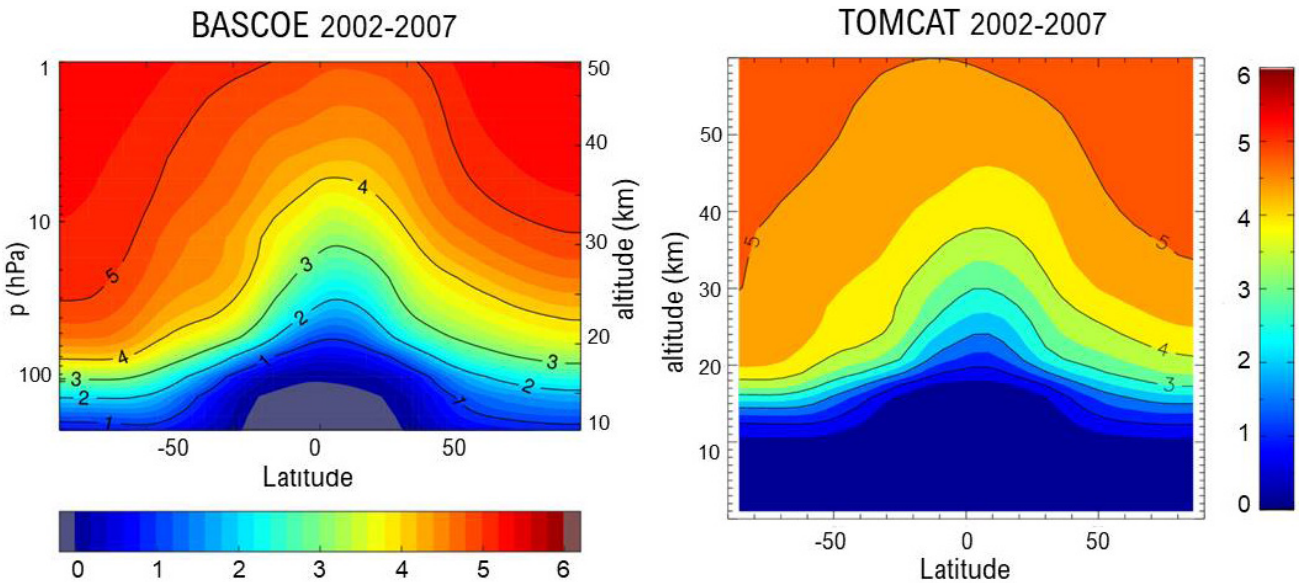

Figure 5.26: *Mean AoA zonal mean averaged over the period 2002 - 2007 from the BASCOE (left) and the TOMCAT (right) simulations driven by ERA-Interim reanalysis.*

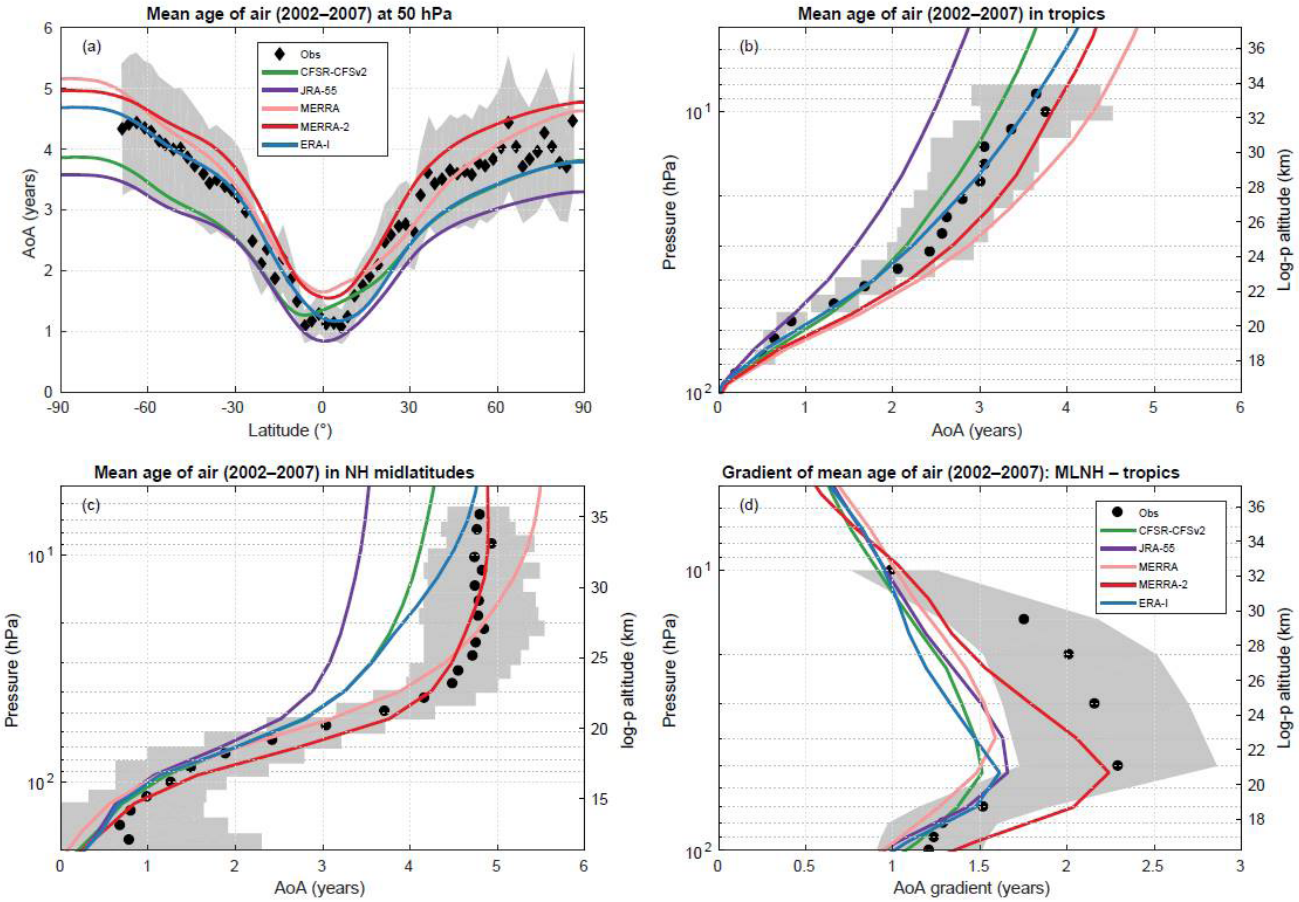

Figure 5.27: *Mean AoA in 2002 - 2007 by the BASCOE TM driven by five reanalyses (colour solid lines) versus in-situ observations (symbols) with their 1σ uncertainties (grey shading). The five reanalyses are ERA-I (blue), MERRA-2 (red), MERRA (pink), JRA-55 (purple) and CFSR (green). The modeled AoA fields are corrected so that mean age = 0 at the tropical tropopause (100 hPa). (a) AoA at 50 hPa with aircraft observations of CO_2 (Andrews et al., 2001); (b) AoA in the tropics(10° N - 10° S) with aircraft observations (Andrews et al., 2001); (c) AoA in the northern mid-latitudes (35° N - 45° N) with balloon observations s (Engel et al., 2009) and (d) AoA gradient between the northern mid-latitudes and tropics (Chipperfield et al., 2014; Neu et al., 2010). Figure from Chabrillat et al. (2018).*

A comparison of five different reanalyses (CFSR, JRA-55, MERRA, MERRA-2 and ERA-Interim) over this shorter period (2002 - 2007) obtained with BASCOE is shown in **Figure 5.27**. The AoA distribution is averaged for this period to remove seasonal and quasi-biennal oscillation signals. The figure shows the intercomparison of AoA zonal means at 50 hPa, vertical profiles over the tropics and over northern midlatitudes, and the gradient profile between these two latitudinal bands.

The intercomparison at 50 hPa (**Figure 5.27.a**) shows large disagreement between the five model simulations. JRA-55 yields the youngest AoA at all latitudes, with values ranging from 0.8 years at the equator to 3.6 years at the South Pole, while MERRA and MERRA-2 give the oldest AoA with 1.6 years at the equator and around 5 years at the South Pole. CFSR and ERA-Interim yield intermediate results with nearly identical values in the northern extratropics but different latitude gradients in the tropics and SH. In the SH, CFSR results in mean AoA values nearly as young as JRA-55 while ERA-Interim reaches larger values much closer to observations. Overall, the spread between the five simulations at 50 hPa is larger than the 1-σ observational uncertainties in the tropics, and nearly as large in the extratropics.

The AoA difference between the tropics and mid-latitudes (**Figure 5.27.d**) is directly related to the inverse of the tropical upwelling velocity and is independent of quasi-horizontal mixing: a smaller AoA latitudinal gradient indicates faster tropical ascent (*Linz et al.*, 2016). The agreement among reanalyses for the considered period is better for the AoA latitudinal gradients than for the AoA profiles. The spread between the four reanalyses (MERRA-2 excluded) reaches a maximum of 0.2 years at 30 hPa. While there is good agreement with the observation-based latitudinal gradient from 10 - 60 hPa, the four reanalyses significantly underestimate the AoA for the pressure range in between those two levels. This indicates an overestimation of the tropical upwelling obtained with ERA-Interim, CFSR, JRA-55 and MERRA in the LS region. MERRA-2 shows an outlying vertical profile of mean AoA for the latitudinal gradient; it underestimates the tropical upwelling in the lowermost stratosphere (100 - 60hPa), agrees well with observations at 50 hPa and joins the results of the four other reanalyses above that level.

The zonal cross-section of mean AoA for the period 2002 - 2007 from the BASCOE simulation driven by ERA-Interim (**Figure 5.26** left panel) shows the expected hemispheric asymmetry with a stronger latitudinal gradient in southern mid-latitudes and polar regions than in the NH. It also shows old air masses reaching lower altitudes over the Antarctic than over the Arctic. The corresponding mean AoA distributions obtained with the other four reanalyses (**Figure 5.28**) are significantly different. JRA-55 and CFSR are the "younger reanalyses" with AoA not exceeding 5 years in the polar upper stratosphere; MERRA is the "oldest reanalysis" with maximum AoA values as large as 6.5 years; ERA-Interim shows intermediate values (5.8 years in the same regions). MERRA-2 shows upper stratospheric values similar to those with ERA-Interim but very different latitudinal gradients. Also the hemispheric asymmetry is more evident with ERA-Interim than with any other reanalysis, *e.g.*, the 3 and 4-year isolines with JRA-55 and CFSR respectively, or the 5-year isoline with MERRA-2 and MERRA, reach nearly the same level above the North Pole than above the South Pole. MERRA-2 stands out in the middle stratosphere with nearly vertical isolines, *i.e.*, very small vertical gradients. Regarding differences in the mean AoA values themselves (bottom row in **Figure 5.28**) the largest relative differences with respect to ERA-Interim are found in the tropical lower stratosphere for all reanalyses, except for JRA-55 which shows the largest differences in the middle and upper stratosphere.

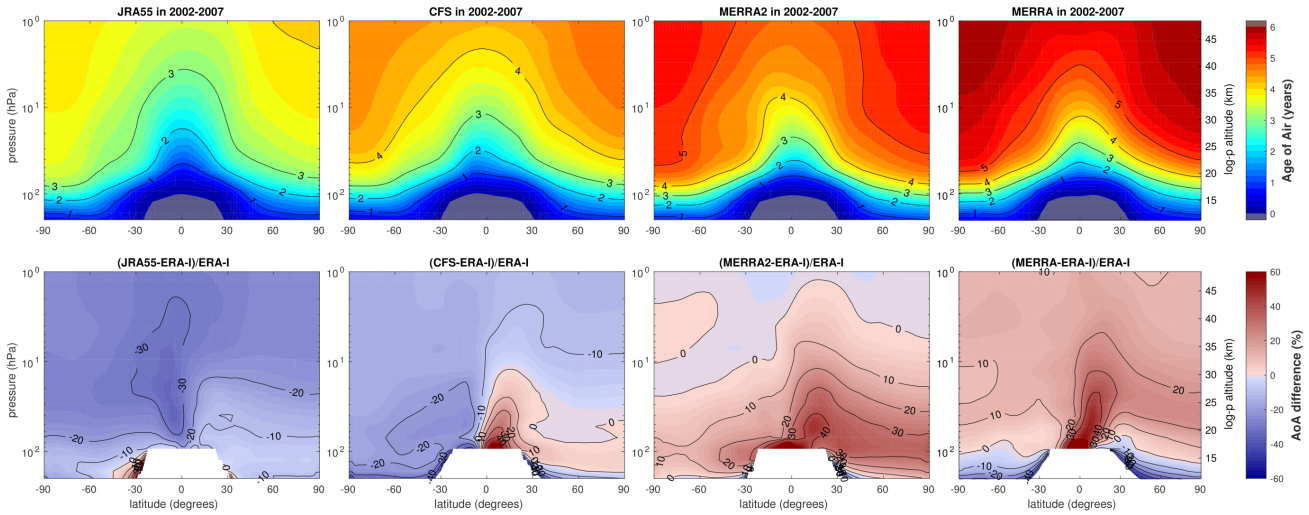

Figure 5.28: *Latitude-pressure distribution of mean AoA averaged over the period 2002 - 2007 from BASCOE simulations driven by all reanalyses but ERA-Interim (top row). The reanalyses are, from left to right, JRA-55, CFSR, MERRA-2 and MERRA. The bottom row shows corresponding relative differences with respect to the mean AoA from the ERA-Interim-driven simulation in* ***Figure 5.26****; darker blue indicates more negative differences and darker red more positive differences. Figure from Chabrillat et al. (2018).*

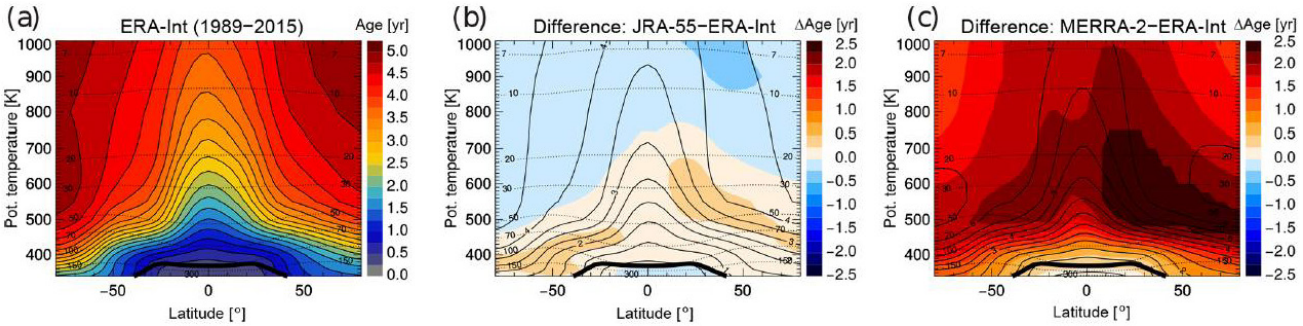

Figure 5.29: *Mean age climatology (1989 - 2015) obtained from the CLaMS simulation with for ERA-Interim (left), and differences for the corresponding simulations with JRA-55 (middle), and MERRA-2 (right). Thin solid black lines highlight particular mean age contours, thin dashed black lines show pressure levels in hPa, and the thick black line is the (lapse rate) tropopause (calculated from each reanalysis following WMO, 1957). Figure from Ploeger et al. (2019).*

JRA-55 is younger than ERA-Interim for all latitudes and altitudes, while MERRA-2 is older than ERA-Interim for all regions. MERRA and MERRA-2 exhibit a similar differences pattern but MERRA show younger values than ERA-Interim in the LS extratropics. The overall classification of mean AOA values from BASCOE simulations for the period 2002 - 2007 also holds for the whole 1989 - 2015 period (as discussed later in *Section 5.5.2.5*): MERRA and MERRA-2 result in the oldest mean AoA, JRA-55 and CFSR the youngest, ERA-Interim lays in between.

Figure 5.29 shows the climatological average for the zonal mean AoA obtained with the CLaMS offline model with ERA-Interim, and the differences for the corresponding simulations with JRA-55 and MERRA-2, for the period 1989 - 2015. For the CLaMS simulations, JRA-55 is older than ERA-Interim in the LS region for all latitudes (as opposed to BASCOE results), and younger above 700 K (as found with BASCOE). The mean age from MERRA-2 is older than ERA-Interim by more than two years throughout most of the stratosphere (similarly to BASCOE results, but with CLaMS we see differences larger than two years).

Figure 5.30 shows the zonally averaged mean AoA for December to February (DJF) and June to August (JJA) seasons as obtained with the CLaMS model; these CLaMS simulations compare results obtained with ERA-Interim, JRA-55 and MERRA-2 and are averaged over the period 1980 - 2015. The global patterns in the mean age distribution are robust for the three reanalyses. However, the exact mean age values are sensitive to the dataset used. Overall, JRA-55 shows the youngest stratospheric mean age, MERRA-2 the oldest mean age, and ERA-Interim is in between. In particular, MERRA-2 shows the largest differences compared to the other two reanalyses, with mean AoA values about two years older in many regions of the stratosphere, consistent with the corresponding differences in the annual climatology (**Figure 5.29**).

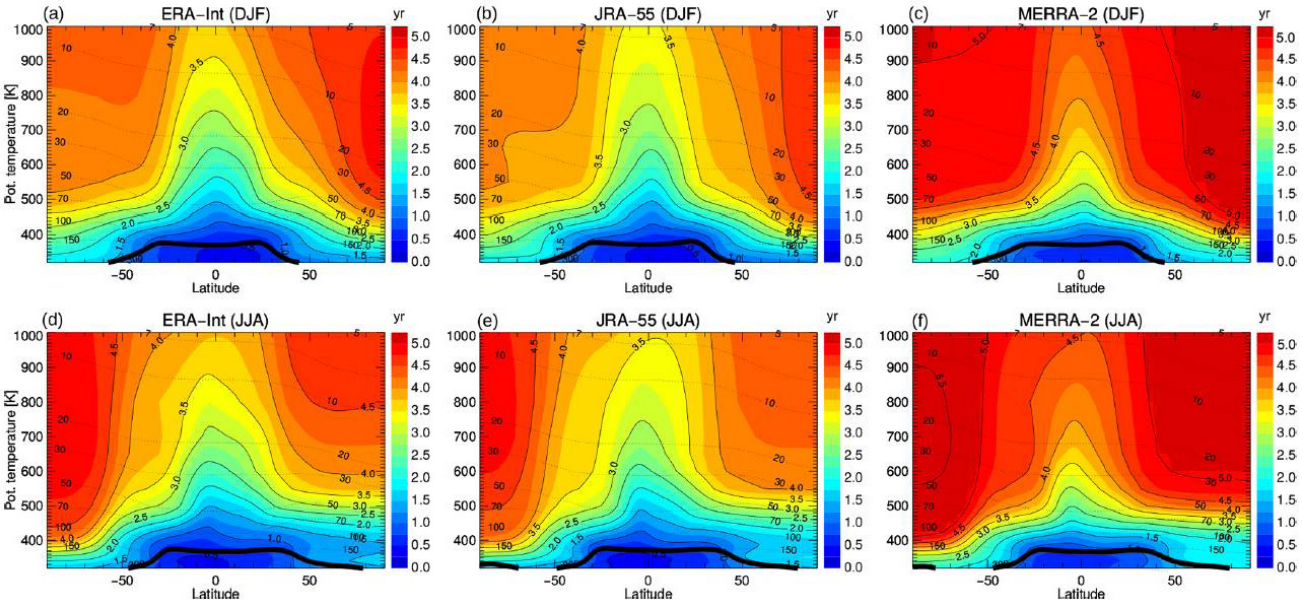

Figure 5.30: *Mean age climatology (1989 - 2015) for DJF (top) and JJA (bottom) for ERA-Interim (left), JRA-55 (middle), and MERRA-2 (right). Thin solid black lines highlight particular mean age contours, thin dashed black lines show pressure levels in hPa, and the thick black line is the (lapse rate) tropopause (calculated from each reanalysis following WMO, 1957). Figure from Ploeger et al. (2019).*

These differences in mean AoA are consistent with differences in the diabatic heating rate fields that drive the vertical transport in CLaMS; heating rates in **Figure 5.20** showed a stronger tropical upwelling for JRA-55 than for ERA-Interim for the period 1989 onwards, and a weaker tropical upwelling for MERRA-2. Despite the different ways in which both transport models operate, the classification of older and younger reanalysis obtained with CLaMS agrees with that obtained from BASCOE, which provides robustness to this classification result. The only exception is the difference between JRA-55 and ERA-Interim in the lower stratosphere.

