Time varying changes in the simulated structure of the Brewer **Dobson Circulation**

Chaim I Garfinkel¹, Valentina Aquila^{2,3,4}, Darryn W Waugh², and Luke D Oman⁴

¹The Fredy and Nadine Herrmann Institute of Earth Sciences, Hebrew University, Jerusalem, Israel.

² Department of Earth and Planetary Science, Johns Hopkins University, Baltimore, MD, USA

³ Goddard Earth Science Technology and Research, Greenbelt, Maryland, USA

⁴ NASA Goddard Space Flight Center, Greenbelt, MD, USA.

Correspondence to: Chaim I. Garfinkel (chaim.garfinkel@mail.huji.ac.il)

Abstract. A series of simulations using the NASA Goddard Earth Observing System Chemistry-Climate Model are analyzed in order to assess changes in the Brewer-Dobson Circulation (BDC) over the past 55 years. When trends are computed over the past 55 years, the BDC accelerates throughout the stratosphere, consistent with previous modeling results. However, over the second half of the simulations (i.e. since the late 1980s), the model simulates structural changes in the BDC as the temporal evo-

- lution of the BDC varies between regions in the stratosphere. In the mid-stratosphere in the mid-latitude Northern Hemisphere, 5 the BDC does not accelerate in the ensemble mean of our simulations despite increases in greenhouse gas concentrations and warming sea surface temperatures, and even decelerates in one ensemble member. This deceleration is reminiscent of changes inferred from satellite instruments and in-situ measurements. In contrast, the BDC in the lower-stratosphere continues to accelerate. The main forcing agents for the recent slowdown in the mid-stratosphere appear to be declining ODS concentrations
- 10 and the timing of volcanic eruptions. Changes in both age of air and the tropical upwelling of the residual circulation indicate a lack of recent acceleration. We therefore clarify that the statement that is often made that climate models simulate a decreasing age throughout the stratosphere only applies over long time periods, and is not necessarily the case for the past 25 years when we have most tracer measurements.

1 Introduction

15 The global circulation in the stratosphere - the Brewer-Dobson circulation (BDC) - consists of air-masses rising across the tropical tropopause, moving poleward, and sinking into the extratropical troposphere (Holton et al., 1995; Waugh and Hall, 2002; Butchart, 2014). Since the BDC and its changes have important implications on both stratospheric and tropospheric climate as well as stratospheric ozone chemistry (SPARC-CCMVal, 2010; World Meteorological Organization, 2011, 2014; Manzini et al., 2014), it is important to assess the factors that lead to simulated BDC changes and whether historical changes in the BDC as simulated by models are inconsistent with available observational constraints. 20

The BDC has historically been deduced either from the residual circulation or from the average time for an air parcel to travel from the tropical troposphere to a given stratospheric sampling region (i.e. the mean age of air or mean age). While these two diagnostics are clearly related, differences can arise due to isentropic mixing and recirculation (Waugh and Hall,

2002; Strahan et al., 2009; Li et al., 2012). Specifically, the vertical component of the residual circulation ($\overline{w*}$) measures the instantaneous advection whereas mean age is an integrated measure of the total transport. It would not be a surprise if the two metrics had a different evolution.

Chemistry climate models robustly predict a strengthened BDC under climate change in the middle and lower strato-

sphere of approximately 1 – 5% per decade (the precise rate depends on the level considered and varies among models; Butchart and Scaife, 2001; Butchart et al., 2006; Garcia and Randel, 2008; Li et al., 2008; Waugh, 2009; Shepherd and McLandress, 2011; Garcia et al., 2011; Lin and Fu, 2013; Butchart, 2014; Oberländer-Hayn et al., 2015). The model used in this study, the Goddard Earth Observing System Chemistry-Climate Model, Version 2 (GEOSCCM), predicts a trend quantitatively similar to those in other models both for the historical period and for the future (Oman et al., 2009; Waugh, 2009; Butchart et al., 2010;

10 Li et al., 2012).

It has been argued that observational estimates of historical changes do not agree with the simulated acceleration trend. Specifically, the analysis of historical tracer data does not provide evidence for an acceleration trend in the mid-stratosphere Northern Hemisphere (NH), where mean age actually appears to have increased or remained unchanged(Engel et al., 2009; Bönisch et al., 2011; Stiller et al., 2012; Hegglin et al., 2014; Ray et al., 2014). In particular, the mean age evolution in the

- 15 figures of Engel et al. (2009) and Ray et al. (2014) indicates aging since the late 1980s, with earlier changes less clear. ERAinterim reanalysis data also suggests aging of NH mid-stratosphere air since the late 1980s (Diallo et al., 2012; Monge-Sanz et al., 2013; Ploeger et al., 2015). While these observational and reanalysis based studies disagree about the sign of changes in other regions of the stratosphere, they all indicate aging of air in the NH midlatitude mid-stratosphere. Ray et al. (2014) argue that the large uncertainty estimates on the trends presented by Engel et al. (2009) are overly conservative, and that this aging trend
- 20 is statistically significant even after taking into account the nonlinear growth rates of these trace gases (i.e. the arguments of Garcia et al., 2011). Several of the aforementioned studies suggest that these changes in the mean age imply a redistribution of the BDC, and specifically a slowdown of the deep (i.e. mid-stratospheric) NH branch of the BDC and/or less mixing of fresh tropical air into this region. However, it is not clear what forcings (if any) could be responsible for this redistribution and also whether models of the kind used in WMO-Ozone Assessments (World Meteorological Organization, 2011, 2014) can capture
- 25 such a slowdown given known forcings.

Oman et al. (2009) conducted time-slice and transient simulations using a previous version of the model used in this study, and they found a trend towards younger mean age everywhere in the stratosphere between 1960 and 2004, though the decrease in mean age in the Southern Hemisphere (SH) is larger due to Antarctic ozone depletion. Oman et al. (2009) also found that ozone recovery would lead to a slowdown of the BDC if not for warming SSTs due to increasing greenhouse gas emissions.

30 Oberländer-Hayn et al. (2015) recently presented differences in mean age and the residual circulation in time-slice simulations using the ECHAM/MESSy Atmospheric Chemistry Model. The changes in tropical upward mass flux indicate a strengthening of the BDC between 1960 and 2000 in the NH winter season in the lower stratosphere and a weakening in the upper stratosphere with a change in sign at 10hPa. Changes in mean age show a decrease of about 0.13yr/decade in the lower and middle stratosphere and a slight increase in the Arctic upper stratosphere and lower mesosphere. While there is some hint of a structural change in the properties of the BDC, the changes occur higher in the stratosphere and at more polar latitudes than is suggested by available observations.

The motivation for studying historical changes in the BDC in free running climate simulations is not to form a best estimate of the actual historical evolution; for that purpose, reanalyses and/or nudged experiments are far better (though care must be

- exercised when computing trends). Rather, the motivation is two-fold: one, future projections of the BDC can only be produced 5 by free running climate simulations, and these projections are of limited value if a model's simulation of the past is inconsistent with observational constraints; two, assuming the model is capable of following the observed evolution, the forcings that caused these historical changes can be systematically diagnosed by sequentially adding these forcings. Combined, the goal of this study is to understand whether historical forcings could have led to the structural changes in the BDC that have been inferred
- by observational studies. 10

We show that over the full duration of the experiments (i.e for a start-date in 1960), we recover the result from previous modeling studies: anthropogenic climate change leads to acceleration of the BDC throughout the stratosphere. However, our model can simulate statistically significant aging of air in the mid-latitude NH near 20hPa between the early 1990s and the present as suggested by available observations. This suggests that historical forcings caused structural changes of the BDC

since the late 1980s, whereby the BDC accelerated in the lower stratosphere, but decelerated in the mid-stratosphere, in both 15 the tropics and in the NH. Mean age and the residual circulation (as measured by tropical $\overline{w*}$) change in unison. The cause of this mid-stratospheric deceleration trend is a combination of forcings - ODS recovery and the timing of volcanic eruptions that together outweighed greenhouse gas induced acceleration since the late 1980s. We therefore emphasize that if one wishes to capture observed historical changes, careful attention must be paid to the start and end dates used for trend calculation and the forcings included in a model simulation. 20

2 Methods

The model focused on for this study is GEOSCCM, an aerosol and chemistry focused version of the GEOS-5 Earth system model. GEOSCCM couples the GEOS-5 (Rienecker et al, 2008; Molod et al., 2012) atmospheric general circulation model to the comprehensive stratospheric chemistry module StratChem described in Pawson et al. (2008), and the Goddard Chemistry,

Aerosol, Radiation, and Transport Model (GOCART) described in Colarco et al. (2010). The model has 72 vertical layers, with 25 a model top at 0.01 hPa, and all simulations discussed here were performed at 2° latitude x 2.5° longitude horizontal resolution. Previous versions of GEOSCCM have been graded highly in the two phases of the Chemistry-Climate Model Validation (Eyring et al., 2006; SPARC-CCMVal, 2010). Improvements to the model since then are described in Oman and Douglass (2014) and Aquila et al. (2016). Note that this model version includes a spontaneous QBO, and that the QBO phase differs among the three ensemble members.

A series of simulations of the period from January 1960 to December 2014 have been performed in order to understand the past evolution of the stratosphere. These simulations were presented in Aquila et al. (2016), where the focus was on changes in temperatures. Here we examine changes in the BDC. We start with an ensemble in which the only time-varying forcing is

³⁰

changing SSTs and sea ice, and then sequentially add the following forcings: greenhouse gases, ozone depleting substances, volcanic eruptions, and solar variability. More specifically, these simulations are grouped into the following five experiments:

- 1. **SST**, which uses time-varying observed sea surface temperatures (SSTs) and sea ice up to November 2006 from the MetOffice Hadley Centre observational dataset (Rayner et al., 2006) and from Reynolds et al. (2002) and updates since
- 5
- MetOffice Hadley Centre observational dataset (Rayner et al., 2006) and from Reynolds et al. (2002) and updates since then (Figure 1a). GHGs and ODS concentrations are fixed at 1960-levels. Volcanic eruptions are not included in this experiment, and the solar forcing is held constant;
- SSTGHG, which includes observed SSTs and increasing GHG concentrations (Figure 1b). GHG concentrations are from observations up to 2005 and from the Representative Concentrations Pathway 4.5 after 2005 (Meinshausen et al., 2011);
- SSTGHGODS, which includes observed SSTs, increasing GHGs, and changing ODS concentrations following World Meteorological Organization (2011, Figure 1c);
 - SSTGHGODSVOL, which includes observed SSTs, increasing GHGs, changing ODS, and volcanic eruptions, specified after Diehl et al. (2012) from 1979 to December 2010 and Carn et al. (2015) from January 2011 to December 2014 (Figure 1d). The only eruption before 1979 included is Mt. Agung in 1963;
- 5. SSTGHGODSVOLSOL (or All-forcing), which includes observed SSTs, increasing GHGs, changing ODS, volcanic eruptions and changes in solar flux as in Lean (2000) and subsequent updates (Figure 1e). Note that neither Oberländer-Hayn et al. (2015) nor Oman et al. (2009) considered volcanic forcings and solar variability.

All simulations used emissions of tropospheric aerosol and aerosol precursors following Granier et al. (2011). We focus on transient simulations as it is difficult to compare historical tracer observations to time-slice simulations.

- Each experiment is composed of three ensemble members initialized with different initial conditions from a 1960 time-slice simulation. Because we have three members for each forcing combination, we can also assess at least partially the range of internal atmospheric variability. This internal variability is not a source of model uncertainty; rather it is an inherent part of the climate system. If the BDC in one ensemble member, but not in the other two, evolves consistently with observational constraints, one can reasonably conclude that models can capture the observed trends if part of the observed trend was due to
- 25 internal variability and was not forced (cf. Santer et al., 2008).

We assume that BDC perturbations induced by each forcing agent adds linearly to the others, as previous work focusing on forcing agents for lower stratospheric mass flux and mid-stratospheric mean age suggests that non-linearities are small (e.g. Oman et al., 2009; McLandress et al., 2010). The model version used to perform the integrations described in this paper is no longer supported, and hence we cannot explicitly test this assumption.

30 The use of observed SSTs in our simulations, rather than internally calculated by the model, produces a climate state closer to the observed one. However, partitioning trends into an SST driven component and a component from other radiative or chemical forcings is somewhat artificial, as the prescribed SST changes occur in response to and in tandem with the changing



Figure 1. Forcing applied in the simulations. a) 10S-10N average sea surface temperatures, with a smoothed version of the curve bolded; b) Atmospheric concentrations of CO₂ following RCP4.5; c) equivalent effective stratospheric chlorine (EESC Newman et al., 2007, equation 1 with a = 1 and α for $Br_y = 60$, and using 3-year mean age); d) ensemble mean of the aerosol optical thickness from explosive volcanic eruptions, resulting from prescribed injections of volcanic SO2; e) tot**g** solar irradiance.

direct atmospheric forcing; however such a partitioning is an effective tool for disentangling the physical mechanisms leading to changes in the atmospheric circulation (Deser and Phillips, 2009). Specifically, Oman et al. (2009) found that the mean tropospheric warming associated with rising SSTs had a bigger impact on mean age than the direct radiative impacts of CO_2 . Note that interannual and decadal variability in SSTs drives changes in the BDC that likely have nothing to do with climate

5 change, and hence we include a smoothed version of the SST variations in Figure 1a (all experiments were forced with the full time evolving SST fields).

A passive tracer is used to derive the mean age. The mixing ratio of this tracer increases linearly with time, and the time lag in tracer concentrations between a certain grid point in the stratosphere and a reference point in the troposphere provides an estimation of mean age at this stratospheric grid point (Waugh and Hall, 2002). We adopt a reference point of 200hPa at

10 the equator. Mean age was first initialized in a 10 year time-slice control run with 1960 conditions before we branch off for each experiment, and so mean age can be defined at the beginning of the experiments. Note that diagnostic output necessary to compute the full age spectrum was not saved for these model experiments, and hence we are limited in our ability to quantify mixing changes.

The aging of NH mid-stratospheric air in observations is pronounced mainly after the late 1980s (see figures in Engel et al.,

- 15 2009; Ray et al., 2014), and several of the reanalysis-based studies begin their analysis in the late 1980s as they need to initialize trajectories for 10 years before computing mean age. Hence, in our presentation and discussion of the results, we consider trends and variability both over the full period of integration and also since 1988. Motivated by our results, we also discuss trends for a start-date of 1992, as the recent slowdown of the deep branch of the BDC is most pronounced (and locally statistically significant) for this start date due to the eruption of Mt. Pinatubo. Finally, global mean age profiles as retrieved by satellites are
- only available since 2002 (Stiller et al., 2012; Haenel et al., 2015), and hence we show trends since 2002 separately as well. The trends are calculated with a linear least squares fit. Statistical significance of the trends in individual ensemble members of GEOSCCM are computed using a 2-tailed Student's t-test, and the reduction in degrees of freedom due to autocorrelation of the residuals is taken into account with the formula $N(1-r_1)(1+r_1)^{-1}$, where N is the number of years and r_1 is the lag-1 autocorrelation (eq. 6 of Santer et al., 2008). In computing the ensemble mean response, we first averaged the mean age among
- the three ensemble members for each season/year (in order to damp the internal, unforced, atmospheric variability), and then compute the trend based on the ensemble mean mean age.

