Reply to Anonymous Referee #1 (ACP-2020-947)

We thank referee #1 for the constructive comments on our manuscript and for his/her new ideas to improve this work. According to the referee's suggestions we now include some additional analysis to this paper (for details see below). Through these, we gained additional insight into the processes determining the ozone trends in the LS. Due to the new results we also reorganized the structure of the manuscript in the last sections (see new section 3.4 and 3.5). Moreover, note that we do not consider interannual correlations anymore, as we felt that not too much was learned here.

Below this reply, you find the revised manuscript considering all questions and comments. Additionally, we highlighted the changes of the manuscript for the pages 1-19, and attached them to the reply. From Section 3.4 on, the changes were so comprehensive, that it doesn't make sense to highlight them.

Major issues:

1.The discussion of a forced signal (driven by GHG vs. ODS changes) presented in Section 4 is somewhat lacking, especially given the facility with which the authors can bring in the results not just of the fGHG simulations that they have analyzed but also the fODS simulations that were also performed for the REF-C2 scenario. Both, for example, were employed in the study by Abalos et al. (2019) cited below. By explicitly looking at these simulations, in addition to the two others considered here, the authors can more quantitatively address the relative roles of ODS vs. GHG (and the linearity between their interactions). This is a reasonable request especially given that ODS themselves can alter the stratospheric circulation (see the impacts on upwelling documented in the second study listed below) and given that these experiments have already been performed

-> Thank you for this helpful idea. We now included the fODS simulations to Figure 8a and 8b. Note that the ozone MMM trend has slightly changed compared to the last version of Fig. 8, as we now only took the simulations into account that provide ozone for both the fGHG and as well as the fODS simulations. Moreover, we changed the structure of the text, and included a new Section (Sec. 3.5) on the forced trends and their attribution (which was previously a part of the discussion).

2. The discussion of the mechanism underlying the different ozone trends is a bit unsatisfying. Of course, most of this derives from focusing on a multi-model comparison for which it is (understandably) difficult to do a detailed budget analysis for each model. However, the authors have more information than they may realize. In particular, I would strongly encourage the authors to consider analyzing the "age of air" or "e90" tracers that were also carried in these integrations as these provide a description of the actual transport circulation changes simulated in the models (which may, or may not, be directly related to changes in upwelling). The lack of any passive tracer diagnostic is a bit discouraging and I think it's incorporation would add substantially to the discussion.

-> Thank you for this excellent suggestion. Following your comment, we now included analysis of the intermodel correlations between local ozone trends to AoA trends, and their correlation is

indeed very strong in the tropical to mid-latitude lower stratosphere (new Fig. 7a). Besides that, in order to distangle the different transport mechanism, we also include the correlation to residual circulation transit time (RCTT) and Aging by mixing trends (see Fig. 7b+c). The revised Section 3.4 describes the results and the additional insights from those analysis.

3. In contrast to the previous sections, I find much of the material in Section 4 to be qualitative and speculative. For example, it is, of course, true that intermodal differences in internal variability (contributed from the QBO and ENSO) can contribute to the spread in trends among the models. However, this is never explicitly shown (only described in generalities) and I think a basic analysis needs to be done by which, for example, the authors select two models with very different ozone trends over midlatitudes and then show their ozone composites with respect to different phases of ENSO and the QBO. How does the ozone variance contributed from these two modes vary across models? Is it large? This would be an easy calculation to do and could be provided as a supplementary figure. Without a more quantitative analysis, though, it is not clear what exactly is gained from this discussion, besides raising issues that have been discussed in previous studies.

-> The reviewer is right, in that the discussion here is only qualitative. However, it is not the scope of our study to investigate the role of natural variability (e.g. QBO or ENSO) for the spread in the LS ozone trends in detail – we only want to discuss our results in the light of what is known from literature. Therefore, this is part of the discussion section. To make the discussion character of section 4 clearer, we included to the text the following sentence: '' While it is not easily possible to test which of the above explanations is correct, in the following we will discuss their possible contributions to the diagnosed disagreement in the light of our results and of what is known from literature." (see p. 29, line 8). Moreover, we also added in the discussion paragraph "Representation of natural variability in models", that we leave the assessment of the representation of natural variability and its effects on ozone to future studies (p.32, line1).

4. It appears that one of the main results from this study is, per the conclusions, the fact that "in midlatitudes the observational trends are a rather extreme value of the models' distribution." I agree with the authors that this is an important conclusion and I think this is a nice finding from this study. However, I think the authors need to acknowledge that this was also the conclusion made in Orbe et al. (2020). …..

-> Agreed. The Orbe et. al 2020 paper was published just before we submitted our draft, thus we missed to include it at several instances. It is now cited at several places.

5. Quite a bit of attention is paid to the correlation between tropical ozone trends and midlatitude ozone trends. This is understandable, given that the two are plausibly connected, but Figure 3 does not really seem to support this. The correlation seems very small, no? I think the reader would find this relationship more convincing if the authors showed a figure showing this relationship for, say, a given model. In particular, does this relationship manifest by just considering interannual variability? What does the correlation between midlatitude and tropical ozone look like for individual years

within a given model? Without a stronger case it just seems like Figure 3 is exhibiting a very weak relationship.

-> We agree in that the relationship between tropical and mid-latitude ozone trends is weak. Indeed, as the reviewer stated, there was a certain expectation to find a relation (which is backed up by the inter-annual correlations we showed in the previous version of the manuscript, see old Section 3.5 and in particular old Table 3 that did provide the correlation of inter-annual variability for individual models). Given our new insights into the role of different transport pathways, and the insights from the single-forcing experiments (see new Sec. 3.4 und 3.5), we realized that the expectation of anti-correlation between tropical and extratropical ozone might be misleading. Therefore, we strongly de-emphasized this point throughout the manuscript, including also the removal of old Section 3.5. Instead we focus on the trends and their correlation to transport measures (see above, new Sec. 3.4). Nevertheless, we decided to keep Fig. 3 as it nicely illustrates the mutual distribution of tropical and mid-latitude trends, but revised the text accordingly (see p. 14, lines 6-11).

6. Page 18, Lines 7-26: A lot of ambiguity and potential for intermodel differences is described here as stemming from differences in the latitudinal extent of upwelling/downwelling between models. I certainly agree with this comment. However, there is a very straightforward solution. One could compare w* between models in such a way that accounts for intermodel differences in the turnaround latitudes of the BDC. In particular, it is possible that the fixed lattiude boxes considered here do not span the region of mean downwelling in every model owing to differences in the meridional extent of the BDC. Not accounting for this information, therefore, would lead to the misleading conclusion that the models somehow underestimate downwelling but, actually, this may not be the case since the model may simply have downwelling occurring at different latitudes. What happens when you redo your analysis to be more dynamically consistent in this regard?

->Thank you, that's a good point – we were aware of that problem and therefore decided to follow the reviewer's suggestion to define a dynamically consistent mid-latitude box by averaging the LS ozone column from the turnaround latitudes of the BDC to 50°N . We re-calculate the LS ozone trend for this box (see Tab. 2) and moreover we indeed find a stronger inter-model correlation of LS mid-latitude ozone trends to up- and downwelling trends for this dynamical box (for details see Section 3.4 and Fig. 6).

<u>Minor Points:</u>

-Page 6, Line 20: Are different ensemble members treated the same/given the same weight as different models? Shouldn't they be weighted in such a way that distinguishes between ensemble members versus distinct models? Perhaps that is what has been done but it is not clear in the present text, however.

-> See p.10, line 30: "Note that for the calculation of the MMM trend, we chose to weight all 31 simulations equally (i.e., not considering that some models have multiple ensemble members) because "the trend variations among ensemble members are as large as among the different models over this period ". However later on in Section 3.4 the MMM is now calculated as average

of ensemble-means from each model, as for the longer time-periods the forced trends outweight variability, so that the above argument does not hold anymore.

-Page 7, Line 13: “dynamical linear modeling” needs to be described here.

-> We added a brief description of the “dynamical linear modeling” now – see p.9 line 1-4.

-Page 8, Line 13: The Orbe et al. (2020) study also showed this discrepancy in the LS ozone trend between the observations and the models.

-> We include Orbe et al. 2020 to the citations.

-Figure 5: Can you add the observed trends in upwelling as well? This seems important. Of course, there may be differences between reanalyses but you can add, for example, estimates from MERRA-2 and ERA-Interim. This should be easy to do as you can use the TEM residual circulation estimates from the SPARC Reanalysis Intercomparison Project

-> We now provide the upwelling trends of ERA5 in Figure 2 and Figure 5 and Tab. 2.

-Page 24, Line 6: It is not clear to me what the discrepancy is here that you are claiming between the GEOSCCM results presented in that study compared to the ones the authors show in Figure 1. Please explain in more detail.

->As we restructured the text, we exclude that sentence now.

-The language throughout could be improved at various places. I have noted a few grammatical errors below but there are many others. I strongly encourage the main author to have all co-authors check for lingering language issues/typos.

-> We revised the language to the best of our capabilities, and took all comments below into account.

Technical Points:

-Page 1, Line 7: Please indicate a reference for CCMI

-> Done.

-Page 2, Line 6: “results from” -> “result from”

-> Done.

-Various paragraphs throughout are not indented which renders the formatting a bit awkward (e.g. Page 12, Line 5). Please fix.

-> Done.

-Page 12, Line 3: “depending on” -> “dependent on”

-> Done.

-Page 12, Line 11: Do you need to remove “not” in front of significant? This is confusing given that the next sentence implies that the trends are significantly related.

-> Done.

-Page 15, Line 3: The sentence starting with “We will show. . .” is not complete. -Page 15, Line 15: “evolve” -> do you mean “simulate”?

-> Done.

Analysis of recent lower stratospheric ozone trends in chemistry climate models

Simone Dietmüller[1], Hella Garny[1,2], Roland Eichinger[2,1,3], and William T. Ball[4,5,6]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Ludwig Maximilians Universität, Faculty of Physics, Institute for Meteorology, Munich, Germany
[3]Charles University, Department of Atmospheric Physics, Faculty of Mathematics and Physics, Prague, Czech Republic
[4]Institute for Atmospheric and Climate Science, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland
[5]Physikalisch-Meteorologisches Observatorium Davos World Radiation Centre, Dorfstrasse 33, 7260 Davos Dorf, Switzerland
[6]Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, TU Delft, Stevinweg 1, 2628 CN Delft, the Netherlands

Correspondence to: S. Dietmüller (simone.dietmueller@dlr.de)

Abstract. Recent observations show a significant decrease of lower stratospheric (LS) ozone concentrations in tropical and mid-latitude regions since 1998. By analyzing 31 chemistry climate model (CCM) simulations performed for the Chemistry Climate Model Initiative (CCMI, Morgenstern et al. (2017)), we find a large spread in the 1998–2018 trend patterns between different CCMs and between different realizations performed with the same CCM. The latter, in particular, indicates that natural variability strongly influences LS ozone trends. However none of the model simulations reproduces the observed ozone trend structure of coherent negative trends in the LS. In contrast to the observations, most models show a LS trend pattern with negative trends in the tropics (20°S-20°N) and positive trends in the northern mid-latitudes (30°-50°N) or vice versa. To investigate the influence of natural variability on historical LS ozone trends we analyze the sensitivity of observational trends and the models' trend probability distributions for varying periods with start dates between 1995 to 2001 and end dates between 2013 to 2019. Generally, modeled and observed LS trends remain robust for these different periods, however observational data show a change towards weaker mid-latitude trends for certain periods, likely forced by natural variability. Moreover we show that in the tropics the observed trends agree well with the models' trend distribution, whereas in the mid-latitudes the observational trend is typically an extreme value of the models' distribution. We further investigate the LS ozone trends for extended periods reaching into the future and find that all models develop a positive ozone trend at mid-latitudes and the trends converge to constant values by the period that spans 1998–2060. Intermodel correlations between ozone trends and measures of transport circulation trends confirm the dominant role of greenhouse gas (GHG)-driven tropical upwelling enhancement on the tropical LS ozone decrease. Mid-latitude ozone, on the other hand, appears to be influenced by multiple competing factors: an enhancement in the shallow branch decreases ozone, while an enhancement in the deep branch increases ozone and, furthermore, mixing plays a role here too. Sensitivity simulations with fixed forcing of GHGs or ozone depleting substances (ODSs) reveal that the GHG-driven increase in circulation strength do not lead to a net trend in LS mid-latitude column ozone. Rather, the positive ozone trends simulated consistently in the models in this region emerge from the decline of ODSs, i.e. the ozone recovery. Therefore, we hypothesize that next to the influence of natural variability, the disagreement of

modeled and observed LS mid-latitude ozone trends could indicate a mismatch in the relative role of the response of ozone to ODS versus GHG-forcing in the models.

1 Introduction

Stratospheric ozone is essential for protecting the Earth's surface from ultra violet radiation, which is harmful for plants, animals and humans. Human-made ozone depleting substance (ODS) emissions significantly reduced ozone concentrations for some decades after 1960. After controlling the use of ODSs by the 1987 Montreal protocol and later adjustments, however, ODS concentrations started to decline in the mid-to-late 1990s (e.g. Newman et al., 2007; Chipperfield et al., 2017). As a consequence, total stratospheric ozone is expected to recover in the future. Dhomse et al. (2018) have analyzed the recovery of stratospheric ozone mixing ratios of the CCMI-1 (Chemistry Climate Model Intercomparison project part 1) climate projection simulations. They found that the ozone layer is simulated to return to a pre-1980 ODS level between 2030 and 2060, depending on the region. However, they discovered a large spread among the individual models, which shows that there are many uncertainties in these projections. The evolution of stratospheric ozone in the 21st century does not only result from a decrease in ODS concentrations but also from an interplay between changes in both the atmospheric composition and the circulation (World Meteorological Organization (WMO) 2014). Increasing anthropogenic greenhouse gas (GHG) emissions (CO_2, CH_4, N_2O) leads to enhanced tropical upwelling and thereby to an acceleration of tracer transport along the stratospheric overturning circulation (e.g. Butchart, 2014; Eichinger et al., 2019). On the other hand, increasing GHGs also slows down ozone depletion through GHG-induced stratospheric cooling (e.g. Jonsson et al., 2004; Oman et al., 2010; Bekki et al., 2013; Dietmüller et al., 2014; Marsh et al., 2016) and emissions of CH_4 and N_2O additionally impact ozone through chemical processes (e.g. Ravishankara et al., 2009; Kirner et al., 2015; Revell et al., 2012; Winterstein et al., 2019).

In the recent years, a number of studies have analyzed observational records to identify ozone trends in the stratosphere (e.g. Harris et al., 2015; Steinbrecht et al., 2017; Weber et al., 2018). These studies consistently report an ozone recovery in the upper stratosphere after the turnaround of the ODS concentrations around the year 1998. In the lower stratosphere (LS), however, most observed ozone trends are not statistically significant for such a relatively short period due to large internal variability and instrumental difficulties (e.g. Steinbrecht et al., 2017). Subsequently, Ball et al. (2018) analyzed LS ozone trends from satellite data between 1998 and 2016 in detail making use of a dynamical (multiple) linear regression analysis. They identified a statistically significant decline of LS ozone between 60°S and 60°N in that period, of approximately 2 DU in the LS below 24 km of altitude. The implication was that the stratospheric ozone column was continuing to decline, because the LS ozone reduction more than offsets the positive trend in the upper stratosphere. Shortly afterwards Wargan et al. (2018) studied ozone trends in the reanalysis products MERRA-2 and GEOS-RPIT. In the tropics they detected a positive ozone trend in a 5 km layer above the tropopause and a negative trend at 7-15 km above the tropopause. Nevertheless, in the northern and southern mid-latitude LS they detected a negative ozone trend. As such, there are some similarities to the findings of Ball et al. (2018), but there are also quantitative differences, for example the positive trend in the 5 km layer or a missing overall statistically significant decrease in the column integrated ozone. Wargan et al. (2018) suggested that the negative mid-latitude trend might

be explained by enhanced isentropic transport between the tropical and mid-latitude LS. However, the recent study of Orbe et al. (2020) explicitly demonstrated that in the NH this mid-latitude ozone decrease is primarily associated with large scale advection. Furthermore, they showed that the observed changes in advection and in ozone are well within the range of model variability (gauged from one CCM). By means of using a chemistry transport model (CTM) and extending the analysis period to the year 2017, Chipperfield et al. (2018) suggested that the negative LS ozone trends are only a result of large natural variability. They showed that there was a strong positive ozone anomaly in 2017 which is driven by short term dynamical transport of ozone, and concluded that this points to large year-to-year variability rather than to an ongoing downward trend. However, an update of the data set which was used in Ball et al. (2018) showed that the large interannual variability alone cannot explain the entire trend in Chipperfield et al. (2018) (see Ball et al., 2019): the larger year-to-year variability in the SH was implicated to result from a non-linear interaction between the quasi-biennal-oscillation (QBO) and seasonal variability and despite this large variability the observed negative LS ozone trend remains.

To improve confidence in future projections of the ozone layer it is important to evaluate the skill of chemistry climate models (CCMs) in simulating the observed ozone trends over recent decades. A direct comparison between the CCM multi-model-mean (MMM) values and observational data showed that the ozone trend profiles of modeled MMM data agree well with observations, except in the lowermost mid-latitude stratosphere (SPARC CCMVal, 2010; WMO, 2018). The most recent study of Ball et al. (2020) investigated LS ozone trends of the 1998–2016 period in merged satellite data and compared them to the ozone trends in CCMs using the climate projection simulations of the CCMVal2 project. Similar to the observations, the CCMs showed a decline in LS ozone in the tropics, likely due to enhanced tropical upwelling, following from an increase in greenhouse gases (see e.g. Randel et al., 2008). In contrast to the observations, however, models do not show a decrease, but rather an increase in LS mid-latitude ozone. Ball et al. (2020) argue that these discrepancies in the LS between models and observations can possibly be explained by differences in the horizontal two-way mixing between the tropics and mid-latitudes, though they did not provide explicit evidence from the models (see also Wargan et al., 2018). The study suggested that the negative mid-latitude observational trend is caused by an intensification of two-way mixing (by analyzing effective diffusivity in reanalysis data). On the other hand enhanced downwelling of ozone-rich air to the mid-latitudes could consequently lead to a positive trend in the mid-latitudes. Apparently, the processes that determine mid-latitude LS ozone in models and observations are not understood so far.

In the present study, we seek to quantify whether the observed LS ozone trends lie within the suite of modeled trends. If yes, this would imply that the observed trend is just one realization of possible trends given within the large year-to-year variability. If not, this would imply that either models do not represent year-to-year variability correctly, or that there is a forced trend in the real world that is not adequately represented in the models. In contrast to the study of Ball et al. (2020) we are using the simulation data of a more recent inter-model comparison project (namely the Chemistry Climate Model Initiative, phase 1, CCMI-1) and analyze the ozone trends for a wider range of updated current state-of-the-art CCMs including all their ensemble simulations.

A brief description of the model simulations, of the observational data sets and of the methods used is presented in Section 2. In Section 3 we show our results. We provide a detailed comparison of ozone trends over the years 1998–2018 in different

CCM simulations and observations (Section 3.1). Here we focus on LS ozone trends, and we investigate how natural variability influences these LS ozone trends (Section 3.2 and 3.3). We link LS ozone trends with stratospheric transport trends (Section 3.4) and we investigate how ozone trends are forced by GHG and ODS emissions (Section 3.5). A discussion of the reasons for the disagreement in the LS mid-latitude ozone trends between models and observations and the conclusions follow in Sections 4 and 5, respectively.

2 Data and Methods

2.1 Models and Simulations

In the present study, we analyze the model output from 18 state-of-the-art CCMs from the Chemistry Climate Model Initiative phase 1 (CCMI-1, Morgenstern et al. (2017)). Tab. 1 lists all these CCMs together with their references, the forcing that underlies the sea surface temperatures (SSTs), and the simulation type considered. A detailed overview of all models that participated in CCMI-1 can be found in Morgenstern et al. (2017). We mainly evaluate the long term 'free running' simulations of CCMI-1 (REF-C2), as they span the time period 1998–2018. We do not use REF-C1 'free running' simulations of the recent past or the specified dynamics simulations (REF-C1SD), as they only span the period from 1998 to 2010. Moreover we want to point out that the specified dynamics simulations performed for CCMI do not represent stratospheric circulation better than the 'free running' simulations: Chrysanthou et al. (2019) compared stratospheric residual circulation among specified dynamic (SD) simulations and found that the spread in these simulations is even larger than in REF-C2. Furthermore Ball et al. (2018) showed poor agreement with the observed ozone trend for some selected SD simulations of CCMI. For the REF-C2 model simulations used in our study, all available ensemble members of the individual models are taken into account. The ensemble size of a certain simulation (if ensemble simulations were performed) is also given in Tab. 1 (brackets after simulations). Thus for the REF-C2 simulations, 18 models performed a total of 31 realizations (six models performed multiple ensemble members simulations). The REF-C2 simulations include hindcast and forecast periods spanning 1960–2100. They are all 'free running' simulations, thus each model simulation has its own internal variability. Note that REF-C2 simulations use a variety of different SSTs and SICs (sea ice concentrations), either prescribed climate model SST fields from offline model simulations (of the same or of a different model) or they are coupled to an interactive ocean and sea ice module. Moreover the representation of the QBO is different across the CCMs, with models having an internally generated QBO (e.g. MRI, EMAC-L90), nudged QBO (e.g. NIES, WACCM, SOCOLv3, EMAC-L47, EMAC-L47-o) or no QBO (e.g. CMAM, LMDZ). REF-C2 reference simulations follow the WMO (2011) A1 scenario for ODSs and the RCP 6.0 scenario (Meinshausen et al., 2011) for other greenhouse gases, tropospheric ozone precursors, and aerosol and aerosol precursor emissions. For anthropogenic emissions, the CCMI recommendation was to use MACCity (Granier et al., 2011) until 2000, followed by RCP 6.0 emissions. Besides the REF-C2 simulations we also consider the 11 sensitivity simulations with fixed greenhouse gases (fGHG) and with fixed ODSs (fODS) in our analysis. These sensitivity scenarios are both based on the REF-C2 simulation. However in case of the fGHG simulations CO_2, CH_4, N_2O, and other non-ozone depleting GHGs are held at their 1960 value, and so we are able to study the impact due to ODS concentration changes only (i.e. in the absence of GHG-induced climate change). In the case of the fODS simulations

the ODS concentrations are fixed to the 1960 level throughout the simulation. All models providing both of these sensitivity simulations are given in Tab 1.