Simulations with the TRACZILLA Lagrangian model also confirm this overall classification of mean AoA values obtained with these reanalysis datastes (**Figure 5.31**). TRACZILLA, as CLaMS, calculates the mean AoA values from the age spectrum distributions. **Figure 5.31** displays the zonal-mean mean AoA obtained by TRACZILLA when using ERA-Interim, MERRA and JRA-55 reanalyses. The youngest values correspond to JRA-55, MERRA shows the oldest values and ERA-Interim is in between but much closer to JRA-55 than to MERRA. The figure also shows a comparison of the effect of mass correction and of different correction techniques applied to the tail of the age spectrum for the different datasets.

The effect of the mass correction is smaller for ERA-Interim than for the other two reanalysis's, and for this dataset it acts making mean AoA values overall younger; a similar effect is true for JRA-55. However, in the case of the MERRA dataset, applying the mass correction makes mean AoA overall older. The clipping techniques have an effect on the mean AoA which to a large extent depends on the tail of age distribution. The slope of this distribution is much flatter for MERRA than for ERA-Interim and JRA55, which makes MERRA the most sensitive dataset to these corrections.

In general, uncorrected ages not accounting for the tail (first row in **Figure 5.31**) are too young, *e.g.,* for ERA-Interim the tail correction accounts for an increase in AoA of up to 1.8 years in polar regions. Compared to observation-based AoA distributions (*e.g., Section 5.5.2.2* above), the uncorrected and unclipped ages are too young for all reanalysis but the effect of applying tail correction varies according to the chosen clipping in a different reanalyses. The strong interplay between mass correction and clipping in the ERA-Interim suggests that uniform mass correction is probably inappropriate for ERA-Interim.

From these mean AoA distributions, MERRA data seem to provide much weaker tropical upwelling than the other reanalyses used by TRACZILLA. When applying the tail clipping correction techniques, MERRA provides an unrealistically old AoA compared to the other reanalyses.

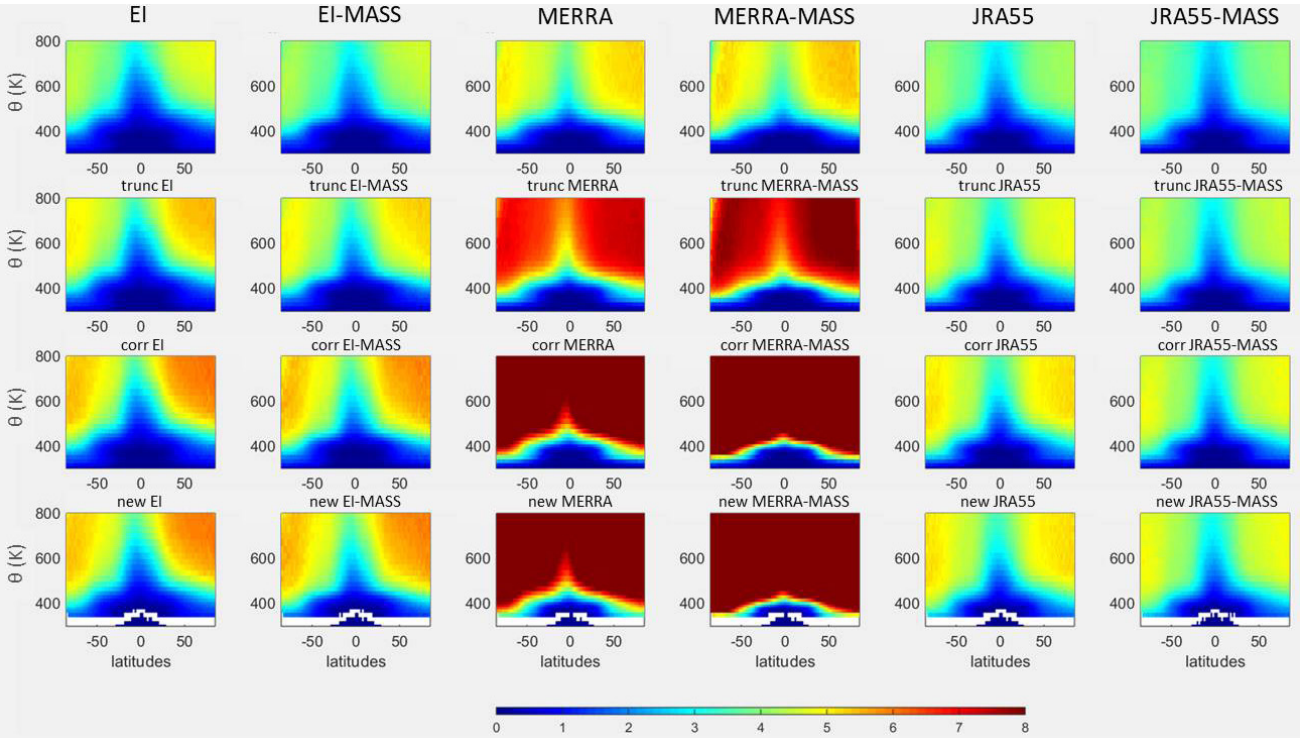

Figure 5.31: *Comparison of the mean AoA for ERA-Interim (two left columns), MERRA (two middle columns) and JRA-55 (two right columns) from TRACZILLA simulations for the period 1989 - 2010. For each reanalysis the left column shows the case without mass correction, and the right column shows the case with mass correction. The first row is without tail correction, the second row uses a correction by setting all the old parcels to 10 years. The third and the fourth rows are two different versions of the spectrum tail correction applied according to Scheele (2005), more details in the main text.*

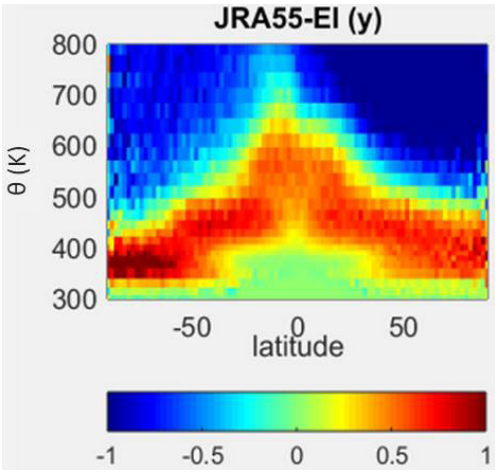

Figure 5.32: *Difference in mean AoA from the TRACZILLA run with JRA-55 minus the run with ERA-Interim. The configuration is that giving the best choice for ERA-Interim, i.e., with tail correction, clipping at 0.5 hPa, and no mass correction.*

The results from these TRACZILLA simulations clearly evidence that the same correction technique that makes one particular reanalysis dataset perform more realistically may not have the same effect with a different reanalysis and, therefore, such correction techniques need to be applied with caution.

Using the combination of corrections and clipping that gives the best results with ERA-Interim, **Figure 5.32** shows the difference in mean age-of-air from the TRACZILLA runs with JRA-55 and ERA-Interim. In agreement with CLaMS results (**Figure 5.29b**), JRA-55 is older than ERA-Interim in the lower tropical stratosphere and the extratropical lowermost stratosphere while it is younger at higher altitudes, especially in the NH. This suggests that the BDC favours the lower branch in ERA-Interim and the upper branch in JRA-55.

5.5.2.4 Age spectrum

The enormous advantage of the mean AoA diagnostic is the possibility of comparing it with actual tracers observations; however, a more complete picture of stratospheric transport in models can be obtained from the age spectrum diagnostic. Age spectrum distributions have been computed by the CLaMS and the TRACZILLA Lagrangian models with ERA-Interim for the period 2000 - 2010, showing a remarkable agreement of the spectra between the two models (Figure not shown); although the median and mean ages were overall older in TRACZILLA than in CLaMS. This is due to differences in the tail distribution of ages, in agreement with the differences in mean AoA distributions shown earlier for these two models.

CLaMS has performed age spectrum calculations with ERA-Interim, JRA-55 and MERRA-2 for the period 1989 - 2013 (**Figure 5.33**). The 400 K isentrope has been chosen as a representative level for the shallow BDC branch, while the spectra at 600 K represent levels at which the deep BDC branch dominates. The CLaMS age spectra show similar variability between ERA-Interim, JRA-55 and MERRA-2. In particular, multiple peaks in lower stratospheric age spectra are a common and robust feature for the three reanalysis. Effects of mixing as shown by changes in the spectrum tail are more sensitive to the reanalysis data used. For MERRA-2, the transition between tropical and extratropical age spectra is less pronounced, for age values older than 2 years, indicating stronger exchange between tropics and middle latitudes in the LS region. This stronger exchange likely causes a stronger recirculation of extratropical older air masses into the tropics, resulting in the older AoA mean ages values shown in previous sections.

Age spectra results from TRACZILLA simulations, for the period 1979 - 2010, are displayed in **Figure 5.34** for ERA-Interim, JRA-55 and MERRA. The annual modulation shown is due to the more intense BDC in the NH winter. It is visible that the amplitude decays much faster as a function of age in MERRA but has also a flatter tail, similar to what CLaMS has found with MERRA-2 data (**Figure 5.33**). JRA-55 is the reanalysis with the strongest annual modulation of the spectrum in TRACZILLA simulations. With TRACZILLA the three reanalysis, but especially MERRA, display reduced age and reduced modulation of the cycle in years following the Pinatubo eruption in June 1991. This effect appears to propagate across most of the 1990's, although the post Pinatubo transition coincides with the introduction of AMSU satellite observations in the reanalyses, and both effects can be confused. It is worth noting that none of the reanalyses considered explicitly includes the effects of the Pinatubo aerosols injection.

For the same TRACZILLA simulations, **Figure 5.35** shows that the horizontal distribution of the spectrum displays fairly similar patterns in the three reanalyses in the lower stratosphere, but at higher altitudes the two-lobe pattern clearly exhibited by ERA-Interim and JRA-55 is replaced by a one-lobe pattern in MERRA. This is an indication of a more leaky tropical pipe in MERRA, which is consistent with the distribution of young air in the tropical region for this dataset (**Figure 5.31**).

5.5.2.5 Mean AoA time evolution

Time series of mean AoA in the middle stratosphere, averaged between 30 hPa and 5 hPa, are displayed in **Figure 5.36**. These have been obtained with BASCOE with the five reanalyses shown for the SH and the NH. This figure shows the large disagreements among the five reanalyses over the long-term period 1989 - 2015. In the SH, MERRA and MERRA-2 values decrease quickly until 1995 and increase after 2007 while ERA-Interim values follow an opposite pattern.

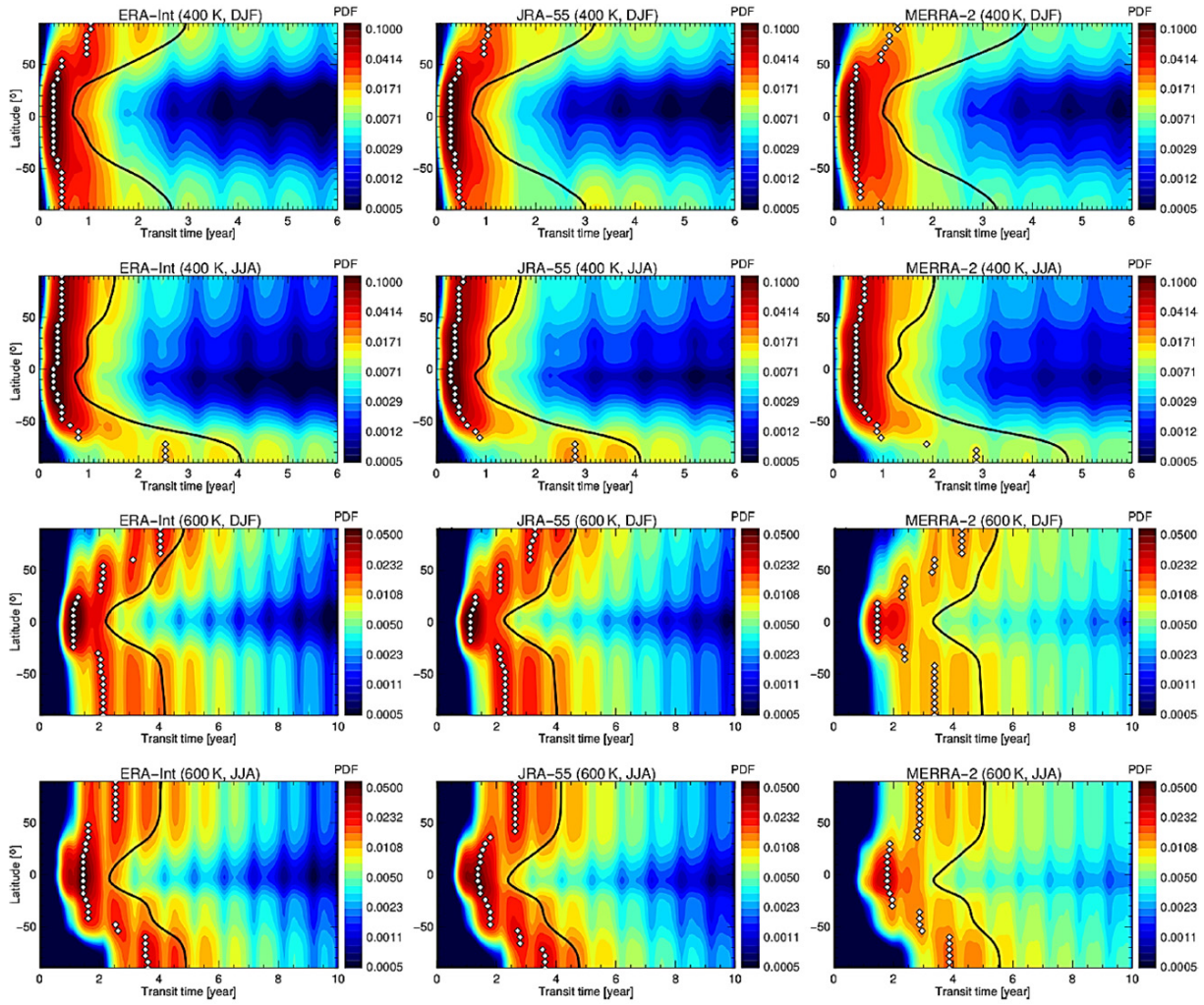

Figure 5.33: *Age spectrum from CLaMS simulations for 1989 - 2013 at 400 K (top two rows) for Dec-Feb (DJF) and Jun - Aug (JJA); and the same at the 600 K potential temperature isentrope (bottom two rows). Results correspond to the simulations using ERA-Interim (left), JRA-55 (middle), and MERRA-2 (right) reanalyses. The black line shows the mean of the AoA spectrum in each case, while the white symbols show the modal age. Figure adapted from Ploeger et al. (2019).*

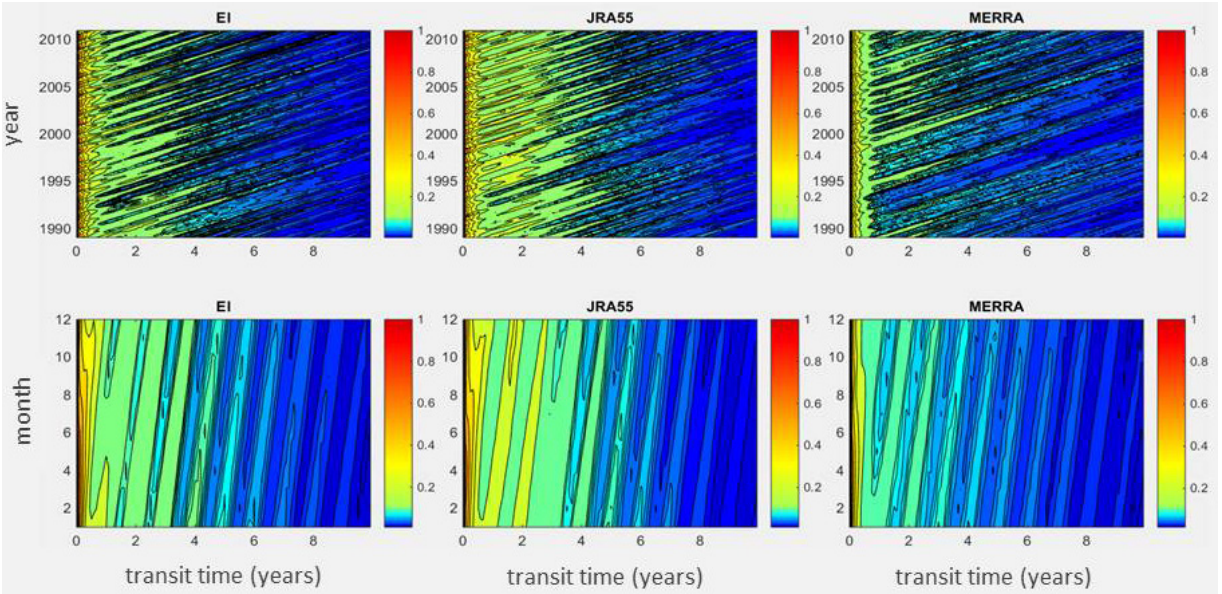

Figure 5.34: *Age spectrum average over the whole stratosphere below 800 K as obtained by TRACZILLA for the period 1989 - 2010 (upper panels) and the age spectrum mean annual cycle in the lower row. The reanalyses used, left to right, are ERA-Interim, JRA-55 and MERRA.*

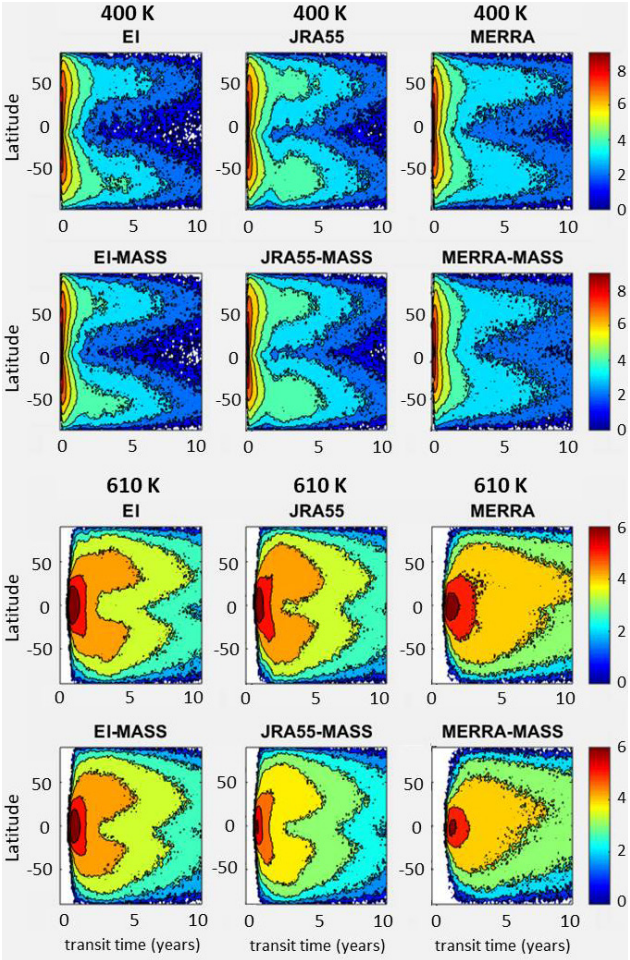

Figure 5.35: *Horizontal sections of the age spectrum as a function of latitude for two different levels: 400 K (top), and 610 K (bottom), as obtained from TRACZILLA simulations with ERA-Interim (left), JRA-55 (centre) and MERRA (right), without mass correction (upper row) and with mass correction (lower row).*

The long-term evolution of AoA in this region is very different with JRA-55, which shows a gradual decrease until 2002 followed by a slight recovery and stabilization after 2005, and differs also from CFSR, which shows no trend before 1997 and a rapid increase during 1997 - 2003.