3 Results

30

We now examine how the BDC as simulated by GEOSCCM has changed in structure over the past 55 years. We show that BDC changes (or "trends") vary with period considered and location. These structural changes are associated with several distinct forcings, and these forcings transiently drive changes in the BDC. When combined with internal variability, it is possible that these forcings can drive, over the 27 year period of 1988-2014, a deceleration trend in the NH mid-stratosphere.

3.1 Changes in the All-forcing ensemble

We begin with changes in the BDC for the all forcing ensemble. We first consider changes in the residual circulation and in mean age as a function of latitude and pressure over the full duration of the experiment in the ensemble mean (Figure 2). The BDC accelerates throughout the stratosphere as mean age decreases and the residual circulation accelerates. Hence, changes over the

- 5 full integration period in our experiments are consistent with previous work (e.g. Oman et al., 2009). Statistical significance of the trends are considered in Figures 3 and 4, which show the 95% confidence bounds on the trends for a range of start dates for mean age in select region and for tropical mean upwelling. The changes in both metrics of the BDC are statistically significant at the 95% level throughout the stratosphere for a trend start-date of 1960, and all three ensemble members indicate quantitatively similar acceleration trends.
- 10 However, the ensemble mean acceleration trend weakens (and even reverses locally) and its robustness goes away as we consider more recent periods. To demonstrate this, we start by showing trends since 1988 in Figure 5. In the lower stratosphere (i.e. the shallow branch of the BDC), the BDC continues to accelerate in all three ensemble members, and this change is statistically significant at the 95% level in each ensemble member individually and in the ensemble mean (e.g. 3def and 4b). In the mid- and upper- stratosphere, however, trends are not robust across the various ensemble members. One of the three
- 15 ensemble members shows decreasing mean age and an accelerated residual circulation (Figure 5a and gray lines in Figure 3abc and 4a), while another shows the opposite (Figure 5e and red lines in Figure 3abc and 4a). In this ensemble member with aging air, upwelling decreases throughout the tropics, such that both the residual circulation and mean age diagnostics suggest deceleration of the BDC. None of the mean age trends in the mid- and upper- stratosphere in Figure 5ace are statistically significant at the 95% level (e.g. figure 3abc). Note that if the start-date for the trend is advanced to 1992 then none of the
- 20 three ensemble members indicate acceleration of the BDC in the NH midlatitude mid-stratosphere (figure 3a), but we prefer to demonstrate that even for a start date of 1988, aging can be simulated given the large amount of internal atmospheric variability.

These differences in trends since 1988 among ensemble members can be reconciled with the changes in the wave forcing of the BDC. Previous modeling and theoretical work has demonstrated that changes in the wave forcing directly force changes in the residual circulation (Rosenlof, 1995; Butchart, 2014). Specifically, enhanced wave convergence (i.e. deceleration of the

25

mean flow) leads to enhanced upwelling on the tropical flank of the enhanced wave convergence and downwelling on the poleward flank (Haynes et al., 1991). As both resolved waves and gravity waves are important for the total wave driving, we evaluate their combined impact in common units of acceleration of the mean flow (m/s/day/decade); positive values indicates less wave convergence and deceleration of the residual circulation, while negative values indicates enhanced wave convergence and acceleration of the residual circulation. The right column of Figure 5 demonstrates that the difference between the ensemble

30 member with a weakened SH residual circulation and the one with an accelerated SH circulation is related to differences in the wave driving. For the ensemble member with a weakened residual circulation in the SH (and thus aging of air in the midstratosphere), there is less wave flux converging in the SH (Figure 5f). In contrast, for the member with an accelerated residual circulation in the SH (and thus decreasing mean age in the NH mid-stratosphere), there is enhanced wave flux converging in the SH (Figure 5b). Note that a difference in wave fluxes in the SH can influence mean age in the NH because mean age is an integral measure of transport, and thus changes in the BDC in the tropics due to wave flux changes in the SH can impact transport pathway into the NH. Hence, the difference between aging and freshening of mid-stratospheric air is associated with the internal atmospheric variability associated with wave fluxes. The supplemental material discusses implications of this intra-ensemble difference for trends in temperature and zonal wind. There is reduced wave convergence in the NH stratosphere

in all three ensemble members, and this effect is most pronounced in winter (not shown). The cause of the decrease in NH 5 stratospheric wave convergence is discussed in Garfinkel et al. (2015) and in the next section. Overall, our model simulations indicate that known forcings could have led to a slowdown of the deep branch of BDC since 1988 given the large amount of internal variability in wave fluxes.

We now consider the time evolution of changes in mean age in various regions in Figure 6. As the eruption of Mt. Pinatubo

- had a pronounced impact on the BDC in these simulations, we separately consider the evolution before and after its eruption. 10 Changes in the NH mid-latitudes in the mid-stratosphere are considered in Figure 6a. Note that this is the region where available observations suggest that air has aged over the past several decades. In the All-forcing ensemble (blue line), mean age decreases by 0.45 years between 1960 and 1992 (i.e. 0.15 years per decade, blue in Figure 7a), but then ages by 0.15 years since 1992 (i.e. 0.05 years per decade, blue in Figure 7b). The aging trend since 1992 is statistically significant at the 95%
- level in two of the three ensemble members and in the ensemble mean (figure 3a). Figure 8 considers the evolution of each of 15 the three ensemble members individually. The ensemble spread in mean age in any given year can exceed 5%. One ensemble member simulates anomalously old air after 2010, and for this member aging trends are simulated for a start-date of the trend calculation of either 1992 or 1988. Note that a different ensemble member simulates anomalously younger air relative to the other two after El Chichon, highlighting the large role of internal atmospheric variability in masking the response to climate forcings.
- 20

25

In the tropical and SH mid-stratosphere, mean age also increases from 1992 through the end of the simulation (blue curves in Figure 6bc and Figure 3bc). The large uncertainty in the trends in the SH arise due to the correction of Santer et al. (2008) for the effective number of degrees of freedom associated with residuals with large autocorrelation. In contrast, in the lower stratosphere, mean age continues to decline in the tropics and in the NH, though not in the SH (blue curves in Figure 6def). The decline in mean age is statistically significant at the 95% level in two of three ensemble members and in the ensemble mean in the NH lower stratosphere (Figure 3d). Hence, there are time- and space-varying variations in recent mean age trends, and only in some regions of the stratosphere has mean age continued to decline.

Similar structural changes are evident for the tropical upwelling mass flux between the turnaround latitudes $(\int_{w=0_{SH}}^{w=0_{NH}} \rho \overline{w*} dA)$ where ρ is the density and A is the area of a given zonal band, Figure 4 and 9). In the all-forcing experiment, tropical upwelling

- accelerated until 1992 in both the mid-stratosphere and lower stratosphere (i.e. blue line rises in Figure 9, blue in 10aceg), 30 but since 1992 has decreased at all levels above 70hPa (e.g. Figure 9a and 10bdfh). These changes are reminiscent of those proposed by Ray et al. (2010) and Ray et al. (2014) in order to explain how NH midlatitude air in the mid-stratosphere anomalously ages. Specifically, Ray et al. (2010) concluded that for a tropical leaky pipe model, "the best quantitative agreement with the observed mean age and ozone trends over the past three decades is found assuming a small strengthening of the mean cir-
- culation in the lower stratosphere, [and] a moderate weakening of the mean circulation in the middle and upper stratosphere". 35



Figure 2. Trends in annual averaged BDC from 1960 to 2014 in the ensemble mean of the three all-forcing integrations. Mean age trends are indicated by contours with a contour interval of 6days/decade, and the residual circulation trends are indicated with streamlines. The thickness of the streamline is proportional to the magnitude of the wind speed.

as simulated by GEOSCCM. Overall, it is clear that GEOSCCM can simulate structural changes in the BDC that resemble those inferred from observations.



mean age trend through 2014 for each start-date

All #1 All #2 All #3 All ensemble mean

Figure 3. Trend through 2014 in annual averaged mean age in the NH (left), tropics (middle), and SH (right) for start dates starting in 1960 and ending in 2002. Each line is for one member of the all-forcing ensemble or for the all-forcing ensemble mean. Changes in the mid-stratosphere (deep branch) are shown in (a)-(c), and changes in the lower stratosphere (shallow branch) are shown in (d)-(f). 95% confidence intervals on the calculated trends are indicated; note that in the SH mid-stratosphere, the correction of Santer et al. (2008) for the number of degrees of freedom leads to very few degrees of freedom for the trend, and therefore to high uncertainty.

3.2 Forcing of the Trends

We now consider the forcing mechanisms behind these structural changes in the BDC.

3.2.1 Mid-Stratosphere

As shown above, mean age in the NH mid-latitudes in the mid-stratosphere decreases by 0.45 years between 1960 and 1992
5 in the All-forcing ensemble, but then ages by 0.15 years since 1992 (blue line; Figure 6a and 7). This evolution can be broken down into its various forcing components.

SSTs The red line in Figure 6a shows that SSTs lead to a decrease in mean age of 0.1 years over the course of these 55 years, but with substantial interannual and decadal variability. Specifically, over the last 30 years of the integrations (from 1985 to the end), there is a weak and insignificant aging trend (e.g. red in 7b). The likely cause of this is a tug-of-war



Figure 4. As in Figure 3 but for the upwelling mass flux in the tropics.



Figure 5. (left) As in Figure 2 but for trends in mean age from 1988 to 2014 in the three ensemble members. Note that (a) corresponds to the ensemble member with decreasing mean age and (e) to the ensemble member with aging air in the NH mid-latitude stratosphere. The right column shows the changes in total wave forcing of the BDC (EP flux divergence plus gravity wave drag).



Figure 6. Time series of annual averaged mean age in the NH (left), tropics (middle), and SH (right). Each line is for one ensemble mean. Blue is all forcing, red is SST only, and intermediate colors span the other three experiments performed. Changes in the mid-stratosphere (deep branch) are shown in (a)-(c), and changes in the lower stratosphere (shallow branch) are shown in (d)-(f).

of opposing effects. On the one hand, gradual warming of the oceans in isolation leads to an acceleration of the BDC (figure 1a and Oman et al., 2009). On the other hand, the spatial pattern of recent changes in SSTs has led to a decline in planetary wave flux (especially wave 1) entering the NH stratosphere in midlatitudes (Garfinkel et al., 2015): the vertical component of the Eliassen Palm flux at 100hPa area averaged between 40N-80N declines in all three all forcing and in all three SST only experiments, and the decrease is statistically significant at the 95% level in the ensemble mean in both January through March (the focus of Garfinkel et al., 2015) and in the annual mean. This decline in upward propagating midlatitude planetary waves at 100hPa impacts the deep branch more strongly (Plumb, 2002; Ueyama et al., 2013). Hence, it is not surprising that little change occurred over the last 30 years.

- 2. **Greenhouse gases** The influence of GHG can be deduced from the difference between the red and magenta curves in Figure 6a, as the SST only ensemble is conducted with radiative forcings fixed at 1960 levels. The difference between the curves exceeds 0.1 years towards the end of the integrations.
 - 3. **Declining ODS concentrations** The effect of increasing ODS concentrations can be deduced from the difference between the magenta and green curves in Figure 6a, and it suggests that increasing ODS concentrations led to a decrease in mean age of 0.25 years by the late 1990s (Figure 6a) when the ODS burden was peaking. More recently, declining

10



Figure 7. Trends in midlatitude NH mean age in the ensemble mean of each of the five experiments (a) from 1960 to 1992 and (b) from 1992 to 2014. The trend over the full duration of the experiments is shown in panel (c). The horizontal line indicates the ensemble mean trend, and the vertical line indicates the 95% uncertainty bounds.



Figure 8. Modeled annual averaged mean age in the NH mid-stratosphere between 30hPa and 10hPa in each of the three All-forcing GEOSCCM integrations. Thin lines denote individual integrations, while thick lines denote ensemble means. For clarity, we also include the ensemble mean of the SST+GHG+ODS simulation.

ODS concentrations (Figure 1c) lead to a recovery towards older air; that is, the gap between the magenta and green curves decreases between the late 1990s and the present in Figure 6a. Note that Oman et al. (2009) also found that ozone recovery leads to a slowdown of the BDC in a previous version of the model we use. An ozone-induced acceleration of the residual circulation in the SH and tropics can affect mean age in the NH due to mixing. Finally, we note the caveat that while declining ODS concentrations clearly impact the BDC in these integrations, a statistically significant recovery of ozone has been detected in observations only in the upper stratosphere in mid-latitudes and the tropics (World Meteorological Organization, 2014).

4. **volcances and solar** The influence of volcanic eruptions can be deduced from the difference between the green and cyan curves in Figure 6a, and in our model simulations the eruption of Mt Pinatubo and El Chichon led to a decrease in mean age of 0.2 years, which gradually decayed over four to six years. Minor volcanic eruptions in the past ten years may have led to an additional decrease in mean age of 0.05 to 0.1 years. The net effect is that large eruptions (or lack thereof) can influence decadal variability in mean age in GEOSCCM. Solar influences appear to be relatively minor, and we therefore focus our attention on the other forcings in this paper (Figure 6a).

The net effect is that over the second half of the experiments (when observations are more numerous), decreasing ODS con-15 centrations and the recovery from Pinatubo overcame the influence of rising GHG concentrations and led to aging of 0.15 years.

The same forcings that led to a statistically significant aging trend in the NH since 1992 in the ensemble mean and in two integrations also lead to similar significant aging in the tropics (blue curve in Figure 3b and 6b). Specifically, changes in this region are dominated by ODS concentrations, and since ODS concentrations decrease after the late-1990s (Figure 1c), mean

10



Figure 9. Time series of annual averaged mass flux between the turnaround latitudes at 50hPa (top) and 100hPa (bottom). Each line is for one ensemble mean. Blue is all forcing, red is SST only, and green is for SSTGHGODS. For clarity, we suppress two of the intermediate experiments.

age increases despite rising GHG concentrations. As for the NH, the earliest start-date of significant aging trends is in the early 1990s due to the eruption of Mt. Pinatubo. ODS concentrations also dominate the SH mid-stratospheric mean age evolution (Figure 6c), and since ODS concentrations decrease after the late 1990s, mean age does not change significantly since 1992 in the all-forcing ensemble as well. Note that the influence of the eruption of Mt. Pinatubo is weaker in the SH than in the NH; Mt. Pinatubo is located at 15N, and in our experiments the majority of the aerosols stay in the NH.