Table 1. Overview of the CCMI simulations, analyzed for the present study. For the individual CCMs their reference(s), their SSTs and their available simulations (REF-C2, fGHG, fODS) are given. The numbers in brackets behind the simulations indicate the number of realizations of each REF-C2, fGHG or fODS simulation. Detailed information about the models' SSTs and the models' representation of the QBO are given in the supplement of Morgenstern et al. (2017).

CCMI Model	Reference(s)	SSTs	Analyzed Simulation
CMAM	Jonsson et al. (2004)	prescribed	REF-C2(1), fGHG(1), fODS(1)
	Scinocca et al. (2008)		
CESM1-WACCM	Solomon et al. (2015); Garcia et al. (2017)	interactive	REF-C2(4)*, fGHG(3), fODS(3)
	Marsh et al. (2013)		
EMAC-L90	Jöckel et al. (2010, 2016)	prescribed	REF-C2(1)
EMAC-L47	Jöckel et al. (2010, 2016)	prescribed	REF-C2(1)
EMAC-L47-o	Jöckel et al. (2010, 2016)	interactive	REF-C2(1)**
GEOSCCM	Molod et al. (2012, 2015)	prescribed	REF-C2(1)
	Oman et al. (2011, 2013)		
MRI	Deushi and Shibata (2011)	interactive	REF-C2(1)
	Yukimoto et al. (2011, 2012)		
SOCOLv3	Stenke et al. (2013); Revell et al. (2015)	prescribed	REF-C2(1)
NIWA-UKCA	Morgenstern et al. (2009, 2013)	interactive	REF-C2(5), fGHG(2), fODS(2)
	Stone et al. (2015)		
ULAQ	Pitari et al. (2014)	prescribed	REF-C2(3), fGHG(1), fODS(1)
HadGEM	Walters et al. (2014); Madec et al. (2015)	interactive	REF-C2(1)
	Hunke et al. (2010); Morgenstern et al. (2009)		
	O'Connor et al. (2014); Hardiman et al. (2017)		
UMUKCA	Morgenstern et al. (2009); Bednarz et al. (2016)	prescribed	REF-C2(2)
ACCESS-CCM	Morgenstern et al. (2009, 2013)	prescribed	REF-C2(3), fGHG(1), fODS(1)
	Stone et al. (2015)		
NIES	Imai et al. (2013); Akiyoshi et al. (2016)	prescribed	REF-C2(1), fGHG(1), fODS(1)
UMSLIMCAT	Tian and Chipperfield (2005)	prescribed	REF-C2(1), fGHG(1), fODS(1)
CHASER	Sudo and Akimoto (2007)	interactive	REF-C2(1), fGHG(1), fODS(1)
LMDz-REPROBUS	Marchand et al. (2012); Szopa et al. (2013)	interactive	REF-C2(1)
	Dufresne et al. (2013)		
CESM1-CAM4-Chem	Tilmes et al. (2016)	interactive	REF-C2 (3)

* The fourth ensemble of WACCM (WACCM-4) was provided by M. Abalos; ** EMAC-L47 simulations are not ensembles, as one simulation is with prescribed SSTs and one with interactive ocean

2.2 Observational data

For observations, we make use of the BAyeSian Integrated and Consolidated (BASIC) ozone composite that merges SWOOSH (Davis et al., 2016) and GOZCARDS (Froidevaux et al., 2015) through the BASIC method of Ball et al. (2017). The method was developed to account for artefacts in composite datasets that are a consequence of merging observations from different instruments that each have unique spatial and temporal observing characteristics. As a result, these artefacts can alias in regression analysis and bias, e.g., trend estimates (see examples in Ball et al. (2017)). BASIC composites aim to account for and reduce artefacts using an empirically driven Bayesian inference methodology, but it relies on the availability of already developed ozone composites. Here, BASIC$_{SG}$ has been extended to the end of 2019 using the latest versions of GOZCARDS, v2.20, and SWOOSH, v2.6. As such BASIC$_{SG}$ covers 1985-2019 as monthly mean zonal means on a 10° latitude grid from 60°S–60°N and over a pressure range of 147–1 hPa (∼13–48 km). BASIC$_{SG}$ was presented in Ball et al. (2018), and a sensitivity analysis of trends was applied to it in Ball et al. (2019), with examples of data artefacts that it addresses in the accompanying appendix and supplementary materials, respectively.

To obtain an observationally constrained estimate of tropical upwelling and extratropical downwelling mass fluxes, we use ECMWF's fifth generation of atmospheric reanalysis data, ERA5 (Hersbach et al., 2020). The mass fluxes are calculated from 6-hourly data on the reduced set of pressure levels.

2.3 Statistical Methods

In some parts of our analysis, and to make a robust comparison between multiple models and a single 'real-world' realization, i.e. observations, we form probability distributions to estimate the combined probability of the ozone trends from all REF-C2 models. To do so, we calculate the linear trend and the associated uncertainty using a least squares method for every simulation. Then, to build the trend probability distribution of the models, first one of the 18 CCMI models is randomly selected, assuming that the models are randomly uniformly distributed. In case the selected CCM provided ensemble member simulations, in a second step one of these members is randomly chosen, thus taking into account that ensemble members are treated differently than individual models. In the next step, the trend estimate ($t^{M_{i,k}}$) of the specific randomly selected CCMI model M_i with ensemble member k is calculated by randomly choosing an ozone trend value from the trends associated and assumed normal distribution $\mathcal{N}$, which is based on the mean $\mu_{M_{i,k}}$ and standard deviation $\sigma_{M_{i,k}}$ of the simulations linear trend. Thus we can write the trend estimate of the selected model simulation as: $t^{M_{i,k}} = \mathcal{N}(\mu_{M_{i,k}}; \sigma_{M_{i,k}})$. In order to take into account the uncertainty of the single observational dataset (σ_{obs}), we also add to the calculated model trend estimate a random estimate of the observational noise by taking the observational standard deviation of the linear regression coefficient. We repeat the above described procedure 50 000 times. With that we have a large sample of model trends and can build up a robust probability density function (PDF) of the REF-C2 ozone trends. From these estimated PDFs we can then estimate the probability of a given trend relative to the models. We derive a "probability of disagreement" between the observational and the modeled trend distribution by taking the central interval of the models' trend distribution with the observed trend value as threshold of this interval. To calculate this central interval we order the the 50 000 values from the REF-C2 trend distribution according to

their probability values and then sum up the ordered probability values until the value of the observed trend is reached. This probability value indicates our estimate of whether the observations agree with the models, i.e. high probability values indicate that a disagreement between models and observations is less likely due to chance.

2.4 Analysis Methods

We here provide a short description of our methodology to analyze transport processes, which follows the studies of Dietmüller et al. (2018) and Eichinger et al. (2019). Stratospheric mean AoA is defined as the mean residence time of an air parcel in the stratosphere (Hall and Plumb, 1994; Waugh and Hall, 2002). In the CCMs, the AoA tracer is implemented as an inert tracer with a mixing ratio that linearly increases over time as lower boundary condition. AoA is then calculated as the time lag between the local mixing ratio at a certain grid point and the current mixing ratio at a reference point.

The residual circulation transit time (RCTT) is the hypothetical age that air would have if it only followed the residual circulation, thus without processes such as eddy mixing or diffusion. RCTTs are calculated by backward trajectories on the basis of the Transformed Eulerian Mean (TEM) meridional and vertical velocities (referred to as residual velocities) with a standard fourth-order Runge-Kutta integration (Birner and Bönisch, 2011). The RCTT is then the time that these backward trajectories require to reach the tropopause from their respective starting point in the stratosphere. The RCTT differs from AoA because of resolved and unresolved mixing. In the stratosphere, this is due to the mixing of air between branches and the in-mixing of air from the mid-latitudes into the tropical pipe, which leads to recirculation of old air around the BDC branches. In global model studies, this effect has been named aging by mixing (AbM) and is interpreted as the difference between AoA and RCTT (e.g. Garny et al., 2014).

3 Results

3.1 Ozone trends over the period 1998–2018 in CCM simulations and observations

In this section we analyze the ozone trends of all 'free running' CCMI-1 simulations (REF-C2), including all ensemble realizations of each model, for the period 1998–2018 together with the observational data, BASIC$_{SG}$. We chose the period 1998–2018 to be consistent with the observational trend estimate in the ozone recovering phase as presented by Ball et al. (2018). Note that ODSs are declining in this period as a result of the Montreal Protocol and its amendments. By using the REF-C2 simulations we include a wide spectrum of SST variability in the different CCMs, as they use either an interactive ocean or prescribed SSTs from a coupled ocean-atmosphere model simulation (see Tab.1). Ozone trends are calculated by simple linear regression, using the monthly deseasonalized ozone time series. We refrain from excluding sources of variability such as QBO, ENSO (El Nino Southern Oscillation), solar cycle or volcanic eruptions in the regression analysis to capture the full range of variability of ozone trends over the given period. Hence our trend estimates have to be interpreted as resulting from both forced trends (e.g., via GHG increases and ODS decreases) as well as from natural and internal climate variability. In the following we compare the calculated ozone trend from the observational data to the trends presented in Ball et al. (2018, 2019, 2020) that used a

dynamical linear modeling (DLM) approach, which attempts to take natural sources of variability into account. In a nutshell, DLM has many similarities with ordinary least squares multiple linear regression (MLR), using predictor variables to account for some of the variability in the timeseries (e.g. solar variability, the QBO). Where DLM primarily differs from MLR is in allowing for a non-linear trend to be estimated and for the seasonal cycle to evolve with time, and therefore the shape of these terms is not predefined. For more details, see Laine et al. (2014) and Ball et al. (2018).

The panels of Fig. 1 show a latitude-pressure cross-section of the ozone trend for observations (first panel of Fig. 1) and all free running CCMI model simulations. Generally, the linear trend fit we perform on the BASIC$_{SG}$ data yields similar spatial patterns and magnitudes to those estimated in Ball et al. (2018) with the DLM approach (see their Fig. 1f). There are a few small differences, e.g., our linear trend fit results in larger positive trends in the upper stratosphere over the southern tropics of $\sim 1\%$, a slightly less negative trend in the northern hemisphere middle stratosphere ($<1\%$), and consistently large and negative trends close to 100 hPa in the tropics as opposed to a smaller and insignificant trend at around 10°S and over 100-80 hPa in the DLM estimate as shown by Ball et al. (2019). Most notably, linear trend calculations result in small positive trends (up to $\sim$3%) in the southern mid-latitude lower stratosphere, as opposed to overall negative but insignificant trends reported by Ball et al. (2019) in that region. However, the comparison reveals that the overall magnitude and trend pattern is also captured by the simple linear regression, i.e. it is not dependent on the exact method used to calculate the trends. Therefore, we proceed with using a linear fitting approach for the comparison between observations and CCMs, though the above caveats should be kept in mind when comparing with a full regression analysis using DLM (Ball et al., 2019).

Overall, large inter-model variability of the trends derived from the individual REF-C2 simulations (including all ensemble members) is revealed in Fig. 1. Nevertheless, a number of features can be identified that are consistent over most models and all their ensemble members. In the upper stratosphere (1-10 hPa) nearly all simulations consistently show an overall positive ozone trend. This ozone increase can be explained by the decrease of ODSs (see e.g. WMO, 2018) and by a slow down in ozone destruction rates as the stratosphere cools from GHG increases (see e.g. Portmann and Solomon, 2007), as will be further discussed in Sec. 3.5. This upper stratospheric ozone trend has been found for climate model simulations and for observational data in several studies before (e.g. SPARC CCMVal, 2010; Harris et al., 2015; Steinbrecht et al., 2017; Ball et al., 2018; WMO, 2018; Ball et al., 2020). However, in the lower stratosphere (30-100/150 hPa) we find a wide spread in the ozone trends among the CCM simulations over recent decades. Many REF-C2 simulations exhibit negative trends in the tropical LS, and they are comparable to the observational trend in magnitude and structure. In agreement with earlier studies (e.g. WMO, 2018; Orbe et al., 2020), we will show in Section 3.4 that this tropical ozone decrease is related to enhanced tropical upwelling in a warmer climate. However, there are also simulations showing a positive LS ozone trend in the tropics (i.e., GEOSCCM, SOCOLv3, NIWA-1, WACCM-3/4, CAM4-1/2, LMDZrepro, HadGEM; note that the number of the ensemble run is denoted with -1, -2 and so on). At northern and southern mid- and high-latitudes most simulations exhibit a positive trend, but with a pronounced intermodel spread. Only a few simulations show negative trends in either northern or southern mid-latitudes (e.g. GEOSCCM, WACCM-3, WACCM-4), but it is important to point out here that none of the 31 simulations reproduces the observed negative ozone trend pattern with an ozone decrease covering the tropical belt and extending to the mid-latitude (50°S-50°N), as shown in the upper-left panel and previously in Ball et al. (2018, 2019). This discrepancy in the

LS ozone trend between observations and models has been reported before (e.g. ozone trends, based on CCMI simulations, (WMO, 2018; Orbe et al., 2020), and in comparison to CCMVal-2 simulations (Ball et al., 2020)). For CCMs that provide multiple ensemble members (WACCM, NIWA, ULAQ, ACCESS, CAM4 and UMUKCA), we also identify a large ensemble spread in the simulated LS ozone trends. For example in WACCM two ensemble members simulate positive tropical ozone trends, while the two other members simulate negative tropical ozone trends. In WACCM (as well as in NIWA and CAM4), the coupled ocean allows for differences in the SST variability between the ensemble members, possibly explaining the large spread in tropical ozone trends. However, as is also the case for models with prescribed SSTs (ACCESS, ULAQ, UMUKCA) that exhibit a large spread between the simulations, the SST variability is not the only reason for the different trend pattern, as was similarly reported and discussed by Ball et al. (2020) for CCMVal-2 models. The large spread in LS ozone trends between ensemble members is further in agreement with the study of Stone et al. (2018). They used a nine member ensemble of a 'free running' CCM simulation (CESM1-WACCM) and showed that LS ozone trends over the years 1998–2016 are characterized by large internal variability, with e.g. the LS ozone trend ranging from +6% to -6% per decade. But note, again, that none of these ensemble members showed the coherent decrease in ozone in the tropics and extratropics as found in observations (Ball et al., 2020).

Following this qualitative discussion on the spread in the ozone trend pattern between the CCM simulations, we now turn to the LS ozone trends with a more quantitative comparison of the apparent inconsistencies between observations and CCMs. We calculate the trends of the deseasonalized LS ozone columns for the period 1998–2018 in two regions: the inner tropics (20°N-20°S) and in the northern mid-latitudes (30°N-50°N). We choose the northern mid-latitude band 30°N-50°N for direct comparability with the study of Ball et al. (2020). The pressure range of the lower stratosphere was taken to be 30–100 hPa for the tropics and 30–150 hPa for the mid-latitudes, to take into account the differences in latitudinal tropopause heights. Trends and their uncertainties (represented by the 90% confidence interval of the linear slope) are shown for each of the 31 available REF-C2 simulations of 18 different CCMs in Fig. 2. We decided to focus on the northern mid-latitudes here, because the SH mid-latitude trends are likely more strongly influenced by the large chemical depletion of ozone within the polar vortex. We will come back to the LS ozone trends of the southern mid-latitudes in Section 3.5.

In the tropics about half (42%) of the REF-C2 simulations show a significant decrease, about the same (42%) show a non-significant change, and about 15% a significant increase in the integrated tropical LS ozone column. Note that significance is defined as the non-overlap of the error bars (90% confidence interval) with the zero trend. The resulting MMM ozone trend (see red bar on right of Fig. 2) is negative (-0.37 DU/dec), but it is insignificant due to the considerable spread among the different models. The 25th-75th quantile of the distribution ranges from -1.12 to 0.20 DU/dec (see edges of box on the right of Fig. 2). Note that for the calculation of the MMM trend, we choose to weight each of the 31 simulations equally (i.e., not taking into account that some models have multiple ensemble members) because the trend variations among ensemble members are as large as among the different models over this period.

The observed tropical LS ozone trend of -1.07 DU/dec is statistically significant at the 90% level. Thus the observed tropical trend is more strongly negative than the MMM trend, but lies within the 90% confidence interval of the MMM trend ([-1.76 DU/dec; 1.03 DU/dec]).

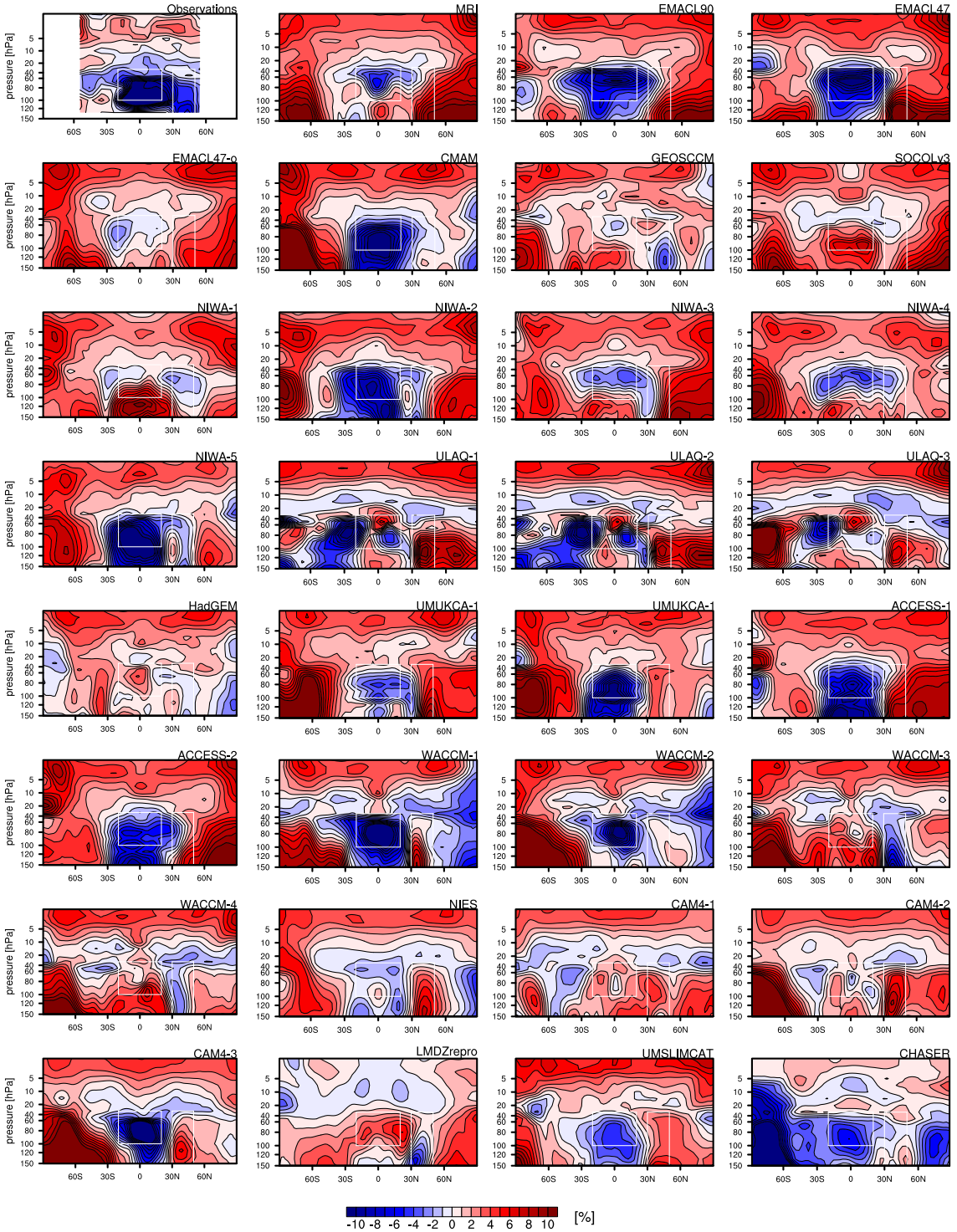

Figure 1. Latitude-pressure cross section of the ozone trend over the period 1998–2018 for the observational data set BASIC$_{SG}$ and for all CCMI REF-C2 simulations. Trends are given as relative ozone changes over the whole time period. Boxes illustrate the regions selected to integrate ozone in the LS for trend comparisons later in this study, i.e. in the tropics (20°N-20°S, 30–100hPa) and in the northern mid-latitudes (30°N-50°N, 30–150 hPa).

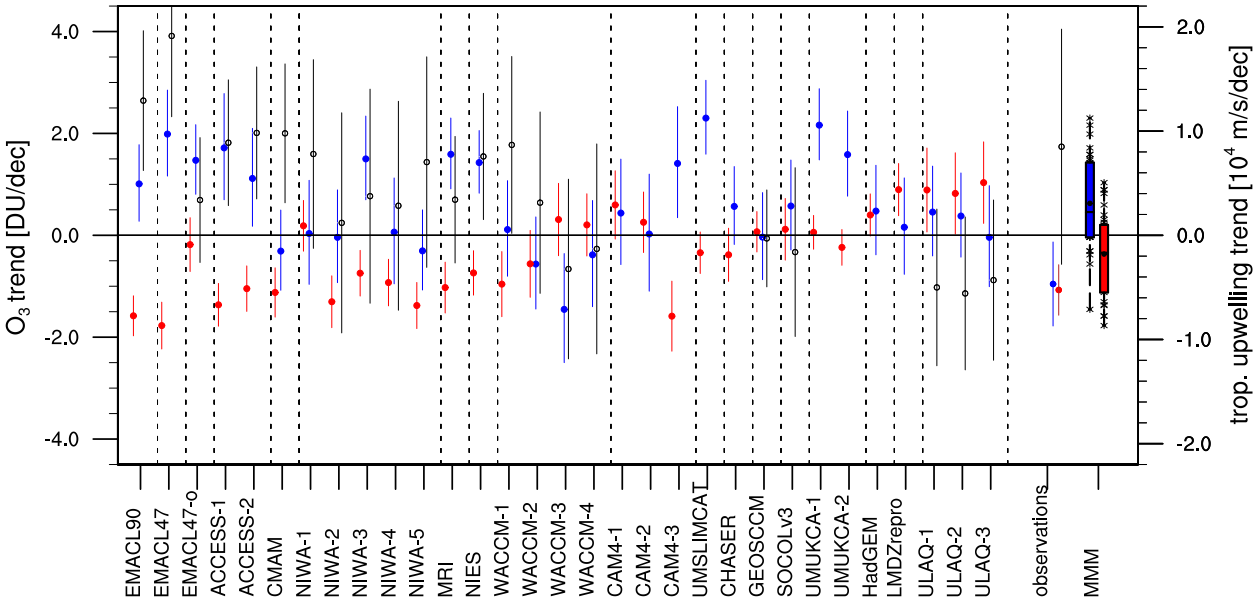

Figure 2. LS ozone trends and their uncertainties in the tropics (20°N–20°S, red dots) and northern mid-latitudes (30°N-50°N, blue dots) together with tropical upwelling trend (black circles, for all simulations providing TEM diagnostics) for the period 1998–2018 for all REF-C2 simulations. Dashed lines separate the individual models. Moreover, observational trends (1998–2018) and multi-model mean trends are given. Observational data for ozone are taken from BASIC$_{SG}$, and for tropical upwelling from ERA5 reanalysis. Error bars associated with each LS ozone trend represent the 90% confidence intervals. The multi-model mean trends are shown as boxplots: the black solid line in the box indicates the median, the black point the MMM and the colored box ranges from the 25th to the 75th percentile of the trends. Crosses denote trends of individual model simulations not lying within the box.