Thin lines in **Figure 5.36** allow a qualitative comparison of faster variations in the five time series. The seasonal signal dominates in all cases, and all reanalyses show similar phases: AoA is older in autumn and younger in spring. The seasonal amplitudes in the SH are very dependent on the particular year but also on the considered reanalysis. It can be seen that some reanalyses, in particular MERRA and ERA-Interim, exhibit a stronger modulation of the seasonal cycle by the QBO than the others; for these two reanalyses the seasonal amplitude during easterly QBO years (*e.g.,* 2006, 2008) is half of that during westerly QBO years (*e.g.*, 2005, 2009). For the NH, **Figure 5.36** (right panel) compares the BASCOE model results with the balloon observations from *Engel et al.* (2009; 2017). The spread between the five simulations is as large as the observational uncertainties, highlighting again the magnitude of the disagreements between the five reanalyses. ERA-Interim delivers a small positive trend over the period 1989 - 2015, in agreement with the balloon observations.

5.5.2.6 Mean AoA trends

In the late 2000s, *Engel et al.* (2009), based on CO_2 and SF_6 observations, suggested that the widespread result from climate models predicting increasing strength of the BDC (younger mean AoA values) was not holding over the NH midlatitude stratosphere for recent past decades. **Figure 5.22** (updated from **Figure 3** in *Engel et al.*, 2009) shows the time evolution of mean AoA between 24 - 35 km altitude from SF_6 and CO_2 in-situ measurements from aircrafts and balloons taken from 1975. The *Engel et al.* (2009) study was based on sparse mean AoA observation-based values, and the trend obtained was not statistically significant compared to the observations' uncertainties.

But in 2012 new published studies gave robustness to this apparent discrepancy between climate models and observations. Based on MIPAS global satellite observations, *Stiller et al.* (2012) found a region in the middle stratosphere over NH midlatitudes where mean AoA trends were positive during the MIPAS period (2002 - 2012); this region coincided with the one considered in *Engel et al.* (2009). At the same time, *Monge-Sanz et al.* (2012) was the first model study to show the dipole structure in the mean AoA trend, using offline simulations of the Eulerian TOMCAT CTM driven by ERA-Interim reanalyses covering a 20-year period (1990 - 2009). This model study found a statistically significant (at 95 % confidence level) positive trend in the mean AoA between 25 - 40 km altitude over the NH, in overall agreement with the results derived from MIPAS observations by *Stiller et al.* (2012). A parallel study using the Lagrangian transport model TRACZILLA (*Diallo et al.*, 2012) also showed an heterogeneous structure in the mean AoA trend using ERA-Interim meteorological fields.

These early studies with ERA-Interim prompted an active research debate on the causes for discrepancies between observations and what climate models had been predicting. Increasing our knowledge on this issue has been one of the scientific objectives of the work done by the different model scientists involved in this SRIP Chapter. This section summarises results we have found when computing AoA trends with the different reanalyses.

Figure 5.37 is an updated version of **Figure 3** in *Monge-Sanz et al.* (2012), showing the zonal cross-section of the linear trend in the mean AoA from the offline TOMCAT simulations driven by ERA-Interim for the period 1990 - 2013. The dipole structure in the mean AoA displays maximum positive values over the NH middle stratosphere midlatitudes of up to 0.24 years/decade, and minimum values of up to - 0.14 years/decade. The figure shows that this trend over the NH

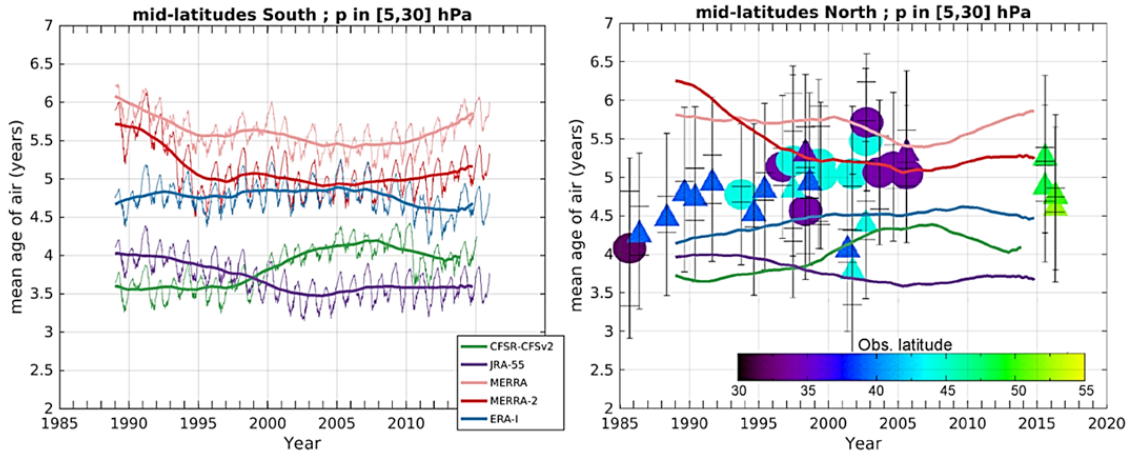

Figure 5.36: *Time evolution of AoA averaged from 30 hPa to 5 hPa (approximately 24 km to 36 km) in the southern (50° S - 40° S, left) and northern mid-latitudes (40° N - 50° N, right). Solid lines show model output with color codes according to the legend shown in the left panel. Thin lines (left panel only; omitted from right panel for clarity) show instantaneous model output every 5 days while thick lines are smoothed with a one-year running mean. Northern mid-latitude symbols (right panel) represent values derived from balloon observations of SF_6 (circles) and CO_2 (triangles) with color code showing the latitude of the measurements and outer error bars including sampling uncertainties (Engel et al., 2017). Adapted from Chabrillat et al. (2018).*

and SH middle stratosphere is statistically significant. The equivalent figure for the TOMCAT trend over the MIPAS period is displayed in **Figure 5.38**, which shows an intensification of the dipole with maximum values of up to + 0.50 years/decade over the NH and - 0.50 years/decade over the SH. This intensification in the AoA trend with ERA-Interim is consistent with the hypothesis in *Miyazaki et al.* (2016). They suggested that the increased eddy transport in the subtropics, and the weakened mean poleward motion in the middle stratosphere found with ERA-Interim during the period 2000 - 2012, would translate into larger increasing trends in the NH compared to the previous 20 years (1979 - 2000).

Figure 5.39 compares the latitude-pressure distributions of AoA trends across five reanalyses for the early (1989 - 2001), recent (2002 - 2015) and overall (1989 - 2015) periods as obtained from BASCOE simulations. It is important to note that the trends over the early and overall periods should be considered with more caution because of the beneficial impact of assimilation of new datasets in later years (*e.g.*, the AMSU dataset from 1998).

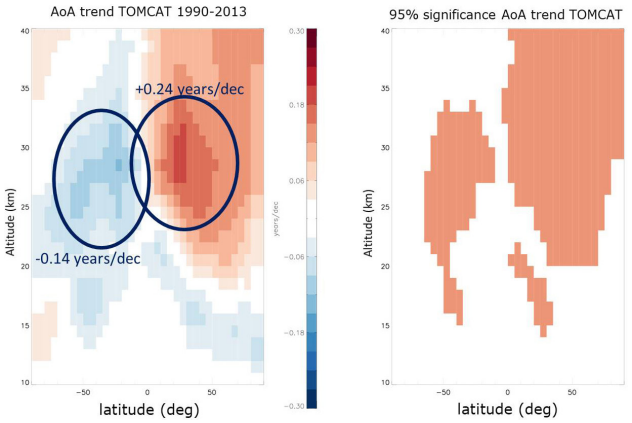

Figure 5.37: *Cross-section of the linear trend (years per decade) of the mean AoA for the period 1990 - 2013 from the TOMCAT simulation with ERAInterim fields (left); red colours indicate positive trends and blue colours negative trends. Regions where the trend is significant at least to the 95% confidence level are shown by the shaded areas in the right panel. The dipole structure in the mean AoA displays maximum positive values over the NH middle stratosphere midlatitudes of up to 0.24 years/decade, and minimum values of up to -0.14 years/decade. Updated from Monge-Sanz et al. (2012).*

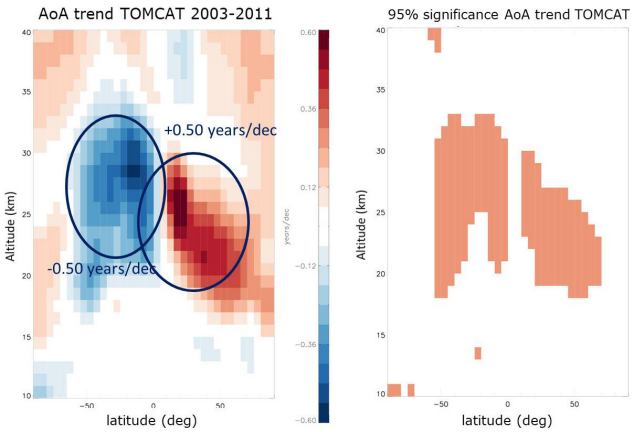

Figure 5.38: *Cross-section of the linear trend (years per decade) of the mean AoA for the period 2003 - 2011 from the TOMCAT simulation with ERA-Interim (left); red colours indicate positive trends and blue colours negative trends. Regions where the trend is significant at least to the 95 % confidence level are shown by the shaded areas in the right panel. The dipole structure in the mean AoA displays maximum positive values over the NH middle stratosphere midlatitudes of up to 0.50 years/decade, and minimum values of up to - 0.50 years/decade. (From Monge-Sanz et al., in prep).*

The AoA trends derived from ERA-Interim wind fields during the early period (upper left) show unexpected growth in both hemispheres, except in the northern lowermost stratosphere. During the recent period, the dipole structure derived from ERA-Interim (**Figure 5.39** upper middle) is similar to, but less clear than, over the shorter period 2002 - 2012 (**Figure 11** in *Chabrillat et al.*, 2018), with weaker increases in the NH which remain significant only in the polar lower stratosphere. The trend for the overall period 1989 - 2010 (**Figure 5.39** upper right) does not show a dipole structure but positive trends in the middle stratosphere, which are statistically significant over the NH region with positive trends during the 1989 - 2001 period, and significantly negative trends in the lowermost stratosphere at all latitudes (except the SH polar latitudes).

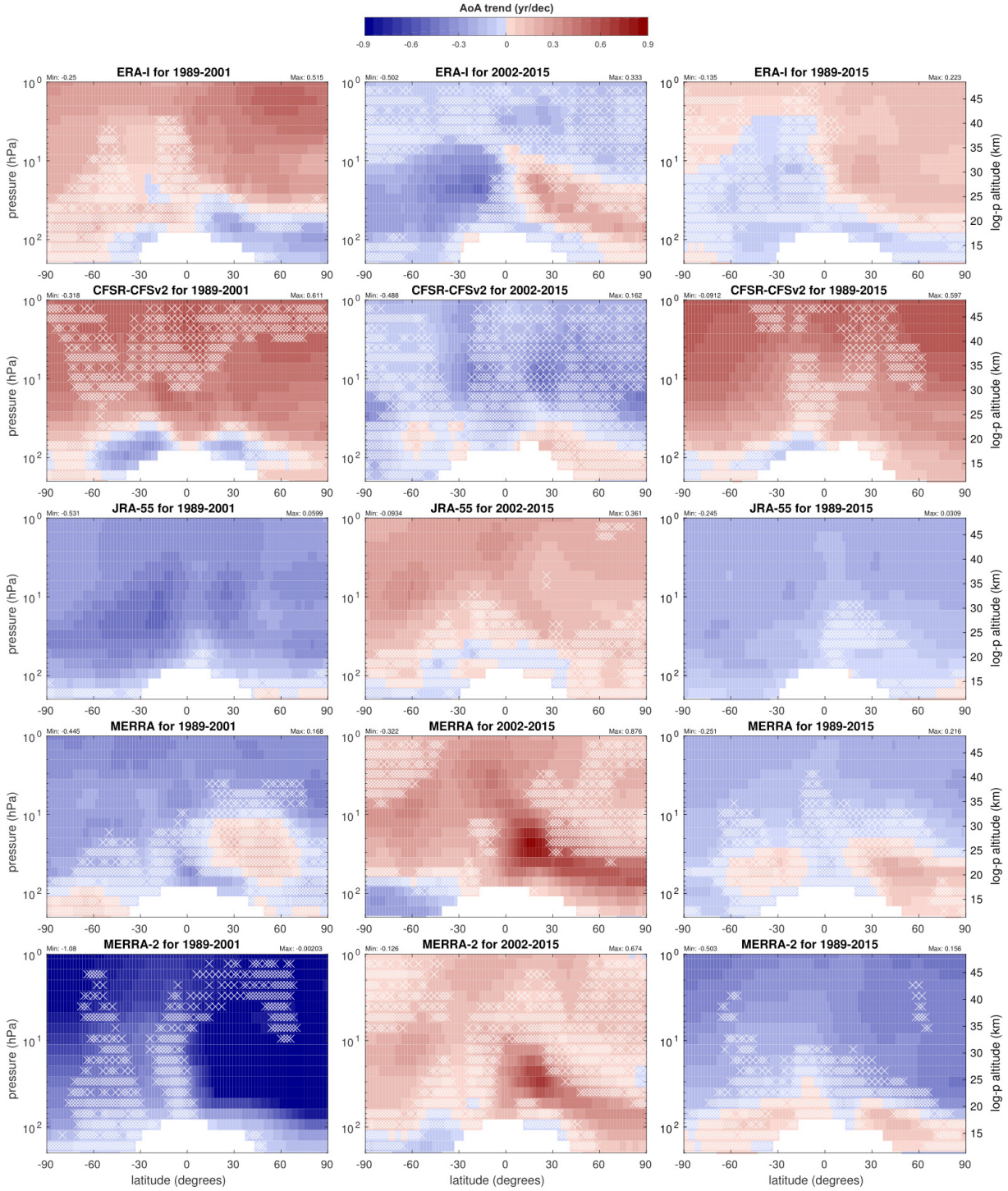

Figure 5.39: *Latitude-pressure distributions of AoA trends (years/decade) over 1989 - 2001 (left column), 2002 - 2015 (middle column) and 1989 - 2015 (right column) using the five reanalyses (from top to bottom: ERA-I, CFSR, JRA-55, MERRA, MERRA-2). White crosses indicate where the sign of the trend is not significant at the 95 % confidence level. Darker blues indicate more negative trends and darker reds more positive trends. Figure from Chabrillat et al. (2018).*

Diallo et al. (2012), using the diabatic Lagrangian transport model TRACZILLA driven by ERA-Interim for the period 1989 - 2010, found negative AoA trends in the lower stratosphere and positive trends in the mid-stratosphere, suggesting that the shallow and deep BDCs may be evolving in opposite ways. *Monge-Sanz et al.* (2012) with the Eulerian TOMCAT model showed significant positive trends over the NH middle stratosphere and negative trends in practically all other regions (see also **Figure 5.37**), although the negative trends were significant only in the LS region and the SH middle stratosphere. The BASCOE transport model simulations, using only wind fields and surface pressure from ERA-Interim, show a similar finding to the previous studies with TOMCAT and TRACZILLA for a similar period.

Comparing the BASCOE trend results obtained with ERA-Interim with those from other reanalyses, there is general agreement between ERA-Interim and CFSR (**Figure 5.39**, first and second rows) while JRA-55, MERRA and MERRA-2 (third to fifth rows in **Figure 5.39**) exhibit overall opposite trends for all periods. A remarkable result in **Figure 5.39** is the overall reversal of trends between the early (1989 - 2001) and recent (2002 - 2015) periods. This reversal is found for all five reanalyses in all regions of the stratosphere (first and second columns in **Figure 5.39**). This period separation for the AoA trend is in agreement with the findings of *Cook and Roscoe* (2009; 2012) for BDC trends over the Antarctic based on polar observations of NO_2.

For the early period, there is very good agreement between ERA-Interim and CFSR (**Figure 5.39**, first and second row) while MERRA shows almost exactly opposite trends, except in the LS where MERRA agrees with CFSR and ERA-Interim. Both JRA-55 and MERRA-2 show negative trends in the whole stratosphere for this period. During the recent period MERRA and MERRA-2 show good agreement. Therefore, the sign of the trend and their statistical significance strongly depends on the input reanalysis. ERA-Interim stands out as the only reanalysis showing a dipole structure in the mean AoA trend for the period 2002 - 2015, in overall agreement with trend values derived from observations. **Figure 5.39** also shows the strong dependence of the trend on the particular period considered, with values above 10 hPa varying between approximately -0.4 and 0.4 years per decade for the same reanalysis, within the same range of values of the interannual variability exhibited by the curves in **Figure 5.36**.

Figure 5.40 shows the linear trend of mean AoA derived from the KASIMA Eulerian model simulations with ERA-Interim for two periods, the overall period 1979 - 2012 (left panel) and the MIPAS period 2002 - 2012 (right panel). The linear trend has been obtained from an idealized linear tracer (T1) with a multi-linear regression analysis including additional annual and semi-annual harmonics and the two QBO indices (*Reddmann et al.*, 2001). The results for the overall period show a positive trend over the NH middle stratosphere of up to 0.3 years/decade, and no significant trend elsewhere. For the MIPAS period, the dipole structure emerges, with more confined positive trend values over the NH low and middle stratosphere between 20 - 30 km of altitude (up to 0.10 years/year) and a negative trend region over the SH low and middle stratosphere (up to - 0.10 years/year).

With the KASIMA simulations we can assess the impact of the mesospheric sink of SF_6 on mean AoA trends. The KASIMA model has used an additional SF_6 tracer (T3) that includes the effects of chemical loss as described in *Reddmann et al.* (2001). **Figure 5.41** shows the cross section of the mean AoA trend with ERA-Interim, for the overlapping MIPAS period 2002 - 2012, when including mesospheric SF6 chemical loss. The general pattern in the low to mid latitude stratosphere is preserved showing the dipole structure in the trend, between 20 - 30 km for both hemispheres, but especially in the SH high latitudes the trend is clearly affected by the chemical loss of SF_6.