3.2.2 Lower Stratosphere

5

In the tropical and NH lower stratosphere (Figure 6d-e), increasing GHG concentrations and warming SSTs drive a decrease in mean age throughout the period of the All-forcing ensemble integrations. In agreement with the modeling results of Lin et al. (2015), warming SSTs (cf. figure 1a) impact the shallow branch more strongly than the deep branch. As midlatitude planetary

10 waves are less important for the shallow branch than for the deep branch (Plumb, 2002; Ueyama et al., 2013; Abalos et al., 2014), it is reasonable to expect that a reduction in their strength has a relatively smaller impact on the shallow branch. Volcanic eruptions have a weaker effect in the lower stratosphere as compared to the middle stratosphere (compare the difference)



Figure 10. As in Figure 7 but for tropical mean upwelling at a variety of pressure levels.

between the blue and green curves for the years 1963/1964, 1983/1984 and 1991/1992 between Figure 6a and Figure 6d). A possible explanation is that longwave and near-IR heating due to volcanic aerosols occurs at the level of the aerosols (not shown), and in our integrations the aerosols are quickly lofted higher in the stratosphere (compare the volcanic influence on temperature as a function of time in the various levels of Aquila et al., 2016). This allows for stronger and longer-lasting

- 5 changes mainly in the "deep" branch of the BDC. As for volcanoes, changing ODS concentrations impact the deep branch of the BDC more strongly (except in the SH where ozone depletion is strongest), and this effect is consistent with the idealized modeling results of Gerber (2012). The colder vortex that follows ozone depletion creates a waveguide higher into the stratosphere, raising the breaking level of Rossby waves and deepening the BDC. Hence, it is to be expected that ozone depletion and recovery has a disproportionate impact on the deep branch. Overall, the aging in the deep branch since 1992 does not
- 10 extend to the shallow branch because the two factors that led to aging (recovery from the Pinatubo eruption and declining ODS concentrations) preferentially impact the deep branch, while the two factors that led to freshening (GHG increases and SST warming) preferentially impact the shallow branch.

In the SH lower stratosphere (Figure 6f), ODS concentrations are the dominant forcing, but the gradual decline in ODS concentrations is balanced out by rising GHG concentrations and mean age is flat since 1995 in the blue All-forcing curve. It

15 is known that ozone-depletion induced polar cooling can directly modulate extratropical wave propagation down to the troposphere (Chen and Held, 2007; Oman et al., 2009; Garfinkel et al., 2013), though future work is needed in order to understand how this influence led to an accelerated shallow branch.

3.2.3 Residual Circulation

The same forcings led to structural changes in tropical upwelling (Figure 4 and 9). In the all-forcing experiment, tropical upwelling accelerated until 1992 in both the mid-stratosphere and lower stratosphere (i.e. blue line rises in both panels of Figure 9), and this acceleration is driven largely by rising greenhouse gas concentrations, warming SSTs, and ozone depletion (Figure 10aceg). Since 1992, however, the residual circulation has decelerated at 50hPa and 70hPa (Figure 4a, 9a, and 10bd). The recent deceleration of we at 50hPa comes about due to competition between changing SSTs and declining ODS concentrations: changing SSTs lead to continual acceleration (i.e. the red curve continues to rise in Figure 9a and trends in Figure 10bd are weak), but ODS recovery leads to a slight deceleration (i.e. the gap between the red curve and green curve gradually decreases after 2000 in Figure 9a). Because of the eruption of Pinatubo, the deceleration trends starts in 1992 in the all forcing experiment (blue curve in Figure 9a) rather than in the late 1990s when ODS concentrations began to decrease (Figure 1a). For a start date of 1991, the deceleration trend at 50hPa is statistically significant at the 95% level in two ensemble members and in the ensemble mean (Figure 4a). At 100hPa (Figure 9b), on the other hand, the dominant forcing is SSTs, and we

30 continues to increase throughout the experiment (Figure 4b). The implications of these changes for ozone and temperature in the lower stratosphere are discussed in Polvani et al. (2016). Overall, the same forcings that control the age of mid-stratospheric air also control the residual circulation, and these forcings can explain the lack of acceleration of the deep branch of the BDC since 1992.

3.2.4 Summary of Key Forcing Agents

In summary, from 1960 to the late 1980s (the first half of the experiments), ozone depletion, rising GHG, and warming SSTs all led to a decrease in mean age in all regions of the stratosphere. Over the second half of the experiments (since the early 1990s), rising GHGs continue to lead to decreasing mean age, though the decrease is more prominent in the lower stratosphere.

5 However, declining ODS concentrations and the proximity of the start-date to the eruption of Pinatubo lead to an aging trend that is most prominent in the mid-stratosphere. The degree of compensation between these forcings is region-specific, and for the NH-midlatitude mid-stratosphere the volcanic effects and declining ODS concentrations dominate while in the lower stratosphere the SSTs and GHGs dominate. Hence, structural changes occurred in the BDC in our simulations.

4 Discussion of Observed Changes

10 There are no direct measurements of historical changes in the BDC. However, its past evolution can be deduced from trace gas measurements or from satellite data, and here we consider whether the modeled evolution of the BDC in GEOSCCM is consistent with these constraints.

4.1 Comparison with BDC Changes inferred from in-situ CO₂ and SF₆ concentrations since 1975

Balloon measurements of CO₂ and SF₆ concentrations are available from 1975, and this data does not provide evidence for
an acceleration trend in the mid-stratosphere Northern Hemisphere (NH), where mean age actually appears to have increased
(Engel et al., 2009; Ray et al., 2014). In particular, the mean age evolution in the figures of Engel et al. (2009) and Ray et al.
(2014) indicates aging since the late 1980s, with earlier changes less clear. As discussed in 3.1, a similar evolution is present in our simulations. In order to make the comparison more precise, we sub-sample the simulated mean age on the day of each flight analyzed by Ray et al. (2014). As daily three-dimensional fields of mean age were not archived from the simulations
we cannot map the simulated age into equivalent latitude space (as done by Ray et al., 2014). However, we expect that for the summertime midlatitude mid-stratosphere (the season of most of the balloon flights) the difference between equivalent latitude and geometric latitude is small, though the veracity of this assumption should be tested for future work. We show all three GEOSCCM members in order to estimate the internal variability in the model simulated mean age (see Figure 11).

GEOSCCM captures the mean age averaged over this period accurately: the difference between the observations and model for these datapoints is three months. A similar three month offset is evident when comparing GEOSCCM to the mean ages reported by Engel et al. (2009), which falls within the 6-month uncertainty in the observations (Engel et al., 2009). The value of mean age in other regions also agrees well with satellite-based estimates presented in Stiller et al. (2008).

GEOSCCM mean age lies within the error bar for most measurements. While the weak (non-significant) aging trend noted in observations since 1975 is not present in GEOSCCM, observed and modeled trends agree within the 95% uncertainty level.

30 Note that if we use the wider uncertainties reported by Engel et al. (2009), mean age in GEOSCCM agrees with all balloon flights and trends agree within the 90% uncertainty level. Over the recent period (since 1992), one of the three ensemble

members simulates a trend in close agreement with the observed trend $(0.12\pm0.1 \text{ years/decade for GEOSCCM and } 0.14\pm0.14 \text{ years/decade for observations})$. That being said, there is apparently less subseasonal and QBO variability in GEOSCCM than in the observations (and also the tropical leaky pipe model of Ray et al., 2014), and the trend towards older air is weaker in the GEOSCCM ensemble mean than in the observations. (Changes in mixing cannot be diagnosed from these simulations, but it

5 is conceivable that mixing could account for some of the difference). For future work, we will consider whether GEOSCCM is consistent with more recent tracer measurements.

4.2 Comparison with BDC Changes since 2002

10

While extreme caution must be exercised in interpreting a trend over such a short period due to the large stochastic variability in the atmospheric circulation, we now assess whether BDC changes since 2002 in GEOSCCM are consistent with observational constraints.

Vertically and latitudinally resolved changes in satellite measured SF_6 are available since 2002, and Haenel et al. (2015) infer mean age trends from this data (their figure 6). They find that mean age declines in the tropical lower and mid stratosphere south of the equator, and increases in the NH mid-latitudes and in the SH polar stratosphere. We show changes in annual averaged mean age from January 2002 to December 2011 in Figure 12. The model simulates younger mean age in the lower

- 15 stratosphere in all three ensemble members, but changes higher in the stratosphere are not robust among the various ensemble members. These intra-ensemble differences highlight the fact that one should not base any conclusions on the long-term behavior of the BDC on 10-year trends, as trends over one decade are strongly influenced by unforced (internal) variability. While not one of the ensemble members capture the inter-hemispheric dipole in the trends as suggested by satellite data (though individual integrations separately capture half of the dipole), we suggest that such a correspondence should not necessarily
- 20 be expected as the wave forcing of the BDC differs in any realization of the atmospheric state (e.g. see the discussion in Santer et al., 2008).

Aschmann et al. (2014) deduce changes in the tropical mass upwelling from changes in ozone, and they infer a lack of acceleration in mass upwelling above 70hPa since 2002 (no significant changes) and an acceleration below 70hPa of 11% per decade. Quantitatively similar behavior is evident in Figure 9 - upwelling increases at 100hPa by approximately 10% between

- 25 2000 and 2014, but decreases at 50hPa. Hence, the trend towards younger air in the NH lower stratosphere in GEOSCCM is consistent with observational constraints. Aschmann et al. (2014) further speculate that this slowdown of the upwelling above 70hPa is associated with the La Nina-like sea surface temperature trends over this period. This conjecture is supported by our modeling results: in the SST-only experiment, $\overline{w*}$ at 100hPa continues to increase over this period, but is largely flat at 50hPa (and also at 70hPa and 30hPa, not shown). Garfinkel et al. (2015) also noted that recent changes in SSTs (including the La
- 30 Nina-like sea surface temperature trends) lead to less planetary wave heat flux entering the stratosphere in midlatitudes, and changes in midlatitude planetary waves will impact the deep branch more strongly.

4.3 **Response to the eruption of Mt. Pinatubo**

The eruption of Mt. Pinatubo led to younger mean age throughout the stratosphere and enhanced tropical upwelling in our model. Figure 2 of Garcia et al. (2011) suggests that similar behavior is present in WACCM. Similar behavior is evident in the mid-stratosphere in SOCOL, though not in the lower stratosphere (Muthers et al., 2016). Diallo et al. (2012) also infer older mean age in the lower stratosphere following Pinatubo using ERA-interim data. However, changes in the residual vertical

5 velocity following Pinatubo differ among reanalysis product and for varying methodologies used for computing the residual vertical velocity (Abalos et al., 2015), and hence the actual response of the BDC to Pinatubo cannot be constrained by existing reanalysis data. Specifically, Diallo et al. (2012) use diabatic heating rates in ERA-interim to define $\overline{w*}$, and tropical diabatic heating rates show cooling after Pinatubo in the reanalyses (as the reanalyses do not assimilate aerosol burden, cf. figure 1 of Abalos et al., 2015) but warm in GEOSCCM due to increased shortwave heating from the aerosol plume (not shown). Future

10 work is needed in order to better constrain the response of the BDC to volcanic eruptions using observations.

It is worth noting that the response to the eruption of Mt. Pinatubo is stronger in the NH mid-stratosphere as compared to the SH (cf. Figure 6). The likely cause of this is as follows: Mt. Pinatubo is located at 15N, and in our experiments the majority of the aerosols stay in the NH. Therefore, the increased shortwave heating is stronger in the NH. Future work is needed in order to explore sensitivity to the details of the prescribed volcanic forcing in the model.

15 4.4 Summary of the Comparison of GEOSCCM to Observations

In conclusion, the evolution of the BDC in GEOSCCM is qualitatatively, and by most measures considered here quantitatively, consistent with observational constraints. There is a transition between declining mean age throughout the stratosphere before the late 1980s and regionally-specific changes in mean age afterwards (including the possibility of aging in the mid-latitude mid-stratosphere in the NH). The statement that is often made that climate models simulate a decreasing age throughout the stratosphere only applies over long time periods, and is not the case for the past 25 years when we have most tracer measurements.

5 Conclusions

20

The Brewer-Dobson Circulation (BDC) and its changes have important implications for both stratospheric and tropospheric climate as well as stratospheric ozone chemistry (SPARC-CCMVal, 2010; World Meteorological Organization, 2011, 2014;

25 Manzini et al., 2014). Hence, it is crucial to understand 1) the structure of historical changes in the BDC, and 2) the factors that lead to these changes. It is also important, for predicting future changes, to know how well models can simulate historical changes of the BDC as given by available observational constraints.

Analysis of a series of chemistry-climate model experiments of the period January 1960 through December 2014 yielded the following conclusions:

Over the full duration of the experiments (i.e for a start-date in 1960), we recover the result from previous modeling studies: anthropogenic climate change leads to acceleration of the BDC throughout the stratosphere. Ozone depletion, rising GHG concentrations, and warming SSTs all led to declining mean age in all regions of the stratosphere.



Figure 11. Mean age estimates in the data from figure 7 of Ray et al. (2014) in the NH mid-stratosphere between 20km and 25km and in the three all forcing integrations for the same days. The trend from 1992 to 2012 is included. Note that in computing the trends we do not include any information as to the difference in uncertainty among the measurement for the Ray et al. (2014) data because all sample in GEOSCCM have the same uncertainity, and we prefer to use the same statistical test for both data sources for consistency. This leads to overly conservative estimates on the uncertainity for the Ray et al. (2014) data.

- 2. Since the late 1980s, structural changes occurred in the BDC. The BDC accelerated in the lower stratosphere in the NH and tropics, but not in the mid-stratosphere. Specifically, since 1992, mean age has increased by 0.15 ± 0.1 years (95% confidence intervals) in the mid-stratosphere of the midlatitude NH and tropical mass upwelling has slowed down by 2% (statistically significant in two ensemble members). Hence, there is no inconsistency in trends between our model and available observations.
- 5

10

3. The source of this structural change is the time varying evolution of the forcing factors. While warming SSTs and rising greenhouse gas concentrations both lead to acceleration of the BDC (consistent with previous work), their influence is stronger in the lower stratosphere. In contrast, volcanic eruptions and ODS concentrations generally impact the deep branch more strongly. Declining ODS concentrations and the proximity of the start of declining ODS concentrations to the eruption of Pinatubo led to an aging trend since the early 1990s in the midlatitude NH mid-stratosphere. If internal atmospheric variability is taken into consideration, then the start-date of an aging trend in the midlatitude NH mid-stratosphere can be pushed back to 1988.

In light of these results, we wish to emphasize that if one wishes to capture observed historical changes in the BDC, careful attention must be paid to the start and end dates used for trend calculation and the forcings included in a model simulation.

15 In addition, it should not be expected that any single model experiment should capture the precise observed trend or neces-



Figure 12. As in Figure 2 but for changes in the BDC from Januar 32002 to December 2011 in the annual mean of the three all-forcing ensembles. Note that the contour interval differs from Figure 2.



Figure 13. mean age estimates in the data from Engel et al. (2009) in the NH mid-stratosphere between 30hPa and 5hPa and in the three all forcing integrations for the same locations and months (results are similar if we use 30hPa to 10hPa). The uncertainty for the observational estimates is taken from Engel et al. (2009), and the uncertainty of the model simulated mean age can be deduced from the intra-ensemble spread. The GEOSCCM mean age is offset by 3 months, i.e. the bias in the mean age (which is less than the 6 month uncertainty in the observed mean age as quoted by Engel et al., 2009).

sarily resemble a second integration using that same model, as the wave forcing of the BDC differs in any realization of the atmospheric state.