In the northern mid-latitudes less than half (40%) of the REF-C2 simulations show an increase in the LS ozone column, while the remaining 60% of the simulations show a non-significant change (either positive or negative). There is only one simulation (WACCM-3) that shows a significant decrease in the mid-latitude LS ozone column, and in this simulation the tropical ozone trend is positive (but not significant). The resulting MMM trend in the northern mid-latitudes is positive (+0.63 DU/dec) with a high inter-model spread: the 25th to 75th quantile of the distribution ranges from -0.04 to 1.42 DU/dec. Note here, that the observational trend (-0.96 DU/dec) lies outside the 90% confidence interval of the MMM trend in the mid-latitudes ([-0.91 DU/dec; 2.16 DU/dec]).

Fig. 2 also reveals that over the years 1998–2018 more than half of the model simulations have a dipole trend pattern in the LS ozone column, i.e. the sign of the tropical ozone trend is opposite to that in mid-latitudes. This trend pattern with

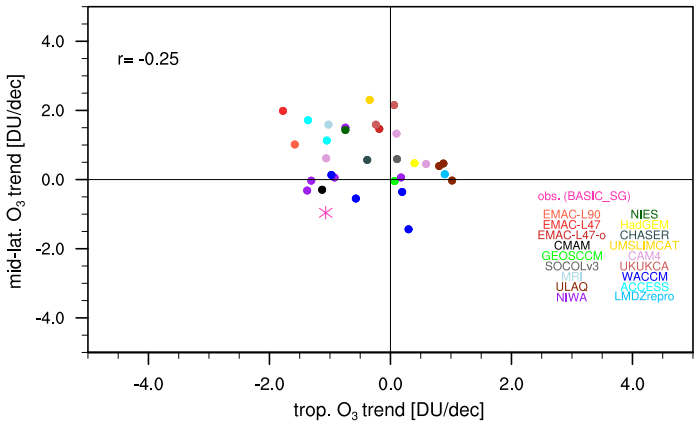

Figure 3. Inter-model correlation between tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trends, calculated over the period 1998–2018 for 31 CCMI REF-C2 simulations. All ensemble members of a particular model are shown in the same color. The observational ozone trends (BASIC$_{SG}$) are also included in here as a star.

.

negative LS ozone trends in the tropics and positive LS ozone trends in the northern mid-latitudes can be found for almost half the simulations (45%), and a trend pattern with a positive ozone trend in the tropics and negative trend in the northern mid-latitudes is found in 13% of the simulations. The remaining simulations do not show this dipole, but either both have a positive trend in the tropics and the mid-latitudes (29%), or a negative trend in both tropics and mid-latitudes (13%, i.e. 3 simulations, namely NIWA-5, CMAM, WACCM-2). Only 3 out of 31 simulations simulate negative, but not significant, trends both in the tropics and northern extratropics, and thus they show a similar behavior to observations (see right of Fig. 2 and Ball et al. (2019)). However, their zonal trend patterns (see Fig. 1) reveal, that none of these three simulations reproduces the observed trend pattern with consistent negative trends from 50°S-50°N in the LS. Consequently it is important to keep in mind that the results of these (averaged) trends depend on the choice of the latitude-pressure box, as the integration over a wider latitude band can lead to a cancellation of opposing trends.

Next, we analyze whether a systematic relationship between the LS tropical and mid-latitude trends exists in the CCM simulations. For this, the simulated northern mid-latitude LS ozone trends are plotted against the simulated tropical LS ozone trends over the time period 1998–2018 for all 31 REF-C2 simulations and for the observed data-set BASIC$_{SG}$ in Fig. 3. As discussed above, in the LS the majority (45%) of the models have a negative ozone trend in the tropics and a positive trend in the northern mid-latitudes. Moreover this illustration again highlights that the trends estimated from observational data are lying on the outer edge of the model trends distribution. The inter-model correlation between the tropical to mid-latitude trends is negative with a low correlation coefficient (-0.25). Thus, for the chosen period the tropical ozone trends are only weakly linked to mid-latitude ozone trends in the models. However, we expected that the two trends are highly (negatively) correlated,

as from our understanding increased tropical upwelling leads to decreased tropical ozone, and this upwelling increase should be linked to an increased mid-latitude downwelling, which would enhance ozone in the mid-latitudes. However Fig. 3 does not support this. Also slightly varying the period (i.e. looking at the periods 1999-2019, 2000-2020, 2001-2021) reveals very low negative or near zero correlations (not shown here). To get a better understanding of the processes leading to the given LS ozone trend patterns we will investigate the relationship of LS ozone trends to stratospheric transport trends in Section 3.4.

Overall we can conclude from the analysis of ozone trends in the suite of CCMI models (see Fig. 1, 2 and 3), that the LS ozone trends exhibit a considerably large spread across both the different models, but also across ensemble members from a single model, in particular in the mid-latitudes. This indicates that ozone variability considerably influences the LS trends, in agreement with the recent studies by Chipperfield et al. (2018) and Stone et al. (2018). However, even when considering the high variability of possible trends in CCM simulations, the observational trends emerge as an unlikely realization of the simulations over the period 1998–2018. In the next section, we will analyze the robustness of this finding by varying the period of the trend calculation, and providing an in-depth statistical analysis of the likelihood of the observed trend lying within the suite of modeled trends.

3.2 Robustness of lower stratospheric ozone trends

In the previous section we found that the observed LS negative ozone trend in the mid-latitudes together with a simultaneous negative trend in the tropics is unlikely based upon the suite of CCM simulations. To further establish the robustness of this result, we here test whether this also holds for time periods that are slightly different to the period 1998–2018 we considered before. Thus, in this section we first want to investigate how variability influences the ozone trends, and second we want to quantify the likelihood of the observed trend being a realization of the distribution of the modeled trends. To answer those questions, we calculate the LS ozone trends by varying the start and end years of the time period. In Fig. 4 (a) and (b), the observed tropical and mid-latitude ozone trend in the LS is shown for start years varying from 1995–2001 (y-axes) and end years from 2013–2019 (x-axes). Both tropical as well as mid-latitudes LS ozone trends are consistently negative for all chosen periods in the observations (top row). This is in line with the results of Ball et al. (2019), who found that the observed negative sign of the tropical and mid-latitude trends remain insensitive to changing the end year. In the tropics, observational LS ozone trends are consistently negative with values between -0.64 and -1.24 DU/dec for all possible start end year combinations. In the mid-latitudes the trends are also negative for all shown time periods, but are more variable than in the tropics (values range between -0.11 and -1.22 DU/dec). In particular at mid-latitudes, the strongest negative trends are found for start years of 1996 to 1998, and a sudden decrease in the trend magnitude is found for the start year 1999 and 2000. Thus, the analysis in Ball et al. (2018, 2019) and in the preceding section focused on a period with particularly strong negative mid-latitude ozone trends. Possible reasons for the sudden change in the trend, such as the strong ENSO event in 1998, are discussed in Section 4. Note that the trend magnitude increases again for the start year 2001, which again suggests that interannual variability influences the observational mid-latitude trends.

Fig. 4 (c) and (d) display the tropical and mid-latitude trends as a function of start and end year derived from the model simulations. To do so, a robust estimate of the trend probability distribution considering all model simulations was derived (see

Section 2.3) and from this distribution the most likely trend is shown (see peak in the models' trend probability distributions of Fig. S1 and S2 in the supplement). In the tropics the ozone trends derived from the REF-C2 simulations are negative and range from -0.74 to +0.02 DU/dec. In the mid-latitudes the trends are positive for all possible start/end year combinations, with values ranging from +0.4 to +1.48 DU/dec. In contrast to the sudden change in the mid-latitude observational trend for start years 1999 and 2000, in the REF-C2 simulations no such systematic change can be found. The estimated probability distributions of the trends from the REF-C2 simulations (see Figs. S1 and S2 in the supplement) are typically symmetric around their maximum value and show a single, central peak. The width of the distribution changes when varying the start/end year combination, with narrower distributions for longer time periods. Moreover, visual inspection of the distribution implies that the tropics (Fig. S1) generally have Gaussian-like distributions, whereas the mid-latitudes (Fig. S2) often show a more peaked structure, i.e. with heavier tails. Nevertheless, as an estimate of the width of the models trend distribution, we show in Fig. 4 (e) and (f) the standard deviation of the models distribution (in DU/dec) in the tropics and mid-latitudes, respectively. For longer time periods (values in lower right corner) the standard deviation of the models' trend is smaller, i.e. the distribution is narrower. This indicates that the influence of natural variability is less important for longer time periods, as should be expected.

Given the distributions representing the combined trends of the models, we can now quantify the disagreement between the observational trend estimate and the models' trend probability distributions for each start/end year combination. In Figs. 4 (g) and (h) the "probability of the disagreement" between observational and modeled LS ozone trends is given for the tropics and the mid-latitudes. The value of the "probability of disagreement" is calculated by the central interval of the models' probability distribution when taking the observed trend value as threshold of this interval. Thus, a probability value of 90% indicates that the observed trend falls within the inner 90% of the distribution, i.e., only 10% of the distribution is more extreme than the observed trend: the smaller the given "probability of disagreement" value, the higher is the probability that the observed trend lies within the models' distribution. In the tropics, the observed LS ozone trend falls within the 13% to 73% interval of the modeled probability distribution, i.e. the observed trends are generally likely representations of the models' trends. The agreement is best for short time periods (values in diagonal in Fig. 4 (g)), mostly because of the broader distribution (see Fig. 4 (e) and Fig. S1). Also for early start years (in particular 1995) and end years ranging from 2013 to 2018, the disagreement is small, because model trends are strongly negative for this period (see Fig. 4 (c)). In the mid-latitudes, the observed trend generally lies at more distant parts of the models' trends distribution (73% to 96%), i.e. the observed trend is a more extreme value in the models' distribution. The disagreement is smallest for both the earlier periods (lower left, start years 1995–1997 and end years 2013–2015) and the later periods (upper right, start years 1999–2001 and end years 2017–2019). This coincides with the generally smaller negative trends in those periods in observations (see Fig. 4 (b)) and rather constant trend distributions in the models (see Fig. 4 (d)). For the periods with the strongest negative observed trend (start years 1996–1998), the observed trend lies within the central 90% or higher of the models' distribution, i.e. is an unlikely representation from the modeled trends. The sudden decrease in the observed trend magnitude for start years 1999 (Fig. 4 (b)) is reflected by a decrease of the central interval to about 75%. In general, one might have expected that longer periods lead to better agreement of the observed and modeled trend due to the smaller influence of variability (see Figs 4 (e) and (f))- as we do in the models- however, we do not find this to be true for either the tropics or the mid-latitudes.

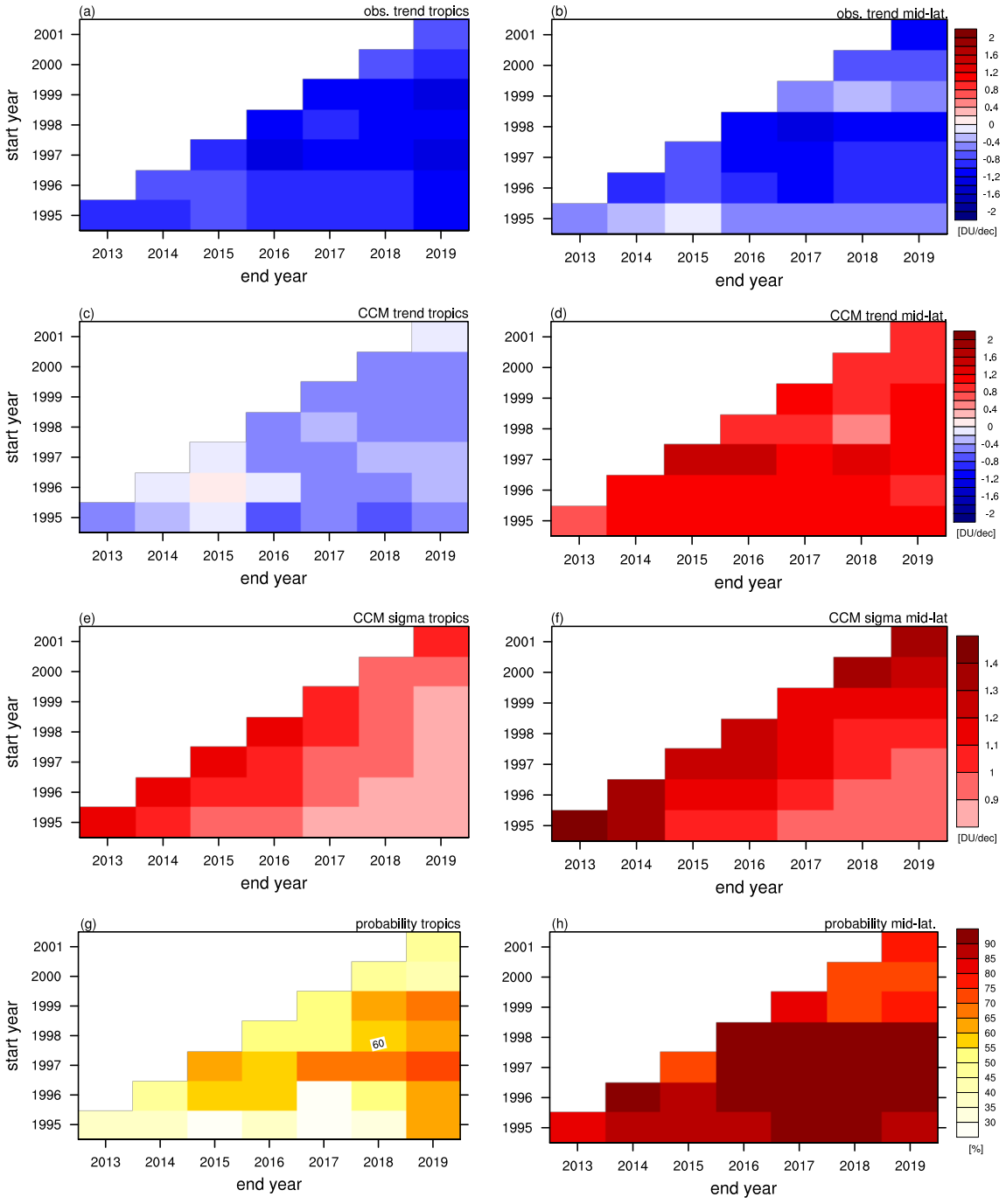

Figure 4. Tropical (right) and mid-latitude (left) LS ozone trends (in DU/dec) as function of different periods for the observational trend of BASIC_{SG}((a)+(b)), the most likely trend of the modeled REF-C2 probability distribution ((c)+(d)) and the 1-σ standard deviation (in DU/dec) of the mean obtained from the probability distribution ((e)+(f)). The panels (g)+(h) show the "probability of disagreement" (in %) between observed trends and the REF-C2 trend probability distribution. In all panels the x-coordinate denotes the different end years (2013-2019) and the y-coordinate the different start years (1995-2001).

3.3 Convergence of future lower stratospheric ozone trends

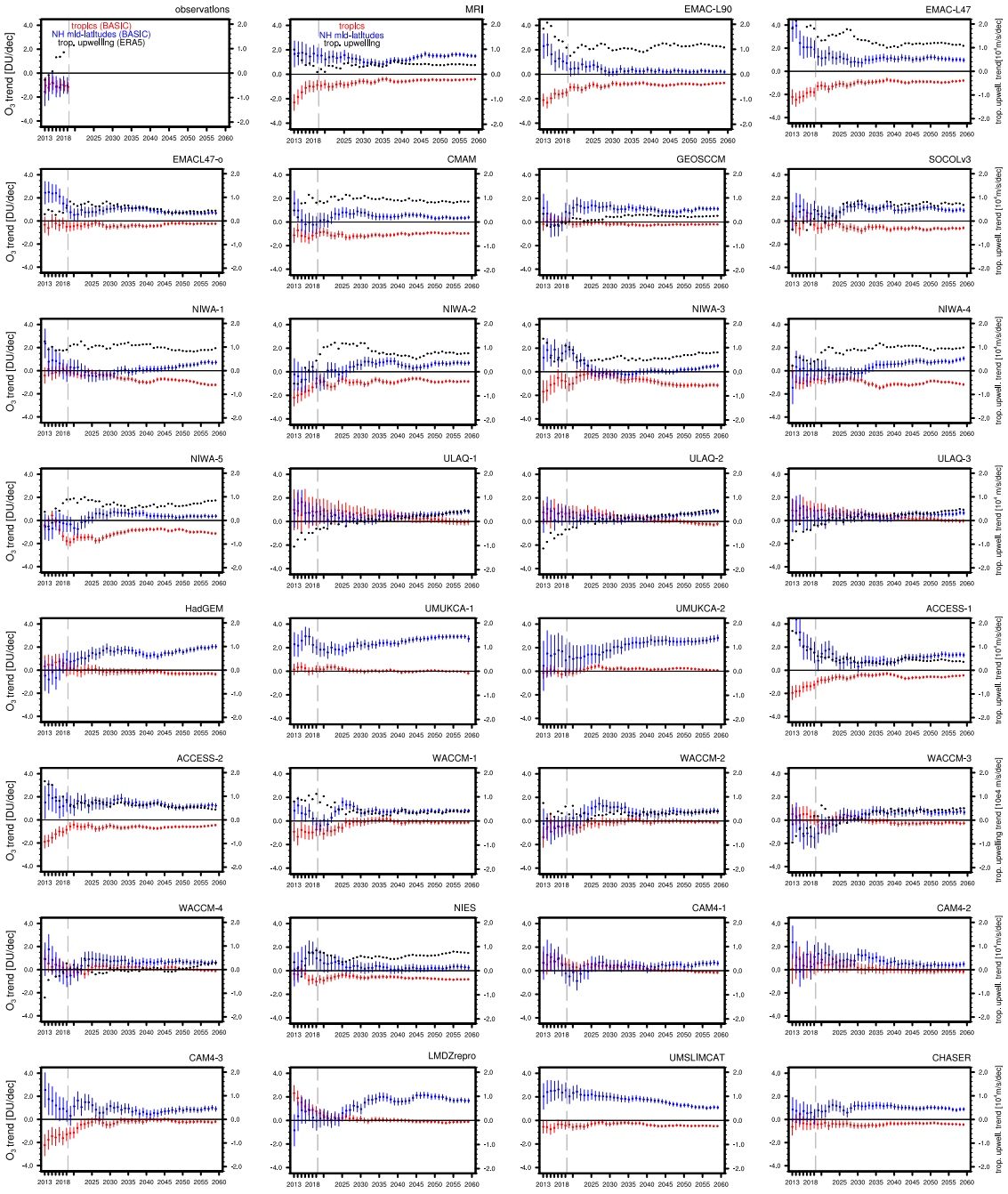

Figure 5. Tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trend and their uncertainties (in DU/dec) of observations (BASIC$_{SG}$) and REF-C2 simulations as a function of the end year (red and blue dots, respectively). Tropical upwelling trends are included for all REF-C2 simulations, where TEM diagnostics was available (black dots); observational tropical upwelling is taken from ERA5 reanalysis. The end year varies from 2013 to 2019 for observational data and from 2013 to 2060 for REF-C2 simulations. Error bars associated with each trend represent the 90% confidence intervals.

In the previous section, the ozone trend robustness was analyzed for time periods of up to 25 years. We will show in the following that, as the considered time periods are extended, the influence of natural variability decreases and the trends converge to the trend forced by long-term GHG and ODS concentration changes. To analyze the timing and the values of the trends' convergence, we extend the period for the trend calculation into the future for all REF-C2 simulations.

Fig. 5 shows the tropical and northern mid-latitude LS ozone trends together with the tropical upwelling trend (black; if available) for periods with the fixed start year 1998, and the end year varying from 2013 up to 2060, by extending the time period by steps of one year. For reference, the observational trends of ozone (from BASIC$_{SG}$) and tropical upwelling (from ERA5) are shown in the upper left panel of Fig. 5, with the last available end point in the year 2019. As shown in the last section, the trends derived from observational data are consistently negative both in the tropics and in the northern mid-latitudes.

As discussed in Sec. 3.1, the ozone trends exhibit a strong inter-model spread for the observational time periods. Both tropical and mid-latitude ozone trends in the individual model simulations vary considerably for different end point years within the observational period (left of the vertical dashed gray lines). The northern mid-latitude trend is generally more variable than the tropical trend. For longer time periods extending into the future, the uncertainties in the LS ozone trends decline and the trends converge in all simulations. All model simulations consistently simulate persistent negative or near-zero trends in the tropics and positive or near-zero trends in the northern mid-latitudes. However, the timing of convergence of the trends to this trend pattern is rather different in the simulations, as can be inferred from Fig. 5, i.e. the convergence appears to be model dependent. For some models, the trends vary little for end years after 2020 (e.g. MRI in Fig. 5), while in other models, the trends still vary considerably until end years around 2030 to 2040 (e.g. the four WACCM ensemble members in Fig. 5). The timing of the convergence is controlled by the ratio of the year-to-year variability to the strength of the forced trends. The relative forcing by ODS versus GHG changes over time, and thereby the forced ozone trends vary over the time periods as well, making it difficult to quantify an exact date of convergence. Still, the trend estimates for the entire period 1998 to 2060 do converge to stable values for almost all models, thus representing the forced trend for this time period. The trend magnitudes over this long period vary strongly between the models, from -0.10 to -1.32 DU/dec in the tropics and from +0.39 to +2.00 DU/dec in the mid-latitudes. Comparing this to the model range of the shorter time period 1998–2040, we see that the tropical trend (+0.06 to -1.12 DU/dec) has not converged to the end point values of 2060, yet. The mid-latitude trend (+0.54 to +2.15 DU/dec) is however close to the 2060 values.