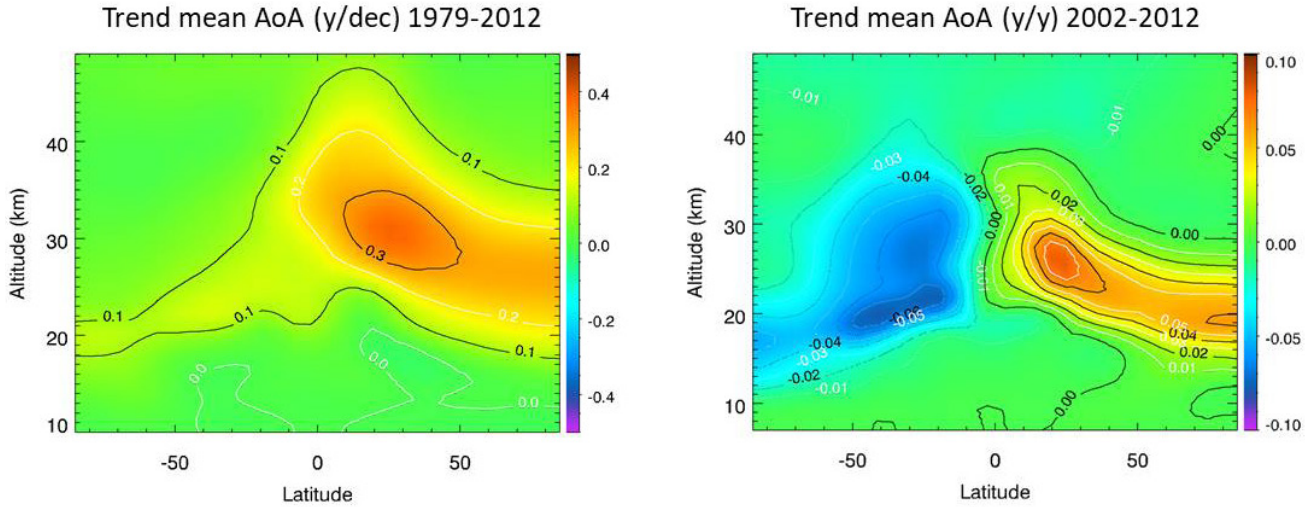

Figure 5.40: *Cross-section of the linear trend in the mean AoA from the linear tracer T1 from the KASIMA model simulation with ERA-Interim. Two different periods are shown: 1979 - 2012 (left) and the MIPAS period 2002 - 2012 (right). Note the different colour scales in both panels.*

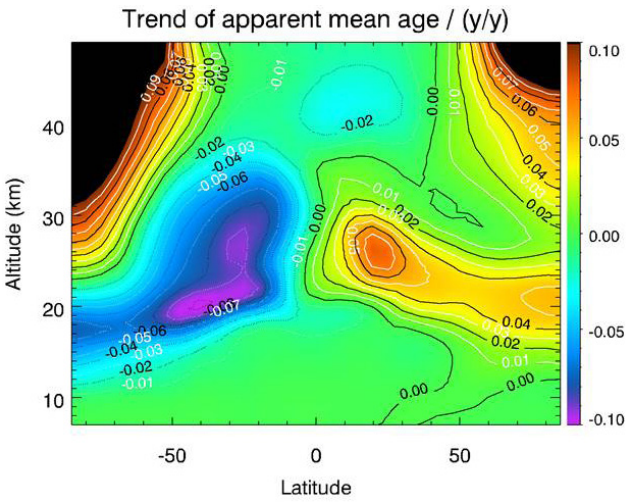

Figure 5.41: *Trend of the apparent mean AoA (expressed as the lag time) of the SF_6 tracer T3 in the KASIMA simulation with ERA-Interim (2002 - 2012), when including mesospheric loss for SF_6.*

The derived trend pattern agrees well with the results of the SF_6 trend features from MIPAS observations in the upper stratosphere (*Haenel et al.*, 2015; *Stiller et al., 2012*). Whereas tracer T3 provides the most realistic results from KASIMA's simulations compared with SF_6 observations, one needs to be cautious as the loss mechanism of SF_6 is subject to significant uncertainties.

Figure 5.42 shows the effect of mass-correction in the mean AoA trend values obtained with the TRACZILLA model driven by ERA-Interim, JRA-55 and MERRA renalayses. The priod 1989 - 2010 has been used in this simulations. There are large differences between reanalyses: for the non-corrected reanalyses fields, both ERA-Interim and JRA-55 show a decrease of mean AoA in the lower stratosphere, but ERA-Interim shows a positive trend in the higher levels in the extratropics, while JRA-55 shows a general negative trend in the middle and upper stratosphere. MERRA shows a similar pattern to JRA-55 but negative values are larger and an area of significant positive trends appears centred over 50 °N in the LS. This agrees with the overall trend structure found with BASCOE (**Figure 5.39** right column panels) when comparing these three reanalyses, however mean AoA trends from CLaMS simulations only agree with BASCOE and TRACZILLA trends for the ERA-Interim reanalysis (**Figure 7** in *Ploeger et al.*, 2019).

When using the mass-corrected TRACZILLA simulations, **Figure 5.42** shows a similar overall structure as with the non-corrected fields but differences are also evident: i) positive trends in ERA-Interim become stronger while negative trends in JRA-55 become weaker, and much weaker for MERRA; ii) for the three reanalyses results become non statistically significant in a larger area of the LS and the tropical pipe.

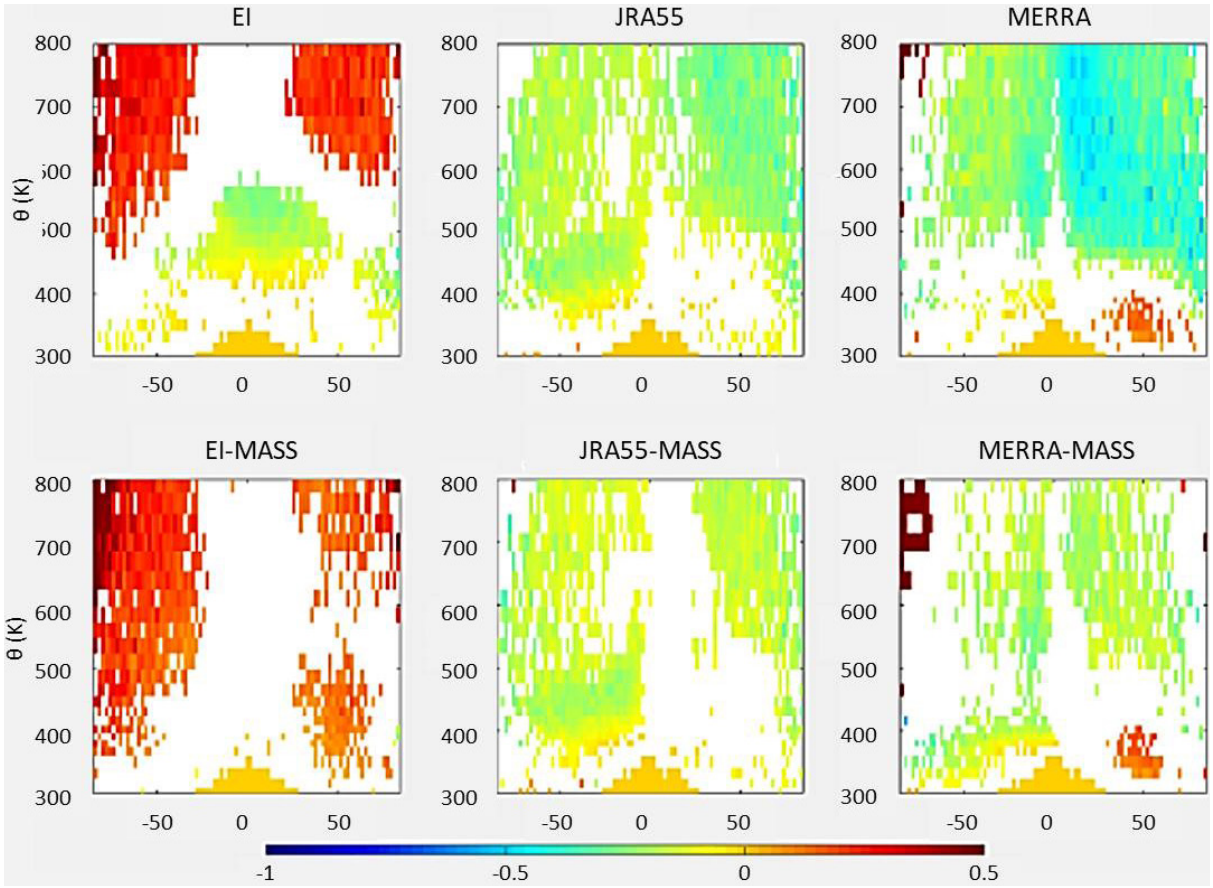

Figure 5.42: *Trends (years/decade) from the TRACZILLA simulations with ERA-Interim (left), JRA-55 (middle) and MERRA (right), with mass correction (bottom) and without (top) mass correction. They have been obtained over the period 1989-2010. Green and blue colours show negative trend values, orange and red colours show positive trends. Non-significant areas are white.*

In addition, an area of positive trend values appears for ERA-Interim around the 400 K isentropic level for NH mid-latitudes, similar to the one featured with MERRA. These differences show the strong effect the mass correction can have on the mean AoA diagnostic, and therefore on chemical tracers distributions obtained by CTMs driven by these reanalyses.

In this section we have shown that mean AoA trends are dependent, not only on the reanalysis used, but also on the exact dates used to calculate such trends; this is also in agreement with recent CCM model studies (*e.g.*, *Garfinkel et al.*, 2017; *Hardiman et al.*, 2017). However, a robust feature emerging from the previous trend distributions is also that, during the period covered by MIPAS observations, ERA-Interim simulations are in significantly better agreement with observations than simulations driven by the other reanalyses; and the trend observed during this period contributes to explain other observed trends in atmospheric tracers (*e.g.*, *Mahieu et al.*, 2014), which adds robustness to this feature.

In the offline tracer simulations that we have examined, as in the real atmosphere, the trends in mean AoA are due to the combined changes in mean-meridional circulation (MMC) and eddy mixing processes. A few recent studies have dealt with ways to quantify the separate contribution of both effects to mean AoA model distributions: *Garny et al.* (2014) quantified the effect of age by mixing in a climate model as the difference between the mean AoA distribution and the corresponding RCTT distribution; *Ploeger et al.,* (2015) do the same with CLaMS ERA-Interim simulations to quantify the two contributions (residual circulation and mixing) to the AoA trend for the MIPAS period (2002 - 2012); and *Miyazaki et al.,* (2016) performed a thorough comparison of MMC and eddy mixing in six reanalyses (ERA-Interim, JRA-55, CFSR, and their predecessor versions ERA-40, JRA-25 and NCEP) and discussed this comparison results also in the context of expected impacts on corresponding AoA distributions. Overall, for the periods and reanalyses they considered, *Miyazaki et al.* (2016) found more consistency among reanalyses regarding mixing processes than MMC.

5.5.2.7 Impact of other processes on the AoA

Quasi-Biennial Oscillation

A point on which the reanalysis strongly disagree is the amplitude and pattern of the correlation of AoA with the quasi-biennial oscillation (QBO).

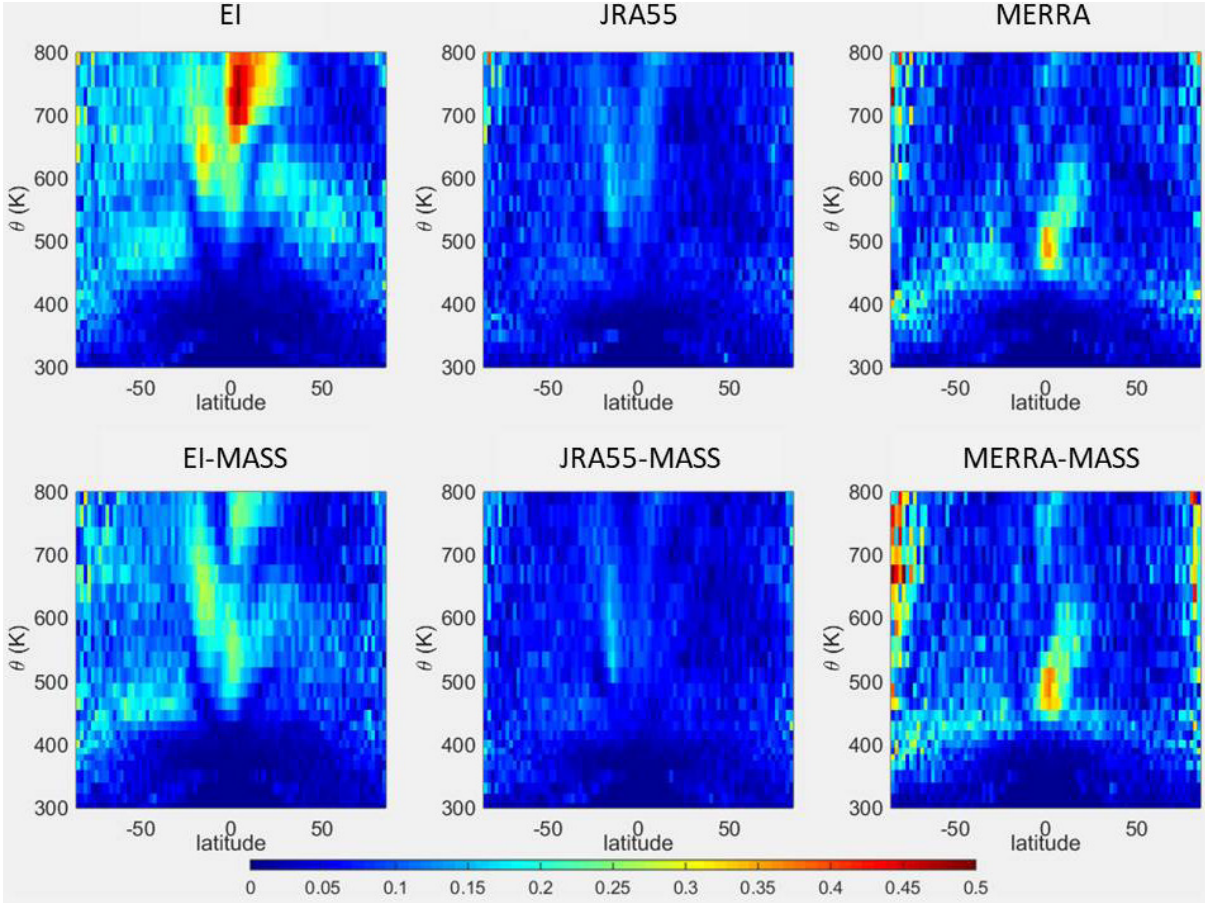

Figure 5.43: *Cross-sections of the amplitude of the correlation of the mean AoA with the QBO signal (defined at the 30 hPa level) in the TRACZILLA model for the ERA-Interim (left), JRA-55 (middle) and MERRA (right) reanalysis without (top) and with (bottom) mass correction. This simualtions cover the period 1989 - 2010.*

Figure 5.43 shows the correlation between the mean AoA and the QBO signal for the TRACZILLA simulations without mass correction (upper rows) and with mass correction (lower rows) for ERA-Interim, JRA-55 and MERRA. ERA-Interim and JRA-55 display approximately the same pattern; however, the amplitude is much stronger for ERA-Interim, reaching 0.5 correlation values over the tropical high stratosphere, while for JRA-55 correlation values stay between 0.0 - 0.2 for all locations. The QBO influence is stronger above the 600 K isentrope, especially over the NH. MERRA also shows a distinct tropical maximum for the correlation, stronger than for JRA-55 and weaker than for ERA-Interim, but the tropical maximum of MERRA is located between 450 K and 500 K at a much lower altitude than for the two other reanalysis. Applying the mass correction has little influence on the pattern of the correlation but reduces its amplitude, especially for ERA-Interim. It is worth noting that mass correction is more important for ERA-Interim than it is for MERRA or JRA-55 which are better balanced.

Results in **Figure 5.43** are not only due to the differences in AoA but they also point towards differences in the representation of the QBO signal among the different reanalyses. These results agree with other AoA studies looking into QBO effects on AoA, *e.g., Diallo et al.* (2012) or *Chabrillat et al.* (2018, **Figure 10** in their paper, and **Figure 5.36** in this Chapter). A full assessment of the QBO representation in all reanalyses can be found in *Chapter 9* of this Report.

Volcanic aerosols effects

The effects of increases in the stratospheric aerosol loading due to volcanic eruptions on the BDC has been estimated by CLaMS using modelled mean AoA and trends. For this estimation a multiple regression technique accounting for observed stratospheric aerosol has been used (calculation details in *Diallo et al.*, 2017). We have used observed stratospheric aerosol optic depth (AOD) timeseries averaged from 50 °S - 50 °N over the 1989 - 2012 time period for merged satellites datasets GISS, and SAGE II + GOMOS(525-nm) + CALIPSO (532-nm).

Figure 5.44 shows averaged timeseries of these stratospheric AOD satellite observations, deseasonalised mean AoA timeseries from CLaMS using ERA-Interim and JRA-55, and residual of the multiple linear regression with and without removal of the AOD signal. It can be seen that for both reanalyses there is a strong positive signal in the mean AoA following the Pinatubo eruption for both reanalyses. For the more recent extratropical volcanic eruptions after 2008, the signal is much smaller and the time lag from the eruption is longer. Therefore, a substantial contribution to decadal variability in the stratospheric circulation, as represented by variability in mean age of air, is caused by volcanic aerosol injections. As shown by *Diallo et al.* (2017), this mean AoA increase after a major volcanic eruption is significantly affected by corresponding induced mixing effects after the eruption. This increase we see in mean AoA is linked on the one hand to an increase in mixing, and on the other hand to a change in the upwelling strength at different levels. *Diallo et al.* (2017) also show that part of the mean AoA positive trend found over the NH for the recent past can be attributed to the minor volcanic eruptions that have taken place after 2008.