Many questions as to the historical changes in the BDC are left unanswered by this study. Diagnostic output necessary to compute the full age spectrum was not saved for these model experiments and hence we are limited in our ability to quantify
mixing changes, but it is conceivable that mixing changes contributed to recent observed mean age trends (Ray et al., 2014; Ploeger et al., 2015). Future work is needed in order to better constrain the response of the BDC to volcanic eruptions using observations, and to explore sensitivity to the details of the prescribed volcanic forcing in the model. Finally, a more quantitative comparison of model mean age from a range of modeling centers to recent in-situ observations is needed, in order to more firmly diagnose areas of agreement and disagreement between models and observations.

10 Author contributions. ???????

Acknowledgements. The work of CIG was supported by the Israel Science Foundation (grant number 1558/14) and by a European Research Council starting grant under the European Union's Horizon 2020 research and innovation programme (grant agreement No 677756). The work of DWW is supported, in part, by grants of the US National Science Foundation to Johns Hopkins University. VAA and LDO thank the NASA MAP program for their support. We also thank Eric Ray for providing data from Figures 7 and 8 of Ray et al. (2014) and for

5 help in interpreting balloon data and its uncertainties, and the four anonymous reviewers for their constructive criticsm. We also thanks those involved in model development at GSFC-GMAO, and Steven Pawson for initially suggesting the suite of GEOSCCM simulations analyzed here. High-performance computing resources were provided by the NASA Center for Climate Simulation (NCCS). Correspondence and requests for data should be addressed to C.I.G. (email: chaim.garfinkel@mail.huji.ac.il).

References

5

15

- Abalos, M., Randel, W. J., and Serrano, E.: Dynamical forcing of subseasonal variability in the tropical Brewer–Dobson circulation, Journal of the Atmospheric Sciences, 71, 3439–3453, 2014.
- Abalos, M., Legras, B., Ploeger, F., and Randel, W. J.: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979–2012, Journal of Geophysical Research: Atmospheres, 120, 7534–7554, 2015.
- Aquila, V., Swartz, W. H., Colarco, P. R., Pawson, S., Polvani, L. M., Stolarski, R. S., and Waugh, D. W.: Attributing changes in global stratospheric temperatures using model integrations with incrementally added single forcings, Journal of Geophysical Research: Atmospheres, in press, doi:10.1002/2015JD023841, 2016.
- Aschmann, J., Burrows, J., Gebhardt, C., Rozanov, A., Hommel, R., Weber, M., and Thompson, A.: On the hiatus in the acceleration of
 tropical upwelling since the beginning of the 21st century. Atmospheric Chemistry and Physics, 14, 12803–12814, 2014.
 - Bönisch, H., Engel, A., Birner, T., Hoor, P., Tarasick, D. W., and Ray, E. A.: On the structural changes in the Brewer-Dobson circulation after 2000, Atmospheric Chemistry and Physics, 11, 3937–3948, 2011.
 - Butchart, N.: The Brewer-Dobson circulation, Reviews of Geophysics, 52, 157–184, doi:10.1002/2013RG000448, 2014.
 - Butchart, N. and Scaife, A. A.: Removal of chlorofluorocarbons by increased mass exchange between the stratosphere and troposphere in a changing climate. Nature, 410, 799–802, doi:10.1038/35071047, 2001.
- Butchart, N., Scaife, A. A., Bourqui, M., de Grandpre, J., Hare, S. H. E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D., and Sigmond, M.: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation, Clim. Dyn., 27, 727–741, doi:10.1007/s00382-006-0162-4, 2006.
 - Butchart, N., Cionni, I., Eyring, V., Shepherd, T., Waugh, D., Akiyoshi, H., Austin, J., Brühl, C., Chipperfield, M., Cordero, E., et al.:
- 20 Chemistry-climate model simulations of twenty-first century stratospheric climate and circulation changes, Journal of Climate, 23, 5349– 5374, doi:10.1175/2010JCLI3404.1, 2010.
 - Carn, S., Yang, K., Prata, A., and Krotkov, N.: Extending the long-term record of volcanic SO2 emissions with the Ozone Mapping and Profiler Suite nadir mapper, Geophysical Research Letters, 42, 925–932, doi:10.1002/2014GL062437, 2015.

Chen, G. and Held, I. M.: Phase speed spectra and the recent poleward shift of Southern Hemisphere surface westerlies, Geophys. Res. Lett.,

- 25 34, L21 805, doi:10.1029/2007GL031200, 2007.
 - Colarco, P., da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, Journal of Geophysical Research: Atmospheres (1984–2012), 115, 2010.
 - Deser, C. and Phillips, A. S.: Atmospheric circulation trends, 1950-2000: The relative roles of sea surface temperature forcing and direct atmospheric radiative forcing, Journal of Climate, 22, 396–413, doi:10.1175/2008JCLI2453.1, 2009.
- 30 Diallo, M., Legras, B., and Chédin, A.: Age of stratospheric air in the ERA-Interim, Atmospheric Chemistry and Physics, 12, 12 133–12 154, 2012.
 - Diehl, T., Heil, A., Chin, M., Pan, X., Streets, D., Schultz, M., and Kinne, S.: Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and SO 2 from 1980 to 2010 for hindcast model experiments, Atmospheric Chemistry and Physics Discussions, 12, 24 895–24 954, doi:10.5194/acpd-12-24895-2012, 2012.
- 35 Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., et al.: Age of stratospheric air unchanged within uncertainties over the past 30 years, nature geoscience, 2, 28–31, doi:10.1038/ngeo388, 2009.

- Eyring, V., Butchart, N., Waugh, D. W., Akiyoshi, H., Austin, J., Bekki, S., Bodeker, G. E., Boville, B. A., Brühl, C., Chipperfield, M. P., Cordero, E., Dameris, M., Deushi, M., Fioletov, V. E., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Grewe, V., Jourdain, L., Kinnison, D. E., Mancini, E., Manzini, E., Marchand, M., Marsh, D. R., Nagashima, T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Shepherd, T. G., Shibata, K., Stolarski, R. S., Struthers, H., Tian, W., and Yoshiki, M.:
- 5 Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, Journal of Geophysical Research (Atmospheres), 111, doi:10.1029/2006JD007327, 2006.
 - Garcia, R. R. and Randel, W. J.: Acceleration of the Brewer–Dobson Circulation due to Increases in Greenhouse Gases, J. Atmos. Sci., 65, 2008.

Garcia, R. R., Randel, W. J., and Kinnison, D. E.: On the determination of age of air trends from atmospheric trace species, Journal of the Atmospheric Sciences, 68, 139–154, doi:10.1175/2010JAS3527.1, 2011.

10

- Garfinkel, C. I., Waugh, D. W., and Gerber, E. P.: The Effect of Tropospheric Jet Latitude on Coupling between the Stratospheric Polar Vortex and the Troposphere, Journal of Climate, 26, doi:10.1175/JCLI-D-12-00301.1, 2013.
- Garfinkel, C. I., Hurwitz, M. M., and Oman, L. D.: Effect of Recent Sea Surface Temperature Trends on the Arctic Stratospheric Vortex, J. Geophys. Res., 120, doi:10.1002/2015JD023284, 2015.
- 15 Gerber, E. P.: Stratospheric versus Tropospheric Control of the Strength and Structure of the Brewer–Dobson Circulation, J. Atmos. Sci., 69, 2857–2877, doi:http://dx.doi.org/10.1175/JAS-D-11-0341.1, 2012.
 - Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., Van Der Gon, H. D., Frost, G. J., Heil, A., Kaiser, J. W., Kinne, S., Klimont, Z., et al.: Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period, Climatic Change, 109, 163–190, doi:10.1007/s10584-011-0154-1, 2011.
- 20 Haenel, F., Stiller, G., von Clarmann, T., Funke, B., Eckert, E., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., et al.: Reassessment of MIPAS age of air trends and variability, Atmospheric Chemistry and Physics, 15, 13 161–13 176, 2015.
 - Haynes, P. H., McIntyre, M. E., Shepherd, T. G., Marks, C. J., and Shine, K. P.: On the 'Downward Control' of Extratropical Diabatic Circulations by Eddy-Induced Mean Zonal Forces., Journal of Atmospheric Sciences, 48, 651–680, doi:10.1175/1520-0469(1991)048, 1991.
- 25 Hegglin, M., Plummer, D., Shepherd, T., Scinocca, J., Anderson, J., Froidevaux, L., Funke, B., Hurst, D., Rozanov, A., Urban, J., et al.: Vertical structure of stratospheric water vapour trends derived from merged satellite data, Nature Geoscience, 7, 768–776, doi:10.1038/ngeo2236, 2014.
 - Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L.: Stratosphere-troposphere exchange, Reviews of Geophysics, 33, 403–439, doi:10.1029/95RG02097, http://dx.doi.org/10.1029/95RG02097, 1995.
- 30 Lean, J.: Evolution of the Sun's spectral irradiance since the Maunder Minimum, Geophysical Research Letters, 27, 2425–2428, doi:10.1029/2000GL000043, 2000.
 - Li, F., Austin, J., and Wilson, J.: The Strength of the Brewer Dobson Circulation in a Changing Climate: Coupled Chemistry Climate Model Simulations, Journal of Climate, 21, doi:10.1175/2007JCL11663.1, 2008.
 - Li, F., Waugh, D. W., Douglass, A. R., Newman, P. A., Strahan, S. E., Ma, J., Nielsen, J. E., and Liang, Q.: Long-term changes in stratospheric
- 35 age spectra in the 21st century in the Goddard Earth Observing System Chemistry-Climate Model (GEOSCCM), Journal of Geophysical Research: Atmospheres (1984–2012), 117, doi:10.1029/2012JD017905, 2012.
 - Lin, P. and Fu, Q.: Changes in various branches of the Brewer–Dobson circulation from an ensemble of chemistry climate models, Journal of Geophysical Research: Atmospheres, 118, 73–84, 2013.

- Lin, P., Ming, Y., and Ramaswamy, V.: Tropical climate change control of the lower stratospheric circulation, Geophysical Research Letters, 42, 941–948, 2015.
- Manzini, E., Karpechko, A. Y., Anstey, J., Baldwin, M., Black, R., Cagnazzo, C., Calvo, N., Charlton-Perez, A., Christiansen, B., Davini, P., et al.: Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling,
- 5 Journal of Geophysical Research: Atmospheres, 119, 7979–7998, 2014.

10

McLandress, C., Jonsson, A. I., Plummer, D. A., Reader, M. C., Scinocca, J. F., and Shepherd, T. G.: Separating the dynamical effects of climate change and ozone depletion. Part I: Southern Hemisphere stratosphere, Journal of Climate, 23, 5002–5020, doi:10.1175/2010JCLI3586.1, 2010.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M., Lamarque, J., Matsumoto, K., Montzka, S., Raper, S., Riahi, K., et al.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic change, 109, 213–241, 2011.

Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., and Eichmann, A.: The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna, Technical Report Series on Global Modeling and Data Assimilation, 28, https://gmao.gsfc.nasa.gov/pubs/docs/Molod484.pdf, 2012.

Monge-Sanz, B., Chipperfield, M., Dee, D., Simmons, A., and Uppala, S.: Improvements in the stratospheric transport achieved by a chem-

- 15 istry transport model with ECMWF (re) analyses: identifying effects and remaining challenges, Quarterly Journal of the Royal Meteorological Society, 139, 654–673, 2013.
 - Muthers, S., Kuchar, A., Stenke, A., Schmitt, J., Anet, J. G., Raible, C. C., and Stocker, T. F.: Stratospheric age of air variations between 1600-2100, Geophysical Research Letters, pp. n/a–n/a, doi:10.1002/2016GL068734, http://dx.doi.org/10.1002/2016GL068734, 2016GL068734, 2016.
- 20 Newman, P., Daniel, J., Waugh, D., and Nash, E.: A new formulation of equivalent effective stratospheric chlorine (EESC), Atmospheric Chemistry and Physics, 7, 4537–4552, doi:10.5194/acp-7-4537-2007, 2007.
 - Oberländer-Hayn, S., Meul, S., Langematz, U., Abalichin, J., and Haenel, F.: A chemistry-climate model study of past changes in the Brewer-Dobson circulation, Journal of Geophysical Research: Atmospheres, 120, 6742–6757, doi:10.1002/2014JD022843, 2015.

Oman, L., Waugh, D. W., Pawson, S., Stolarski, R. S., and Newman, P. A.: On the influence of anthropogenic forcings on

- changes in the stratospheric mean age, Journal of Geophysical Research: Atmospheres, 114, n/a–n/a, doi:10.1029/2008JD010378, http://dx.doi.org/10.1029/2008JD010378, 2009.
 - Oman, L. D. and Douglass, A. R.: Improvements in total column ozone in GEOSCCM and comparisons with a new ozone-depleting substances scenario, Journal of Geophysical Research: Atmospheres, doi:10.1002/2014JD021590, 2014.

Pawson, S., Stolarski, R. S., Douglass, A. R., Newman, P. A., Nielsen, J. E., Frith, S. M., and Gupta, M. L.: Goddard Earth Observing

- 30 System chemistry-climate model simulations of stratospheric ozone-temperature coupling between 1950 and 2005, Journal of Geophysical Research (Atmospheres), 113, D12103, doi:10.1029/2007JD009511, 2008.
 - Ploeger, F., Abalos, M., Birner, T., Konopka, P., Legras, B., Müller, R., and Riese, M.: Quantifying the effects of mixing and residual circulation on trends of stratospheric mean age of air, Geophysical Research Letters, 42, 2047–2054, 2015.

Plumb, R. A.: Stratospheric transport., J. Meteorol. Soc. Jpn., 80, 793–809, doi:10.2151/jmsj.80.793, 2002.

35 Polvani, L., Wang, L., Aquila, V., and Waugh, D.: The impact of ozone depleting substances on tropical upwelling, as revealed by the absence of cooling trends since the late 1990ss, Journal of Climate, submitted, 2016.

- Ray, E. A., Moore, F. L., Rosenlof, K. H., Davis, S. M., Boenisch, H., Morgenstern, O., Smale, D., Rozanov, E., Hegglin, M., Pitari, G., et al.: Evidence for changes in stratospheric transport and mixing over the past three decades based on multiple data sets and tropical leaky pipe analysis, Journal of Geophysical Research: Atmospheres (1984–2012), 115, doi:10.1029/2010JD014206, 2010.
- Ray, E. A., Moore, F. L., Rosenlof, K. H., Davis, S. M., Sweeney, C., Tans, P., Wang, T., Elkins, J. W., Bönisch, H., Engel, A., et al.:
- 5 Improving stratospheric transport trend analysis based on SF6 and CO2 measurements, Journal of Geophysical Research: Atmospheres, 119, 14–110, 2014.
 - Rayner, N., Brohan, P., Parker, D., Folland, C., Kennedy, J., Vanicek, M., Ansell, T., and Tett, S.: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset, Journal of Climate, 19, 446–469, doi:10.1175/JCLI3637.1, 2006.
- 10 Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An improved in situ and satellite SST analysis for climate, Journal of climate, 15, 1609–1625, doi:10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2, 2002.
 - Rienecker et al, M. M.: The GEOS-5 Data Assimilation System Documentation of Versions 5.0.1, 5.1.0, and 5.2.0, Technical Report Series on Global Modeling and Data Assimilation, 27, http://gmao.gsfc.nasa.gov/pubs/docs/Rienecker369.pdf, 2008.
 - Rosenlof, K. H.: Seasonal cycle of the residual mean meridional circulation in the stratosphere, J. Geophys. Res., 100, 5173–5192, doi:10.1029/94JD03122, 1995.