Overall, the mid-latitude trends converge to positive values in the majority of the model simulations (about 85%) by 2030. Thus, if both the year-to-year variability and the forced response of the models is simulated realistically, we should expect the emergence of positive mid-latitude trends from observational records within the next decade.

3.4 Influence of transport processes on LS ozone trends

In this section we aim at improving our understanding of how transport processes control the LS ozone trends in the models. As is well known from earlier studies, tropical upwelling significantly influences stratospheric ozone in the tropics (e.g. Oman et al., 2010). Enhanced tropical upwelling leads to more transport of tropospheric ozone-poor air into the tropical LS. Moreover, a faster removal of ozone in the tropical pipe reduces the residence time in the LS. To analyze how tropical and mid-latitude LS

ozone trends are influenced by transport processes, we show in Fig. 2 the tropical upwelling trends (20°N-20°S, 70 hPa) for all simulations providing TEM diagnostics. This shows that models with strong positive tropical upwelling trends also have large negative tropical ozone trends. However, for the mid-latitude trend it is difficult to visually detect a clear relation with tropical upwelling trends.

Therefore we analyze the relation of tropical upwelling and extratropical downwelling trends to LS ozone trends in terms of a correlation analysis. Fig. 6a shows the inter-model correlation between the tropical upwelling mass flux trends at different stratospheric levels and tropical LS ozone column trends over a sub-set of 20 REF-C2 simulations. Additionally the correlation of the northern mid-latitude downwelling mass flux trends at different levels and LS ozone column trends is provided in Fig. 6b. As above we calculate the trends over the period 1998-2018 and tropical ozone trends are averaged over 20°N-20°S and mid-latitude ozone trends over 30°-50°N.

The correlation profiles between tropical ozone column trends and tropical upwelling trends (red line in Fig. 6a) show significant high negative correlations (r≈-0.8) at all levels between 30 and 100 hPa. Thus, as expected, changes in tropical upwelling at all levels below 30 hPa highly influence LS tropical ozone. This is in line with previous studies (e.g. Oman et al., 2010; SPARC CCMVal, 2010). Between 10 and 30 hPa, the correlation decreases with altitude and becomes insignificant. The correlation values of tropical ozone trends to downwelling trends are positive and also rather high (Fig. 6b). This is clear, as upwelling is directly linked to downwelling, however the negative sign of downwelling causes a sign reversal of the correlation coefficients.

For ozone trends in the northern mid-latitudes (30°-50°N), the correlation of LS ozone to tropical upwelling trends varies in altitude from about -0.2 to +0.4 (solid blue lines in Fig. 6a): It is weakly negative up to 100 hPa; above, the correlation turns to positive values (r≈ 0.4 at 70 hPa). Compared to the relation of upwelling trends to tropical ozone trends, these correlations are quite low and not significant at the 95% level - moreover these correlations are not robust when slightly varying in the period (not shown). The same is true for correlations between mid-latitude ozone trends and downwelling trends (see solid blue lines in Fig. 6b). A possible reason for the non-robust and not significant correlations might be the choice of the mid-latitude averaging region from 30°-50°N. This region can partly include regions of upwelling at some pressure levels, and the location of the turnaround latitude is model dependent. Not accounting for a dynamically consistent averaging region might obscure the correlation analysis. Therefore, we additionally define a dynamically more consistent mid-latitude region by averaging the LS ozone column from the turnaround latitudes of the BDC to 50°N. For each month the averages were taken by calculating the position of the residual streamfunction maximum at each level, and then averaging the LS ozone column from this turnaround latitude to 50°N. It was further ensured that tropospheric air is not included in the averages (which could happen at levels below the tropical tropopause) by using only the region above the tropopause.

The ozone trends in this dynamically defined box are slightly higher compared to the fixed latitudinal region between 30° and 50°N, but given the large spread in trends this difference is not significantly different (see Table 2; the same is true for the longer period 1998–2040, not shown). The correlation profiles for LS ozone trends within this dynamically defined mid-latitude box is included in Figs. 6a and b (see dashed blue line): Due to the dynamical consistency of mid-latitude ozone and the downwelling region, the correlations increase in absolute number compared to the correlations with ozone trends in the fixed

boxes, and the correlations are more robust across different periods (not shown). In particular, the correlation of ozone trends in the dynamically defined averaging box to downwelling peaks at 100hPa with a significant correlation coefficient (r≈0.5). Up- and downwelling at around 100 hPa reflects the shallow branch of the BDC (see e.g. **?**Dietmüller et al., 2018). Thus, the significant positive correlation of downwelling trends around this level to mid-latitude ozone trends suggests that an enhanced shallow branch leads to a decrease in ozone in this region. This would be consistent with enhanced horizontal advection via the shallow branch that transports tropical ozone poor air to the mid-latitudes. The fact that correlations decrease to insignificant correlation values above (and correlations to tropical upwelling even change sign) likely reflects the relation of mid-latitude ozone trends to downward transport of ozone via the deep branch. Thus, overall the correlation analysis suggests that the two competing transport processes of shallow horizontal versus deep vertical advection influence ozone in the mid-latitude LS.

In general, the weaker correlations of mid-latitude ozone to up/downwelling compared to tropical ozone suggests that mid-latitude ozone changes are controlled by a variety of processes, possibly also including two-way mixing. Furthermore, not only changes in the transport strength, but also in the background ozone gradients, can lead to changes in the transport of ozone. For example, the increase in upper stratospheric ozone mixing ratios could lead to enhanced downward transport of ozone despite an unchanged downwelling strength.

To better elucidate the role of different transport processes in the different regions, we additionally analyze the local correlation of AoA trends to the ozone trends for a subset of 9 REF-C2 simulations that provide the necessary diagnostics (namely EMAC-L90, EMAC-L47-1, ACCESS-1, WACCM-1, CMAM, GEOS, SOCOL, MRI, NIWA-1). As shown in Fig. 7a, in the middle stratosphere the correlation coefficients are relatively weak, consistent with the expectation that chemical processes play an important role there. In the LS, we find very high correlations (larger than 0.8) between ozone and AoA trends in the tropics and extending to about 40°N. Thus, inter-model differences in ozone trends are highly controlled by differences in transport trends in this region. Negative correlation values can be found in the LS mid-latitudes north of about 40°N and above 80 to 60 hPa. Interestingly, in the SH correlations are positive throughout the LS. To analyze the role of different transport processes, we separate AoA into the components of residual transport (RCTT) and aging by mixing (AbM; for details see section 2.4). The inter-model correlations between ozone trends and RCTT and AbM trends, respectively, are shown in Fig. 7 b and c. In the LS, RCTT trends are highly positively correlated to ozone trends between 40°S-40°N, whereas for latitudes poleward of 40° the correlation coefficients turn to negative values. AbM trends and ozone trends correlate strongly (and positively) in the LS for latitudes poleward of 30°. This again underlines that in the tropical LS residual transport changes largely control the ozone trends: negative RCTT trends (indicating faster upwelling) are associated with negative ozone trends. This is also in line with the findings of Fig. 6a. In the LS mid-latitudes, on the other hand, both changes in residual transport (RCTTs) and in mixing (AbM) have an impact on ozone trends, leading to the non-homogeneous correlation structure with AoA trends (Fig. 7a). In the region of our interest, i.e. 30°N-50°N, the different transport processes of residual transport with its deep and shallow branch and of two-way mixing appear to influence ozone trends: the RCTT correlations (Fig. 7b) suggest that an enhancement of the meridional component of the residual circulation (shallow branch) leads to an ozone decrease up to 40°N by enhanced transport of tropical ozone-poor air to mid-latitudes. This is in line with the significant positive correlation of models' LS ozone and downwelling trends, that we presented in Fig. 6b. The negative correlations between RCTT and ozone

trends north of 40°N indicate that ozone trends are driven by vertical downwelling (from the deep branch) here: enhanced downwelling (lower transit time) is associated with transport of ozone-rich air from above. Moreover mixing processes play a role in the mid-latitude region. The correlation of AbM trends with ozone trends is positive (r≈ 0.6) north of 30°N in the LS, indicating that mixing is strongly influencing ozone trends in this region as well. Overall Fig. 7 reveals that transport processes in the LS mid-latitudes are complex, as this region is influenced by many competing transport processes. We will discuss this issue further in Section 4.

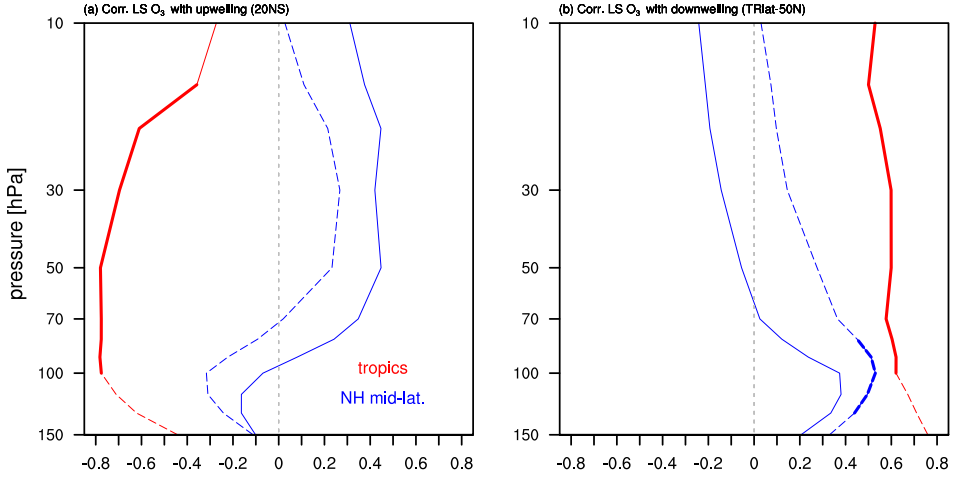

Figure 6. Vertical profile of the inter-model correlation coefficients for (a) tropical upwelling (20°N-20°S) trends (kg/s/dec) to tropical (red line) and northern mid-latitude (blue line) LS ozone column trends and for (b) downwelling mass flux (between the turnaround latitudes and 50°N) trends (kg/s/dec) to tropical and northern mid-latitude LS ozone column trends. Correlations are calculated for upwelling and downwelling trends between 10 and 150 hPa. Mid-latitudes ozone trends are averaged over the latitude band of 30°N-50°N (solid blue line) and also over the dynamical defined latitude band between the turnaround latitudes to 50° (dashed blue line). Trends are calculated over the period 1998–2018 for a subset of 20 REF-C2 simulations. Correlation coefficients which are significant on the 95% level are highlighted in bold.

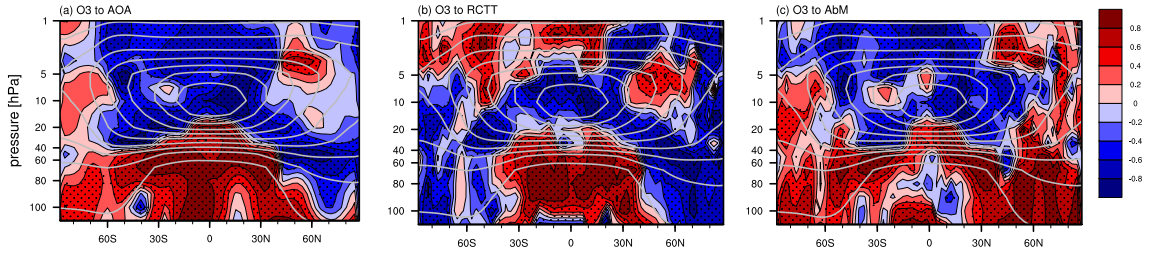

Figure 7. Inter-model correlation coefficients between local ozone trends and (a) local AoA trends, (b) local RCTT trends and (c) local AbM trends. Trends are calculated over the period 1998–2018 for a subset of 9 REF-C2 model simulations. White contours show the MMM ozone climatology, and the stippled regions mark where correlation coefficients are significant on the 95% level.

Table 2. MMM and observational ozone trends, calculated over the period 1998–2018 for tropical upwelling at 70 and 100 hPa, for extra-tropical downwelling at 70 and 100 hPa, for the LS tropical ozone column and for the northern mid-latitude ozone column. Note that LS mid-latitude ozone trends are averaged over the fixed latitude band of 30°N-50°N and also over the dynamical defined latitude band between the turnaround latitudes to 50°. MMM trends and their standard deviation are given over a subset of 20 REF-C2 simulations. Observational based data for up- and downwelling are taken from ERA5 reanalysis and observational data for ozone from BASIC$_{SG}$.

	MMM	**observations**
trop. upwelling trend (70 hPa) [kg/sec/dec]	0.78 10E+7 $\pm$ 1.92 10E+7	1.53 10E+7
trop. upwelling trend (100 hPa) [kg/sec/dec]	1.62 10E+7 $\pm$ 2.21 10E+7	3.14 10E+7
downwelling trend (70 hPa) [kg/sec/dec]	-0.22 10E+7 $\pm$ 1.19 10E+7	-0.35 10E+7
downwelling trend (100 hPa) [kg/sec/dec]	-0.69 10E+7 $\pm$ 2.12 10E+7	-0.15 10E+7
trop. ozone trend [DU/dec]	-0.53 $\pm$ 0.91	-1.07
mid-lat. (fixed) ozone trend [DU/dec]	0.47 $\pm$ 0.87	-0.96
mid-lat. (dyn) ozone trend [DU/dec]	0.78 $\pm$ 0.91	–

3.5 Forced ozone trends in models

In the previous sections we analyzed the ozone trends of the recent 20 year period in detail and found that modeled and observed ozone trends disagree, especially in the northern mid-latitude LS. Assuming the observational data are correct, the question that arises from our results is whether the disagreement stems from the influence of natural variability, or whether the forced response to GHG or ODS concentrations is not captured correctly in the models. Thus in the following, we will investigate the relative role of GHG versus ODS forcing on the ozone trends in the models for the observational period and periods extending into the future. Figs. 8a and 8b show upper and lower stratosphere MMM ozone trends in the tropics (20°N-20°S), in the northern mid-latitudes (30°-50°N) and in the southern mid-latitudes (30°-50°S) for the REF-C2 simulations as well as for the sensitivity simulations with fixed ODS (fODS) and with fixed GHG (fGHG) concentrations (for a detailed description of these sensitivity simulations see Section 2.1). These MMM ozone trends are calculated for the recent time period (1998–2018), for a time period, which extends into the future (1998–2040) and for a future time period (2050–2100). We also include the respective observational trends for 1998–2018. Note that for the calculation of the MMM trends only 10 model simulations are taken into account, as the fODS and fGHG simulations have a smaller sample size than the REF-C2 simulations (see Tab. 1). Moreover we exclude ULAQ for the MMM calculation, as its values are clear outliers compared to other models, such that it would shift the MMM to lower absolute values. Note further that the MMM ozone trends are calculated as the average of the ensemble-means from each model. This ensures that models are weighted equally regardless their ensemble size, which is desirable here as we aim to extract the forced trends, in particular for the longer time periods. Next to the trends averaged over the tropics and mid-latitudes, Fig. 9 shows the latitudinal distribution of the ozone column trends in the upper and lower stratosphere over the period 1998-2040 for the REF-C2, fODS and fGHG simulations. Note that we show the trend over the period 1998–2040 here, as we expect the forced signal to emerge more clearly for this period compared to the shorter observational period.

In the upper stratosphere, the MMM ozone trends over the periods 1998–2018 and 1998–2040 are positive and of the same magnitude in tropical and mid-latitude regions (Fig. 8a). The 1998–2018 MMM trends are more than twice as strong as the observed trends (dots in Fig. 8a), with only one model simulation having lower trend values (in the tropics and NH). Even for the short period of 20 years, the ozone trends are consistently positive for both the models and the observations, indicating that the upper stratosphere MMM trend is robust to inter-annual variability. Therefore, this likely is the forced signal driven by GHG and ODS changes. The analysis of the models' latitudinal distribution in upper stratospheric ozone column trends shows no considerable latitudinal variation (see Fig. 9a). The positive upper stratospheric MMM trend can be explained by the combined effect of still decreasing ODS concentrations at the beginning of the trend periods 1998-2018 and 1998–2040 and by rising GHG concentrations causing stratospheric cooling. The contribution of these two effects is quantified by comparing fGHG, fODS and REF-C2 simulations. In fGHG, the GHG-driven increase of the stratospheric circulation (resulting mostly from the increase in SSTs), as well as GHG induced stratospheric cooling is excluded. In fODS, the chemical ozone destruction via ODS concentrations is excluded. Upper stratospheric ozone trends in fGHG and fODS are positive, but considerably lower than in REF-C2, with trends in fODS having the lowest values. This is in particular true for the extended period 1998–2040,

where we expect clearly forced trends. The weaker upper stratospheric ozone trend in the fGHG simulations can be explained by the missing additional ozone increase due to GHG-induced stratospheric cooling, as ozone is photochemically controlled in these upper regions. The weaker trend in the fODS simulations can be explained by the missing additional increase via the recovery from ODS destruction. The comparison of fODS and fGHG trends over the period 1998–2040 reveals that about 2/3 of the REF-C2 upper stratospheric trend is due to the ODS forced trend. The upper stratospheric trends over the second half of the century (2050–2100) reveal that the ceasing influence of ODS forcing manifests in decreasing ozone trends in the fGHG simulations. However, the ODS forcing still contributes to the ozone increase by about as much as the GHG forcing.

For the LS, Fig. 8b highlights that ozone trends are highly variable in particular for the shorter period of about 20 years and that the MMM ozone trends over the period 1998–2018 and 1998–2040 are negative in the tropics and positive in the mid-latitudes in the REF-C2 simulations. In general, the mid-latitude ozone trends are very variable both in the northern and southern mid-latitudes, but the southern mid-latitude trends are somewhat lower (and negative in some models) for the shorter period. Also in observations, the SH mid-latitude trend is more uncertain and variable (compare observational estimates in Fig. 8b, and see Ball et al., 2019).

In order to attribute modeled LS ozone trends to GHG and ODS changes, we compare the ozone trends of the REF-C2 to fGHG and fODS simulations in Fig. 9b (see also MMM trends in Table S1 of the supplement). For the short time period of about 20 years we find that the MMM mid-latitude ozone trends are positive and overall similar between the fGHG and the REF-C2 simulations. The fODS simulations, in contrast, show a negative MMM mid-latitude trend, but with a very high intermodel spread. Compared to the REF-C2 simulations, the tropical LS trends are less negative in the fGHG simulations and more negative in the fODS simulations. This it what we expect from the missing influence of the GHG concentration rise on tropical upwelling. But note that trends of fODS and fGHG are not significantly different from the REF-C2 simulation. The small, mostly not significant differences (not shown) with its high intermodel spread in the fGHG, fODS and REF-C2 trends over the quite short observational period (1998–2018) again underlines the conclusion that variability strongly impacts LS ozone trends.

For the longer time period (1998–2040), the MMM fGHG trend in the tropical LS is near zero (see Fig. 8b and Table S1). In contrast to the trends over the short time period it can be clearly distinguished from the negative REF-C2 trend and also from the negative MMM fODS trend which is comparable to the REF-C2 trend. This can be explained by the absence of GHG-induced enhancement of tropical upwelling, which strongly influences tropical LS ozone trends. The latitudinal distribution in Fig. 9b shows in more detail the tropical LS ozone column trends in the individual fGHG simulations (red thin lines): most models show trends near zero in the tropical region. The slightly negative ozone trends in the tropics in two models are a bit surprising. However, they probably can be explained by the fact that the upper stratospheric ozone increase can reduce the UV radiation reaching the LS, and thus less ozone is produced there chemically (see e.g. Meul et al., 2014). In the mid-latitudes, the MMM trend in the fGHG simulations is positive and only slightly smaller than the REF-C2 trend, whereas the fODS MMM trend is near zero (see Fig. 8b). This indicates that enhanced downwelling associated with the strengthened circulation plays a minor role in this selected region and is consequently not responsible for the positive trend found in REF-C2. This weak influence of downwelling trends on mid-latitude ozone trends is consistent with the results presented in Section 3.4. There, we found that

downwelling mass flux via the deep branch and mid-latitude ozone trends are only weakly related in REF-C2. Moreover, the near zero ozone trend in the fODS simulations underlines that the mid-latitude ozone trends are strongly influenced by ODS recovery. This might be through decreased local ozone destruction (as ODSs are still decreasing), or through ozone transport from upper or polar regions, where ozone is increasing strongly because of the ”closure of the ozone hole“. Thus, ozone increases in the mid-latitudes, even without an enhanced transport circulation.

To better understand the fact that the mid-latitude fODS trend is near zero, although we expect transport-induced changes in the LS, we show in Fig. 9 (blue thick line) the latitudinal distribution of the MMM fODS LS ozone partial column trend. Here we see that the LS mid-latitude band between 30°N-50°N lies just within a region where ozone trends are shifting from negative to positive values. The MMM trend is negative between 30°N-40°N and positive between 40°N-50°N, explaining the near zero mid-latitude trend over the total latitude band. We suppose that the negative trend 30°-40°N can be explained by enhanced advection through the shallow branch and/or two-way mixing and the positive trend between 40°N-50°N by enhanced downwelling, as suggested by the correlations with RCTTs (Fig. 7b). However, the individual models show quite noisy behaviour in the latitudinal distribution of LS mid-latitude ozone trends, mainly in the NH (thin blue lines in Fig. 9b), indicating that the relative role of trends in the different transport processes might differ in models. The trends in the fGHG simulations are near zero in the inner tropics and positive at all other latitudes, indicating that the recovery from ODSs leads to an increase in ozone almost everywhere throughout the LS. The latitudinal distributions thus indicate that the GHG-driven circulation changes would induce a decrease in ozone from the tropics up to 40°N/S (leading to a near zero trend in the region 30°N-50°N), but due to the recovery of ozone from ODSs, the trend is essentially shifted to positive values, so that the average trend over 30°N-50°N is positive.