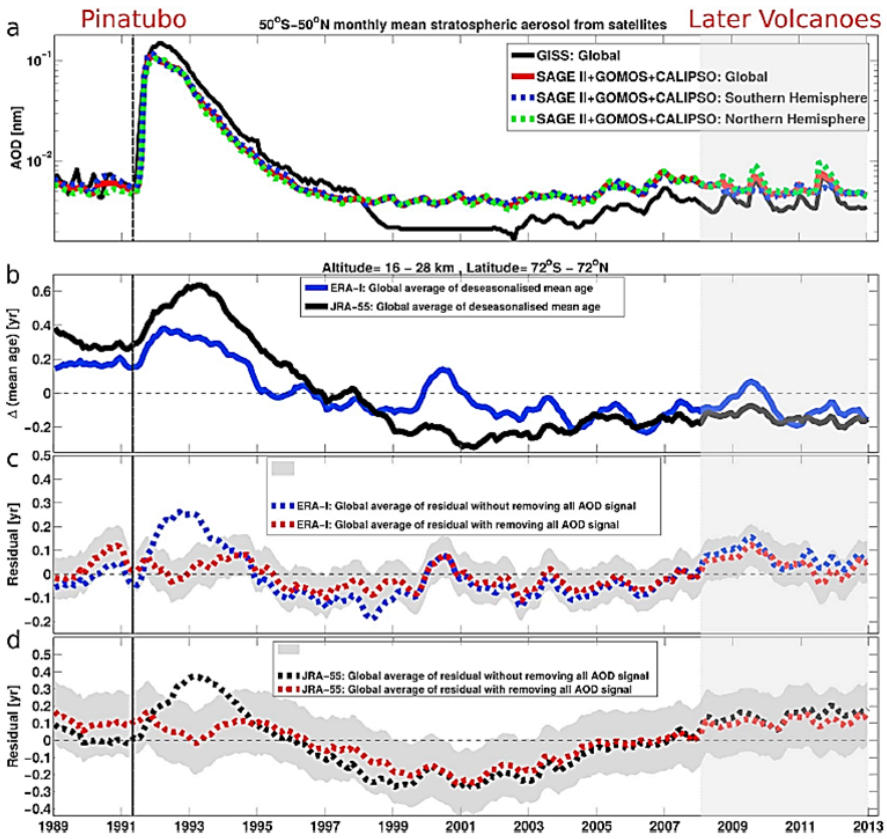

Figure 5.44: *Globally averaged timeseries of the stratospheric AOD, deseasonalised mean AoA and residual of the multiple linear regression with and without removal of all AOD signal. (a) Stratospheric AOD timeseries is averaged from 50° S - 50° N over the 1989 - 2012 time period and is shown for merged satellites datasets (GISS: black and SAGE II+GOMOS(525-nm)+CALIPSO (532-nm): red, blue and green). (b) The deseasonalised mean AoA driven by ERA-Interim and JRA-55 reanalyses is globally averaged between 72° S - 72° N and 16 - 28 km. (c, d) The residual of the multiple linear regression with (red-dashed line) and without (black-dashed line) removing the AOD signal from the deseasonalised mean age (b). The gray shading area indicates the standard deviation. Figure from Diallo et al. (2017). ©American Geophysical Union. Used with permission.*

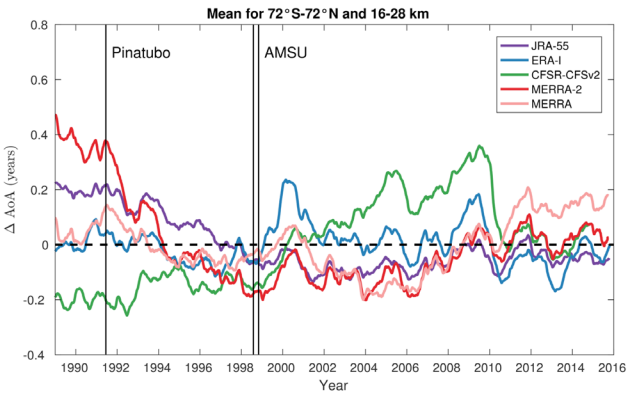

Figure 5.45: *Time evolution of the globally averaged (72° S - 72° N) anomalies of AoA with respect to their mean (1989 - 2015) annual cycles, between 16 km and 28 km, using the five reanalyses with same colour codes as in* ***Figure 5.36****. The black vertical lines highlight the start of the Pinatubo eruption and the first assimilation of AMSU data. From Chabrillat et al. (2018).*

Therefore, the representation of volcanic aerosols is an important element for reanalyses to correctly capture the time evolution of the stratospheric circulation.

Figure 5.45 shows the deseasonalized time series of mean AoA in the extra-polar LS, between 72° S-72° N and 16 - 28 km of altitude. The impact of the Pinatubo eruption is not evident in these BASCOE simulations, while *Diallo et al.* (2017) showed a very clear Pinatubo signal in the AoA time series from CLaMS simulations with ERA-Interim and JRA-55 (**Figure 5.44**). These differences between models can be partly explained by the fact that BASCOE is a kinematic model while CLaMS is a diabatic model. In BASCOE the vertical motion comes from the wind velocity fields while in CLaMS it comes from the diabatic heating rates. Since BASCOE was run in a purely advective mode, it did not take any temperature information from the reanalysis fields. Therefore, this comparison between BASCOE and CLaMS puts into evidence that for transport models to capture the signal from volcanic aerosols using reanalyses fields, radiative or temperature information is explicitely required from the reanalyses, as such signal is not fully present in the reanalyses wind fields. The comparison of results in **Figure 5.44** and **Figure 5.45** therefore shows that ERA-Interim and JRA-55 reanalyses include some volcanic aerosols information in the temperature field, but that wind fields do not contain sufficient information on volcanic signals. Also worth noting that future further investigation comparing volcanic responses in CCMs and CTMs will be needed, as some studies (*e.g.*, *Pitari et al.*, 2016; *Garfinkel et al.*, 2017) have shown different BDC volcanic response in CCM simulations compared to the offline simulations driven by the reanalyses we have considered. Future comparison assessments including CTM results with ERA5 will also be able to provide further information on the impacts of including volcanic aerosol forcing in the model used to produce the reanalysis.

ENSO signal effects

Using a multiple regression method applied to Aura MLS observations and CLaMS model simulations driven by ERA-Interim and JRA-55 reanalysis, we analyse the impact that the El Niño Southern Oscillation (ENSO) signal has on the BDC. **Figure 5.46** shows the zonal mean distribution of the ENSO impact on monthly-mean young and old air mass fractions from CLaMS simulations.

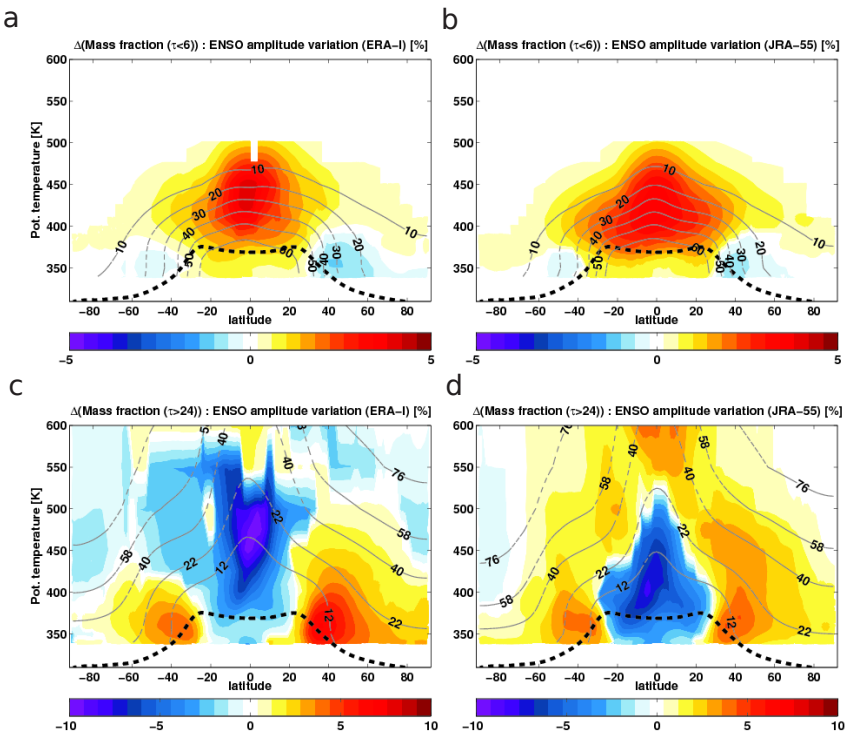

Figure 5.46: *Zonal mean distribution of the ENSO impact on monthly-mean young and old air mass fraction from CLaMS simulations driven by ERA-Interim (left column) and JRA-55 (right column) reanalyses. The amplitude of the air mass fraction variations attributed to ENSO is calculated by using MEI index from the multiple regression fit for the 1981 - 2013 period. (a, b) show the ENSO amplitude variation of the young air mass fraction with transit times shorter than 6 months. (c, d) show ENSO amplitude variation of the old air mass fraction with transit time longer than 24 months. Contours are the climatology values over the 1981 - 2013 period. The black dashed line indicates the tropopause location from reanalyses. Figure from Diallo et al. (2019).*

The amplitude of the air mass fraction variations attributed to ENSO is calculated by using the Multivariate ENSO index (MEI) from the multiple regression fit for the 1981 - 2013 period (*Wolter et al.*, 1998). The young air mass fraction is defined as that with transit times shorter than 6 months, while the old air mass fraction corresponds to transit times longer than 24 months. Looking into these two fractions gives information on the separate effect the ENSO has on the shallow and the deep branches of the BDC.

During El Niño conditions, the mass fraction of young air increases over the tropical lower stratosphere (up to 4 % increase) while there is a smaller decrease over the extratropical LS region. The structure and amplitude of these changes are in good agreement for ERA-Interim and JRA-55. The changes in the old air mass fraction (lower panels in **Figure 5.46**) show a strong decrease over the tropical tropopause, with a maximum decrease of up to 10 %, located between 450 K - 500 K for ERA-Interim, and of up to 7.5 % for JRA-55 located at lower altitude right above the tropopause at 400 K. The decrease region is much more confined for JRA-55 than for ERA-Interim. In the ECMWF reanalysis the effect of the ENSO signal makes the old-air mass fraction decrease also over middle and high latitudes above 450 K, while in JRA-55 the old-air mass fraction increases everywhere, except for the polar latitudes and the tropics below 500 K. In the extratropical LS region both reanalyses agree, showing regions where the mass fraction of old-air increases, especially over the NH subtropics and midlatitudes.

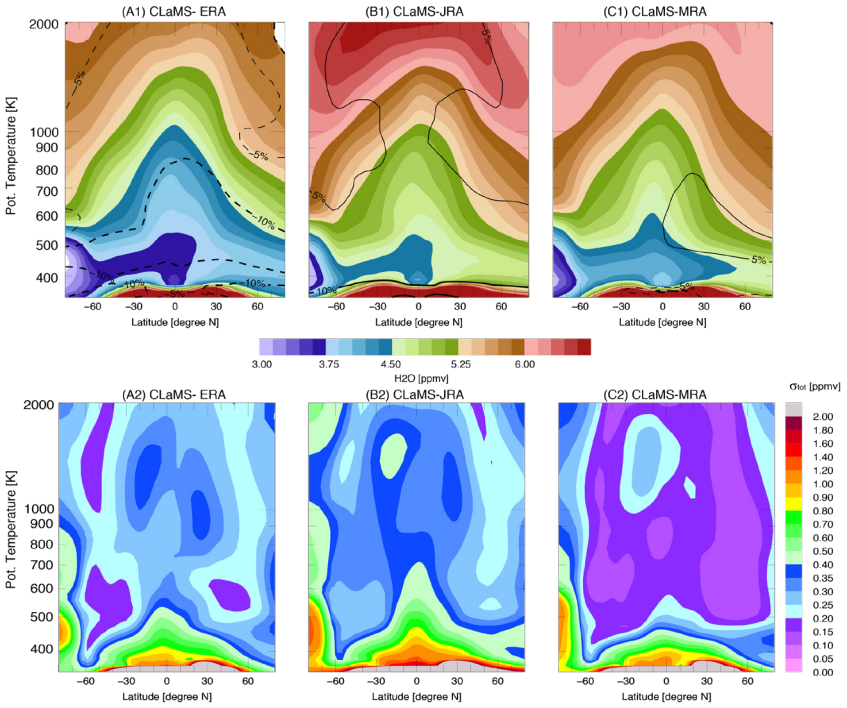

Figure 5.47: *The zonal and annual mean of water vapor (ppmv) from reanalysis-driven CLaMS simulations, averaged over the period 1980 - 2013 (top panel). In the bottom panel, the total variances (relative to the climatology) of respective monthly means are shown. The black contours show the differences of each CLaMS run relative to the means of (A1), (B1) and (C1). The reanalyses used are ERA-Interim (left), JRA-55 (middle) and MERRA-2 (right).*

The ENSO influence on the BDC for ERA-Interim and JRA-55 is more evident for the LS region, below 600 K (~ 24 km), thus it affects the transition and shallow circulation branches of the BDC. During El Niño, the transition branch weakens, while the shallow branch strengthens. Opposite changes occur during La Niña (not shown here). A detailed discussion of these ENSO effects can be found in *Diallo et al.* (2019). Similar patterns are found for ERA-Interim and JRA-55 but the intensity of the effects is different for each reanalysis.

5.5.2.8 Stratospheric water vapour tracer

The zonal annual mean of stratospheric water vapour (SWV) is shown in **Figure 5.47** for CLaMS simulations driven by ERA-Interim, JRA-55 and MERRA-2. These distributions of SWV have been obtained by averaging the model results over the period 1980 - 2013, the total variances with respect to the climatology are also shown in the figure. The overall structure of the climatological annual mean is well captured by the three reanalyses, however there are also several differences among the three simulations. The driest stratosphere corresponds to the simulation with ERA-Interim, the moistest one to JRA-55 and MERRA-2 shows values in between the other two reanalyses: overall, ERA-Interim is 0.75 ppmv drier than JRA-55 and 0.5 ppmv drier than MERRA-2 for all locations. The corresponding total variance distributions for the three simulations show a similar pattern structure, but the magnitude of the variance differs among reanalyses. JRA-55 shows the largest variances, MERRA-2 the lowest ones and ERA-Interim shows in between values more similar to JRA-55 in the NH and to MERRA-2 in the SH. The differences in SWV concentrations are not only due to differences in the stratospheric circulation but also to the entry rates through the TTL, hence to differences in TTL temperatures and mixing processes. From **Figure 5.47** one can see that ERA-Interim already shows the lowest SWV at the tropical tropopause.

Figure 5.48 shows the stratospheric tape-recorder signal based on SWOOSH SWV observations (top panel) and SWV values from the three CLaMS runs, averaged over 20 °N-20 °S for the period 1980 - 2013.

Upward propagation of the tape-recorder signal between 450 K and 600 K is 0.5 - 1.5 months faster in the ERA-Interim and JRA-55 simulations compared to SWOOSH, and the MERRA-2 simulation is 1 - 1.5 months slower than in SWOOSH. Similarly, the amplitude of the tape-recorder signal is systematically stronger than SWOOSH in the ERA-Interim and JRA-55 simulations, but weaker above 450 K in the one with MERRA-2. These differences are partly attributable to the slower upwelling in MERRA-2 (weaker heating rates as shown in **Figures 5.18** and **5.19**). Slower upwelling not only delays the propagation of the signal but also allows more time for horizontal advection and mixing of middle latitude air into the tropics, which tend to damp the signal.

Tao et al. (2019) also show the strong contribution of CH_4 oxidation in the CLaMS MERRA-2 run, indicated by the blue and red contour lines in **Figure 5.48**. This contribution to the tape-recorder signal is substantially larger than in the other two runs. This feature is a

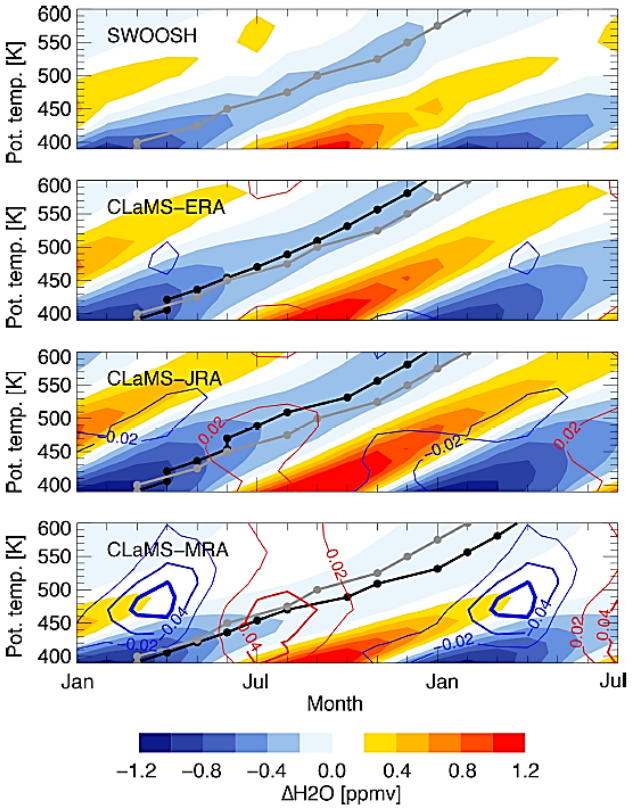

Figure 5.48: *Structure of the stratospheric tape-recorder signal based on SWOOSH observations (top panel) and the three CLaMS runs, averaged over the period 1980 - 2013. The tape-recorder is defined as anomalies in tropical (20° S-20° N) mean H_2O relative to the climatological mean at each level (color shading). The phase of upward propagation (solid black line and circles) is defined by the largest correlation with the layer below. For convenience, propagation based on SWOOSH is included in each panel (grey line). Red and blue contours indicate positive and negative contributions of CH_4 to H_2O anomalies (in units of ppmv, at intervals of 0.02 ppmv). Figure from Tao et al. (2019).*

secondary effect of the slow tropical upwelling (in addition to more in-mixing from the extratropics), resulting in a relatively pronounced seasonal cycle in H_2O/CH_4 in CLaMS driven by MERRA-2 with a maximum amplitude of 0.05 ppmv near the 450 K isentrope. The amplitude of H_2OCH_4 in the MERRA-2 run is twice as large as that in the JRA-55 one. The run with ERA-Interim on the other hand, shows virtually no anomalies in H_2O/CH_4 at these levels due to relatively rapid rates of ascent in the lower branch of the BDC.

Figure 5.49 shows the timeseries of the tropical anomalies (averaged between 10° N - 10° S) for water vapour at the 400 K level. Timeseries have been obtained from CLaMS simulations driven by ERA-Interim and JRA-55, and for the overlapping periods are also compared to satellite observations from Halogen Occultation Experiment (HALOE, *Harries et al.*, 1996) and from the Microwave Limb Sounder (MLS, *Waters et al.*, 2006). Both reanalysis products resolve well the subseasonal variability of H_2O fluctuations at the tropical tropopause. The variability on a time scale of 1-3 years (QBO; shaded regions in **Figure 5.49** correspond to easterly QBO phases), as well as on a time scale of 4 - 8 years (ENSO), is better represented with ERA-Interim, especially during the HALOE period (see also *Tao et al.*, 2015).

The lower panel in **Figure 5.49** shows the corresponding mean AoA anomalies in the CLaMS simulations. The decadal variability shows larger differences between ERA-Interim and JRA-55, both for water vapour and mean AoA; JRA-55 shows no trend along the 1979 - 2013 period, while ERA-Interim shows a negative trend for this tropical tropopause region. ENSO and stratospheric volcanic aerosols have been shown to modulate both the tropical ascending branch of the BDC (*e.g.*, **Figure 5.44**; *Diallo et al.*, 2017, 2019) and tropical tropopause temperatures (*e.g.*, *Holton and Gettelman*, 2001; *Mitchell et al.*, 2015), consequently affecting the distribution and evolution of SWV concentrations in the stratosphere.

5.6 Discussion

We have examined how well five modern reanalyses represent the stratospheric Brewer-Dobson circulation (BDC). For this, we have looked into dynamics diagnostics from the reanalyses data and into transport tracers from offline simulations driven by the reanalyses data. Results from both dynamics diagnostics and offline tracers show significant improvements in modern reanalyses compared to previous reanalysis products. This significant improvement in the representation of the BDC in recent reanalysis products reflects the fact that the corresponding agencies have been paying more continuous attention to improve the representation of stratospheric processes (*Fujiwara et al.*, 2017). Our results also show room for future improvement and need for further attention as we discuss later in this section.