15

- Santer, B. D., Thorne, P., Haimberger, L., Taylor, K., Wigley, T., Lanzante, J., Solomon, S., Free, M., Gleckler, P., Jones, P., et al.: Consistency of modelled and observed temperature trends in the tropical troposphere, International Journal of Climatology, 28, 1703–1722, 2008.
- Shepherd, T. G. and McLandress, C.: A robust mechanism for strengthening of the Brewer-Dobson circulation in response to climate change: critical-layer control of subtropical wave breaking, Journal of the Atmospheric Sciences, 68, 784–797, 2011.
- 20 SPARC-CCMVal: SPARC Report on the Evaluation of Chemistry-Climate Models, SPARC Report, 5, WCRP-132, WMO/TD-No. 1526, http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/, 2010.
 - Stiller, G., Clarmann, T. v., Höpfner, M., Glatthor, N., Grabowski, U., Kellmann, S., Kleinert, A., Linden, A., Milz, M., Reddmann, T., et al.: Global distribution of mean age of stratospheric air from MIPAS SF 6 measurements, Atmospheric Chemistry and Physics, 8, 677–695, 2008.
- 25 Stiller, G., Clarmann, T. v., Haenel, F., Funke, B., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., et al.: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, Atmospheric Chemistry and Physics, 12, 3311–3331, 2012.
 - Strahan, S., Schoeberl, M., and Steenrod, S.: The impact of tropical recirculation on polar composition, Atmospheric Chemistry and Physics, 9, 2471–2480, doi:10.5194/acp-9-2471-2009, 2009.
- 30 Ueyama, R., Gerber, E. P., Wallace, J. M., and Frierson, D. M.: The role of high-latitude waves in the intraseasonal to seasonal variability of tropical upwelling in the Brewer–Dobson circulation, Journal of the Atmospheric Sciences, 70, 1631–1648, doi:10.1175/JAS-D-12-0174.1, 2013.

- 35 World Meteorological Organization: Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project Rep. No. 52, 2011.
 - World Meteorological Organization: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project Rep. No. 55, 2014.

Waugh, D.: Atmospheric dynamics: The age of stratospheric air, Nature Geoscience, 2, 14–16, doi:10.1038/ngeo397, 2009.

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

Dear Editor

We have seriously considered all of the comments from the reviewers, and have made substantial changes to the manuscript which we will upload as soon as instructed. These major changes include:

- **1.** A new supplemental section which compares the zonal wind and temperature trends in the different all-forcing members to trends in reanalysis data, in an attempt to ascertain whether the all-forcing integration showing a recent slowdown of the BDC is more consistent with other observational constraints.
- 2. Four new figures that highlight the statistical significance of the trends, explore sensitivity to the start date used for computing trends, and better quantify the contribution of each forcing factor to the NH mid-stratospheric mean age trends and tropical upwelling trends at a range of levels.
- 3. Removal of large portions of sections 4.2 and 4.3, and specifically the parts that were qualitative and could not be made more quantitative.

We hope that all of the concerns of the reviewers will be addressed. Below are point-by-point responses to the reviewers' comments.

Sincerely, Chaim Garfinkel (on behalf of the co-authors)

Interactive comment on "Time varying changes in the simulated structure of the Brewer Dobson Circulation" by Chaim I. Garfinkel et al.

Anonymous Referee #1

Received and published: 18 July 2016

The paper presents an assessment of the trends in stratospheric circulation in model simulations of GEOSCCM, and using specific forcing simulations attributes the trends separately to GHG, ozone depletion and volcanoes. The main points are 1) the model trends are highly sensitive to the period chosen, in particular they change

between before and after the late 1980s and 2) the model shows aging trends in the NH for the latter period, in qualitative agreement with observational estimates, which can be attributed to volcanic emissions and the end of ozone depletion. The paper is clearly written, the analysis is very relevant and well-timed and the results are interesting. Nevertheless, the paper would improve if the authors carefully address the following minor comments.

General comments:

1- The choice of the year separating the early and late periods is not clearly explained, and the authors change it throughout the paper somewhat arbitrarily (1988 in Fig. 3, 1992 in other parts). In particular, considering the trends from 1992 makes the starting point right at the peak influence of the Pinatubo eruption (Fig. 5). Since there are no observational trends reported from this year, I do not see the point in taking this year as a reference, besides enhancing the NH mid-stratosphere aging trends.

There are two complementary ways of defining the start date for the trends. The first is to blindly adopt the start date used in previous work (motivating 1988 as the start date), and the second is to choose the start date based on the physical mechanisms that underlie the evolution (motivating 1992 as the start date). We believe that both are justifiable.

That being said, we did jump back-and-forth between these two in the original version of the manuscript, and never explicitly justified our choice of 1992 as one of the possible start dates. We now explicitly list a start date of 1992 in the methods section, and more consistently use 1992 after we first start using it.

The net effect is that in the revised manuscript, we start with 1988 as the start date, but motivated by our results, we transition over to 1992 towards the beginning of the results section.

We also include two new figures where we explicitly explore dependence on the start-date for the trend calculation.



Figure R1: Trend in tropical mass upwelling at 50hPa and 100hPa for each allforcing ensemble member separately and for the ensemble mean for start-dates ranging from 1960 through 2002. The vertical line corresponds to the 95% uncertainty range for each trend.



mean age trend through 2014 for each start-date

All #1 All #2 All #3 All ensemble mean

Figure R2: Trend in mean age in the NH, tropics, and SH for each all-forcing ensemble member separately and for the ensemble mean for start-dates ranging from 1960 through 2002. The vertical line corresponds to the 95% uncertainty range for each trend. The high uncertainty in panel (c) arises because of the large reduction in the effective degrees of freedom due to the high autocorrelation of the residuals (cf. Santer et al 2008).

2. The age of air trends shown in Fig. 3 are quite similar to those reported for the ERA-Interim reanalysis and also consistent with balloon-derived measurements, but only for simulation 3. The other two members give completely different patterns in the mid-stratosphere. Hence, a major point to extract from Fig. 3 is that there is very large inter-member spread, due to the strong interannual variability and the relatively short time series. In his sense, the sentence in the abstract on P1 L5-6 is misleading, as it leads to think that the model is robustly producing positive trends in the NH mid-stratosphere since the late 1980s.

We agree that there is very large inter-model spread; the revised version of the paper now emphasizes this point in the conclusion section with the new sentence "In addition, it should not be expected that any single model experiment should capture the precise observed trend or necessarily resemble a second integration using that same model, as the wave forcing of the BDC differs in any realization of the atmospheric state." We also changed the sentence in the abstract to:

"does not accelerate in the ensemble mean, and even decelerates in one ensemble member"

Specific comments:

-P1 L 23 – P2 L2: and even there it is not fully true, near the tropopause there is strong isentropic mixing.

We changed this to "There is no reason to expect the two metrics to indicate a similar time-evolution"

-P2 L14-15: "Reanalysis data": this has only been shown for the ERA-Interimreanalysis, not for others.

This has been clarified

- P6 L1-2: The reanalysis studies are limited in their initial date because they need to run trajectories for 10 years before they output age of air.

This has been clarified

-P6 L9-11: Is this similar to averaging the trends of each member?

The mean trend is similar, but the uncertainty differs.

-P6 L31-32: Are the trends in the lower stratosphere significant?

Trends at 100hPa and 85hPa between 20degrees and 60degrees are significant in all three ensemble members. The new figure R2 shows this explicitly, and we plan on including this figure in the revised paper.

-P17 L 9-13: The trends derived from tracer observations are positive from 1975, not just from the late 1980s. This difference between the model and the observations should be mentioned around this part.

If you look closely at the figures of Engel et al and Ray et al, there is a short-lived freshening trend from the start of the datasets up until the mid-1980s. The trend using the Ray et al data from 1975-2012 is 0.04years/decade while for the shorter period 1992-2012 it is 0.14 years/decade. Calculating the trend starting in 1975 mixes together two very different regimes of the BDC (i.e. an ozone depletion induced acceleration with a slow deceleration).

-P 9 L27: This is directly shown for sub-seasonal variability in Abalos et al. (2014) JAS.

A citation has been added.

-P10 L3-4: I do not think the ozone depletion impacts only the deep branch of the circulation. For instance Abalos et al. 2015 JGR (Fig. 15) show an impact of ozone depletion in the lower stratosphere BDC in reanalyses, linked to enhanced EP flux propagation and convergence around 50 hPa in DJF. In fact, I do not see a significant difference between the All-forcing curves in the middle and lower stratosphere (Figs. 4c and 4f).

Yes, the reviewer is correct. We noted in the next paragraph that for the SH lower branch, ozone depletion matters, but we agree the language here was imprecise. This has been corrected

-Fig. 6: Why not show the same latitudinal range and levels as in Fig. 4 for consistency?

We pick 32S to 32 N because of its proximity to the turnaround latitudes, however results are similar though much noisier if we focus on 10S-10N like for figure 4 (see below). We show 50hPa and 100hPa as these are below the pressure levels we show for mean age. We show 100hPa instead of 85hPa as the acceleration of w* is more pronounced at 100hPa than at 85hPa (see below).

wstar, mm/s



Figure R3: Evolution of annual averaged w* in the tropics at 100hPa and 85hPa for a range of different definitions of the tropics.

In any event, in the revised version of the paper, we will show the evolution as the total mass flux between the turnaround latitudes (see figure R4). We also have added a new figure showing the trend at a range of pressure levels (see figure R5a below).


Figure R4: Annual averaged upward tropical mass from in the SST-only, SSTGHGODS, and All forcing ensemble means at 50Pa and 100hPa.



Figure R5a: trend in tropical upwelling mass flux in the lower stratosphere for the full duration of the experiments and for the period after 1992, for the ensemble mean of each experiment.

-P16 L2-3: This important point in the context of interpreting model trends and comparing them to observations, and could be further emphasized.

We agree, the parenthesis have been removed and we now repeat this point in the conclusions section

P17 L11: (including aging in the mid-latitude mid-stratosphere in the NH): this is only seen in one out of three members (see general comment 1).

We have changed to "including the possibility of aging"

- Fig. 8: None of the simulations present an inter-hemispheric dipole in the trends as the observations suggest (e.g. MIPAS). This should be pointed out.

We now point out this difference with MIPAS explicitly

Technical corrections: -P4 L21: has ! have corrected -P7 L28: simulate ! simulates corrected

-P7 L 30: relative to the other two?

Yes, relative to the other two, corrected

-P7 L30-31: this sentence is not clear, please rephrase it.

We removed this sentence

-P8 L 29: GHG gases ! GHG

corrected

-P9 L20: recover ! decrease

corrected

-P10 L4 remove "since"

corrected

- P15 L2: It is impossible to directly measure ! There are no direct measurements

corrected

Interactive comment on "Time varying changes in the simulated structure of the Brewer Dobson Circulation" by Chaim I. Garfinkel et al.

Anonymous Referee #3

Received and published: 3 August 2016

The study by Garfinkel et al uses a comprehensive set of hindcast simulations with a chemistry-climate model to assess past trends in the Brewer-Dobson-Circulation. They show that the circulation speeds up over longer time scales (i.e. 1960 to present), in agreement with past studies, but trends differ in different regions when considering only the last few decades (since the late 1980s). In particular, they show that their model is able to simulate positive AoA trends in the NH mid-stratosphere, which would be in agreement with tracer measurements there. They further show how different forcings contribute to the different trends, and emphasize that the flat or positive trends in the mid-stratosphere since 1990 are due to the

timing of volcanic eruptions and the recovery of ozone (declining ODS concentrations).

Overall, the paper presents interesting new results that contribute significantly to our understanding of the evolution of AoA over the past decades. The statements are supported by the results shown, and the paper is overall well written. However, the paper could be strengthened significantly by clearer and more quantitative description and presentation of the results, as detailed below. I suggest that the authors revise the paper according to the following comments.

General comments:

1.Overall, the paper would benefit from a more quantitative assessment of the results. The large number and design of the experiments is very well suited for the purpose, but it would be great to see more quantitative measures of the trends and contributions of different forcings. In particular, the availability of 3 ensemble members for each experiment is a perfect basis to estimate the robustness of trends, and I would strongly suggest that this is done.

Some examples: Fig.3: It is hard to deduce quantitative changes in the residual circulation from this Figure. I would suggest to include e.g. a Figure of the trend in tropical upwelling as a function of height to justify the statements in the text better. Also, while the text mentions at many points that trends are "significant", please indicate regions of significance in the Figures (e.g. in Fig. 3). Fig. 4: The ensemble members of the experiments can be used to assess the influence of internal variability (e.g. as seen in Fig. 5 for the all-forcing experiment). I agree that it might be too messy to add all the ensemble members in Fig. 4, but for example adding a shaded region to indicate the variability for each experiment would allow better to distinguish between forced differences and variability. Also, the trends discussed from Fig. 4 should be assessed more quantitative rather than only "by eye" - for example calculate the trends for each region as function of start year for each ensemble member and for the ensemble mean (see below).

We agree that we need to do a better job of conveying statistical significance. Note that our attempts at including significance information on the existing figures would lead to unreadable figures, and hence we will include four new figures in the revised manuscript. Figures R1 and R2 shown above will be included in the revised manuscript. These figures explicitly follow your last suggestion.

Figure R5a above follows another of your suggestions, and we plan to include it as well in the revised manuscript. It shows the trend in upwelling as a function of height to better justify the statements in the text. Figure R5b is similar to R5a, but for mean age in the NH mid-latitudes. It too helps quantify aspects in the text that previously were discussed mostly qualitatively.



Figure R5b: trend in tropical upwelling mass flux in the lower stratosphere for the full duration of the experiments and for the period after 1992, for the ensemble mean of each experiment.

Finally, we made figures showing the range of trends for all three integrations for all five experiments, but as the reviewer intuited there is quite a lot of internal variability especially for trends over the past few decades. See below for an example. We believe this figure isn't particularly helpful, and decided not to include it.



Figure R6: trend in mean age in the mid stratosphere in the NH for the full duration of the experiments and for the period after 1992, for each experiment. We do not intend to include this figure in the revised manuscript, as internal variability swamps out the forced signal for 23 year trends. This point is already made in Figures R1 and R2, so this figure is redundant.

2.In general, the idea to reconcile the current results with previous studies (Section 4) is a good addition to the paper. However, the discussion is mostly qualitative, and I was left puzzled at the end whether the agreement is good or not. I suggest to make the comparison more quantitative (e.g. add the trend as function of start year from the Ray/Engel observations to the Figure suggested above) where possible, and shorten the other parts.