The LS ozone trends calculated over the period 2050–2100 confirm the role of ODSs in influencing the mid-latitude ozone trends: despite a strong increase in tropical upwelling in this period (not shown), which drives the strong decrease of tropical ozone in the REF-C2 and likewise the fODS simulations, mid-latitude MMM ozone trends are essentially zero (or slightly negative in the NH) in the fGHG simulation. The effects of an ODS recovery on mid-latitude ozone are smaller in this period due to the declining influence of ODSs, but in the SH mid-latitudes this still leads to a robust positive ozone trend.

Overall our analysis of the fODS and fGHG simulations suggests that the recovery from ODSs is a dominant player for LS mid-latitude ozone trends. GHG-induced circulation strengthening also impacts LS mid-latitude ozone trends, but the competing transport effects via shallow and deep branches lead only to small transport induced trends when averaged over the region from 30°N-50°N.

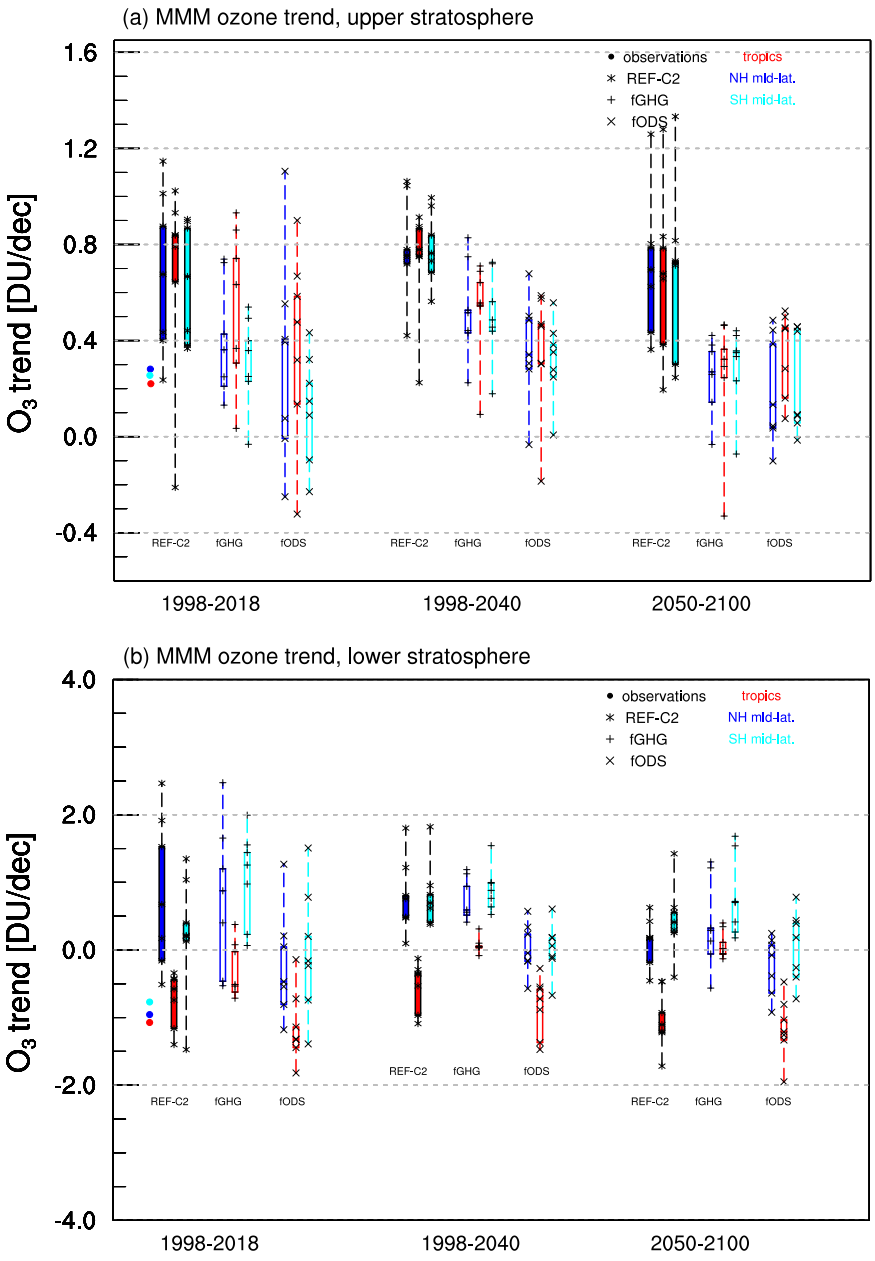

Figure 8. MMM ozone column trends in the tropics (red, 20°N-20°S), in the northern mid-latitudes (blue, 30°-50°N) and in the southern mid-latitudes (cyan, 30°-50°S) for thee different periods (i.e. 1998–2018, 1998–2040, 2050–2100) for (a) the upper stratosphere (1-10 hPa) and (b) the LS (30-100/150 hPa). The boxes extend from the lower to upper quartile of the data with a line for the median and with whiskers to show the minimum and maximum values of the LS MMM ozone trends. MMM trends are given for REF-C2 simulations (filled boxes) as well as for fGHG and fODS simulations (not-filled boxes). Note here that for the estimate of MMM trends only 10 model simulations are taken into account, as fixed GHG simulations have this smaller sample size, and we want to ensure that all three simulation types include the same models for the MMM trend estimate. Individual model trends are denoted by black stars for REF-C2, by black pluses for fGHG and by black crosses for fODS. Observational data are included for the trends over the period 1998–2018 (red, blue, cyan points, respectively).

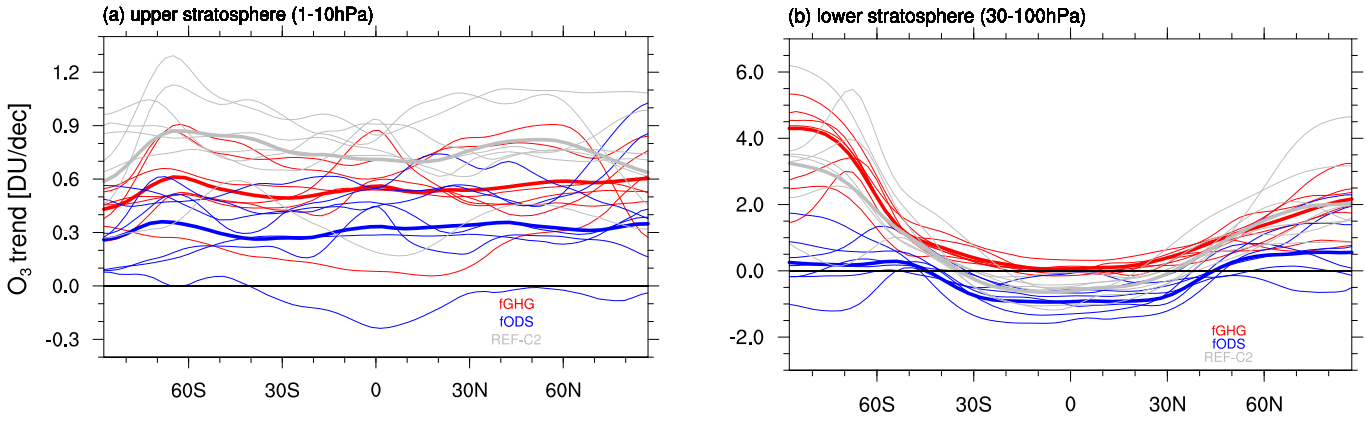

Figure 9. Latitudinal distribution of ozone column trends over the period 1998–2040 for all REF-C2 (gray lines), fGHG (red lines), fODS simulations (blue lines) for (a) the upper stratosphere (1-10 hPa) and (b) the LS (30-100 hPa). Thick lines indicate the MMM ozone trends.

4 Discussion

In the previous sections we analyzed ozone trends over periods spanning the past two decades (i.e., 1998–2018) in detail. We found that modeled and observational ozone trends agree well in the tropical lower stratosphere, but in the northern mid-latitude LS the observed ozone trend represents an extreme value in the distribution of model trends.

In the following, possible reasons for the discrepancy between the mid-latitude ozone trends in the model simulations and the observations will be discussed. One possible reason for the disagreement between modeled and observed LS ozone trends could be issues with the satellite records. For example, instrument biases and drifts can lead to large uncertainties in the observations, particularly in the lower stratosphere. The effect can manifest as steps in the data when instruments, which have different vertical resolutions, are added that can influence trend estimates (REFS as here - I think Gaudel is not relevant here). For a thorough discussion on this topic, see Harris et al. (2015); Ball et al. (2017); Petropavlovskikh et al. (2019). However, for the sake of this discussion we will assume that the observational data record is correct. Hence, the question that arises from our results is whether the disagreement stems from the influence of natural variability, or whether it is related to the forced trend, or more specifically:

- The mean value of the modeled trend distributions might be incorrect. In other words, the forced trend might not be captured correctly by the models.

– If we assume that modeled trend distributions are correct, the observed ozone trend as an unlikely representation might emerge due to very anomalous conditions during the considered periods. This may be caused by extrema in natural variability in the beginning of the time series (late 1990s), and/or in the end of the time series (late 2010s).

– The modeled trend distribution constructed from the REF-C2 simulations might be biased because natural variability (e.g. QBO and ENSO) is not represented adequately in the models. This could lead to an overly narrow trend distribution, and thus would make the observed trend seem more unlikely than it is.

While it is not easily possible to test which of the above explanations is correct, in the following we will discuss their possible contributions to the diagnosed disagreement in the light of our results and what is known from literature.

Representation of forced trends

Based on the CCMI-1 data, we confirmed previous studies in that the decrease in tropical LS ozone is strongly related to the GHG-driven increase of tropical upwelling. The tropical upwelling trend derived from reanalysis (ERA5) lies in the range of the upwelling trends simulated by the models, but on the upper end of the range. This is consistent with tropical ozone trends, which are on the stronger (more negative) end of the trend range simulated by the models as well. Circulation trends derived from reanalysis bear considerable uncertainty (e.g. Abalos et al., 2015), however reanalyses tend to agree better in the recent decades (T. Birner, personal communication, S-RIP report). Therefore, the upwelling trend derived over the period 1998–2018 from ERA5 is likely better constrained compared to earlier periods.

In the mid-latitudes (30°N-50°N), we find that the GHG-driven circulation changes do not lead to a net trend in ozone. This is evident from the fODS simulations (see Section 3.5) and from the vanishing mid-latitude LS ozone trends over the period 2050–2100, when the influence of ODSs cease. The correlation analysis in Sec. 3.4 revealed that competing processes influence ozone trends in this region: an enhanced shallow branch in the LS can decrease ozone due to enhanced horizontal advection, while enhanced downwelling in the deep branch increases ozone (see correlation to RCTTs, Fig. 7b). In the fODS simulations, those competing influences lead to negative LS ozone trends equatorward 40°N/S, and to positive ozone trends poleward 40°N/S (see Fig.9). Thus, this leads to nearly vanishing ozone trends in the mid-latitude region defined as 30°-50°N. The consistent simulation of positive ozone trends in the mid-latitude LS in the REF-C2 MMM for the recent past and the coming decades is thus a result of the ODS concentration decline rather than of GHG-driven circulation changes. The effects of declining ODS concentrations on LS mid-latitude ozone can either be related to the chemical recovery of ozone, leading to local increases in ozone, or maybe more importantly to enhanced ozone transport into this region. Another effect can be induced by the circulation changes due to ODS-driven ozone changes, that have been shown to have had a strong impact on AoA trends in the past (Polvani et al., 2019; Abalos et al., 2019). However, future circulation changes due to this effect are shown to be weak (Polvani et al., 2019). Furthermore, ozone-induced circulation changes are stronger in the SH, not consistent with approximately symmetric ozone trends in the mid-latitudes of both hemispheres.

Given that the positive mid-latitude ozone trends in models are driven by ODSs rather than by GHG changes, the discrepancy to the observed trend could indicate a mismatch in the relative role of the response of ozone to ODS versus GHG forcing. This

means that either the GHG-driven circulation change in the models could be underestimated or differ in structure, or the ODS-driven ozone increase in the mid-latitude LS could be overestimated in the models. As for the latter, we showed that upper stratospheric ozone increases more strongly in the models than in the observational data (see Fig. 8a). Thus, one hypothesis would be that the ODS-driven recovery of stratospheric ozone in the period since the late 1990s is generally overestimated in the models, which would then make negative ozone trends in the mid-latitude LS unlikely in the models. As for the effects of the GHG-driven circulation changes, we mentioned earlier that the MMM tropical upwelling trend is weaker compared to the estimate from ERA5 reanalysis (see Tab. 2). However, the generally consistent tropical ozone trends between models and observations rule out a vast underestimation of tropical upwelling changes. Rather, structural circulation trend differences could contribute to the disagreement in the mid-latitudes. An indication of which is the lower mid-latitude downwelling trend diagnosed from ERA5 as opposed to the models (see Tab. 2). This is also consistent with the finding of poleward shifted turnaround latitudes by Orbe et al. (2020), as discussed below. While it is a likely explanation that structural circulation trends or anomalies contribute to the observed ozone trends, it is not easily possible to separate the role of natural variability in forming those structural circulation trends (see discussion on natural variability below).

In general, since LS mid-latitude ozone trends are driven by competing transport processes (see Section 3.4), the mis-match of trends in this region between models and observations might also indicate a mis-representation of transport processes in the models. We show that ozone trends in the LS correlate well with trends in the passive AoA tracer, indicating that the differences in ozone trends between models are transport-driven. While there is a long-standing discrepancy of AoA trends derived from observations and models in the mid-stratosphere, AoA trends in the mid-latitude LS tend to agree well between models and observations (see e.g. Chapter 5 of WMO, 2018). On the other hand, climatological mean AoA in the suite of CCMI models used in this study varies considerably between models, and it was shown that this is due to differences in mixing effects on AoA (Dietmüller et al., 2018).

The studies of Wargan et al. (2018) and Ball et al. (2020) argue that the LS mid-latitude ozone decrease in observational data is possibly linked to enhanced two way mixing. Ball et al. (2020) used effective diffusivity (Haynes and Shuckburgh, 2000) as diagnostic for horizontal mixing and found that in reanalysis data (JRA-55, ERA-Interim) mixing is enhanced in the 1998–2018 period. In an earlier study, also Ray et al. (2010) showed a substantial increase in effective diffusivity under a changing climate for CCMs and reanalyses data (JRA-25, ERA-40). Recently, Orbe et al. (2020) used the TEM budget analysis of an idealized short lived tracer (that covaries with ozone on interannual and decadal time scales) in 10 free-running ensemble member simulations with the GEOSCCM model, in order to identify the mechanism that is driving the negative LS ozone trends. In contrast to the studies of Ball et al. (2019) and Wargan et al. (2018), the study by Orbe et al. (2020) showed that the mixing effect is not as important for the LS mid-latitude ozone trend. Rather they found a poleward expansion of the residual circulation in the LS with weaker downwelling in the subtropics, and stronger downwelling over the mid-latitudes, leading to negative LS trends in the NH. However, as discussed in Orbe et al. (2020), mixing must be considered in the context of the specific tracer that is analyzed (i.e. short lived tracers are less sensitive to mixing). As such, the analysis of the TEM budget for the tracer ozone could be a focus in further investigations.

Overall, the LS ozone trends are strongly affected by variability over the short period, making it difficult to infer whether the forced trends in models and observations agree. For the models, we extended the time period into the future to investigate for which period length the trends converge. We find that the inter-model spread of the ozone trends substantially diminishes for the longer time period (1998–2040), but to a different extent for different regions (see Fig. 8). In the upper stratosphere, MMM trends are significantly positive already for the shorter period 1998–2018. In the LS, the MMM ozone trends consistently show positive trends in the mid-latitudes for the period 1998–2040, with a comparably low inter-model spread. Thus the question arises as to whether we can expect observational data to also show a positive ozone trend in the mid-latitudes in the future. If the forced model trends are assumed to be correct, we should expect this positive trend to emerge by about 2030 to 2040 (compare Fig. 5).

Influence of natural variability on the observed trend

Sources of natural variability that strongly influence LS ozone are volcanic eruptions, the QBO and ENSO. No major volcanic eruption occurred during the analyzed period, so we will disregard this source of variability. The influence of the QBO and ENSO on the hemispheric mean mid-latitude ozone is of the same magnitude, and thus they can both impact LS ozone trends, as shown by the study of Olsen et al. (2019).

We know from earlier studies that the QBO has a strong dynamical effect on the sub-tropical and mid-latitude LS ozone (e.g. Randel and Wu, 2007). Moreover it was recently shown that ozone trends in the mid-latitudes are directly linked to the QBO, as the QBO induces a secondary circulation (see e.g. Ball et al. (2019) and A. Stenke personal communication, EGU 2020). In 2016, the typical QBO phasing was disrupted, and this has been shown to be associated with negative LS ozone anomalies in the tropics (Kusuma et al., 2019). These negative anomalies at the end of the time period would lead to a strengthened negative ozone trend, and our analysis indeed shows slightly stronger negative tropical ozone trends for the end year 2016 compared to 2015 (see Fig. 4a). The mid-latitude ozone trend is also stronger for the end year 2016, which however does not fit expectations (QBO-induced anomalies are of a different sign in tropics and extratropics, see e.g. Randel and Wu, 2007). Another way, in which the QBO could lead to decadal scale variability in ozone, and thus influence the trends, was recently reported (J. Neu personal communication, AGU 2018): since the QBO's influence on tropical upwelling depends on the season, the timing of the QBO phases is crucial for its influence on trace gas concentrations. Similarly, Ball et al. (2019) pointed out that non-linear attribution may be required to capture the QBO's impact.

One of the strongest warm ENSO events on record occurred in late 1997 (Jensen et al., 1998). By using CCM (WACCM) simulations with prescribed SSTs from observations, Calvo et al. (2010) showed that this strong ENSO event was associated with low ozone values in the tropics and high values in the mid-latitudes. This is in line with observational results by Randel et al. (2009). Consequently, mid-latitude ozone trends should be more negative when beginning the time period with this warm ENSO year. This is consistent with the strong mid-latitude trends in the BASIC$_{SG}$ data-set for the start years 1998 (and less so for 1996-1997, see Fig.4 (b)). However, as the tropical trend is not associated with weaker negative trends for the start year 1998, this explanation is again not fully consistent.

As stated earlier, we have refrained from applying a multiple linear regression (MLR), which potentially would take at least part of the named sources of variability into account. If the trend strengths and patterns are strongly influenced by anomalous natural variability events, one might argue that removing this variability via an MLR method would have a large impact on the trends. However, the trend estimates by Ball et al. (2018), that take ENSO and QBO variability into account differ only in details from our trend estimates. Note that an MLR method might not fully account for the induced signals by QBO or ENSO, because, as mentioned above, their influence is likely non-linearly dependent on the signal strength and the signal timing. Thus, an MLR analysis cannot conclusively clarify the role of natural variability for the observed trends.

Overall, the sudden systematic change in the magnitude of the mid-latitude observational trend (Fig. 4 (b)) indicates that natural variability (in particular the strong ENSO event in 1997) influenced the observed trends over the analyzed periods, and contributed to the particularly strong disagreement of observed and modeled mid-latitude trends for the relevant time periods. However, the expected effects of QBO and ENSO events on the trends are not entirely consistent between tropics and mid-latitudes. Possibly an exceptional combination of different factors led up to the particular observed trend pattern, causing the mid-latitude trends to be more anomalous than the tropical trends in comparison to the trend distribution derived from the models.

Representation of natural variability in models

Above, we argued that natural variability likely influenced the observed ozone trends, and that might partly explain that trends over the observed period disagree with the trends in model simulations. However, how large this disagreement is, depends on the underlying trend distribution derived from the models. For example, if the influence of natural variability is underestimated in the models, the trend distribution is too narrow.

The QBO is represented differently in the individual CCMs: some models generate a QBO internally, some models nudge winds towards a given QBO, and in some models, the representation of the QBO is missing entirely (for more details see Morgenstern et al., 2017). Thus, over the whole suite of models, this could cause an underestimation of ozone variability in the models and therewith consequently a too narrow trend distribution. Moreover, as the QBO signal is treated differently across the REF-C2 model setups, we can also expect that the inter-model differences in the QBO representation contributes to the spread in ozone trends over recent decades.

The analyzed 'free running' REF-C2 simulations either use an interactive ocean model, or use SSTs from other model simulations that are coupled to an ocean model. However, these coupled models still have biases with respect to the simulation of ENSO (Bellenger et al., 2014), thus ENSO-related variability in LS ozone might also be underrepresented.

Further, even if the QBO and ENSO are represented with the correct signal strength (e.g., by nudging the QBO, and prescribing observed SSTs), the induced circulation anomalies might not be captured entirely by the models. Hence, even if hindcast simulations with prescribed observed SSTs are used, it is not guaranteed that the effects of natural variability on ozone trends are fully captured. It would be interesting to compare the modeled trend distributions from the REF-C2 simulations to such hindcast simulations (REF-C1), however, in CCMI-1 the data of those hindcast simulations are only available until 2010. The

assessment of the representation of natural variability and its effects on ozone would require a more in-depth analysis, which we leave for future studies.

5 Conclusions

In the present study, we analyzed in detail lower stratospheric ozone trends for the recent period 1998–2018, and variations of this period, using a total of 31 simulations of different state-of-the-art chemistry climate models and compared them to the observation based dataset BASIC$_{SG}$. Moreover, we linked the ozone trends to stratospheric circulation trends and discussed the reasons for the differences in the LS ozone trends between models and observations. The main findings of our study are summarized in the following.

1) LS ozone trends over the period 1998–2018 vary strongly across different models and among different ensemble members of the same model. Therefore, internal variability strongly influences the LS ozone trends over this short time period. But even if this high variability is taken into account, none of the model simulations reproduces the pattern of observational ozone trends with negative values extending from the southern to the northern mid-latitudes. Thus the observed LS ozone trend pattern is a rather unlikely realization in state-of-the-art CCM simulations.

2) The models' LS ozone trend (given as the most likely values of the models' trend probability distribution) remains negative in the tropics and positive in the mid-latitudes for variations in the time period between 1995 and 2019. Although there is quite a large spread in the magnitude of model trends, the trends do not show a systematic change for the different periods. For observations, LS trends remain negative in both the tropics and the mid-latitudes for all these periods. In contrast to the models' consistent trend we find a systematic shift in the trend magnitude towards less negative mid-latitude trends for the start years 1999 and 2000, which is likely associated with natural variability.