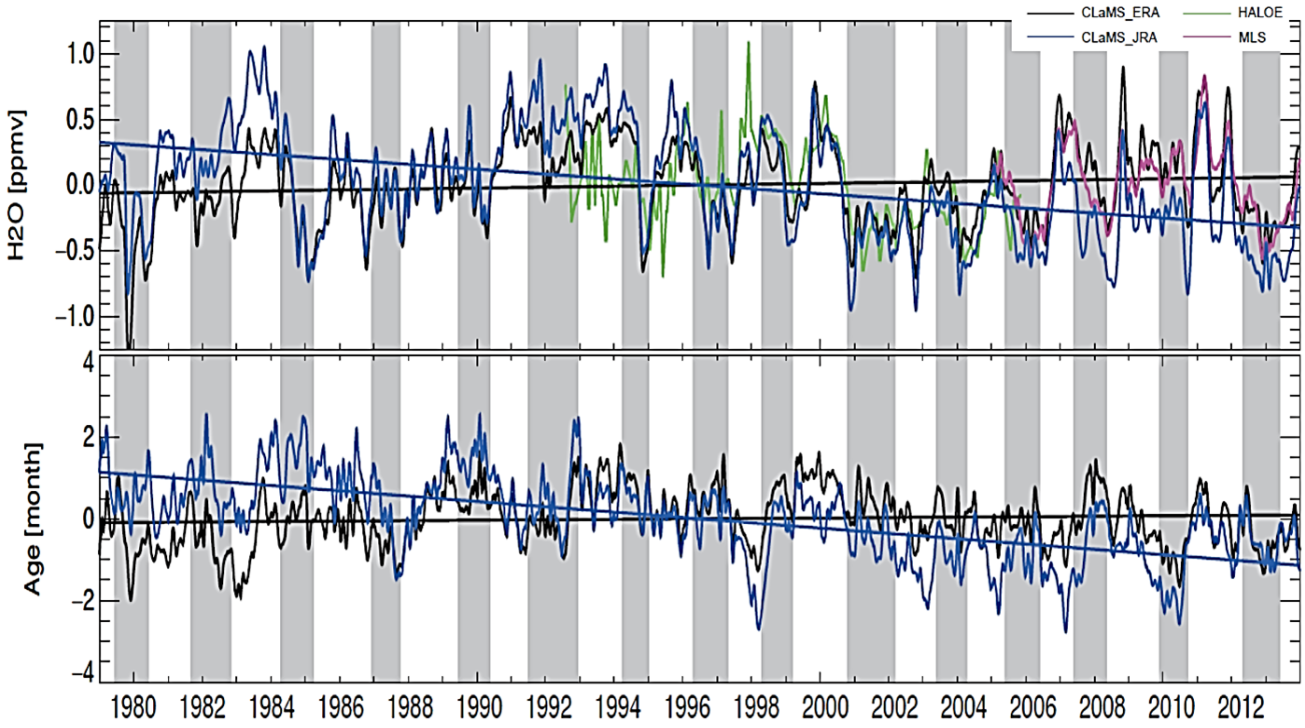

Figure 5.49: *Water vapour (ppmv) tropical anomalies timeseries at 400 K (upper panel) and mean AoA tropical anomalies at 400 K (lower panel) for the period 1979 - 2013 derived from CLaMS simulations driven by ERA-Interim (blue) and JRA-55 (black). Anomalies have been deseasonalised with respect to the 1979 - 2013 climatology and averaged over the tropics (10° N - 10° S) at the 400 K level. Satellite observations are also shown for HALOE (green curve) and MLS (magenta). The corresponding linear trends for the model results are also plotted (straight lines). Grey shading corersponds to the easterly phases of the QBO.*

Our dynamics diagnostics have shown close agreement in terms of climatologies for many derived metrics, such as total tropical upwelling (**Figure 5.8**), although some metrics still show strong disagreement even amongst the most recent products (*e.g.*, upwelling at the equator, **Figure 5.5**). Long-term trends in conventional metrics of BDC strength, such as tropical upwelling, still show disagreement across even the most modern products (**Figure 5.14**). reanalysis products tend to be best constrained in regions and for diagnostics that rely on fundamental balance relations, such as geostrophically balanced flow that couples wind and temperature fields. The mean meridional overturning circulation by definition uses the ageostrophic components of the flow and may therefore be viewed as more prone to uncertainties. In addition, mass conservation is not necessarily strictly fulfilled in reanalysis products due to data assimilation. Our results indicate that the more sophisticated data assimilation schemes employed by modern reanalysis products are less prone to such issues. Nevertheless, most aspects of climatological wave driving, as well as climatological circulation strength and structure are in close agreement (*e.g.*, **Figures 5.4**, **5.11**), especially among the most recent reanalysis products, for which older products showed larger spreads.

An important practical issue for end users of reanalysis products is the vertical resolution of the standard output in the region of the shallow BDC branch. In particular, at least one more output level between 100 hPa and 70 hPa, *i.e.*, the region of strongest vertical gradients in circulation strength, would be necessary to derive more meaningful diagnostics of the shallow BDC branch (such as "outwelling" strength). Note that all modern reanalysis products have at least one model level between 100 - 70 hPa.

Our results from offline simulations have shown that modern reanalyses produce mean AoA in much better agreement with observations than the previous generation of reanalyses (*e.g.*, ERA40). There are however remaining significant discrepancies among reanalyses, and differences with existing observations that imply there is still room for significant improvement in the way reanalyses represent the stratospheric BDC. This means that reanalyses have advanced significantly in the last decades and can still do so in coming ones.

In this Section we discuss possible causes for such discrepancies and point to aspects that need further attention in reanalyses to achieve further improvements in the representation of the BDC. To the extent possible, in the case of diagnostics obtained with CTMs, we also point to CTM model differences that can be causing differences in the results, but since this is not the scope of this Report we do this briefly and refer the interested reader to a more in-depth study we are conducting on this topic (*Monge-Sanz et al.*, in prep.).

All our offline model simulations show decreasing AoA values (strengthening BDC) in the LS region, in agreement with climate models.

However, our offline simulations depict a complex heterogeneous AoA trend in the stratosphere, in agreement with observations and not with most previous climate models studies. There is very good overall agreement between ERA-Interim and JRA-55 but they also show differences, especially in the representation of long-term trends. MERRA and MERRA-2 exhibit too slow vertical transport over the tropics, *e.g.*, as already reflected by the tropical upwelling diagnostic (**Figure 5.8**) and the diabatic heating rates (**Figures 5.18 - 5.20**). This is further shown by the tracer simulations with the offline models, both diabatic and kinematic ones (*e.g.*, **Figures 5.28** and **5.29**), which indicates that the slow BDC bias in the MERRA system is not only related to the radiation budget. The RCTT diagnostic also shows longer residence times for the MERRA datasets but to a much lower extent than the AoA differences, which means that aging by mixing also plays a significant role (*e.g.*, see **Figure 13** in *Ploeger et al.*, 2019). This fact points towards differences among reanalyses in mixing processes across latitudinal barriers (*Stiller et al.*, 2017; *Ploeger et al.*, 2015; *Garny et al.*, 2014).

The best overall agreement with mean AoA observation-based values, both for the climatological value and for trends, is shown using ERA-Interim (*e.g.*, **Figures 5.27**, **5.36)**. This reanalysis dataset is also the only one showing a dipole structure in the mean AoA trend obtained with offline simulations for the MIPAS period (*e.g.*, **Figure 5.38**, **5.40**). This dipole structure is in agreement with the MIPAS satellite observations we have used (*Haenel et al.*, 2015; *Stiller et al.*, 2012), and consistent with some studies explaining other observed tracers' recent past trends (*e.g., Mahieu et al.*, 2014). However we have also shown that AoA trends are very sensitive to the exact period considered and, therefore, future long-term global observations like MIPAS will be essential to understand the evolution of the BDC.

The volcanic signal is not equally present in all reanalyses, and in all simulations. In particular the comparison we have done between BASCOE and CLaMS simulations (**Figures 5.44**, **5.45**) highlights the fact that the volcanic information in the reanalyses is mainly contained in the temperature field, and not in the wind fields, which creates an unrealistic dynamical mismatch among different fields in one same reanalysis dataset. This result points towards the need of a more interactive representation of volcanic aerosols in the reanalyses. ERA5 includes a more realistic treatment of volcanic aerosols than previous reanalyses and it will be necessary to compare the results from offline simulations included in this Chapter with equivalent ones driven by ERA5 fields, to assess associated improvements in the BDC representation.

Here we summarise several possible causes for the discrepancies we have found among reanalyses, and therefore aspects that require further attention in future reanalyses:

Clouds and convection:

The different ways in which reanalyses include the radiative effects of clouds and the parameterisation of convection has also an impact on the tropical entry rates and tropical upwelling of the BDC. MERRA and MERRA-2 have strong cooling during summer in the TTL that tends to block transport, while in ERA-Interim diabatic motion is too fast due to the heating effect of cirrus clouds (*Tegtmeier et al.*, 2019). Deep convection also impacts the tropical UTLS wave activity and therefore the modelled BDC. A detailed comparison of clouds and convection treatment in all major reanalyses and their impact on the TTL is included in *Chapter 8* of this Report. Also a relevant study was conducted with several ECMWF reanalyses and operational analyses (*Feng et al.*, 2011) and should be further investigated with reanalyses from other Centres regarding their impact on wave activity and the BDC.

Gravity wave drag:

ERA-Interim, JRA-55 and CFSR all neglect non-orographic gravity wave drag (except for CFSv2, *i.e.*, CFSR after 2010) and each one uses its own parameterisation of orographic gravity wave drag. MERRA and MERRA-2 use the same parameterisation for orographic gravity wave drag and both take non-orographic gravity wave drag into account. In all the CTM studies we have shown here, MERRA and MERRA-2 provide significantly older AoA than the three other reanalyses. Different parameterisations of gravity wave drag are therefore a possible modelling cause for the disagreements in the stratospheric circulation diagnostics (*e.g.*, *Dharmalingam et al.*, 2019; *Podglajen et al.*, 2016). Since the recent ERA5 reanalysis includes non-orographic gravity wave drag, future comparisons using ERA5 driven simulations will provide further insight on related impacts on the representation of the BDC.

Heat budgets and radiation schemes:

Differences in heat budgets in the tropical region have substantial implications for the representation of transport and mixing in the LS region (*e.g.*, *Wright and Fueglistaler*, 2013). *Abalos et al.* (2015) evaluated the vertical component of the advective BDC in ERA-Interim, MERRA and JRA-55 and found large differences between direct (*i.e.*, kinematic) estimates and indirect estimates derived from the thermodynamic balance (*i.e.,* using diabatic heating rates). TRACZILLA and CLaMS simulations shown in this Chapter have used the reanalyses diabatic heating rates, and their differences in mean AoA are consistent with the differences in the diabatic heating rates fields. Younger AoA values are linked to larger diabatic heating rate values, and viceversa, and also the differences in the amplitude of the annual cycle in AoA follow the differences in the diabatic heating rates annual cycle shown in **Figure 5.18**.

However, the differences among reanalyses are also clearly displayed by offline simulations with kinematic transport models (*e.g.*, BASCOE), indicating that differences are not only coming from differences in the heating rates field. Different radiation schemes and treatment of stratospheric radiative species, as well as differences in the assimilated observations, produce differences in the reanalyses temperature field. Differences in temperature distribution and latitudinal gradients result in differences in the stratospheric wind fields. This will affect offline simulations of the BDC even for simulations that do not use the temperature field from reanalyses, *e.g.,* BASCOE kinematic simulations.

Ozone and Water vapour:

One reason why the temperature field differs among different reanalysis models and radiation codes is the different treatment of stratospheric ozone and water vapour. *Fueglistaler et al.* (2009) already showed that unrealistic or oversimplified ozone descriptions in the reanalysis systems lead to unrealistic radiative heating rates. *Chapter 4* in this Report and *Davis et al.* (2017) provide a thorough comparison of the ozone and water vapour distributions provided by the different reanalyses and gives an overview of the way these two components are treated in the different reanalyses radiation codes. ERA-Interim uses an ozone climatology, JRA-55 uses time-varying ozone fields from an external CCM and MERRA-2 uses interactive ozone. We recommend an assessment of the impacts that different ozone and water vapour modelling approaches in the reanalysis systems have on the representation of the stratospheric circulation. A study looking into how different treatments of stratospheric ozone impact stratosphere-troposphere processes in the ECMWF system has been recently carried out (*Monge-Sanz et al.*, 2020); extending this type of study to other major reanalysis systems would provide useful information.

Resolution and resolved mixing, and top of the model:

For differences in the results between reanalyses, we also need to keep in mind that the original grids of the reanalyses are different, and that interpolating to the CTMs' resolution has different numerical effects for each reanalysis. This will also affect mixing processes and their impact on mean AoA values differently for each reanalysis. Additionally, the altitude of the top of the model and the treatment of the top boundary sponge layer is different among reanalysis systems; this also has an effect on the BDC and on offline simulations for the stratosphere. Extending the altitude of the top of the model and including mesospheric processes into the reanalysis systems would improve the representation of the BDC. We also note that different top boundary conditions imposed in the offline models can be partly causing differences in the age-of-air values obtained with CLaMS (which imposes a top boundary condition to match MIPAS AoA values in the top level) and TRACZILLA (which uses removal of trajectories above a certain potential temperature level or an age limit to trajectories).

QBO representation:

The Quasi-Biennial Oscillation (QBO) signal is not equally captured and represented in the different reanalyses. Therefore the way the QBO links with, and influences the meridional circulation is different in each dataset. In our TRACZILLA simulations we have seen that the QBO correlation with the age-of-air diagnostic largely differs among datasets. These differences may well be linked to differences in the parameterization of non-orographic gravity wave drag. This deserves further investigation, especially in the case of MERRA-2, which shows difficulties representing correctly the QBO before 1995. A comprehensive analysis of the QBO representation in the different reanalyses is found in *Chapter 9* of this Report.

Volcanic influence:

How the different reanalyses capture the influence of large volcanic eruptions is linked to the different representation of aerosols, and to what information goes into the assimilated fields. In ERA-Interim and JRA-55 the effects of stratospheric volcanic aerosol are only included by the assimilation of observed temperature and wind data, as discussed in more detail by *Diallo et al.* (2017), whereas MERRA-2 additionally assimilates aerosol optical depth (*Fujiwara et al.*, 2017). Our offline simulations have shown that the analysed temperature field contains information on the volcanic signal, but that wind fields do not carry enough information about this signal (*Chabrillat et al.*, 2018; *Diallo et al.*, 2017). This fact points to potential dynamical mismatches between temperature and winds in the reanalyses, probably due to high assimilation increment values associated to the volcanic eruption effects. In addition, persistent imbalance will generate spurious gravity waves that artificially strengthen the BDC in the models.

To quantify how much each of these differences contributes to the discrepancy among reanalyses, and how much it contributes to disagreement with observations, tailored experimental datasets from Reanalyses Centres would be needed that do not exist at present. For the ECMWF reanalysis system, one study was conducted using tailored datasets to evaluate different aspects of the Data Assimilation system (assimilation window length, assimilation technique) and the model resolution (*Monge-Sanz et al.*, 2012).

Apart from the processes we have discussed above, there are of course other major processes, in different parts of the Earth System, that influence the BDC, including the ENSO signal or the stratospheric polar dehydration (*Chapter 10* in this Report). And we need to keep in mind that all the mentioned processes actually interact with each other, some of the interaction mechanisms are known while others are still a matter of international investigation efforts. In order to achieve a BDC representation that is more realistic, Reanalysis Centres and models will need to continue to move to a representation of the Earth System that is more complete and more coupled in coming years.

5.7 Conclusions and recommendations

In this Chapter we have analysed different diagnostics for the Brewer-Dobson circulation (BDC) for major reanalyses participating in S-RIP.

We have performed a direct comparison of dynamical diagnostics from the reanalyses datasets, including EP-flux divergence, tropical upwelling and outwelling, and residual circulation trajectories (RCTTs). We have also performed transport tracers simulations with different offline chemistry-transport models (CTMs) driven by the reanalyses, and assessed distribution of several tracers, mean age-of-air (AoA) and age spectrum diagnostics.

5.7.1 Conclusions from dynamics diagnostics

The dynamical diagnostics indicate that the BDC is much more consistent in the more recent reanalysis products, with much reduced spread in the respective climatologies compared to the older products. Furthermore, the BDC is generally less strong in more recent products compared to their older versions. However, even these recent products show significant differences in basic climatological diagnostics in some fields (*e.g.*, shallow branch wave driving, tropical upwelling structure and seasonality, upwelling strength below 70 hPa). Nevertheless, for the dynamical diagnostics analysed here the reanalysis products also show overall remarkable agreement with current chemistry-climate models (CCMs).

Time series of annual mean tropical upwelling mass flux at 70 hPa, a common measure of BDC strength used in many modelling studies (*e.g.*, Butchart et al. 2010), show a fairly strong degree of co-variability amongst the recent products (correlation coefficients between 0.65 - 0.82), except for CFSR. This and time series of other dynamical diagnostics suggests spurious fluctuations in CFSR; this product should therefore not be used for long-term trend or interannual variability analyses (consistent with the transport diagnostics in *Section 5.7.2*, see below).

Although MERRA-2, JRA-55, ERA-Interim and ERA5 agree with regards to co-variability on interannual time scales, there is inconsistency with regards to their long-term trend estimates of tropical upwelling at 70 hPa. MERRA-2 and JRA-55 show acceleration, while ERA-Interim shows deceleration, and ERA5 does not show a statistically significant trend. This also holds true at other pressure levels throughout the tropical lower stratosphere. A similar picture emerges for the poleward mass transport through the turnaround latitudes ("tropical outwelling"), although ERA-Interim in this case does not show a statistically significant opposing trend to MERRA-2 and JRA-55 (which both show a long-term strengthening of the circulation). However, the co-variability on interannual time scales is strongly reduced for this metric compared to upwelling, with correlation coefficients only in the range 0.23 - 0.68 (ERA-Interim among the lowest values). This is perhaps due to large sensitivity to structural differences (including those due to GWD) and suggests that the shallow branch of the BDC is not well constrained, even in modern products.

The RCTT diagnostic offers an integrated view of the circulation strength, possibly more robust to inconsistencies and uncertainties amongst products. The global mean RCTT at 50 hPa, a common reference level used for AoA comparisons, does show a high degree of co-variability among modern products (correlation coefficients between 0.53 - 0.85), but also shows large offsets in total values especially in the 1980's. Long-term trend values in this metric qualitatively agree with those obtained from tropical upwelling, including the disagreement between ERA-Interim and MERRA-2/JRA-55. An inspection of the latitudinally and vertically resolved RCTT trends shows that, by and large, RCTTs decrease (consistent with acceleration of the BDC), except for some regions/data sets. The main exception to this general behaviour are the RCTT trends corresponding to the deep branch of the BDC in both ECMWF reanalyses (ERA-Interim, ERA5). However, even these ECMWF products show primarily negative RCTT trends in the lowermost stratosphere, consistent with a strengthening of the shallow branch of the BDC.

5.7.2 Conclusions from transport tracers simulations

Although the dynamical diagnostics allow a clear comparison among reanalyses, they cannot be compared against observed quantities. We have also performed transport tracers simulations with different offline chemistry-transport models (CTMs) driven by the reanalyses. These sets of simulations have allowed us to compare results against observation-based data for the mean age-of-air (AoA) and stratospheric water vapour (SWV). For these diagnostics we have compared mean distributions as well as time series and evolution of trends for the different reanalysis products.