In order to quantify the comparison to Engel/Ray, we would need 3D daily mean age. Unfortunately, this data wasn't saved (only daily lat vs pressure and 3D monthly mean). We deliberated whether to include this section at all in the paper given these data limitations, and decided to still include it. Namely, we don't expect the differences between 3D daily when mapped to equivalent latitude (as done by Ray et al) to differ too drastically from a conventional zonal average at least in midlatitudes in summer (the season for most of these flights). Given these limitations, we are somewhat reluctant to be too quantitative.

That being said, we now include the trend and the uncertainties for the period 1992 to the end of the balloon flights for all three ensemble members and for the two papers directly on the figure. One of the ensemble members agrees quite closely with the balloon measurements $(0.12\pm0.10 \text{ years/decade}$ for one of the ensemble members, and $0.14\pm0.14 \text{ years/decade}$ for the balloon measurements).



We have also significantly shortened sections 4.2 and 4.3, as suggested.

3. The authors argue that the recovery from the influence of ODS contributes to positive trends in AoA in the NH mid-stratosphere, but in the tropics and in particular in the SH, ODS leads to AoA "being flat" (page 9, line 18-22). As the influence of ODS concentrations on ozone and subsequently on dynamics is far stronger in the SH, I don't understand this result. If anything, I would expect that the effect of declining ODS is apparent first in the SH. Also, the "positive trend" in the NH appears mainly due to rising AoA in the last few years (from 2005), and this positive trend is already visible (though weaker) in the SSTE and SSTGHGE experiments, so I would think that this might have to do with anomalous SSTs, as you actually mention in Section 3.2.1. (1. SSTs). So why do you conclude that ODS are important for the positive trends, when SSTs might contribute to at least a flattening of the trend? Furthermore, as mentioned before, this discussion is somewhat qualitative. A suggestion would be to calculate trends for each experiment (and ensemble member, to get a measure of uncertainty due to internal variability) as function of start date (see above), to gain a more quantitative assessment of the contributions of different forcing to the trend.

It is worth noting that in the mid-stratosphere in the tropics and SH, there is a trend toward older air too. See figures R1 and R2. Specifically, there isn't a difference in mean age trends between the NH and tropics. In terms of the SH, there are two factors leading to non-significant trends in mean age. First, the correction of Santer et al 2008 for the autocorrelation of residuals leads to huge error bars. Second, the eruption of Pinatubo appears to have impacted the NH more strongly than the SH. Namely, the gap between the green and cyan curves is far larger in figure 4a than in figure 4b and 4c. Mt. Pinatubo is located at 15N and most of the aerosols stay in the NH in our experiments, so the impact on mean age in the NH is stronger than in the tropics or SH. This is now noted. That being said, we also note that SSTs alone lead to aging in the NH mid-stratosphere.

Figures R1, R2, R5a, and R5b, which will be included in the revised manuscript, will give a more quantitative assessment of the contribution of different forcings to the trend. As discussed above, we elect not to show each ensemble member for the 4 experiments that are more idealized than all-forcing, as the internal variability is quite large.

Finally, the last paragraph of section 3.2.1 was confusing and misleading, and it has been heavily revised:

"The same forcings that led to a statistically significant aging trend in the NH since 1992 in the ensemble mean and in two integrations also lead to similar significant aging in the tropics (blue curve in Figure 3b and 6b). Specifically, changes in this region are dominated by ODS concentrations, and since ODS concentrations decrease after the late-1990s (Figure 1c), mean age increases despite rising GHG concentrations. As for the NH, the earliest start-date of significant aging trends is in the early 1990s due to the eruption of Mt. Pinatubo. ODS concentrations also dominate the SH mid-stratospheric mean age evolution (Figure 6c), and since ODS concentrations decrease after the late 1990s, mean age is flat since 1992 (i.e. not significantly different from no trend) in the all-forcing ensemble as well. Note that the influence of the eruption of Mt. Pinatubo is weaker in the SH than in the NH; Mt. Pinatubo is located at 15N, and in our experiments the majority of the aerosols stay in the NH."

Specific comments:

Abstract, line 6: I would rather emphasize that the trend in NH mid-stratospheric AoA are flat/positive in the ensemble mean than that trends are positive in "a simulation" (=one ensemble member?), as the latter result is far less compelling.

We changed this to does not accelerate in the ensemble mean, and even decelerates in one ensemble member

Abstract, line 9: Please make the statement that changes in AoA and trop. upwelling "are similar" more quantitative (e.g. both have similar relative trends?)

We now clarify that both indicate a lack of recent acceleration

Abstract, line 11: ".. and is not necessarily the case.." (as it is in some regions).

"Necessarily" has been inserted

Page 2, line 2: "...increased or remained unchanged" (as Engel report insignificant changes).

"remain unchanged" has been inserted

Section 2 (Methods): I agree that it is not necessary to specify all details of the model (simulation) if described elsewhere. However, at least a comment on the resolution of the model and the upper model level would be desirable. Furthermore, are the boundary conditions and design of the simulations as those specified e.g. for CCMVal2 or CCMI?

We now specify the resolution, upper boundary, and number of layers (2x2.5, 0.01hPa, and 72 levels). The boundary conditions used for GHG and ODS are clarified in the paper as well – we use the standard CCMI and RCP4.5 forcings).

page 4, line 25: is the version of the model used in this study is no longer supported, or the one used in Oman 2009? Please specify.

The model version used for this paper is no longer supported. The NASA supercomputer facilities no longer support the older intel fortran compiler.

page 4, line 30: "long" = X years ? Please specify.

10 years, added

page 6, line 30: The upwelling trends are hard to detect from Fig. 3. Include e.g. a Figure with upwelling trends as function of height (see above).

Yes, we agree. Such a figure will be included (see figure R5a)

page 6, line 32: Please indicate regions of significance in Fig. 3 (see above).

All of our attempts to add significance directly on the figure led to a very noisy and unreadable figure. Instead of this, we show significance for select regions as a function of start date for the trend in figure R1 and R2, which will be included in the revised paper.

page 7, line 16: The changes in the residual circulation should be consistent with changes in wave driving in a self-consistent free-running model, so it's good to show this, but does not add that much information. It would be interesting to look into the mechanisms of why these changes in the wave flux changes occur, but I understand that this is beyond the scope of the paper. Also, I personally would find it more interesting to see the ensemble mean changes in circulation and wave fluxes for the different experiments rather than the individual ensemble member of the all-forcing simulation - I think more on the mechanisms could be learned from them.

We have looked at differences in the circulation response among the various integrations, but even the circulation response is tightly coupled to the wave fluxes so in the end we didn't learn much. The difference appears to just be unforced differences in wave fluxes. In a new supplemental section (see response to reviewer 5) we now address changes in the circulation associated with these changes in wave fluxes.

We have looked at changes in circulation and wave fluxes in each experiment individually, though the results are indeed beyond the scope of this paper. Aquila et al (2016) describes changes in temperature for each experiment.

page 7, line 30: The argumentation of "anomalies occurring in the same integration.." is very speculative and not very physical. Of course, you can always construct ensemble members that show all sorts of trends - however, the question is how likely this is. Given a certain internal variability plus the forced variations, the likelihood to gain a certain trend over a certain period can be calculated.

We removed this sentence and the associated argument.

Paragraph 4.1: I have to admit I'm left puzzled at the end of this section whether trends/variability of the current study and Ray/Engel are consistent or not. I suggest to add a more quantitative evaluation (see 2. general comment). Furthermore, can you please clarify what the difference between the Ray and Engel time series are?

As discussed above, we are unable to process the data in exactly the same way that Engel and Ray did, and hence we can't directly compare apples to apples. However, over the period 1992 to 2012, the trend in one of the ensemble members agrees quite closely with that from Ray et al. See the revised figure below.

We now clarify in the introduction that Ray et al argues that the large uncertainty estimates on the trends presented by Engel et al are overly conservative, and that this NH mid-stratosphere aging trend is statistically significant.



page 16, line 1-3: The reasoning here is a little weak. I would think what you want to say is that trends over one decade are strongly influenced by internal (unforced) variability, so that a free-running model is not designed to reproduce this trend. You can estimate this uncertainty from your ensemble members (e.g. trend is X+/- Y years/dec) and argue whether the observed trend lies within this uncertainty range. This makes more sense than saying that different aspects of trends are "captured by at least one ensemble member...".

We have removed the remark about at least one of the ensemble members capturing qualitatively the observed trends, and now focus discussion in this paragraph on whether the model can capture the inter-hemispheric dipole. page 16, line 5: Which time period is considered in Bönisch (2011), and do trends only agree in sign or also in magnitude?

They consider changes since 1980 in their paper, but mainly focus on changes that occurred after 2000. The decrease they find is about 1 month/decade over the full period since 1980 which agrees quantitatively with that in our model. However, we decided to remove discussion of the Bonisch paper from this section entirely.

page 16, line 6: "NH lower stratosphere aging trend": here AoA decreases, right? so "aging trend" is maybe not the correct wording?

Corrected

Section 4.2: In general, the discussion in this Section is very qualitative, and for me it is hard to follow which observed aspects are in agreement with the model results and which are not. Please clarify.

We have significantly shortened section 4.2, as including a detailed comparison to all previous studies that focus on the short period since 2000 is really a distraction from the main points of our paper. We have left in the figure showing that trends among the three ensemble members aren't robust, which is sufficient to convey that trends over such a short period are due more to internal variability than to any climate forcing. We now discuss this figure in more detail.

More generally, we believe that there is room for a community paper that quantifies exactly where models and observations disagree and agree, but our paper is not the right location for such a comparison with observations only taken over so short a period. In the final paragraph of the paper, we highlight that we (as a community) need to do a better job of comparing changes in the BDC as inferred from trace gases to those simulated by models.

page 16, line 25: Abalos shows that trends among reanalysis, and in particular among different methods to estimate w* strongly vary. E.g. tropical upwelling estimated from diabatic heating rates in ERA-Interim show a deceleration in the NH (see their Fig. 11, Fig. 14), consistent with Ploeger and Diallo (that use diabatic heating rates from ERA-Interim). Please clarify.

Yes, the reviewer is correct. However, note that Monge-Sanz et al 2012 use the divergence of the horizontal wind and still find aging in the NH mid-stratosphere. In any event, we have removed this paragraph from the paper.

page 17, line 1ff: I think you have to differentiate in which region you look for the response to volcanic eruptions, as the response is likely spatially not homogeneous. Note that while ERA-Interim does not assimilate aerosol data, it does assimilate temperatures and thus might be able to capture the influence of volcanos indirectly.

Indeed we agree that the response is likely not spatially homogeneous. We didn't emphasize this point before, but Pinatubo appears to have impacted the NH midstratosphere more strongly than other regions (at least in our experiments). We now mention this.

The issue with assimilating temperature and not aerosol burden (and thus radiative heating) directly is that the causality in the thermodynamic equation might be reversed. An increase in radiative heating in the stratosphere (as occurred after Pinatubo) can either lead to faster upwelling or to warmer temperatures in order to balance the thermodynamic equation. In GEOSCCM, it leads to both (the temperature response in GEOSCCM is shown in Aquila et al 2016).

In contrast, in a reanalysis product which only assimilates the warmer temperatures but doesn't know that diabatic heating has changed, one could imagine that the balance in the thermodynamic equation will be restored by decreasing w* (i.e. that the warmer temperatures are associated with anomalous descent and thus adiabatic warming). The reanalysis isn't informed that the cause of the warming is an anomaly in diabatic heating, and thus tries to maintain balance incorrectly. This is shown clearly by Abalos et al 2015.

Perhaps more crucially, Diallo et al 2012 use the diabatic heating rate methodology to define w^{*}. Abalos et al 2015 show that depending on which method is used to define w^{*}, the response to Pinatubo is different. In particular, the diabatic heating rate methodology shows deceleration of w^{*} associated with diabatic cooling. In GEOSCCM, there is anomalous diabatic heating after the eruption of Pinatubo (due to anomalous shortwave absorption). We have added this to the text.

The discussion of the effects of the eruption of Pinatubo now reads:

"We find that the eruption of Mt. Pinatubo leads to younger mean age throughout the stratosphere and enhanced tropical upwelling. Figure 2 of Garcia et al 2011 suggests that similar behavior is present in WACCM. Similar behavior is evident in the midstratosphere in SOCOL, though not in the lower stratosphere. Diallo et al 2012 also infer older mean age in the lower stratosphere following Pinatubo using ERA-interim data. However, changes in the residual vertical velocity following Pinatubo differ among reanalysis product and for varying methodologies used for computing the residual vertical velocity (Abalos et al 2015), and hence the actual response of the BDC to Pinatubo cannot be constrained by existing reanalysis data. Specifically, Diallo et al 2012 use diabatic heating rates in ERA-interim to define w*, and tropical diabatic heating rates show cooling after Pinatubo in the reanalyses (figure 1 of Abalos et al 2015) but warm in GEOSCCM due to increased shortwave heating from the aerosal plume (not shown). Future work is needed in order to better constrain the response of the BDC to volcanic eruptions using observations."

It is worth noting that the response to the eruption of Mt. Pinatubo is stronger in the NH mid-stratosphere as compared to the SH. The likely cause of this is as follows: Mt. Pinatubo is located at 15N, and in our experiments the majority of the aerosols stay in the NH. Therefore, the increased shortwave heating is stronger in the NH. Future work is needed in order to explore sensitivity to the details of the prescribed volcanic forcing in the model. "

page 19, line 27 ff: see above (comment on page 16, line 1ff): the likelihood to obtain a certain trend can be calculated when the internal variability is known, which you could estimate from the ensemble members. And/or the trend (quantitative!) plus the uncertainty can be calculated, and you can argue whether the observed trend lies within the modelled trend uncertainty. This would make for a much stronger statement then speculating about other ensemble members.

We removed this sentence

Technical:

page 10, line 4: "the aging since...": delete since

corrected

Fig. 6: 32S to 32N (replace S by N)

corrected

Referee 4

Review of "Time varying changes in the simulated structure of the Brewer Dobson circulation"

By C. Garfinkel et al.

Recommendation: accept after minor revision

This is a useful paper that shows that inferences about changes in the BDC derived from a state of the art numerical model are consistent with recent estimates based on observations of trace species. Further, the study is able to attribute changes in the BDC to various factors (SST, GHG, ODS, volcanoes) by comparing simulations that include one or more of these forcing factors. A few minor suggestions for changes and clarifications are detailed below.

Specific Comments (by page and line number):

(1, 20) "BDC has been deduced from ... average time for air parcel...": The BDC is not deduced from AoA (that is, AoA does not measure the vector circulation (v^*,w^*) ; instead, it is a proxy for the strength of the BDC, which, furthermore, needs careful interpretation.

The vector circulation (v*,w*) only represents the advective part of the BDC, and does not represent the total transport of mass in the stratosphere. This framework misses mixing, a very important element of the BDC. We therefore very intentionally do not wish to define the BDC as the vector circulation (v*,w*). These issues are discussed in the Butchart 2014 review paper. We opt to keep the language as is.