3)In the tropics, the observed trends are a likely representation by the models' trend distribution. However in the mid-latitudes the observational trends represent an extreme value of the models' probability distribution.

4) Tropical LS ozone trends are linked to the GHG-driven increase of tropical upwelling, confirming previous studies. The robust positive mid-latitude LS ozone trends simulated in the models, on the other hand, are found to be driven by changes in ODS- rather than GHG-driven circulation changes. The effects of the latter average to about zero ozone trends between 30° and 50°N, because of competing processes of advection along the shallow versus deep circulation branch, and of two-way mixing.

5) In all models, negative trends in the tropics and positive trends in the mid-latitudes emerge for periods extending into the future (2040), but the models differ in the timing by which trends stabilise. If ozone variability and forced trends would be realistically simulated in the models, we should expect positive mid-latitude ozone trends to emerge in the next 1-2 decades from observational records, too.

Finally we discussed the question as to whether the apparent discrepancy between model and observational trends is due to the misrepresentation of certain processes in the models (e.g., mixing strength, residual circulation strength) or due to inadequate representation of natural variability (ENSO/QBO). Or additionally, the observational trend could just be an extreme (but plausible) realization of the models' trend distribution. Another hypothesis that could emerge from our results is that the discrepancy of mid-latitude ozone trends might stem from an overestimation of ODS-induced ozone recovery in the recent decades in models compared to observations. This effect would be consistent with the weaker upper-stratospheric ozone trends

in the observations compared to models. However, this hypothesis needs further investigation, as does the role of different transport processes for LS ozone trends.

Author contributions. SD performed the data analysis and produced the figures. SD, RE and HG made substantial contributions to conception and design, analysis and interpretation of the data. WTB provided the observational ozone trends and contributed to the interpretation of the results. Moreover all authors participated in drafting the article.

DATA: All data of CCMVal-2 and CCMI-1 used in this study can be obtained through the British Atmospheric Data Centre (BADC) archive (ftp://ftp.ceda.ac.uk). CESM1-WACCM data have been downloaded from http://www.earthsystemgrid.org. For instructions for access to both archives see http://blogs.reading.ac. uk/ccmi/badc-data-access

The authors declare that they have no conflict of interest.

Acknowledgements. This study was funded by the Helmholtz Association under grant VH-NG-1014 (Helmholtz-Hochschul- Nachwuchsforschergruppe MACClim). We acknowledge the modeling groups for making their simulations available for this analysis, the joint WCRP SPARC/IGAC Chemistry-Climate Model Initiative (CCMI) for organizing and coordinating the model data analysis activity, and the British Atmospheric Data Centre (BADC) for collecting and archiving the CCMI model output. We thank Marta Abalos for providing us an additional (forth) ensemble member of WACCM. Moreover, we want to acknowledge that the EMAC simulations were done within the project ESCiMo (Earth System Chemistry integrated Modelling), a national (German) contribution to the ChemistryClimate Model Initiative, and have been performed at the German Climate Computing Centre DKRZ through support from the Bundesministerium fuer Bildung und Forschung (BMBF). W.T.B. was funded by the SNSF project 200020_182239 (POLE)". RE acknowledges support by GA CR under grant nos. 16-01562J and 18-01625S. BASIC$_{SG}$ for 1985-2019 will be available for download from https://data.mendeley.com/datasets/2mgx2xzzpk/4 following review of this paper. GOZCARDS ozone data contributions from L. Froidevaux, R. Wang, J. Anderson, and R. A. Fuller at the Jet Propulsion Laboratory are gratefully acknowledged. Last, we also thank the reviewers for their constructive comments on our manuscript.

References

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.

Abalos, M., Polvani, L., Calvo, N., Kinnison, D., Ploeger, F., Randel, W., and Solomon, S.: New insights on the impact of ozone-depleting substances on the Brewer-Dobson circulation, Journal of Geophysical Research: Atmospheres, 124, 2435–2451, 2019.

Akiyoshi, H., Nakamura, T., Miyasaka, T., Shiotani, M., and Suzuki, M.: A nudged chemistry-climate model simulation of chemical constituent distribution at northern high-latitude stratosphere observed by SMILES and MLS during the 2009/2010 stratospheric sudden warming, Journal of Geophysical Research: Atmospheres, 121, 1361–1380, 2016.

Ball, W. T., Alsing, J., Mortlock, D. J., Rozanov, E. V., Tummon, F., and Haigh, J. D.: Reconciling differences in stratospheric ozone composites, Atmospheric Chemistry and Physics, 17, 12 269–12 302, 2017.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., et al.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmospheric Chemistry and Physics, 18, 1379–1394, 2018.

Ball, W. T., Alsing, J., Staehelin, J., Davis, S. M., Froidevaux, L., and Peter, T.: Stratospheric ozone trends for 1985–2018: sensitivity to recent large variability, Atmospheric Chemistry and Physics Discussions, 2019, 1–27, doi:10.5194/acp-2019-243, https://www.atmos-chem-phys-discuss.net/acp-2019-243/, 2019.

Ball, W. T., Chiodo, G., Abalos, M., Alsing, J., and Stenke, A.: Inconsistencies between chemistry–climate models and observed lower stratospheric ozone trends since 1998, Atmospheric chemistry and physics, 20, 9737–9752, 2020.

Bednarz, E. M., Maycock, A. C., Abraham, N. L., Braesicke, P., Dessens, O., and Pyle, J. A.: Future Arctic ozone recovery: the importance of chemistry and dynamics, Atmospheric Chemistry and Physics, 16, 12 159–12 176, 2016.

Bekki, S., Rap, A., Poulain, V., Dhomse, S., Marchand, M., Lefèvre, F., Forster, P. M., Szopa, S., and Chipperfield, M. P.: Climate impact of stratospheric ozone recovery, Geophysical Research Letters, 40, 2796–2800, 2013.

Bellenger, H., Guilyardi, É., Leloup, J., Lengaigne, M., and Vialard, J.: ENSO representation in climate models: From CMIP3 to CMIP5, Climate Dynamics, 42, 1999–2018, 2014.

Birner, T. and Bönisch, H.: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere, Atmospheric Chemistry and Physics, 11, 817–827, 2011.

Butchart, N.: The Brewer-Dobson circulation, Reviews of Geophysics, 52, 157–184, doi:10.1002/2013RG000448, http://dx.doi.org/10.1002/2013RG000448, 2014.

Calvo, N., Garcia, R., Randel, W., and Marsh, D.: Dynamical mechanism for the increase in tropical upwelling in the lowermost tropical stratosphere during warm ENSO events, Journal of the Atmospheric Sciences, 67, 2331–2340, 2010.

Chipperfield, M. P., Bekki, S., Dhomse, S., Harris, N. R., Hassler, B., Hossaini, R., Steinbrecht, W., Thiéblemont, R., and Weber, M.: Detecting recovery of the stratospheric ozone layer, Nature, 549, 211–218, 2017.

Chipperfield, M. P., Dhomse, S., Hossaini, R., Feng, W., Santee, M. L., Weber, M., Burrows, J. P., Wild, J. D., Loyola, D., and Coldewey-Egbers, M.: On the Cause of Recent Variations in Lower Stratospheric Ozone, Geophysical Research Letters, 2018.

Chrysanthou, A., Maycock, A. C., Chipperfield, M. P., Dhomse, S., Garny, H., Kinnison, D., Akiyoshi, H., Deushi, M., Garcia, R. R., Jöckel, P., et al.: The effect of atmospheric nudging on the stratospheric residual circulation in chemistry–climate models, Atmospheric Chemistry and Physics, 19, 11 559–11 586, 2019.

Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H., Fujiwara, M., and Damadeo, R.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: A long-term database for climate studies, Earth system science data, 8, 461, 2016.

Deushi, M. and Shibata, K.: Development of a Meteorological Research Institute chemistry-climate model version 2 for the study of tropospheric and stratospheric chemistry, Pap. Meteorol. Geophys., 62, 1–46, 2011.

Dhomse, S. S., Kinnison, D., Chipperfield, M. P., Salawitch, R. J., Cionni, I., Hegglin, M. I., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bednarz, E. M., et al.: Estimates of ozone return dates from Chemistry-Climate Model Initiative simulations, Atmospheric Chemistry and Physics, 18, 8409–8438, 2018.

Dietmüller, S., Ponater, M., and Sausen, R.: Interactive ozone induces a negative feedback in CO2-driven climate change simulations, Journal of Geophysical Research: Atmospheres, 119, 1796–1805, 2014.

Dietmüller, S., Eichinger, R., Garny, H., Birner, T., Boenisch, H., Pitari, G., Mancini, E., Visioni, D., Stenke, A., Revell, L., et al.: Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models, Atmospheric Chemistry and Physics, 18, 6699–6720, 2018.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., et al.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, 2013.

Eichinger, R., Dietmüller, S., Garny, H., Šácha, P., Birner, T., Bönisch, H., Pitari, G., Visioni, D., Stenke, A., Rozanov, E., et al.: The influence of mixing on the stratospheric age of air changes in the 21st century, Atmospheric Chemistry and Physics, 19, 921–940, 2019.

Froidevaux, L., Anderson, J., Wang, H.-J., Fuller, R., Schwartz, M., Santee, M., Livesey, N., Pumphrey, H., Bernath, P., Russell III, J., et al.: Global OZone Chemistry And Related Datasets for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H 2 O, and O 3., Atmospheric Chemistry & Physics Discussions, 15, 2015.

Garcia, R. R., Smith, A. K., Kinnison, D. E., Cámara, Á. d. l., and Murphy, D. J.: Modification of the Gravity Wave Parameterization in the Whole Atmosphere Community Climate Model: Motivation and Results, Journal of the Atmospheric Sciences, 74, 275–291, 2017.

Garny, H., Birner, T., Bönisch, H., and Bunzel, F.: The effects of mixing on age of air, Journal of Geophysical Research: Atmospheres, 119, 7015–7034, 2014.

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, 2011.

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

Hardiman, S. C., Butchart, N., O'Connor, F. M., and Rumbold, S. T.: The Met Office HadGEM3-ES chemistry–climate model: evaluation of stratospheric dynamics and its impact on ozone, Geoscientific Model Development, 10, 1209–1232, 2017.

Harris, N. R., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., Steinbrecht, W., Anderson, J., Bhartia, P. K., Boone, C., et al.: Past changes in the vertical distribution of ozone–Part 3: Analysis and interpretation of trends, Atmospheric Chemistry and Physics, 15, 9965–9982, 2015.

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

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., and Elliott, S.: CICE: the Los Alamos Sea Ice Model Documentation and Software User?s Manual Version 4.1 LA-CC-06-012, T-3 Fluid Dynamics Group, Los Alamos National Laboratory, 675, 2010.

Imai, K., Manago, N., Mitsuda, C., Naito, Y., Nishimoto, E., Sakazaki, T., Fujiwara, M., Froidevaux, L., von Clarmann, T., Stiller, G. P., et al.: Validation of ozone data from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES), Journal of Geophysical Research: Atmospheres, 118, 5750–5769, 2013.

Jensen, M. P., Mather, J. H., and Ackerman, T. P.: Observations of the 1997–98 warm ENSO event at the Manus Island ARM site, Geophysical research letters, 25, 4517–4520, 1998.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), Geosci. Model Dev., **3**, 717–752, doi:10.5194/gmd-3-717-2010, 2010.

Jöckel, P., Holger, T. O., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C.A, M., Brinkop, S., Cai, D., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth System Chemistry Integrated Modelling (ESCiMO) with the Modular Earth Submodel System (MESSy, version 2.51), Geosci. Model Dev., **9**, 1153–1200, doi:doi:10.5194/gmd-9-1153-2016, 2016.

Jonsson, A., De Grandpre, J., Fomichev, V., McConnell, J., and Beagley, S.: Doubled CO2-induced cooling in the middle atmosphere: Photochemical analysis of the ozone radiative feedback, Journal of Geophysical Research: Atmospheres, 109, 2004.

Kirner, O., Ruhnke, R., and Sinnhuber, B.-M.: Chemistry–Climate Interactions of Stratospheric and Mesospheric Ozone in EMAC Long-Term Simulations with Different Boundary Conditions for CO2, CH4, N2O, and ODS, Atmosphere-Ocean, 53, 140–152, 2015.

Kusuma, L., Lubis, S. W., and Setiawan, S.: Unprecedented Quasi-Biennial Oscillation (QBO) disruption in 2015-2016: Implications for tropical waves and ozone, in: IOP Conference Series: Earth and Environmental Science, vol. 284, p. 012016, IOP Publishing, 2019.

Laine, M., Latva-Pukkila, N., and Kyrölä, E.: Analysing time-varying trends in stratospheric ozone time series using the state space approach, Atmospheric Chemistry and Physics, 14, 9707–9725, 2014.

Madec, G. et al.: NEMO ocean engine, 2015.

Marchand, M., Keckhut, P., Lefebvre, S., Claud, C., Cugnet, D., Hauchecorne, A., Lefèvre, F., Lefebvre, M.-P., Jumelet, J., Lott, F., et al.: Dynamical amplification of the stratospheric solar response simulated with the Chemistry-Climate model LMDz-Reprobus, Journal of Atmospheric and Solar-Terrestrial Physics, 75, 147–160, 2012.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate change from 1850 to 2005 simulated in CESM1 (WACCM), Journal of climate, 26, 7372–7391, 2013.

Marsh, D. R., Lamarque, J.-F., Conley, A. J., and Polvani, L. M.: Stratospheric ozone chemistry feedbacks are not critical for the determination of climate sensitivity in CESM1 (WACCM), Geophysical Research Letters, 43, 3928–3934, 2016.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M., Lamarque, J.-F., 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, 2011.

Meul, S., Langematz, U., Oberländer, S., Garny, H., and Jöckel, P.: Chemical contribution to future tropical ozone change in the lower stratosphere, Atmospheric Chemistry and Physics, 14, 2959–2971, 2014.

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, 2012.

Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, Geoscientific Model Development, 8, 1339–1356, 2015.

Morgenstern, O., Braesicke, P., O'Connor, F., Bushell, A., Johnson, C., Osprey, S., and Pyle, J.: Evaluation of the new UKCA climate-composition model–Part 1: The stratosphere, Geoscientific Model Development, 2, 43–57, 2009.

Morgenstern, O., Zeng, G., Luke Abraham, N., Telford, P. J., Braesicke, P., Pyle, J. A., Hardiman, S. C., O'Connor, F. M., and Johnson, C. E.: Impacts of climate change, ozone recovery, and increasing methane on surface ozone and the tropospheric oxidizing capacity, Journal of Geophysical Research: Atmospheres, 118, 1028–1041, 2013.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abrams, N. L., Akkyoshi, H., Archibald, A. T., Bekki, S., Butchart, N., Chipperfield, M. P., Deushi, M., Dhomse, S. S., Garcia, R. R., Hardiman, S. C., Horowitz, L. W., Jöckel, P., Josse, B., Kinnison, D., Lin, M. Y., Mancini, E., Manyin, M. E., Marchand, M., Marecal, V., Michou, M., Pitari, L. D. O. N. G. ans Plummer, D. A., Revell, L. E., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T. Y., Tilmes, S., Yamashita, Y., Yoshida, K., and Zeng, G.: Review of the global models used within the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 10, 639–671, doi:10.5194/gmd-10-639-2017, 2017.

Newman, P., Daniel, J., Waugh, D., and Nash, E.: A new formulation of equivalent effective stratospheric chlorine (EESC), 2007.

O'Connor, F., Johnson, C., Morgenstern, O., Abraham, N., Braesicke, P., Dalvi, M., Folberth, G., Sanderson, M., Telford, P., Voulgarakis, A., et al.: Evaluation of the new UKCA climate-composition model–Part 2: The Troposphere, Geoscientific Model Development, 7, 41–91, 2014.

Olsen, M. A., Manney, G. L., and Liu, J.: The ENSO and QBO Impact on Ozone Variability and Stratosphere-Troposphere Exchange Relative to the Subtropical Jets, Journal of Geophysical Research: Atmospheres, 124, 7379–7392, 2019.

Oman, L., Plummer, D., Waugh, D., Austin, J., Scinocca, J., Douglass, A., Salawitch, R., Canty, T., Akiyoshi, H., Bekki, S., et al.: Multimodel assessment of the factors driving stratospheric ozone evolution over the 21st century, Journal of Geophysical Research: Atmospheres, 115, 2010.

Oman, L., Ziemke, J., Douglass, A., Waugh, D., Lang, C., Rodriguez, J., and Nielsen, J.: The response of tropical tropospheric ozone to ENSO, Geophysical Research Letters, 38, 2011.

Oman, L. D., Douglass, A. R., Ziemke, J. R., Rodriguez, J. M., Waugh, D. W., and Nielsen, J. E.: The ozone response to ENSO in Aura satellite measurements and a chemistry-climate simulation, Journal of Geophysical Research: Atmospheres, 118, 965–976, 2013.

Orbe, C., Wargan, K., Pawson, S., and Oman, L. D.: Mechanisms Linked to Recent Ozone Decreases in the Northern Hemisphere Lower Stratosphere, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 631, 2020.

Petropavlovskikh, I., Godin-Beekmann, S., Hubert, D., Damadeo, R., Hassler, B., and Sofieva, V.: SPARC/IO3C/GAW report on Long-term Ozone Trends and Uncertainties in the Stratosphere, 2019.

Pitari, G., Aquila, V., Kravitz, B., Robock, A., Watanabe, S., Cionni, I., Luca, N. D., Genova, G. D., Mancini, E., and Tilmes, S.: Stratospheric ozone response to sulfate geoengineering: Results from the Geoengineering Model Intercomparison Project (GeoMIP), Journal of Geophysical Research: Atmospheres, 119, 2629–2653, 2014.

Polvani, L. M., Wang, L., Abalos, M., Butchart, N., Chipperfield, M., Dameris, M., Deushi, M., Dhomse, S., Jöckel, P., Kinnison, D., et al.: Large Impacts, Past and Future, of Ozone-Depleting Substances on Brewer-Dobson Circulation Trends: A Multimodel Assessment, Journal of Geophysical Research: Atmospheres, 124, 6669–6680, 2019.

Portmann, R. W. and Solomon, S.: Indirect radiative forcing of the ozone layer during the 21st century, Geophysical research letters, 34, 2007.

Randel, W. J. and Wu, F.: A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data, Journal of Geophysical Research: Atmospheres, 112, 2007.

Randel, W. J., Garcia, R., and Wu, F.: Dynamical balances and tropical stratospheric upwelling, Journal of the atmospheric sciences, 65, 3584–3595, 2008.

Randel, W. J., Garcia, R. R., Calvo, N., and Marsh, D.: ENSO influence on zonal mean temperature and ozone in the tropical lower stratosphere, Geophysical Research Letters, 36, 2009.

Ravishankara, A., Daniel, J. S., and Portmann, R. W.: Nitrous oxide (N2O): the dominant ozone-depleting substance emitted in the 21st century, science, 326, 123–125, 2009.

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, 115, 2010.

Revell, L., Tummon, F., Stenke, A., Sukhodolov, T., Coulon, A., Rozanov, E., Garny, H., Grewe, V., and Peter, T.: Drivers of the tropospheric ozone budget throughout the 21st century under the medium-high climate scenario RCP 6.0, Atmospheric Chemistry and Physics, 15, 5887–5902, 2015.

Revell, L. E., Bodeker, G. E., Smale, D., Lehmann, R., Huck, P. E., Williamson, B. E., Rozanov, E., and Struthers, H.: The effectiveness of N2O in depleting stratospheric ozone, Geophysical research letters, 39, 2012.

Scinocca, J., McFarlane, N., Lazare, M., Li, J., and Plummer, D.: Technical Note: The CCCma third generation AGCM and its extension into the middle atmosphere, Atmospheric Chemistry and Physics, 8, 2008.

Solomon, S., Kinnison, D., Bandoro, J., and Garcia, R.: Simulation of polar ozone depletion: An update, Journal of Geophysical Research: Atmospheres, 120, 7958–7974, 2015.

SPARC CCMVal: SPARC Report No 5 (2010) Chemistry-Climate Model Validation, WCRP-132, WMO/TD-No. 1526, 2010.

Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., et al.: An update on ozone profile trends for the period 2000 to 2016, Atmospheric chemistry and physics, 17, 10 675–10 690, 2017.

Stenke, A., Schraner, M., Rozanov, E., Egorova, T., Luo, B., and Peter, T.: The SOCOL version 3.0 chemistry–climate model: description, evaluation, and implications from an advanced transport algorithm, Geoscientific Model Development, 6, 1407–1427, 2013.

Stone, K., Morgenstern, O., Karoly, D., Klekociuk, A., French, W., Abraham, N., and Schofield, R.: Evaluation of the Australian Community Climate and Earth-System Simulator Chemistry-Climate Model., Atmospheric Chemistry & Physics Discussions, 15, 2015.

Stone, K. A., Solomon, S., and Kinnison, D. E.: On The Identification of Ozone Recovery, Geophysical Research Letters, 2018.

Sudo, K. and Akimoto, H.: Global source attribution of tropospheric ozone: Long-range transport from various source regions, Journal of Geophysical Research: Atmospheres, 112, 2007.

Szopa, S., Balkanski, Y., Schulz, M., Bekki, S., Cugnet, D., Fortems-Cheiney, A., Turquety, S., Cozic, A., Déandreis, C., Hauglustaine, D., et al.: Aerosol and ozone changes as forcing for climate evolution between 1850 and 2100, Climate dynamics, 40, 2223–2250, 2013.

Tian, W. and Chipperfield, M. P.: A new coupled chemistry–climate model for the stratosphere: The importance of coupling for future O3-climate predictions, Quarterly Journal of the Royal Meteorological Society, 131, 281–303, 2005.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., Smith, A. K., Neely, R. R., Conley, A., Vitt, F., et al.: Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), Geoscientific Model Development Discussions, 2016.

Walters, D., Williams, K., Boutle, I., Bushell, A., Edwards, J., Field, P., Lock, A., Morcrette, C., Stratton, R., Wilkinson, J., et al.: The Met Office Unified Model global atmosphere 4.0 and JULES global land 4.0 configurations, Geoscientific Model Development, 7, 361–386, 2014.