Our comparison results have shown that recent reanalyses produce mean AoA in much better agreement with observations than the previous generation of reanalysis (*e.g.*, ERA-Interim v. ERA-40), showing the improvement achieved by the reanalysis systems in the representation of the BDC.

However significant discrepancies in AoA and tracers distribution among reanalyses still remain. The spread of AoA obtained with different reanalyses can be as large as among different CCMs (*e.g.*, *Orbe et al.*, 2020).

We have shown that differences in the heating rates field are evident among the reanalyses we have considered, with MERRA reanalyses particularly differing from the rest. Heating rates differences are a major factor affecting the offline simulations of stratospheric tracers with diabatic models. MERRA and MERRA-2 exhibit too slow vertical transport over the tropics, in agreement with the lower values they show for diabatic heating rates compared to the other reanalyses. But the slow tropical transport is shown both by diabatic and kinematic offline simulations, which indicates that the slow BDC bias in the MERRA system is not only related to the radiation budget. The RCTT diagnostic also shows longer residence times for the MERRA datasets.

We have devoted a significant part of the Chapter to quantify mean AoA trends in the stratosphere, to better understand to what extent reanalyses can be used to study changes in the BDC structure and strength. For the overall period (1989 - 2010) our offline results show large spread in values and sign of AoA trends, depending on the reanalysis and on the region of the stratosphere. For the MIPAS period (2002 - 2012) only ERA-Interim is in good agreement with the observed trends, independently of the offline model used. The positive trend in the mean AoA in the NH is a robust feature in our studies and is in agreement with other observed phenomena like HCl observed trends (*Mahieu et al.*, 2014).

Here we need to note that much investigation is still needed on BDC trends, and that trends should be interpreted with caution as many factors affect them, including natural variability and changes in the observation system of assimilated data that make them so sensitive to the particular period chosen (*e.g.*, *Chabrillat et al.*, 2018).

The large spread in AoA results among reanalyses indicate two main aspects: *i)* important differences among the underlying models in the different reanalyses systems, and *ii)* that assimilated observations are not providing a strong constraint for stratospheric transport in reanalyses. As we indicate below in *Section 5.7.3,* we strongly recommend reanalyses centres to invest in model development in order to further improve the representation of the BDC.

We have also discussed in Section *5.5.2.7* how the AoA diagnostic is affected by other Earth System phenomena, not only in the stratosphere like the QBO signal, but also the ENSO and the volcanic signals. This shows the need to include as many Earth System processes as possible in a realistic way to achieve a more accurate BDC representation in future reanalyses.

With one of our offline CTMs (CLaMS) we performed a comparison of SWV distribution using the water model tracer. In this case, the distributions obtained with the different reanalyses showed good overall consistency for climatology and variability in the CTM, but differences against independent observations.

5.7.3 Recommendations to reanalyses users

A summary of the usability of major reanalyses in terms of their representation of the BDC can be found in **Figure 5.50**, where we classify the performance of each reanalysis for the diagnostics we have considered in this Chapter, based on the results and discussions we have included in the above sections.

Although not all the diagnostics we have used can be evaluated against observations, we have decided to assign an evaluation score to all of them. Such value, for those that cannot be compared to observations, reflects their consistency with other processes and our current understanding of the BDC.

In the majority of cases our evaluation is that reanalyses are "suitable with limitations". Such limitations depend on the particular time periods, atmospheric regions and applications. For instance, MERRA-2 is likely not to be a good option for years before 1995. MERRA-2, compared to ERA-Interim and JRA-55, shows difficulties in representing the QBO before 1995 (*Chabrillat et al.*, 2018 and references therein). *Gelaro et al.* (2017) also describe several features in MERRA-2, not present in the other two reanalyses, that can affect stratospheric dynamics, and therefore BDC diagnostics, including the assimilation of Microwave Limb Sounder on the Aura satellite (Aura-MLS) temperatures, from 2004 onwards and above 5 hPa.

Among the recent reanalyses, only in the case of CFSR we have identified several issues that indicate that their use may be problematic for stratospheric BDC studies, especially related to interannular variability and long-term trends. For older reanalyses like ERA-40 and NCEP reanalyses, it had already been shown in numerous published studies that their representation of the BDC, and other stratospheric processes, is unrealistic and, therefore, we also discourage their use for stratospheric studies.

Whenever possible we generally recommend users not to restrict themselves to only one product when it comes to BDC studies. In particular for the period after 2000 a comparison between MERRA-2, JRA-55, and ERA-Interim, together with new products such as ERA5 and JRA-3Q, can help to distinguish robust from non-robust diagnostics results. We also recommend working with reanalyses data on model levels, not only for offline simulations, but also for diagnostics related to the shallow BDC branch as usually no pressure levels are provided between 100 - 70 hPa.

5.7.4 Recommendations to reanalyses centres

From the results and experiences built along this study, this is our list of main recommendations for the development and data release of future reanalysis.

Regarding data availability:

- Provide variables' uncertainty information.
- Provide variables at higher vertical resolution, especially around the UTLS region.
- Provide pressure level data above 1 hPa (important for RCTT calculations).
- Archive data at higher frequencies.
- Archive additional relevant variables by default (e.g. heating rates).

The recently released ERA5 includes most of the above features, although the resolution around the UTLS is still lower than desired.

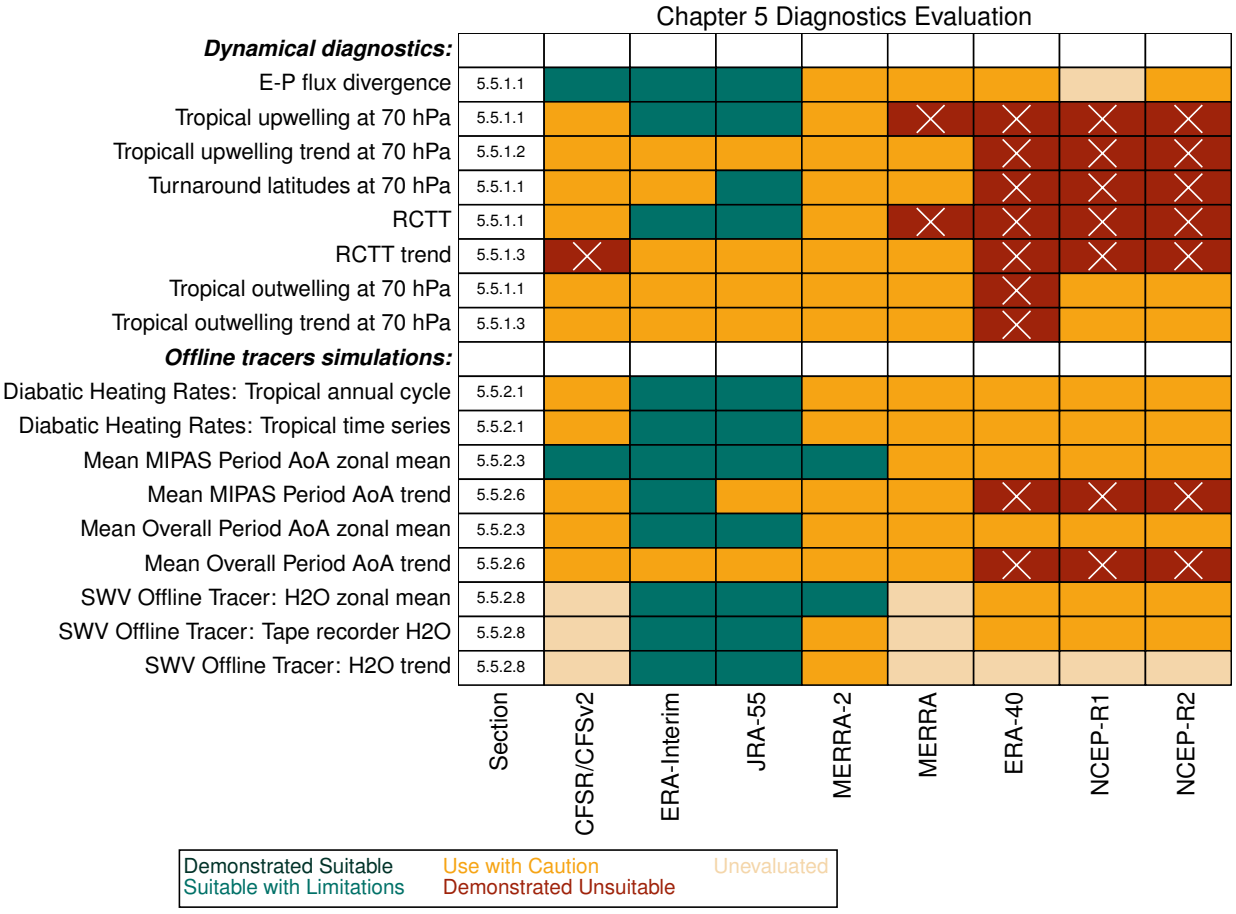

Figure 5.50: *Summary of the BDC diagnostics evaluation. Note that the score corresponding to "demonstrated suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table.*

From early experiences with ERA5, dealing with its huge volume of data requires improved postprocessing strategies and/or more computing/storage power. Interactive communication channels between reanalyses users and producers to improve sustainable solutions will likely become more important in the future as more high volume data products will be available.

Besides continuous assimilation of stratospheric winds as suitable datasets become available (*e.g.,* from the ESA's recent AEOLUS mission), model development stands out from our study as a major recommendation among the actions required to improve the representation of the stratospheric BDC in future reanalyses. Main model aspects that require attention are:

- Gravity wave drag parameterisations
- Representation of radiative gases and aerosols in the stratosphere
- Clouds and convection parameterisations
- Increase of the model resolution in the UTLS
- Extension of the vertical range to incorporate mesospheric processes.

Last but not least, sustained long-term relevant observations platforms are required to monitor any changes in the strength and the structure of the BDC and, therefore, to keep evaluating how well future reanalyses represent stratospheric major circulation patterns. We strongly advocate for the creation of such observation platforms and the necessity to keep them operative for long enough time periods to cover the relevant time scales to validate BDC evolution and trends.

Code & Data availability

Reanalysis data used in this chapter can be obtained from the corresponding reanalyses centres. Observations datasets and offline model data are available upon request and via the referenced publications. The dynamics diagnostics shown are based on the zonal mean dataset produced by *Martineau et al.* (2018) as referenced in the text; code is available upon request.

Acknowledgements

We thank the reviewers of this chapter, whose comments helped to improve the final revised version. BMS acknowledges partial funding from the UK Natural Environment Research Council (NERC) through the ACSIS project (North Atlantic Climate System Integrated Study) led by the National Centre for Atmospheric Science (NCAS); TB acknowledges funding by the National Science Foundation (grant no. AGS-1643167).

Figure 5.21 is adapted from Cook and Roscoe (2009). **Figure 5.22** is adapted from Engel et al. (2017). **Figures 5.23** and **5.24** are adapted from *Haenel et al.* (2015). **Figures 5.27, 5.28, 5.36, 5.39** and **5.45** are adapted from *Chabrillat et al.* (2018). *Figures 5.29, 5.30* and **5.33** are adapted from *Ploeger et al.* (2019). **Figure 5.46** is adapted from *Diallo et al.* (2019). **Figure 5.48** is adapted from *Tao et al.* (2019). All these reproductions are made under a creative commons attribution 3.0 or 4.0 license (**https://creativecommons.org/licenses/by/3.0/** or **https://crertivecommons.org/licenses/by/4.0/**, respectively).

References

Abalos, M., B. Legras, F. Ploeger, and W.J. Randel, 2015: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979–2012. *J. Geophys. Res.*, **120**, 7534 - 7554, doi: 10.1002/2015JD023182.

Abalos M., *et al.*, 2019: New insights on the impact of ozone depleting substances on the Brewer-Dobson circulation. *J. Geophys. Res: Atmospheres*, **124**, 2435 - 2451, doi: 10.1029/2018JD029301.

Andrews, D.G., J.R. Holton, and C.B. Leovy, 1987: Middle Atmosphere Dynamics, Academic Press, doi: 10.1017/S0016756800014333.

Andrews, A.E., *et al.*, 2001: Mean ages of stratospheric air derived from in situ observations of CO_2, CH_4, and N_2O. *J. Geophys. Res.*, **106** (D23), 32 295 - 32 314, doi: 10.1029/2001JD000465.

Birner, T. and H. Bönisch, 2011: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere. *Atmos. Chem. Phys.*, **11**, 817 - 827, doi: 10.5194/acp–11–817–2011.

Boering, K.A., *et al.*, 1996: Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide. *Science* **274**, 1340 - 1343 doi: 10.1126/science.274.5291.1340.

Brewer, A.W., 1949: Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Q. J. R. Meteorol. Soc.* **75**, 351 - 363, doi: 10.1002/qj.49707532603.

Butchart, N., *et al.*, 2006: Simulations of anthropogenic change in the strength of the Brewer–Dobson circulation. *Clim. Dyn.*, **27**, 727 - 741, doi: doi.org/10.1007/s00382-006-0162-4.

Butchart, N., *et al.*, 2010: Chemistry–Climate Model Simulations of Twenty-First Century Stratospheric Climate and Circulation Changes. *J. Climate*, **23**, 5349 - 5374,doi: 10.1175/2010JCLI3404.1

Butchart, N., 2014: The Brewer-Dobson circulation. *Rev. Geophys.*, **52**, 157 -184, doi: 10.1002/2013RG000448.

Chabrillat, S., *et al.*, 2018: Comparison of mean age of air in five reanalyses using the BASCOE transport model. *Atmos. Chem. Phys.*, **18**, 14715 - 14735, doi: 10.5194/acp-18-14715-2018.

Chipperfield, M.P., W. Feng, and M. Rex, M., 2005: Arctic ozone loss and climate sensitivity: Updated three-dimensional model study. *Geophys. Res. Lett.*, **32**, L11813, doi: 10.1029/2005GL022674.

Chipperfield, M.P., 2006: New version of the TOMCAT/SLIMCAT off-line chemical transport model. *Q. J. R. Meteorol. Soc.* **132**, 1179 - 1203, doi: 10.1256/qj.05.51.

Chipperfield, M.P., *et al.*, 2014: Multimodel estimates of atmospheric lifetimes of long-lived ozone-depleting substances: Present and future. *J. Geophys. Res. Atmos.*, **119**, 2555 - 2573, doi: 10.1002/2013JD021097.

Cook, P.A. and H.K. Roscoe, 2009: Variability and trends in stratospheric NO_2 in Antarctic summer, and implications for stratospheric NO_y. *Atmos. Chem. Phys.* **9**, 3601 - 3612, doi: 10.5194/acp-9-3601-2009.

Cook, P.A. and H.K. Roscoe, 2012: Changes in reactive stratospheric gases due to a change in Brewer-Dobson circulation: results from a simple model. *Atmos. Sci. Lett.* **13**, 49 - 54, doi: 10.1002/asl.362, 2012.

Chrysanthou, A., *et al.*, 2019: The effect of atmospheric nudging on the stratospheric residual circulation in chemistry–climate models. *Atmos. Chem. Phys.*, **19**, 11559 - 11586, doi: 10.5194/acp-19-11559-2019.

Davis S.M., *et al.*, 2016: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies. *Earth Syst Sci Data*. **8**(2):461 - 490. doi: 10.5194/essd-8-461-2016.

Davis, S.M., *et al.*, 2017: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP. *Atmos. Chem. Phys.*, **17**, 12 743 - 12 778, doi: 10.5194/acp-17-12743-2017, 2017.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828

Dessler, A.E., *et al.*, 2014: Variations of stratospheric water vapor over the past three decades. *J. Geophys. Res. Atmos.*, **119**, 12,588 - 12,598, doi: 10.1002/2014JD021712.

Dharmalingam, S. *et al.*, 2019: Accuracy of Balloon Trajectory Forecasts in the Lower Stratosphere. *Atmosphere,* **10**, 102, doi: 10.3390/atmos10020102.

Diallo, M., B. Legras, and A. Chédin, 2012: Age of stratospheric air in the ERA-Interim. *Atmos. Chem. Phys.*, **12**, 12133 - 12154, doi: 10.5194/acp-12-12133-2012.

Diallo, M., *et al.*, 2017: Significant Contributions of Volcanic Aerosols to Decadal Changes in the Stratospheric Circulation. *Geophys. Res. Lett.*, **44**, 10,780 - 10,791. doi: 10.1002/2017GL074662.

Diallo, M., *et al.*, 2019: Structural changes in the shallow and transition branch of the Brewer–Dobson circulation induced by El Niño. *Atmos. Chem. Phys.*, **19**, 425 - 446, doi: 10.5194/acp-19-425-2019.

Dietmüller, S., 2018: Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models. *Atmos. Chem. Phys.*, **18**, 6699 - 6720, doi: 10.5194/acp-18-6699-2018.

Dobson, G.M.B., 1956: Origin and distribution of the polyatomic molecules in the atmosphere. *Proceedings of the Royal Society of London* **A236**, 187 - 193, doi: 10.1098/rspa.1956.0127.

Elkins, J.W., *et al.*, 1996: Airborne gas chromatograph for in situ measurements of longlived species in the upper troposphere and lower stratosphere. *Geophys. Res. Lett*. **23**, 347 - 350, doi: 10.1029/96GL00244.

Engel, A., *et al.*, 2009: Age of stratospheric air unchanged within uncertainties over the past 30 years. *Nature Geosci.*, **2**, 28 - 31, doi: 10.1038/ngeo388.

Engel, A., *et al.*, 2017: Mean age of stratospheric air derived from AirCore observations. *Atmos. Chem. Phys.*, **17**, 6825 - 6838, doi: 10.5194/acp-17-6825-2017.

Eyring, V., *et al.,* 2006: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.*, **111**, D22308, doi: 10.1029/2006JD007327.

Feng, W., *et al.*, 2007: Midlatitude ozone changes: studies with a 3-D CTM forced by ERA-40 analyses. *Atmos. Chem. Phys.*, **7**, 2357 - 2369, doi: 10.5194/acp-7-2357-2007.

Feng, W., *et al.*, 2011: Evaluation of cloud convection and tracer transport in a three-dimensional chemical transport model, *Atmos. Chem. Phys.*, **11**, 5783 - 5803, doi: 10.5194/acp-11-5783-2011.

Fisher, M., A. O'Neill, and R. Sutton, 1993: Rapid descent of mesopheric air into the stratospheric polar vortex. *Geophys. Res. Lett.,* **20**, 1267 - 1270, doi: 10.1029/93GL01104.

Fischer, H., *et al.*, 2008: MIPAS: an instrument for atmospheric and climate research. *Atmos. Chem. Phys.*, **8**, 2151 - 2188, doi: 10.5194/acp-8-2151-2008.

Fu, Q., P. Lin, S. Solomon, and D.L. Hartmann, 2015: Observational evidence of strengthening of the Brewer-Dobson circulation since 1980. *J. Geophys. Res. Atmos.*, **120**, 10,214 - 10,228, doi: 10.1002/2015JD023657.

Fueglistaler, S. , *et al.*, 2009: The diabatic heat budget of the upper troposphere and lower/mid stratosphere in ECMWF reanalyses. *Q.J.R. Meteorol. Soc.*, **135**: 21 - 37. doi: 10.1002/qj.361.