(1, 22) "differences": It is not clear how one establishes "differences" between a scalar field (AoA) and a vector field (the BDC). See previous comment.

See comments above.

(2, 14) "pronounced aging": I think this over-states the findings. For example, Engel et al.'s trend estimate is not significantly different from zero.

it is worth noting that the trend in Ray et al is significantly different from zero. Nevertheless, we have removed "pronounced".

(2, 17) "aging of the NH": It might be better to write "increasing age of air in the NH". Certainly, the mid-stratosphere of the NH is not getting older (except insofar as the Earth and all of us upon it are getting older.)

changed

(3, 11) "aging of the mid-latitude NH": Better: "aging of air in the mid-latitude NH". **changed**

(3, 20) "GEOSCCM": Does the model reproduce the QBO? Timing between the QBO and the seasonal cycle can introduce substantial low-frequency, stochastic variability in AoA. This is not relevant to long-term climate change but should be noted, especially if it is present in the model, since low-frequency variability could be misinterpreted as a trend in short records.

The model does spontaneously generate a QBO, though the QBO phase is randomized among the three ensemble members. This has now been noted in the text:

"Note that this model version includes a QBO, and that the QBO phase differs among the three ensemble members."

(4, 21) "interannual and decadal variability in SST": Are you saying that you used the smoothed version of SST in your simulations? This is not clear; and it is not a trivial point, as such stochastic, low-frequency variability will add "noise" to the time series and make it difficult to say much about trends over short periods of time (25 years or less, in my experience.)

We used the full SST fields. We include the smoothed version in the figure solely to highlight that the SST only experiment includes part of the climate change signal, but also includes e.g. ENSO variability. This has now been clarified:

"all experiments were forced with the full time evolving SST fields".

(7, 26) "statistically significant": In general, it would be useful to quote the 2-sigmavalues every time a trend in AoA is quoted. That way one can get a quick idea of the 95% significance of any trends mentioned. I understand why you may not

want to clutter the contour plots by, for example, shading significant regions, but it is easy enough in the text to quote a trend number ± 2-sd.

Figure R2, which we will include in the revised submission, will quantify the trends and their associated uncertainties. We prefer to present this information in the figures, and not in the text, in order to maintain readability of the text as much as possible.

(7,28) simulate →simulates **corrected**

(10, 2) "follows ozone depletion": It is plausible that ozone is responsible for BDC changes in the SH mid-stratosphere. But what about the NH, where ozone changes are minuscule compared to the SH, but where AoA also flattens out after 1990? (cf. Figs. 4a and 4c). This explanation seems incomplete to me.

If ozone depletion changes upwelling anywhere in the tropics, it can affect age of air in the NH, via mixing. This has now been clarified: "An ozone-induced acceleration of the BDC in the SH and tropics will affect mean age in the NH due to mixing."

(10, 24) "the same forcings": Isn't this trivially true? After all, AoA is a proxy for the strength of the circulation. Perhaps you had something more profound in mind, but I do not know what.

It isn't trivially true. One could imagine an exotic change in mixing that affects transport and hence mean age, but has no impact on the residual circulation. (Such a change has been proposed in order to explain part of the historical changes in the BDC.) We prefer to leave the text as is. Given that we don't explicitly diagnose mixing in these simulations, however, we prefer not to explicitly discuss the importance of mixing changes except in a qualitative sense.

(15, 1) "impossible to directly measure changes": It is not clear what this means. If you mean that (v*, w*) (and, therefore, changes in the BDC) cannot be measured, that is correct. But the diabatic BDC can be obtained from the thermodynamic + continuity equations, and given "good enough" data for a "sufficiently long" period, it should also be possible to detect trends in the BDC. I would think this is probably a more precise, and less ambiguous method than looking at trace gases.

We have changed to "There are no direct measurements of historical changes in the BDC."

(15, 19) "aging trend noted in observations": Garcia et al. (2011) have discussed why AoA trends derived from trace species may be misrepresented, even when the trends are corrected for growth rate, so one has to take these trends with a grain (or two) of salt.

The Ray et al 2014 paper adjusts for these nonlinearities in the growth, yet they still find significant aging of air. We now clarify this point in the introduction.

(15, 26) "extreme caution": Trends over 10 years are not very useful. They can be formally computed, but they are more likely to be influenced by stochastic variability than by any real long-term forcing. (In this regard, please clarify whether you used observed SST or smoothed observed SST to drive the model.)
I was going to suggest that you delete this section, but I think it actually serves a useful purpose in illustrating how these trends can be "all over the place", especially above ~70 hPa. Even in the shallow branch, the results are not very consistent among the 3 simulations shown in Fig. 8. So, a useful message from the present exercise is that one should not base any conclusions on the long-term behavior of the BDC on 10-year trends.

We agree, but for better or worse such short trends are published and interpreted as being meaningful. Indeed the entire point of this section was to point out just how "all over the place" such 10-year trends can be. We have now clarified that the reason for extreme caution is the large stochastic variability in the atmospheric circulation, and that one should not base any conclusions on the long-term behavior of the BDC on 10-year trends.

(16, 24) "fraught with danger": This is a bit too dramatic. I would think that inferring trends in the BDC from AoA trends derived from observations is even moreambiguous—yet we do it all the time!

This paragraph has been completely removed from the paper (though we do agree we were a bit too dramatic)

(16, 32) "Pinatubo": Better: "the eruption of Mt. Pinatubo". The volcano itself would be irrelevant, and unknown to most of us, had it not erupted.

changed

(17, 12) "only applies over long periods": This is a very useful point, which we often lose sight of, and I am happy to see it emphasized and illustrated by the results presented here.

thanks!

Interactive comment on "Time varying changes in the simulated structure of the Brewer Dobson Circulation" by Chaim I. Garfinkel et al.

Anonymous Referee #5

Received and published: 15 August 2016

General comment:

This paper presents GEOSCCM model simulations of the Brewer-Dobson circulation (BDC) and mean age over the past decades. The authors consider different ensemble members from their simulation which show a very different evolution of the BDC. One member even shows increasing mean age in the NH since 1988, similar to observed mean age trends. In the lower stratosphere, on the contrary, the BDC continues to accelerate. Hence, structural changes evolve in the BDC after 1988. ODS and volcanic eruptions are identified as the main forcing agents of these structural changes.

This paper addresses a very timely aspect of stratospheric dynamics and transport and will definitely be very interesting to a large readership. However, I have three major concerns which the authors need to consider before I can recommend publication.

First of all, we wish to thank the reviewer for their constructive criticism and very close reading of our manuscript. They have led to substantial improvements.

Major comments:

1) Model vs. real atmosphere:

As mentioned above, it is very interesting to see how strongly simulated mean age trends (on decadal time scales) depend on internal variability, and therefore considering the different ensemble members is a very good idea. However, to me the main question remains: which ensemble member is closest to reality and hence most reliable? And this question is not addressed in the paper.

We wish to clarify that these are free-running experiments run with fixed initial and boundary conditions only. It is not particularly informative to ask which ensemble member is most "reliable", as any difference among them is due to unforced variability. The motivation for studying historical changes in the BDC in free running climate simulations is *not* to form a best estimate of what actually occurred; for that purpose, reanalyses and/or nudged experiments are far better. Rather, the motivation is two-fold: one, future projections of the BDC can only be produced by free running climate simulations, and these projections are of limited value if the models can't generally capture the past evolution; two, the forcings that caused these historical changes can be systematically diagnosed.

That being said, in our response below we do attempt to answer the reviewer's question as to which ensemble member has circulation trends closest to the observed evolution.

Clearly, if started from different initial conditions the system evolves differently. For instance, there will likely be significant differences in the representation of the QBO between the three members. The authors try to discuss dynamical differences based on residual circulation and EP-flux in Sect. 3.1 (Fig. 3) - but in my opinion this discussion needs to be further substantiated. The following two main questions should be answered:

1)What exactly causes the differences in wave flux and residual circulation between the ensemble members (e.g., differences in the background flow)?

There is substantial chaotic, stochastic variability in atmospheric eddies. This large amount of stochastic variability is by definition unforced, and has no "cause" other than the chaos inherent in the Navier-Stokes equations. This large amount of stochastic variability underlies the entire premise of ensemble forecasting in operational use all around the globe.

While there certainly are differences in the background flow among the ensemble members, this needs to be the case if the wave fluxes differ. Hence, analyzing the background state isn't particularly enlightening (though see below for our attempt). Only on very short (submonthly) timescales can one hope to disentangle wave fluxes from the background state associated with them (e.g. Garfinkel et al 2012; Watson et al 2014 in the context of the QBO). 2)Which member has dynamical characteristics closest to the real atmosphere (e.g., a comparison of the QBO with observations could be enlightening)? If the conclusion of the paper should be "models can simulate trends generally consistent with observations" (P19, L12) it needs to be confirmed that also the underlying dynamics is consistent with available observations or reanalysis data (similar statement on P17, L9), such that the resulting age trend is not due to cancelling effects of errors in different modes of variability.

We have removed the words "generally consistent" throughout the paper, as it is vague and not enlightening.

The QBO phase in all three is randomized and does not follow observations in any of the three. Please see the figure below (the period before 1997 is cut off for clarity). For any given 5 year period, one ensemble member mimics observations more than the others, but the "best" ensemble member changes with time. In particular, the main difference among the ensemble members in terms of mean age is in the last five years of the experiments, and over these five years the QBO in none of the ensemble members matches that observed. Note that the red line (all #3) is the ensemble member that bears the strongest trend towards a slowdown of the BDC, though its QBO evolution closely resembles a different ensemble member.



Figure R7: Tropical zonal wind at 50hPa in the three all-forcing integrations and in the MERRA reanalysis.

We now copy and paste from the planned supplementary section, as it follows directly from the reviewers comment

"In the main body we find that one all-forcing ensemble member simulates an aging trend of mid-stratospheric air since 1988 consistent with observational constraints (labeled ALL #3 in Figures 2 and 5), while a second simulates freshening of air and an accelerated mid-stratospheric BDC (labeled ALL #1 in Figures 2 and 5). It is reasonable to ask which ensemble member is closer to reality as measured by other metrics. It is also reasonable to ask whether the difference in the wave fluxes underlying the different BDC trends leads to differences in temperature and zonal wind trends. We now explore these issues by comparing the trend in these two ensemble members to that in MERRA and ERAI reanalysis. As the difference in wave forcing between these ensemble members was pronounced mainly in the Southern Hemisphere, we focus on this hemisphere.

Figure R8 compares the temperature trends over the 1988 to 2014 period in each ensemble member to various reanalysis products in order to assess which integration has a trend in the thermal structure closest to observations.

Consistent with the trend towards more wave convergence in ALL #1, there is a warming trends over the pole. In contrast, the trend towards less wave convergence in ALL #3 (which allows the BDC to slow down) is reflected in cooling over the pole. More relevantly, there is an equator-to-pole gradient in the trends: in ALL #3, the pole cools more than the tropics, reflecting a slowdown of the residual circulation. In contrast, in ALL #1 the tropics cools more than the pole reflecting a speedup of the residual circulation.

These trends can then be compared to those in MERRA and ERAI. First of all, it is worth noting that the two reanalysis products differ as to the magnitude of the trend, though in both there is a tripole pattern in the vertical over the pole, with cooling in the upper and lower stratosphere and warming in the mid-stratosphere. Neither ensemble member captures this tripole pattern. However, both reanalysis products indicate a trend towards cooling upon averaging the trends over the entire SH extratropics, and only ALL #3 captures this feature. As these experiments are all run in climate mode (no nudging), there are differences in the wave fluxes and hence in the temperature trends, so there should not be the expectation that any ensemble member agrees fully with the actual observed trends. Figure R9 compares the zonal wind trends over the 1988 to 2014 period in each ensemble member to various reanalysis products. Consistent with the temperature trends, there is an easterly trend in ALL #1 and a westerly trend in ALL #3 in order to maintain thermal wind balance. ALL #3 compares very favorably to the trends in ERAI, though the large differences between MERRA and ERAI preclude any meaningful quantitative comparison.

The net effect is that ALL #3 (which simulates a slowdown of the BDC from 1988 to 2014) has circulation trends that more closely resemble reanalysis than ALL #1 (which simulates a speedup of the BDC from 1988 to 2014). However, the correspondence between ALL #3 and reanalysis is not full, but it is not reasonable to expect perfect correspondence as the wave fluxes in any given realization of nature differ. "

Figure R8: Trend in annual averaged temperature in the SH from 1988 to 2014 in the (a) MERRA and (b) ERAI reanalyses, and in the all forcing integrations with a trend towards a (c) accelerated BDC and (d) a decelerated BDC.











Figure R9: Trend in annual averaged zonal wind in the SH from 1988 to 2014 in the (a) MERRA and (b) ERAI reanalyses, and in the all forcing integrations with a trend towards a (c) accelerated BDC and (d) a decelerated BDC.

Moreover, I think the wording of the above conclusion is too strong: Only one of the considered ensemble members shows some consistency with observations – and further it is not clear why. If the authors want to make the statement that "BDC in GEOSCCM is generally consistent with observational constraints", both dynamical quantities (as mentioned above) and mean age need to be more thoroughly compared to available observations. Unless these comparisons are done, the main conclusion should be rephrased rather as: "current model uncertainties due to the representation of internal variability are so large that simulations may be consistent with available observations".

We wish to clarify that internal variability isn't a source of model uncertainty. It is an inherent part of the climate system. Even if only one ensemble member shows an evolution consistent with observational constraints, that tells you that models can capture the observed trends given the large amount of internal variability in the climate system, as part of the observed trend may be unforced. We now clarify this when we first introduce the model:

"This internal variability is not a source of model uncertainty; rather it is an inherent part of the climate system. If the BDC in one ensemble member, but not in the other two, evolves consistently with observational constraints, one can reasonably conclude that models can capture the observed trends if part of the observed trend was due to internal variability and was not forced."

More generally, the motivation for studying historical changes in the BDC in free running climate simulations is two-fold: one, future projections of the BDC can only be produced by free running climate simulations, and these projections are of limited value if the models can't generally capture the past evolution; two, the forcings that caused these historical changes can be systematically diagnosed. In contrast, the motivation is *not* to form a best estimate of what actually occurred; for that cause, reanalyses are far better. We realize we never stated this explicitly in the original submission, but we have now added this when we introduce the model as well:

"The motivation for studying historical changes in the BDC in free running climate simulations is *not* to form a best estimate of the actual historical evolution; for that purpose, reanalyses and/or nudged experiments are far better. Rather, the motivation is twofold: one, future projections of the BDC can only be produced by free running climate simulations, and these projections are of limited value if a model's simulation of the past is inconsistent with observational constraints; two, assuming the model is capable of following the observed evolution, the forcings that caused these historical changes can be systematically diagnosed by sequentially adding these forcings. " In this context, I don't agree with the commentary of Sect. 4.3 that "deducing trends in the BDC from reanalyses is fraught with danger", while "the modeled evolution of the BDC in GEOSCCM is consistent with available constraint" (P16, L24ff).