Wargan, K., Orbe, C., Pawson, S., Ziemke, J. R., Oman, L. D., Olsen, M. A., Coy, L., and Knowland, K. E.: Recent decline in extratropical lower stratospheric ozone attributed to circulation changes, Geophysical Research Letters, 2018.

Waugh, D. and Hall, T.: Age of stratospheric air: Theory, observations, and models, Reviews of Geophysics, 40, doi:10.1029/2000RG000101, 2002.

Weber, M., Coldewey-Egbers, M., Fioletov, V. E., Frith, S. M., Wild, J. D., Burrows, J. P., Long, C. S., and Loyola, D.: Total ozone trends from 1979 to 2016 derived from five merged observational datasets–the emergence into ozone recovery, Atmospheric Chemistry and Physics, 18, 2097–2117, 2018.

Winterstein, F., Tanalski, F., Jöckel, P., Dameris, M., and Ponater, M.: Implication of strongly increased atmospheric methane concentrations for chemistry–climate connections, Atmospheric Chemistry and Physics, 19, 7151–7163, 2019.

WMO: Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring ProjectReport, World Meteorological Organization, pp. Report No. 58, 588 pp, 2018.

Yukimoto, S., Yoshimura, H., Hosaka, M., Sakami, T., Tsujino, H., Hirabara, M., Tanaka, T., Deushi, M., Obata, A., Nakano, H., et al.: Development of the Meteorological Research Institute-Earth System Model (MRIESM1)?model description, Tech. rep., Technical reports of the Meteorological Research Institute, 2011.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., et al.: A new global climate model of the Meteorological Research Institute: MRI-CGCM3 model description and basic performance, Journal of the Meteorological Society of Japan. Ser. II, 90, 23–64, 2012.

Supplement of "Analysis of recent lower stratospheric ozone trends in chemistry climate models"

Dietmueller et al.

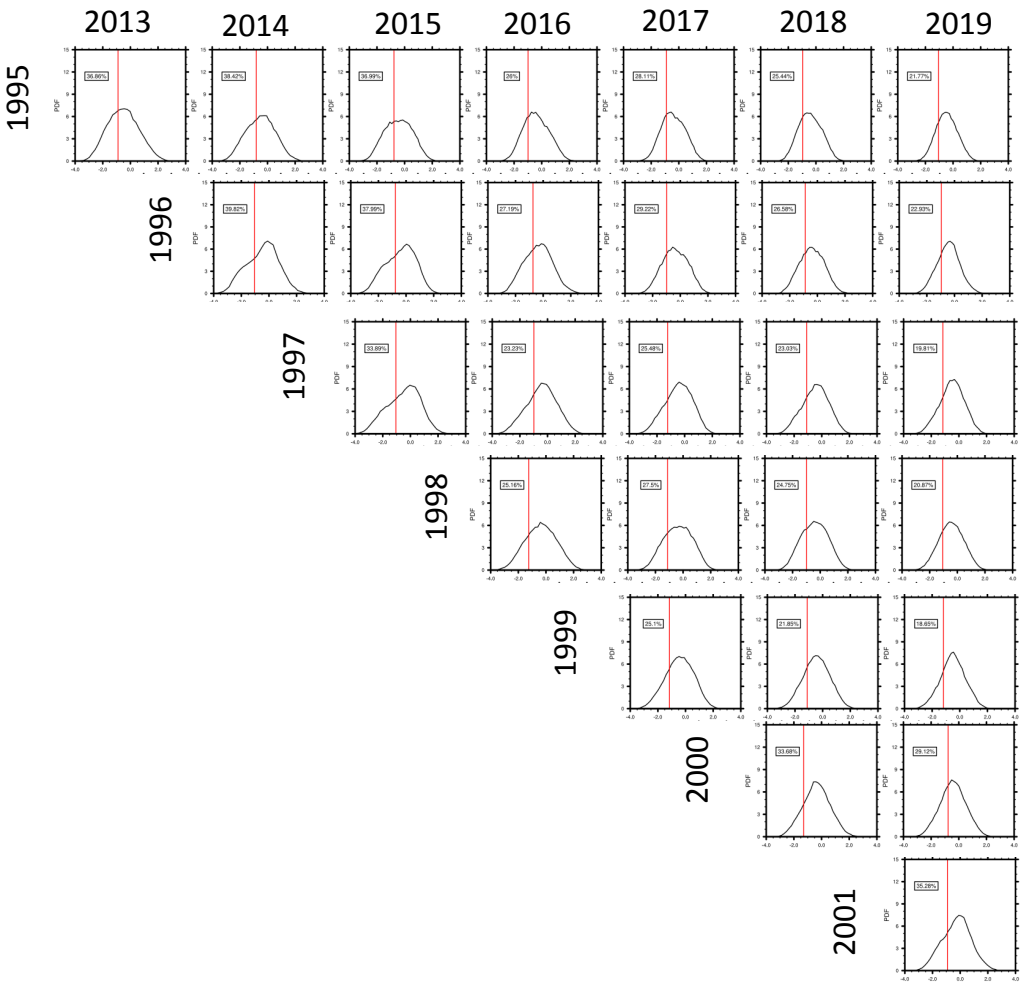

Figure S1. Probability distribution function (PDF) of the models' LS tropical ozone trend (20°S-20°N, 30-100 hPa) as function of different periods. In all panels the x-coordinate denotes the different end years (2013-2019) and the y-coordinate the different start years (1995-2001). The red line indicates the respective observational ozone trend value. Moreover the probability (in %) of the observational trend lying within the models' distribution is given within each panel.

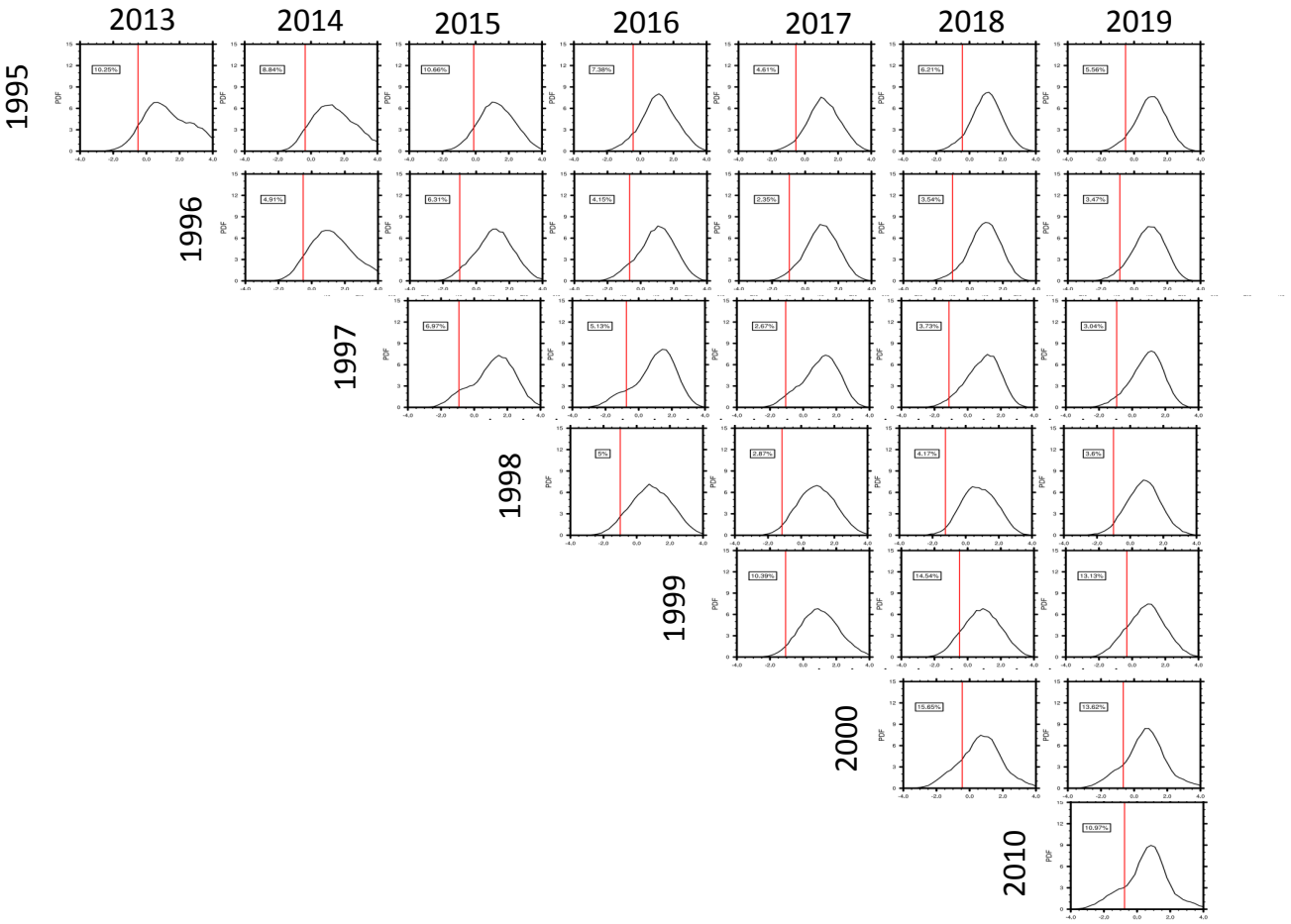

Figure S2. Same as Fig. S1, but PDFs for the models' LS northern mid-latiudes ozone trend (30°N-50°N, 30-150 hPa).

Table S1. MMM ozone trends and their 1-sigma standard deviations [DU/dec] in the lower stratosphere (LS, 30-100/150 hPa) and in the upper stratosphere (US, 1-10 hPa) for the tropical and northern mid-latitudes. Trends are given for the REF-C2 simulation, as well as for the sensitivity simulations with fixed GHG and fixed ODS concentrations (fGHG and fODS). Trends are calculated over the period 1998-2018 and 1998-2040.

MMM trop. O3 trend	**LS**	**US**
REF-C2 (1998-2018)	-0.71±0.40	0.68±0.41
REF-C2 (1998-2040)	-0.55±0.38	0.71±0.25
fGHG (1998-2018)	-0.35±0.41	0.64±0.26
fGHG (1998-2040)	0.03±0.06	0.61±0.07
fODS (1998-2018)	-1.09±0.59	0.51±0.27
fODS (1998-2040)	-0.85±0.48	0.44±0.12
MMM NH mid-lat. O3 trend	**LS**	**US**
REF-C2 (1998-2018)	0.76±1.12	0.64±0.34
REF-C2 (1998-2040)	0.79±0.60	0.75±0.20
fGHG (1998-2018)	0.62±1.11	0.43±0.25
fGHG (1998-2040)	0.68±0.22	0.58±0.16
fODS (1998-2018)	-0.18±0.84	0.41±0.39
fODS (1998-2040)	0.04±0.40	0.43±0.15

Analysis of recent lower stratospheric ozone trends in chemistry climate models

Simone Dietmüller[1], Hella Garny[1,2], Roland Eichinger[2,1,3], and William T. Ball[4,5,6]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Ludwig Maximilians Universität, Faculty of Physics, Institute for Meteorology, Munich, Germany
[3]Charles University, Department of Atmospheric Physics, Faculty of Mathematics and Physics, Prague, Czech Republic
[4]Institute for Atmospheric and Climate Science, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland
[5]Physikalisch-Meteorologisches Observatorium Davos World Radiation Centre, Dorfstrasse 33, 7260 Davos Dorf, Switzerland
[6]Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, TU Delft, Stevinweg 1, 2628 CN Delft, the Netherlands

Correspondence to: S. Dietmüller (simone.dietmueller@dlr.de)

Abstract. Recent observations show a significant decrease of lower stratospheric (LS) ozone concentrations in tropical and mid-latitude regions since 1998. By analyzing 31 chemistry climate model (CCM) simulations performed for the Chemistry Climate Model Initiative (CCMI, **?**), we find a large spread in the 1998–2018 trend patterns between different CCMs and between different realizations performed with the same CCM. The latter, in particular, indicates that natural variability strongly influences LS ozone trends. However none of the model simulations reproduces the observed ozone trend structure of coherent negative trends in the LS. In contrast to the observations, most models show a ~~dipole trend pattern in the LS~~ LS trend pattern with negative trends in the tropics (20°S-20°N) and positive trends in the northern mid-latitudes (30°-50°N) or vice versa. To investigate the influence of natural variability on ~~the~~ historical LS ozone trends we analyze the sensitivity of observational trends and the models' trend probability distributions for ~~slightly varied post-ODS (ozone depleting substances) periods .~~ varying periods with start dates between 1995 to 2001 and end dates between 2013 to 2019. Generally, modeled and observed LS trends remain robust for ~~different post-ODS~~ these different periods, however observational data show a ~~systematic~~ change towards weaker mid-latitude trends for certain periods, likely forced by natural variability~~for certain periods~~. Moreover we ~~can~~ show that in the tropics the observed trends agree ~~quite~~ well with the models' trend distribution, whereas in the mid-latitudes the observational trend is ~~a rather~~ typically an extreme value of the models' distribution. We further investigate the LS ozone trends for extended periods reaching into the future and find that all models develop a ~~dipole trend pattern in the future, i.e. in almost all models the~~ positive ozone trend at mid-latitudes and the trends converge to constant values ~~for the entire period~~ by the period that spans 1998–2060. ~~An investigation of interannual ozone variability also reveals a clear dipole pattern in ozonevariability in all CCMs and in observational data, however it is more pronounced in the models. Thus, although the LS~~ Intermodel correlations between ozone trends and measures of transport circulation trends confirm the dominant role of greenhouse gas (GHG)-driven tropical upwelling enhancement on the tropical LS ozone decrease. Mid-latitude ozone, on the other hand, appears to be influenced by multiple competing factors: an enhancement in the shallow branch decreases ozone, while an enhancement in the deep branch increases ozone and, furthermore, mixing plays a role here too. Sensitivity simulations

with fixed forcing of GHGs or ozone depleting substances (ODSs) reveal that the GHG-driven increase in circulation strength do not lead to a net trend in LS mid-latitude column ozone. Rather, the positive ozone trends simulated consistently in the models in this region emerge from the decline of ODSs, i.e. the ozone ~~variability pattern is similar, the probability of overall negative LS ozone trends simultaneously in the tropics and mid-latitudes is higher in observations. To access the dynamical influence on LS ozone trends, the models' tropical upwelling trends are correlated against their LS ozone trends. For the period 1998–2018 tropical ozone trends are negatively correlated (-0.83) and~~ recovery. Therefore, we hypothesize that next to the influence of natural variability, the disagreement of modeled and observed LS mid-latitude ~~trends are positively correlated (+0.49). However, the correlation in the mid-latitudes is rather weak and not robust for slightly varied time periods, which indicates that other processes like two-way mixing play an important role here, too~~ozone trends could indicate a mismatch in the relative role of the response of ozone to ODS versus GHG-forcing in the models.

1 Introduction

Stratospheric ozone is essential for protecting the Earth's surface from ultra violet radiation, which is harmful for plants, animals and humans. Human-made ozone depleting substance (ODS) emissions significantly reduced ozone concentrations for some decades after 1960. After controlling the use of ODSs by the 1987 Montreal protocol and later adjustments, however, ~~ODSs~~ ODS concentrations started to decline in the mid-to-late 1990s (e.g. **??**). As a consequence, total stratospheric ozone is expected to recover in the future. **?** have analyzed the recovery of stratospheric ozone mixing ratios of the CCMI-1 (Chemistry Climate Model Intercomparison project part 1) climate projection simulations. They found that the ozone layer is simulated to return to ~~pre-ODS~~ a pre-1980 ODS level between 2030 and 2060, depending on the region. However, they discovered a large spread among the individual models, which shows that there are many uncertainties in these projections. The evolution of stratospheric ozone in the 21st century does not only result from a decrease in ~~ODSs~~ ODS concentrations but also from an interplay between changes in both the atmospheric composition and the circulation (World Meteorological Organization (WMO) 2014). Increasing anthropogenic greenhouse gas (GHG) emissions (CO_2, CH_4, N_2O) ~~lead~~ leads to enhanced tropical upwelling and thereby to an acceleration of tracer transport along the stratospheric overturning circulation (e.g. **??**). On the other hand, increasing GHGs also ~~slow~~ slows down ozone depletion through GHG-induced stratospheric cooling (e.g. **?????**) and emissions of CH_4 and N_2O additionally impact ozone ~~by~~ through chemical processes (e.g. **????**).

In the recent years, a number of studies have analyzed observational records to identify ozone trends in the stratosphere (e.g. **???**). These studies consistently report an ozone recovery in the upper stratosphere after the turnaround of the ODS concentrations ~~in the year 1998 (start of the post-ODS period).~~ around the year 1998. In the lower stratosphere (LS), however, most observed ozone trends are ~~statistically not significant for that~~ not statistically significant for such a relatively short period due to large internal variability and instrumental difficulties (e.g. **?**). Subsequently, **?** analyzed LS ozone trends from satellite data between 1998 and 2016 in detail making use of a dynamical (multiple) linear regression analysis. ~~This way, they~~ They identified a statistically significant decline of LS ozone between 60°S and 60°N in that period, ~~by~~ of approximately 2 DU in the LS below 24 km of altitude. The implication was that the stratospheric ozone column was continuing to decline, because

the LS ozone reduction more than offsets the positive trend in the upper stratosphere. Shortly afterwards **?** studied ozone trends in the reanalysis products MERRA-2 and GEOS-RPIT. In the tropics they detected a positive ozone trend in a 5 km layer above the tropopause and a negative trend at 7-15 km above the tropopause. Nevertheless, in the northern and southern mid-latitude LS they detected a negative ozone trend. As such, there are some similarities to the findings of **?**, but there are also ~~major~~ quantitative differences, for example the positive trend in the 5 km layer or a missing overall ~~negative trend in the LS~~statistically significant decrease in the column integrated ozone. **?** suggested that the negative mid-latitude trend might be explained by enhanced isentropic transport between the tropical and mid-latitude LS. However, the recent study of **?** explicitly demonstrated that in the NH this mid-latitude ozone decrease is primarily associated with large scale advection. Furthermore, they showed that the observed changes in advection and in ozone are well within the range of model variability (gauged from one CCM). By means of using a chemistry transport model (CTM) and extending the analysis period to the year 2017, **?** suggested that the negative LS ozone trends are only a result of large natural variability. They showed that there ~~is~~ was a strong positive ozone anomaly in 2017 which is driven by short term dynamical transport of ozone, and concluded that this points to large year-to-year variability rather than to an ongoing downward trend. However, an update of the data set which was used in **?** showed that the large interannual variability alone cannot explain the entire trend in **?** (see **?**): the larger year-to-year variability in the SH was implicated to ~~results~~ result from a non-linear interaction between the quasi-biennal-oscillation (QBO) and seasonal variability and despite this large variability the observed negative LS ozone trend remains. ~~To get more~~

To improve confidence in future projections of the ozone layer it is important to ~~know how well~~ evaluate the skill of chemistry climate models (CCMs) ~~simulate the observed post-ODS peak ozone trends~~ in simulating the observed ozone trends over recent decades. A direct comparison between the CCM multi-model-mean (MMM) values and observational data showed that the ozone trend profiles of modeled MMM data agree well with observations, except in the lowermost mid-latitude stratosphere ~~(??)~~(**??**). The most recent study of **?** investigated LS ozone trends of the 1998–2016 period in merged satellite data and compared them to the ozone trends in CCMs using the climate projection simulations of the CCMVal2 project. Similar to the observations, the CCMs showed a decline in LS ozone in the tropics, likely due to enhanced tropical upwelling, following from an increase in greenhouse gases (see e.g. **?**). In contrast to the observations, however, models do not show a decrease, but rather an increase in LS mid-latitude ozone. **?** argue that these discrepancies in the LS between models and observations can possibly be explained by differences in the horizontal two-way mixing between the tropics and mid-latitudes, though they did not provide explicit evidence from the models ~~, only observations (see also **?**). This~~ (see also **?**). The study suggested that the negative mid-latitude observational trend is caused by an intensification of two-way mixing (by analyzing effective diffusivity in reanalysis data). On the other hand enhanced downwelling of ozone-rich air to the mid-latitudes could consequently lead to a positive trend in the mid-latitudes. Apparently, the processes that determine mid-latitude LS ozone in models and observations are not understood so far.

In the present study, we seek to quantify whether the observed LS ozone trends lie within the suite of modeled trends. If yes, this would imply that the observed trend is just one realization of possible trends given within the large year-to-year variability. If not, this would imply that either models do not represent year-to-year variability correctly, or that there is a forced trend in the real world that is not adequately represented in the models. In contrast to the study of **?** we are using the simulation data of

a more recent inter-model comparison project (namely the Chemistry Climate Model Initiative, phase 1, CCMI-1) and analyze the ozone trends for a wider range ~~and for improved~~ of updated current state-of-the-art CCMs including all their ensemble simulations.

A brief description of the model simulations, of the observational data ~~set~~ sets and of the ~~used statistical methods~~ methods used is presented in Section ~~2.1~~2. In Section 3 we show our results~~: in section 3.1 we~~. We provide a detailed comparison of ozone trends over the years 1998–2018 in different CCM simulations and observations ~~, with a~~ (Section 3.1). Here we focus on LS ozone trends, and we investigate how natural variability influences these LS ozone trends (Section 3.2 and 3.3). ~~Moreover we~~ We link LS ozone trends with stratospheric ~~upwelling~~ transport trends (Section 3.4) ~~, and compare the year-to-year variability in observations and in all CCM simulations~~ and we investigate how ozone trends are forced by GHG and ODS emissions (Section ~~??~~3.5). A discussion of the reasons for the disagreement in the LS mid-latitude ozone ~~trend~~ trends between models and observations and the conclusions follow in Sections 4 and 5, respectively.