Fujiwara, M., et al., 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Garfinkel, C.I., V. Aquila, D.W. Waugh, and L.D. Oman, 2017: Time-varying changes in the simulated structure of the Brewer–Dobson Circulation. *Atmos. Chem. Phys.*, **17**, 1313 - 1327, doi: 10.5194/acp-17-1313-2017.

Garny, H., T. Birner, H. Bönisch, and F. Bunzel, 2014: The effects of mixing on age of air. *J. Geophys. Res. Atmos.*, **119**, 7015 - 7034, doi: 10.1002/2013JD021417.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Hall, T.M. and R.A. Plumb, 1994: Age as a diagnostic of stratospheric transport. *J. Geophys. Res.-Atmos.*, **99**, 1059 - 1070, doi: 10.1029/93JD03192.

Hall, T.M., D:W. Waugh, K.A. Boering, and R.A. Plumb, 1999: Evaluation of transport in stratospheric models. *J. Geophys. Res.,* **104**, 18815 - 18839, doi: 10.1029/1999JD900226.

Haenel, F.J., et al., 2015: Reassessment of MIPAS age of air trends and variability. *Atmos. Chem. Phys.*, **15**, 13 161 - 13 176, doi: 10.5194/acp-15-13161-2015.

Hardiman, S.C., *et al.*, 2017: The influence of dynamical variability on the observed Brewer-Dobson circulation trend. *Geophys. Res. Lett.*, **44**, 2885 - 2892, doi: 10.1002/2017GL072706.

Harnisch, J., R. Borchers, P. Fabian, and M. Maiss, 1996: Tropospheric trends for CF_4 and C_2F_6 since 1982 derived from SF_6 dated stratospheric air. *Geophys. Res. Lett*. **23**, 1099 - 1102, doi: 10.1029/96GL01198.

Harries, J.E., *et al.*, 1996: Validation of measurements of water vapor from the Halogen Occultation Experiment (HALOE). *J. Geophys. Res.*, **101**, 10205 - 10216, doi: 10.1029/95JD02933.

Holton, J.R., 1982: The Role of Gravity Wave Induced Drag and Diffusion in the Momentum Budget of the Mesosphere. *J. Atmos. Sci.*, **39**, 791 - 799, doi: 10.1175/1520-0469(1982)039<0791:TROGWI>2.0.CO;2

Holton, J.R. and A. Gettelman, 2001: Horizontal transport and the dehydration of the stratosphere. *Geophys. Res. Lett.*, **28**, 2799 - 2802, doi: 10.1029/2001GL013148.

Iwasaki, T., H. Hamada, and K. Miyazaki, 2009: Comparisons of Brewer-Dobson Circulations Diagnosed from Reanalyses. *J. Meteor. Soc. Japan*, **87**, 997 - 1006, doi: 10.2151/jmsj.87.997.

Kida, H., 1983: General circulation of air parcels and transport characteristics derived from a hemispheric GCM, Part 2, Very long-term motions of air parcels in the troposphere and stratosphere. *J. Meteorol. Soc. Jpn.* **61**, 510 - 522, doi: 10.2151/jmsj1965.61.2_171.

Kobayashi, S., *et al.*, 2015: The JRA–55 Reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn.*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001.

Konopka, P., *et al.*, 2004: Mixing and Ozone Loss in the 1999–2000 Arctic Vortex: Simulations with the 3-dimensional Chemical Lagrangian Model of the Stratosphere (CLaMS). *J. Geophys. Res.*, **109**, D02315, doi: 10.1029/2003JD003792.

Konopka, P., *et al.*, 2015: Hemispheric asymmetries and seasonality of mean age of air in the lower stratosphere: Deep versus shallow branch of the Brewer-Dobson circulation. *J. Geophys. Res. Atmos.*, **120**, 2053 - 2066. doi: 10.1002/2014JD022429.

Kouker, W., I. Langbein, T. Reddmann, and R. Ruhnke, 1999: The Karlsruhe Simulation Model of The Middle Atmosphere Version 2. *Wiss. Ber.* **FZKA 6278**, Forsch. Karlsruhe, Karlsruhe, Germany, doi: 10.5445/IR/270045162.

Kovács, T., *et al.*, 2017: Determination of the atmospheric lifetime and global warming potential of sulfur hexafluoride using a three-dimensional model. *Atmos. Chem. Phys.*, **17**, 883 - 898, doi: 10.5194/acp-17-883-2017.

Lin, S.-J. and R.B. Rood, 1996: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes. *Mon. Weather Rev.*, **124**, 2046 - 2070, doi: 10.1175/1520-0493(1996)124<2046:mffslt>2.0.co;2.

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. *J. Atmos. Sci.*, **73**, 4507 - 4518, doi: 10.1175/JAS-D-16-0125.1.

Mahieu, E., *et al.*, 2014: Recent Northern Hemisphere stratospheric HCl increase due to atmospheric circulation changes. *Nature*, **515**, 104 - 107, doi: 10.1038/nature13857.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925-2018.

McKenna, D.S., et al., 2002: A new Chemical Lagrangian Model of the Stratosphere (CLaMS): 1. Formulation of advection and mixing. *J. Geophys. Res.*, **107**, 4309, doi: 10.1029/2000JD000114.

Meijer, E.W., B. Bregman, A. Segers, and P.F.J. van Velthoven, 2004: The influence of data assimilation on the age of air calculated with a global chemistry-transport model using ECMWF wind fields. *Geophys. Res. Lett.*, **31**, L23114, doi: 10.1029/2004GL021158.

Ming, A., P. Hitchcock, and P. Haynes, 2016: The Double Peak in Upwelling and Heating in the Tropical Lower Stratosphere. *J. Atmos. Sci.*, **73**, 1889 - 1901, doi: 10.1175/JAS-D-15-0293.1.

Mitchell, D., *et al.*, 2015: Signatures of naturally induced variability in the atmosphere using multiple reanalysis datasets. *Q. J. R. Meteorol. Soc.*, **141**, 2011 - 2031.

Miyazaki, K., *et al.*, 2016: Inter-comparison of stratospheric mean-meridional circulation and eddy mixing among six reanalysis data sets. *Atmos. Chem. Phys.*, **16**, 6131 - 6152, doi: 10.5194/acp-16-6131-2016.

Monge-Sanz, B.M. and M.P. Chipperfield, 2006: Chemical Transport Modelling. *Proceedings of the ECMWF/GEO Workshop on Atmospheric Reanalyses*, UK, 19 - 22 June 2006, ECMWF, pp. 39-44.

Monge-Sanz, B.M., M.P. Chipperfield, A.J. Simmons, and S.M. Uppala, 2007: Mean age of air and transport in a CTM: Comparison of different ECMWF analyses. *Geophys. Res. Lett.*, **34**, L04801, soi: 10.1029/2006GL028515.

Monge-Sanz, B.M., *et al.*, 2012: Improvements in the stratospheric transport achieved by a chemistry transport model with ECMWF (re)analyses: identifying effects and remaining challenges. *Q. J. Roy. Meteor. Soc.*, **139**, 654 - 673.

Monge-Sanz B.M., *et al.*, 2013: On the uses of a new linear scheme for stratospheric methane in global models: water source, transport tracer and radiative forcing. *Atmos. Chem. Phys.*, **13**, 9641 - 9660, doi: 10.5194/acp-13-9641-2013.

Monge-Sanz, B.M., *et al.*, 2020: A stratospheric prognostic ozone for seamless Earth System Models: performance, impacts and future. *Atmos. Chem. Phys. Discuss.*, doi: 10.5194/acp-2020-1261.

Mote, P.W., *et al.*, 1996: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor. *J. Geophys. Res.*, **101**, 3989 - 4006, doi: 10.1029/95JD03422.

Mote, P.W., *et al.*, 1998: Vertical velocity, vertical diffusion, and dilution by midlatitude air in the tropical lower stratosphere. *J. Geophys. Res.* **103**, 8651 - 8666, doi: 10.1029/98JD00203.

Orbe, C., *et al.*, 2020: Description and Evaluation of the specified-dynamics experiment in the Chemistry-Climate Model Initiative. *Atmos. Chem. Phys.*, **20**, 3809 - 3840, doi: 10.5194/acp-20-3809-2020.

Pisso, I. and B. Legras, 2008: Turbulent vertical diffusivity in the sub-tropical stratosphere. *Atmos. Chem. Phys.*, **8**, 697 - 707, doi: 10.5194/acp-8-697-2008.

Pitari, G., *et al.*, 2016: Impact of stratospheric volcanic aerosols on age-of-air and transport of long-lived species. A*tmosphere*, **7**(11), 149, doi: 10.3390/atmos7110149.

Ploeger, F., *et al.*, 2010: Impact of the vertical velocity scheme on modeling transport in the tropical tropopause layer, *J. Geophys. Res.*, **115**, D03301, doi: 10.1029/2009JD012023.

Ploeger, F., et al., 2015: Variability of stratospheric mean age of air and of the local effects of residual circulation and eddy mixing. J. *Geophys. Res. Atmos.*, **120**, doi:10.1002/2014JD022468.

Ploeger, F. and T. Birner, 2016: Seasonal and inter-annual variability of lower stratospheric age of air spectra. *Atmos. Chem. Phys.*, **16**, 10195 - 10213, doi: 10.5194/acp-16-10195-2016.

Ploeger, F., *et al.*, 2019: How robust are stratospheric age of air trends from different reanalyses? *Atmos. Chem. Phys.*, **19**, 6085 - 6105, doi: 10.5194/acp-19-6085-2019.

Plumb, R.A., 2002: Stratospheric transport. *J. Meteor. Soc. Japan*, **80**, 793 - 809, doi: 10.2151/jmsj.80.793.

Podglajen, A., A. Hertzog, R. Plougonven, and B. Legras, 2016: Lagrangian temperature and vertical velocity fluctuations due to gravity waves in the lower stratosphere. *Geophys. Res. Lett.*, **43**, 3543 - 3553, doi: 10.1002/2016GL068148.

Polvani L., *et al.*, 2019: Large impacts, past and future, of ozone-depleting substances on Brewer-Dobson circulation trends: A multi-model assessment. *J. Geophys. Res: Atmospheres*, **124**, 6669 - 6680 doi: 10.1029/2018JD029516.

Pommrich, R., *et al.*, 2014: Tropical troposphere to stratosphere transport of carbon monoxide and long-lived trace species in the Chemical Lagrangian Model of the Stratosphere (CLaMS). *Geosci. Model Dev.*, **7**, 2895 - 2916, doi: 10.5194/gmd-7-2895-2014.

Prather, M.J., 1986: Numerical advection by conservation of second-order moments. *J. Geophys. Res.,* **91**, 6671 - 6681, doi: 10.1029/JD091iD06p06671.

Ravishankara, A.R., S. Solomon, A.A. Turnipseed, and R.F. Warren, 1993: Atmospheric Lifetimes of Long-Lived Halogenated Species. *Science*, **259**, 194 - 199, doi: 10.1126/science.259.5092.194.

Ray, E.A., *et al.,* 1999: Transport into the Northern Hemisphere lower-most stratosphere revealed by in situ tracer measurements. *J. Geophys.Res.*, **104**, 26,565 - 26,580, doi: 10.1029/1999JD900323.

Ray, E.A., *et al.,* 2014: Improving stratospheric transport trend analysis based on SF_6 and CO_2 measurements. *J. Geophys. Res. Atmos.*, **119**, 14,110 - 14,128, doi: 10.1002/2014JD021802.

Reddmann, T., R. Ruhnke, and W. Kouker, 2001: Three-dimensional model simulations of SF6 with mesospheric chemistry. *J. Geophys. Res.*, **106**, 14525 - 14537, doi: 10.1029/2000JD900700.

Roscoe H.K., A.J. Charlton D.J. Fish, and J.G.T. Hill, 2001: Improvements to the accuracy of measurements of NO_2 by zenith-sky visible spectrometers II: errors in zero using a more complete chemical model. *J. Quant. Spectrosc. Ra.*, **68**, 337 - 349, doi: 10.1016/S0022-4073(00)00058-3.

Roscoe, H.K., 2004: A review of stratospheric H_2O and NO_2. *Adv. Space Res.*, **34**, 1747 - 1754, doi: 10.1016/j.asr.2003.01.025.

Rosenlof, K.H., 1995: Seasonal cycle of the residual mean meridional circulation in the stratosphere. *J. Geophys. Res.*, **100**, 5173 - 5191, doi: 10.1029/94JD03122.

Rotman, D.A., *et al.*, 2001: Global Modeling Initiative assessment model: Model description, integration, and testing of the transport shell. *J. Geophys. Res.*, **106**, 1669 - 1691, doi: 10.1029/2000JD900463.

Sato, K. and S. Hirano, 2019: The climatology of the Brewer–Dobson circulation and the contribution of gravity waves. *Atmos. Chem. Phys.*, **19**, 4517 - 4539, doi: 10.5194/acp-19-4517-2019.

Scheele, M.P., P.C. Siegmund, and P.F.J. van Velthoven, 2005: Stratospheric age of air computed with trajectories based on various 3D-Var and 4D-Var data sets. *Atmos. Chem. Phys*. **5**, 1-7, doi: 10.5194/acp-5-1-2005.

Schoeberl, M.R. and A.E. Dessler, 2011: Dehydration of the stratosphere. *Atmos. Chem. Phys.*, **11**, 8433 - 8446, doi: 10.5194/acp-11-8433-2011.

Schlutow, M., E. Becker, and H. Körnich, 2014: Positive definite and mass conserving tracer transport in spectral GCMs. *J. Geophys. Res. Atmos.*, **119**, 11,562 - 11,577, doi: 10.1002/2014JD021661.

Segers, A., van P. Velthoven, B. Bregman, and M. Krol, 2002: On the Computation of Mass Fluxes for Eulerian Transport Models from Spectral Meteorological Fields. In: Sloot P.M.A., Hoekstra A.G., Tan C.J.K., Dongarra J.J. (eds) Computational Science — ICCS 2002. ICCS 2002. *Lecture Notes in Computer Science*, vol **2330**. Springer, Berlin, Heidelberg. doi:10.1007/3-540-46080-2_81.

Skachko, S., *et al.*, 2014: Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model. *Geosci. Model Dev.*, **7**, 1451 - 1465, doi: 10.5194/gmd-7-1451-2014.

SPARC, 2010: SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models. V. Eyring, T. Shepherd and D. Waugh (Eds.), *SPARC Report* **No. 5**, *WCRP*-**30/2010**, *WMO/TD* – **No. 40**, available at **https://www.sparc-climate.org/publications/sparc-reports/**.

Stiller, G.P., *et al.*, 2012: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period. *Atmos. Chem. Phys.*, **12**, 3311 - 3331, doi: 10.5194/acp-12-3311-2012.

Stiller, G.P., *et al.*, 2017: Shift of subtropical transport barriers explains observed hemispheric asymmetry of decadal trends of age of air. *Atmos. Chem. Phys.*, **17**, 11177 - 11192, doi: 10.5194/acp-17-11177-2017.

Tao, M., *et al.*, 2015: Impact of stratospheric major warmings and the quasi-biennial oscillation on the variability of stratospheric water vapor. *Geophys. Res. Lett.*, **42**, 4599 - 4607, doi: 10.1002/2015GL064443.

Tao, M., *et al.*, 2019: Multitimescale variations in modeled stratospheric water vapor derived from three modern reanalysis products. *Atmos. Chem. Phys.*, **19**, 6509 - 6534, doi: 10.5194/acp-19-6509-2019.

Tegtmeier *et al.*, 2020: Temperature and tropopause characteristics from reanalyses data in the tropical tropopause layer. *Atmos. Chem. Phys.*, **20**, 753 - 770, doi: 10.5194/acp-20-753-2020.

Uppala, S. M., *et al.*, 2005: The ERA-40 Re-analysis. *Q. J. R. Meteorol. Soc.,* **131**, 2961 - 3012, doi: 10.1256/qj.04.176.

van Noije, T.P.C., H.J. Eskes, M. van Weele, and P.F.J. van Velthoven, 2004: Implications of the enhanced Brewer-Dobson circulation in ECMWF reanalysis ERA-40 for the stratosphere-troposphere exchange of ozone in global chemistry transport models. *J. Geophys. Res.*, **109**, D19308, doi: 10.1029/2004JD004586.

Waters, J.W., *et al.*, 2006: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite. *IEEE T. Geosci. Remote*, **44**, 1075 - 1092, doi: 10.1109/TGRS.2006.873771.

Waugh, D.W. and T.M. Hall, 2002: Age of stratospheric air: Theory, observations, and models. *Rev. Geophys.*, **40**, 1 - 27, doi: 10.1029/2000RG000101.

Waugh, D.W., 2009: The age of stratospheric air. *Nat. Geosci.*, **2**, 14 - 16, doi: 10.1038/ngeo397.

Wolter, K., and M.S. Timlin, 1998: Measuring the strength of ENSO events: How does 1997/98 rank? *Weather*, **53**, 315 - 324, doi: 10.1002/j.1477-8696.1998.tb06408.x.

Wright, J. and S. Fueglistaler, 2013: Large differences in reanalyses of diabatic heating in the tropical upper troposphere and lower stratosphere. *Atmos. Chem. Phys.*, **13**, 9565 - 9576 doi: 10.5194/acp-13-9565-2013.

Major abbreviations and terms

AoA	Age-of-air
AOD	aerosol optical depth
AMSU	Advanced Microwave Sounding Unit
BAS	British Antarctic Survey
BDC	Brewer-Dobson Circulation
CCM	Chemistry Climate Model
CFSR	Climate Forecast System Reanalysis
CMAM	Canadian Middle Atmosphere Model
CCMI	Chemistry-Climate Model Initiative
CCMVal	Chemistry Climate Model Validation
CTM	Chemistry-transport model
DAS	Data assimilation system
DOE	Department of Energy
DJF	December-January-February
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño Southern Oscillation
EP-flux	Eliassen-Palm Flux
EPFD	Eliassen-Palm Flux Divergence
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA-Interim	ECMWF interim reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF

FFSL	Flux-Form Semi-Lagrangian
GEOSCCM	NASA Goddard Chemistry-Climate Model
GWD	Gravity Wave Drag
HALOE	Halogen Occultation Experiment
HATS	Halocarbons and other Atmospheric Trace Species
JJA	June-July-August
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
KASIMA	Karlsruhe Simulation of the Middle Atmosphere
LS	Lower stratosphere
MEI	Multivariate ENSO index
MERRA; MERRA-2	Modern Era Retrospective Analysis for Research and Applications / Version 2
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MMC	Mean Meridional Circulation
MMM	Multi-Model Mean
MRM	Multi-Reanalysis Mean
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-DOE R2	Reanalysis 2 of the NCEP and DOE
NCEP-NCAR R1	Reanalysis 1 of the NCEP and NCAR
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
OMS	Observations of the Middle Stratosphere
POLARIS	Photochemistry of Ozone Loss in the Arctic Regions in Summer
QBO	Quasi-Biennial Oscillation
RCTT	Residual Circulation Transit Time
REM	Multi-Reanalysis Mean
SH	Southern Hemisphere
SST	Sea Surface Temperature
SSW	Sudden Stratospheric Warming
StratoClim	Stratospheric and upper tropospheric processes for better climate predictions
SWV	Stratospheric Water Vapour
TTL	Tropical Tropopause Layer
TOA	Top of Atmosphere
UKMO	United Kingdom Meteorological Office
UTLS	Upper troposphere and lower stratosphere
WMO	World Meteorological Organization