Our terminology "fraught with danger" was a bit too much, but in any event as you suggest below we have significantly shortened section 4.3 and removed this paragraph. The "consistent with available constraints" comment is no longer present in this section.

Mean age simulations based on reanalysis data have been extensively compared to available balloon borne and MIPAS satellite mean age observations and are consistent within observational uncertainties (e.g., Diallo et al., 2012; Ploeger et al., 2015). Furthermore, dynamical variability in the reanalyses is consistent with available observational constraints by definition. The 2002-2011 trend of ensemble member 3 shows aging in the NH similar to MIPAS (e.g., Haenel et al., 2015, Fig. 9) and reanalysis driven simulations. However, negative mean age trends in the SH as observed by MIPAS and consistently simulated by reanalysis-driven models are not simulated by GEOSCCM. Overall, I don't agree with the statement that GCMs are better for estimating decadal (!) trends than reanalysis and I recommend removing Sect. 4.3 (see also my specific comment regarding the representation of volcanic effects in reanalysis).

We never meant to claim that GCMs are better for estimating decadal trends than reanalysis data - we apologize for the confusing wording in the initial submission. We now discuss explicitly the NH-SH asymmetry in the trends as inferred from MIPAS.

"Vertically and latitudinally resolved changes in satellite measured SF_6 are available since 2002, and Haenel et al 2015 infer mean age trends from this data (their figure 6). They find that mean age declines in the tropical lower and mid stratosphere south of the equator, and increases in the NH mid-latitudes and in the SH polar stratosphere. We show changes in annual averaged mean age from January 2002 to December 2011 in Figure 12. The model simulates younger mean age in the lower stratosphere in all three ensemble members, but changes higher in the stratosphere are not robust among the various ensemble members. These intra-ensemble differences highlight the fact that one should not base any conclusions on the longterm behavior of the BDC on 10-year trends, as trends over one decade are strongly influenced by unforced (internal) variability. While not one of the ensemble members capture the inter-hemispheric dipole in the trends as suggested by satellite data (though individual integrations separately capture half of the dipole), we suggest that such a correspondence should not necessarily be expected as the wave forcing of the BDC differs in any realization of the atmospheric state (e.g. see the discussion in Santer et al 2008)."

Section 4.3 has now been removed.

2) Attribution of mean age trends:

Although the authors explain in the beginning that "mean age is an integrated measure of the total transport" and "only in the tropical lower stratosphere can be thought of as dominated by vertical advection" (P1, L24ff), the following analysis aims to directly link mean age variability with residual circulation (e.g., Fig. 3 and Sect. 3.1). It is known that mixing processes have a strong impact on mean age and its trends (e.g., Neu and Plumb, 1999; Garny at al., 2014) - so what about these effects? I think Fig. 3 can be very misleading in relating the mean age trend just to the residual circulation without including the additional mixing effects, and these likely matter. For instance, why is the residual circulation and wave flux trend in the NH (northward 30N and above 18km) almost the same for ensemble members 1 and 3, but the resulting mean age trends very different (negative vs. positive NH age trends for members 1 and 2)? I think these mixing effects need to be either analyzed or at least a more careful discussion is needed.

Unfortunately, the age spectrum diagnostic was not turned on when these simulations were first performed, and hence we cannot explicitly examine mixing. We therefore are reluctant to conclude anything as to mixing trends.

To answer your question regarding figure 3, we have added " Note that a difference in wave fluxes in the SH can influence mean age in the NH because mean age is an integral measure of transport, and thus changes in the BDC in the tropics due to wave flux changes in the SH can impact transport pathway into the NH."

Specific comments:

P4, L31: What diagnostics are included in the model? If some measure for mixing could be easily added, this would significantly strengthen the analysis presented here (see my Major comment 2).

The age spectrum diagnostic is the best diagnostic for mixing, but it was not turned on when these simulations were first performed. It is not possible to reconstruct this information. w* and v* were saved, as was the EP flux divergence and gravity wave zonal torque.

P8, L20ff: The discussion here is not entirely clear to me: Why is the "no change in mean age" caused by a "reduction in planetary wave flux entering the stratosphere"? Shouldn't this reduction cause a weakening circulation and increasing mean age? Overall, I have the feeling that the paragraph here is more a discussion than belonging to the results section, as it just discusses the presented mean age changes against the background of published literature.

Recent changes in SSTs have caused a tug-of-air of effects. Widespread warming of SSTs causes an acceleration of the BDC (many mechanisms have been presented as to why this must be), but the specific pattern of changes in SSTs over the past 35 years has led to a decline in planetary wave flux (as noted by Garfinkel et al 2015) which in isolation causes a deceleration. The net effect is that there is no change in the BDC in the SST only simulation. We have rewritten this paragraph to enhance clarity.

"The likely cause of this is a tug-of-war of opposing effects. On the one hand, gradual warming of the oceans in isolation leads to an acceleration of the BDC (Oman et al 2009). On the other hand, the spatial pattern of recent changes in SSTs have led to a decline in planetary wave flux (especially wave 1) entering the NH stratosphere in midlatitudes (Garfinkel et al 2015): the vertical component of the Eliassen Palm flux at 100hPa area averaged between 40N-80N declines in all three all forcing and in all three SST only experiments, and the decrease is statistically significant at the 95% level in the ensemble mean in both January through March (the focus of Garfinkel et al 2015) and in the annual mean. This decline in upward propagating midlatitude planetary waves at 100hPa impacts the deep branch more strongly (Plumb 2002, Ueyama et al 2013). Hence, it is not surprising that little change occurred over the last 30 years of the integration."

We seriously considered moving the paragraph to the discussion, but elected to keep it here.

P10, L5: "... factors that led to aging (Pinatubo ...)..." seems not an optimal choice of wording to me. The direct effect of Pinatubo is to decrease mean age (e.g., Fig. 4a). What you mean here is that this decrease of mean age due to Pinatubo causes a stronger aging trend, as Pinatubo is at the beginning of the considered period. Please improve the wording.

Now reads: "recovery from the Pinatubo eruption"

P15, L6ff (section 4.1): The sentence "GEOSCCM mean age lies within the error bar for most measurements, and thus is generally consistent with observations" seems too strong to me. Even if GEOSCCM age lies within the error bars of most of the observations, the model clearly underestimates the mean age after 2000. Further, I think Ray et al. (2014) mapped to 42N equivalent (!) latitude. Is your mean age also sampled at equivalent latitude or just latitude (as I read from the text)? Please clarify.

We acknowledge in the text that "the recent trend towards older air is weaker in GEOSCCM"

The reviewer is correct when saying that we calculate mean age at 42N, while Ray et al calculate at equivalent latitude of 42N. In order to transform to equivalent latitude, one needs 3D daily mean age. Unfortunately, this data wasn't saved (only daily zonal mean and 3D monthly mean). We deliberated whether to include this section at all in the paper given these data limitations, and decided to still include it. Namely, we don't expect the differences between 3D daily when mapped to equivalent latitude (as done by Ray et al) to differ too drastically from daily zonally averaged, especially in summer when most of these balloon flights occurred. However, given these limitations, we focus mainly on qualitative aspects of the agreement. We now explicitly note this difference in data processing. This difference notwithstanding, we have added the trends directly to the figure. Over the period of 1992-2012, one of the ensemble members agrees very closely with reanalysis ($(0.12\pm0.10 \text{ years/decade for}$ one of the ensemble members, and $0.14\pm0.14 \text{ years/decade}$ for the balloon measurements). See the figure below.



We have also significantly shortened sections 4.2 and 4.3, as suggested.

P16, L29: It was shown by Abalos et al. (2015) and Ploeger et al. (2015) that there is NO "apparent inconsistency" between increasing mean age in the NH between 2002- 2012 and an acceleration of the residual circulation in the long-term if mixing effects on mean age and the appropriate time period are taken into account. Please clarify what you mean here.

We agree that changes in mixing could explain this difference. In any event, this paragraph has been removed from the paper.

P17, L1: Diallo et al. (2012) showed increasing mean age after volcanic eruptions only at lower levels in the stratosphere around 19km, and this is indeed consistent with the GCM based results of Muthers et al. (2016) (see their Fig. 3). Hence, there is no "contrast" between the two papers - both are very consistent! The authors are right in saying that ERA-Interim does not assimilate aerosol data. However, parts of the volcanic aerosol effect is included in the reanalysis due to assimilating observed temperatures. And can we be sure that the representation of volcanic aerosol in climate models is correct (e.g., amount of injected aerosol, injection height, ...)? Why, for instance, don't we see an effect of Pinatubo in the SH (Fig. 4c/f) although Pinatubo's effect on temperature appears rather symmetric about the equator – even stronger in the SH (see Fujiwara et al., 2015, Fig. 5)? Hence, I again doubt that decadal trends from climate models are more realistic than from reanalysis-based simulations (see also my Major Comment 1).

We have modified this paragraph to more clearly note that Muthers et al and Diallo et al agree about the lower stratospheric response. We agree that future work needs to explore the sensitivity of the response to Pinatubo to the details of how the aerosol cloud is inserted; we now state this more clearly as well. Note that the eruption of Pinatubo does affect the SH as well, however the influence is stronger in the NH (the reason for this hemispheric asymmetry is likely because Mt Pinatubo is at 15N, and at least in our experiments most of the aerosols stay in the NH.)

That being said, we suggest that the vertical wind product used by Diallo probably isn't ideal if one wants to constrain the response to volcanos. The issue with assimilating temperature and not radiative heating directly is that the causality in the thermodynamic equation might be reversed. An increase in shortwave radiative heating in the stratosphere (as occurred after Pinatubo) can either lead to faster upwelling or to warmer temperatures in order to balance the thermodynamic equation. In GEOSCCM, it leads to both (the temperature response in GEOSCCM is shown in Aquila et al 2016 and the upwelling response is shown here and in Aquila et al 2013).

In contrast, in a reanalysis product that only assimilates the warmer temperatures but doesn't know that diabatic heating has changed, one could imagine that the balance in the thermodynamic equation will be restored by decreasing w* (i.e. that the warmer temperatures are associated with anomalous descent and thus adiabatic warming). The reanalysis isn't informed that the cause of the warming is an anomaly in diabatic heating, and thus tries to maintain balance incorrectly. This point is discussed by Abalos et al 2015.

Perhaps more crucially, Diallo et al 2012 use the diabatic heating rate methodology to define w^{*}. Abalos et al 2015 show that depending on which method is used to define w^{*}, the response to Pinatubo is different. In particular, the diabatic heating rate methodology shows deceleration of w^{*} associated with diabatic cooling. In GEOSCCM, there is anomalous diabatic heating after the eruption of Pinatubo (due to anomalous shortwave absorption). We have added this to the text.

Finally, we agree that un-nudged climate model runs are not as useful for constraining the actual historical changes in the BDC. We never meant to imply this.

The discussion of the effects of the eruption of Pinatubo now reads:

"We find that the eruption of Mt. Pinatubo leads to younger mean age throughout the stratosphere and enhanced tropical upwelling. Figure 2 of Garcia et al 2011 suggests that similar behavior is present in WACCM. Similar behavior is evident in the mid-stratosphere in SOCOL, though not in the lower stratosphere. Diallo et al 2012 also infer older mean age in the lower stratosphere following Pinatubo using ERA-interim data. However, changes in the residual vertical velocity following Pinatubo differ among reanalysis product and for varying methodologies used for computing the residual vertical velocity (Abalos et al 2015), and hence the actual response of the BDC to Pinatubo cannot be constrained by existing reanalysis data. Specifically, Diallo et al 2012 use diabatic heating rates in ERA-interim to define w*, and tropical diabatic heating rates show cooling after Pinatubo in the reanalyses (figure 1 of Abalos et al 2015) but warm in GEOSCCM due to increased shortwave heating from

the aerosal plume (not shown). Future work is needed in order to better constrain the response of the BDC to volcanic eruptions using observations."

It is worth noting that the response to the eruption of Mt. Pinatubo is stronger in the NH mid-stratosphere as compared to the SH. The likely cause of this is as follows: Mt. Pinatubo is located at 15N, and in our experiments the majority of the aerosols stay in the NH. Therefore, the increased shortwave heating is stronger in the NH. Future work is needed in order to explore sensitivity to the details of the prescribed volcanic forcing in the model. "

Figure 1, caption: Include a description of the smoothed curve in (a) in the caption. In the model simulation the version without smoothing is used, right? Please clarify in the text.

The caption has been clarified. The model uses the full SST fields, not the smoothed version. This has also been clarified.

Figure 3: The statistical significance of the trend is only mentioned in the text but not plotted. It would be better to plot the trend and its significance in the same figure (e.g. as additional shading).

Our attempts at adding shading led to an unreadable figure. Instead, we have added new figures that show significance as a function of start-date for each ensemble member for selected regions (see figures R1, R2, R5a, and R5b above).

Figure 6: It would be helpful to see also the SSTGHG case in the figure, to estimate the ODS-effect. And why not showing the net upwelling mass flux averaged between turn-around latitudes here? This should give a much more reliable measure of net upwelling than the flux averaged between fixed latitudes (and as far as I understand, the authors have calculated this already).

Even with three lines, the figure is already somewhat cluttered. We don't want to add a fourth. Figure R5a, which will be included in the revised manuscript, will allow one to infer the ODS effect. All of the w* figures have been updated to the net upwelling mass flux. Figure 7: It would be helpful to include also the linear trend values for both observed and GEOSCCM simulated mean age in the figure (or in the caption).

This has been done, see the figure above

Technical corrections:

P1, L11: I would better say: "...and is not NECESSARILY the case ... "

Changed

P1, L23: The for defining the residual circulation vertical velocity is not raised, it should read w. This occurs several times also at later places in the paper.

We have confirmed that the asterisk in w* is not raised in our source file.

P4, L20: CO₂ should not be in italics.

Changed

P4, L29: I would cite Hall and Plumb (1994) or Waugh and Hall (2002) here, for the calculation of mean age from the linearly increasing tracer.

Citation added

P6, L8: The number of years "n" should be upper case.

Changed

P7, L8: "...resolved waveS..."

Changed

P7, L28: "... one ensemble member simulateS..."

Changed
P8, L7ff Ray et al. (2010) also found that a "moderate increase in the horizontal mixing into the tropics" has to be assumed in their leaky pipe model (this is also related to my Major Comment 2).

We now mention that changes in mixing could be important in section 4, when we revisit observations more directly.

P10, L1: "...concentrations impact..."

Changed

P10, L4: "...aging in the deep branch..."

Changed

P16, L6: "...recent aging trend...' - I think you mean decreasing mean age?

Yes, this has been clarified.

P16, L16: "...but decreases at 70hPa..." I guess you mean 50hPa, right? At least this is the level shown in your Fig. 6.

This has now been changed to 50hPa.

Figure 1, caption: Blank missing before "total solar irradiance".

Changed

Figure 3, caption: "...decreasing mean AGE OF air..."

Changed

Figure 6, title: I guess you mean "32S to 32N", and not "32S to 32S"? Figure 6, caption: Blank missing after "...100hPa"

Fixed

P15, L13: "...age tracer (see Fig. 7)."

Changed