2 Data and Methods

2.1 Models and Simulations

In the present study, we analyze the model output from ~~overall~~ 18 state-of-the-art CCMs from the Chemistry Climate Model Initiative phase 1 (CCMI-1, **?**). Tab. 1 lists all these CCMs together with their references, ~~their underlying~~ the forcing that underlies the sea surface temperatures (SSTs), and the simulation type considered. A detailed overview of all models that participated in CCMI-1 can be found in **?**. We mainly evaluate the long term ~~free running~~ 'free running' simulations of CCMI-1 (REF-C2), as they span the time period ~~we are interested in (namely~~ 1998–2018~~)~~. We do not use ~~the the free running~~REF-C1 'free running' simulations of the recent past or the specified dynamics simulations (REF-C1SD), as they only span the period from 1998 to 2010. Moreover we want to point out that the specified dynamics simulations performed for CCMI do not represent stratospheric circulation better than the ~~free running~~ 'free running' simulations: **?** compared stratospheric residual circulation among specified dynamic (SD) simulations and found that the spread in these simulations is even larger than in REF-C2. Furthermore **?** showed poor agreement with the observed ozone trend for some selected SD simulations of CCMI. For the REF-C2 model simulations used in our study, all available ensemble members of the individual models are taken into account. The ensemble size of a certain simulation (if ensemble simulations were performed) is also given in Tab. 1 (brackets after simulations). Thus for the REF-C2 ~~simulation~~ simulations, 18 models performed a total of 31 realizations (~~6~~ six models performed multiple ensemble members simulations). The REF-C2 simulations include hindcast and forecast periods spanning 1960–2100. They are all ~~free running~~ 'free running' simulations, thus each model simulation has its own internal variability. Note that REF-C2 simulations use a variety of different SSTs and SICs (sea ice concentrations), either prescribed climate model SST fields from offline model simulations (of the same or of a different model) or they are coupled to an interactive ocean and sea ice module. Moreover the representation of the QBO is different across the CCMs, with models having an internally generated QBO (e.g. MRI, EMAC-L90), nudged QBO (e.g. NIES, WACCM, SOCOLv3, EMAC-L47, EMAC-L47-o) or no QBO (e.g. CMAM, LMDZ). REF-C2 reference simulations follow the WMO (2011) A1 scenario for ~~ozone-depleting~~

~~substances~~ ODSs and the RCP 6.0 scenario (**?**) for other greenhouse gases, tropospheric ozone precursors, and aerosol and aerosol precursor emissions. For anthropogenic emissions, the CCMI recommendation was to use MACCity (**?**) until 2000, followed by RCP 6.0 emissions. Besides the REF-C2 simulations we also consider the 11 sensitivity simulations with fixed greenhouse gases (fGHG) and with fixed ODSs (fODS) in our analysis. ~~This sensitivity scenario is~~ These sensitivity scenarios are both based on the REF-C2 simulation~~, but the GHGs~~. However in case of the fGHG simulations CO_2, CH_4, N_2O, and other non-ozone depleting GHGs are held at their 1960 value, ~~thus we can~~ and so we are able to study the impact due to ODS concentration changes only (i.e. in the absence of ~~GHG induced~~ GHG-induced climate change). In the case of the fODS simulations the ODS concentrations are fixed to the 1960 level throughout the simulation. All models providing ~~this~~ both of these sensitivity simulations are given in Tab 1.

Table 1. Overview of the CCMI simulations, analyzed for the present study. For the individual CCMs their reference(s), their SSTs and their available simulations (REF-C2~~and ,~~ fGHG, fODS) are given. The numbers in brackets behind the simulations indicate the number of realizations of each REF-C2~~and ,~~ fGHG or fODS simulation. Detailed information about the models' SSTs and the models' representation of the QBO are given in the supplement of **?**.

CCMI Model	Reference(s)	SSTs	Analyzed Simulation
CMAM	**?**	prescribed	REF-C2(1), fGHG(1), fODS(1)
	?		
CESM1-WACCM	**??**	interactive	REF-C2(4)*, fGHG(~~1~~)3), fODS(3)
	?		
EMAC-L90	**??**	prescribed	REF-C2(1) ~~, fGHG(1)~~
EMAC-L47	**??**	prescribed	REF-C2(1)
EMAC-L47-o	**??**	interactive	REF-C2(1)**
GEOSCCM	**??**	prescribed	REF-C2(1)
	??		
MRI	**?**	interactive	REF-C2(1)
	??		
SOCOLv3	**??**	prescribed	REF-C2(1)
NIWA-UKCA	**??**	interactive	REF-C2(5), fGHG(~~3~~)2), fODS(2)
	?		
ULAQ	**?**	prescribed	REF-C2(3), fGHG(1), fODS(1)
HadGEM	**??**	interactive	REF-C2(1)
	??		
	??		
UMUKCA	**??**	prescribed	REF-C2(2)
ACCESS-CCM	**??**	prescribed	REF-C2(3), fGHG(1), fODS(1)
	?		
NIES	**??**	prescribed	REF-C2(1), fGHG(1), fODS(1)
UMSLIMCAT	**?**	prescribed	REF-C2(1), fGHG(1), fODS(1)
CHASER	**?**	interactive	REF-C2(1), fGHG(1), fODS(1)
LMDz-REPROBUS	**??**	interactive	REF-C2(1)
	?		
CESM1-CAM4-Chem	**?**	interactive	REF-C2 (3)

* The fourth ensemble of WACCM (WACCM-4) was provided by M. Abalos; ** EMAC-L47 simulations are not ensembles, as one simulation is with prescribed SSTs and one with interactive ocean

2.2 Observational data

For observations, we make use of the BAyeSian Integrated and Consolidated (BASIC) ozone composite that merges SWOOSH (**?**) and GOZCARDS (**?**) through the BASIC method of **?**. The method was developed to account for artefacts in composite datasets that are a consequence of merging observations from different instruments that each have unique spatial and temporal observing characteristics. As a result, these artefacts can alias in regression analysis and bias, e.g., trend estimates (see examples in **?**). BASIC composites aim to account for and reduce artefacts using an empirically driven Bayesian inference methodology, but it relies on the availability of already developed ozone composites. Here, BASIC$_{SG}$ has been extended to the end of 2019 using the latest versions of GOZCARDS, v2.20, and SWOOSH, v2.6. As such BASIC$_{SG}$ covers 1985-2019 as monthly mean zonal means on a 10° latitude grid from 60°S–60°N and over a pressure range of 147–1 hPa (∼13–48 km). BASIC$_{SG}$ was presented in **?**, and a sensitivity analysis of trends was applied to it in **?**, with examples of data artefacts that it addresses in the accompanying appendix and supplementary materials, respectively.

To obtain an observationally constrained estimate of tropical upwelling and extratropical downwelling mass fluxes, we use ECMWF's fifth generation of atmospheric reanalysis data, ERA5 (**?**). The mass fluxes are calculated from 6-hourly data on the reduced set of pressure levels.

2.3 Statistical Methods

In some parts of our analysis, and to make a robust comparison between multiple models and a single 'real-world' realization, i.e. observations, we form probability distributions to estimate the combined probability of the ozone trends from all REF-C2 models. To do so, we calculate the linear trend and the associated uncertainty using a least squares method for every simulation. Then, to build the trend probability distribution of the models~~trend~~, first one of the 18 CCMI models is randomly selected, assuming that the models are randomly ~~uniform~~ uniformly distributed. In case the selected CCM ~~has ensemble members,~~ provided ensemble member simulations, in a second step one of these members is ~~then also randomly chosen~~randomly chosen, thus taking into account that ensemble members are treated differently than individual models. In the next step, the trend estimate ($t^{M_{i,k}}$) of the specific randomly selected CCMI model M_i with ensemble member k is calculated by randomly choosing ~~a~~ an ozone trend value from the trends associated and assumed normal distribution $\mathcal{N}$, which is based on the mean $\mu_{M_{i,k}}$ and standard deviation $\sigma_{M_{i,k}}$ of the simulations linear trend. Thus we can write the trend estimate of the selected model simulation as: $t^{M_{i,k}} = \mathcal{N}(\mu_{M_{i,k}}; \sigma_{M_{i,k}})$. In order to take into account the uncertainty of the single observational dataset (σ_{obs}), we also add to the calculated model trend estimate a random estimate of the observational noise by taking the observational standard deviation of the linear regression coefficient. We repeat the above described procedure 50 000 times. With that we have a large ~~amount~~ sample of model trends and can build up a robust probability density function (PDF) of the REF-C2 ozone trends. From these estimated PDFs we can then estimate the probability of a given trend relative to the models~~ in order to compare the model simulations~~. We derive a "probability of disagreement" between the observational and the modeled trend distribution by taking the central interval of the models' trend distribution with the observed trend value as threshold of this interval. To calculate this central interval we order the the 50 000 values from the REF-C2 trend distribution according to

their probability values and then sum up the ordered probability values until the value of the observed trend is reached. This probability value indicates our estimate of whether the observations agree with the models, i.e. high probability values indicate that a disagreement between models and observations is less likely due to chance.

2.4 Analysis Methods

We here provide a short description of our methodology to analyze transport processes, which follows the studies of **?** and **?**. Stratospheric mean AoA is defined as the mean residence time of an air parcel in the stratosphere (**??**). In the CCMs, the AoA tracer is implemented as an inert tracer with a mixing ratio that linearly increases over time as lower boundary condition. AoA is then calculated as the time lag between the local mixing ratio at a certain grid point and the current mixing ratio at a reference point.

The residual circulation transit time (RCTT) is the hypothetical age that air would have if it only followed the residual circulation, thus without processes such as eddy mixing or diffusion. RCTTs are calculated by backward trajectories on the basis of the Transformed Eulerian Mean (TEM) meridional and vertical velocities (referred to as residual velocities) with a standard fourth-order Runge-Kutta integration (**?**). The RCTT is then the time that these backward trajectories require to reach the tropopause from their respective starting point in the stratosphere. The RCTT differs from AoA because of resolved and unresolved mixing. In the stratosphere, this is due to the mixing of air between branches and the in-mixing of air from the mid-latitudes into the tropical pipe, which leads to recirculation of old air around the BDC branches. In global model studies, this effect has been named aging by mixing (AbM) and is interpreted as the difference between AoA and RCTT (e.g. **?**).

3 Results

3.1 Ozone trends over the period 1998–2018 in CCM simulations and observations

In this section we analyze the ozone trends of all ~~free running~~ 'free running' CCMI-1 simulations (REF-C2), including all ensemble realizations of each model~~for the post-ODS~~ , for the period 1998–2018 together with the observational data, BASIC$_{SG}$. We chose the period 1998–2018 to be consistent with the observational trend estimate in the ozone recovering phase as presented by **?**. Note that ODSs are declining in this period as a result of the Montreal Protocol and its amendments. By using the REF-C2 simulations we include a wide spectrum of SST variability in the different CCMs, as they use either an interactive ocean or prescribed SSTs from a coupled ocean-atmosphere model simulation (see Tab.1). Ozone trends are calculated by simple linear regression~~(see section 2.3)~~, using the monthly deseasonalized ozone time series. We refrain from excluding sources of variability such as QBO, ENSO (El Nino Southern Oscillation), solar cycle or volcanic eruptions in the regression analysis to capture the full range of variability of ozone trends over the given period. Hence our trend estimates have to be interpreted as resulting from both forced trends (e.g., via GHG increases and ODS decreases) as well as from natural and internal climate variability. In the following we compare the ~~so~~ calculated ozone trend from the observational data to the ~~trend estimate by using~~ trends presented in **???** that used a dynamical linear modeling (DLM) approach, ~~that does attempt~~ which attempts to take

natural sources of variability into account~~, as presented in ??~~. In a nutshell, DLM has many similarities with ordinary least squares multiple linear regression (MLR), using predictor variables to account for some of the variability in the timeseries (e.g. solar variability, the QBO). Where DLM primarily differs from MLR is in allowing for a non-linear trend to be estimated and for the seasonal cycle to evolve with time, and therefore the shape of these terms is not predefined. For more details, see **?** and **?**.

The panels of Fig. 1 show a latitude-pressure cross-section of the ~~post-ODS~~ ozone trend for observations (first panel of Fig. 1) and all free running CCMI model simulations. Generally, the linear trend fit we perform on the BASIC~~SD~~ ~~SG~~ data yields similar spatial patterns and magnitudes to those estimated in **?** with the DLM approach (see their Fig. 1f). There are a few small differences, e.g., our linear trend fit results in larger positive trends in the upper stratosphere over the southern tropics of $\sim 1\%$, a slightly less negative trend in the northern hemisphere middle stratosphere ($<1\%$), and consistently large and negative trends close to 100 hPa in the tropics as opposed to a smaller and insignificant trend at around 10°S and over 100-80 hPa in the DLM estimate as shown by **?**. Most notably, linear trend calculations result in small positive trends (up to $\sim$3%) in the southern mid-latitude lower stratosphere, as opposed to overall negative but insignificant trends reported by **?** in that region. However, the comparison reveals that the overall magnitude and trend pattern is also captured by the simple linear regression, i.e. it is not dependent on the exact method used to calculate the trends. Therefore, we proceed with using a linear fitting approach for the comparison between observations and CCMs, though the above caveats should be kept in mind when comparing with a full regression analysis using DLM (**?**).

Overall, large inter-model variability of the trends derived from the individual REF-C2 simulations (including all ensemble members) is revealed in Fig. 1. Nevertheless, a number of features can be identified that are consistent over most models and all their ensemble members. In the upper stratosphere (1-10 hPa) nearly all simulations consistently show an overall positive ~~trend in ozone~~ ozone trend. This ozone increase can be explained by the decrease of ODSs (see e.g. **?**) and by a slow down in ozone destruction rates as the stratosphere cools from GHG increases (see e.g. **?**)~~.~~, as will be further discussed in Sec. 3.5. This upper stratospheric ozone trend has been found for climate model simulations and for observational data in several studies before (e.g. **??????**). However, in the lower stratosphere (30-100/150 hPa) we find a wide spread in the ozone trends among the CCM simulations over recent decades. Many REF-C2 simulations exhibit negative trends in the tropical LS, and they are comparable to the observational trend in magnitude and structure. In agreement with earlier studies ~~(e.g. ?)~~(e.g. **??**), we will show in Section 3.4 that this tropical ozone decrease is related to enhanced tropical upwelling in a warmer climate. However, there are also simulations showing a positive LS ozone trend in the tropics (~~e.g.~~ i.e., GEOSCCM, SOCOLv3, NIWA-1, WACCM-3/4, CAM4-1~~, LMDZrepro~~/2, LMDZrepro, HadGEM; note that the number of the ensemble run is denoted with -1, -2 and so on). At northern and southern mid- and high-latitudes most ~~CCMs~~ simulations exhibit a positive trend, but with a pronounced ~~spread between model simulations~~intermodel spread. Only a few simulations show negative trends in either northern or southern mid-latitudes (e.g. GEOSCCM, WACCM-3, WACCM-4), but it is important to point out here that none of the 31 simulations reproduces the observed negative ozone trend pattern with an ozone decrease covering the tropical belt and extending to the mid-latitude (50°S-50°N), as shown in the upper-left panel and previously in **??**. This discrepancy in the LS ozone trend between observations and models has been reported before (e.g. ozone ~~profile~~ trends, based on CCMI

simulations, ~~(?)~~(??), and in comparison to CCMVal-2 simulations (**?**)). For CCMs that provide multiple ensemble members (WACCM, NIWA, ULAQ, ACCESS, CAM4 and UMUKCA), we also identify a large ensemble spread in the simulated LS ozone trends. For example in WACCM two ensemble members simulate positive tropical ozone trends, while the two other members simulate negative tropical ozone trends. In WACCM (as well as in NIWA and CAM4), the coupled ocean allows for differences in the SST variability between the ensemble members, possibly explaining the large spread in tropical ozone trends. However, as is also the case for models with prescribed SSTs (ACCESS, ULAQ, UMUKCA) ~~,~~that exhibit a large spread between the simulations, the SST variability is not the only reason for the different trend pattern, as was similarly reported and discussed by **?** for CCMVal-2 models. The large spread in LS ozone trends between ensemble members is further in agreement with the study of **?**. They used a nine member ensemble of a ~~free running CCM simulation~~ 'free running' CCM simulation (CESM1-WACCM) and showed that LS ozone trends over the years 1998–2016 are characterized by large internal variability, with e.g. the LS ozone trend ranging from +6% to -6% per decade. But note, again, that none of these ensemble members showed the coherent decrease in ozone in the tropics and extratropics as found in observations (**?**).

Following this qualitative discussion on the spread in the ozone trend pattern between the CCM simulations, we now turn to the LS ozone trends with a more quantitative comparison of the apparent inconsistencies between observations and CCMs. We calculate the trends of the deseasonalized LS ozone columns for the ~~post-ODS~~ period 1998–2018 in two regions: the inner tropics (20°N-20°S) and in the northern mid-latitudes (30°N-50°N). We choose the northern mid-latitude band 30°N-50°N ~~in order to be comparable to~~ for direct comparability with the study of **?**. The pressure range of the lower stratosphere was taken to be 30–100 hPa for the tropics and 30–150 hPa for the mid-latitudes, to take into account the differences in ~~the~~ latitudinal tropopause heights. Trends and their uncertainties (represented by the 90% confidence interval of the linear slope) are shown for each of the 31 available REF-C2 simulations of 18 different CCMs in Fig. 2. We decided to focus on the northern mid-latitudes here, because the SH mid-latitude trends are likely more strongly influenced by the large chemical depletion of ozone within the polar vortex. ~~However, we~~ We will come back to the LS ozone trends of the southern mid-latitudes ~~later on (see section 4)~~in Section 3.5.

In the tropics about half (42%) of the REF-C2 simulations show a significant decrease, about the same (42%) show a non-significant change, and about 15% a significant increase in the integrated tropical LS ozone column. Note that significance is defined as the non-overlap of the error bars (90% confidence interval) with the zero trend. The resulting MMM ozone trend (see red bar on right of Fig. 2) is negative (-0.37 DU/dec), but it is insignificant due to the considerable spread among the different models~~, with the~~ . The 25th-75th quantile of the distribution ~~ranging~~ ranges from -1.12 to 0.20 DU/dec (see edges of box on the right of Fig. 2). Note that for the ~~MMM trend~~ calculation of the MMM trend, we choose to weight each of the 31 simulations ~~is weighted equally ,~~ equally (i.e., not taking into account that some models have multiple ensemble members~~,~~) because the trend ~~variation~~ variations among ensemble members are as large as among the different models ~~.~~over this period.

The observed tropical LS ozone trend of -1.07 DU/dec is statistically significant at the 90% level. Thus the observed tropical trend is more strongly negative than the MMM trend, but lies within the 90% confidence interval of the MMM trend (~~being~~ [-1.76 DU/dec; 1.03 DU/dec]).

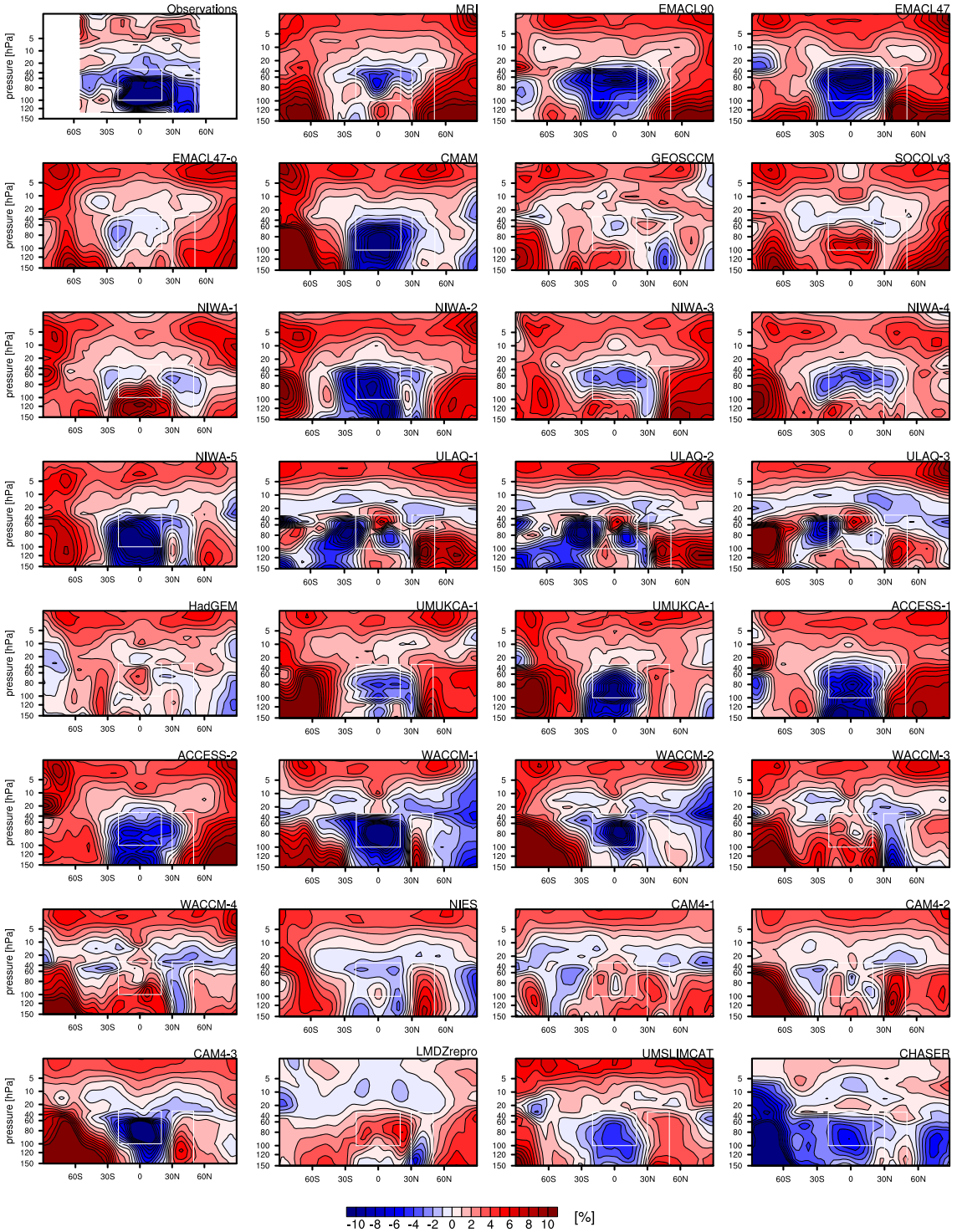

Figure 1. Latitude-pressure cross section of the ozone trend over the ~~post-ODS~~ period 1998–2018 for the observational data set BASIC$_{SG}$ and for all CCMI REF-C2 simulations. Trends are given as relative ozone changes over the whole time period. Boxes illustrate the regions selected to integrate ozone in the LS for trend comparisons later in this study, i.e. in the tropics (20°N-20°S, 30–100hPa) and in the northern mid-latitudes (30°N-50°N, 30–150 hPa).

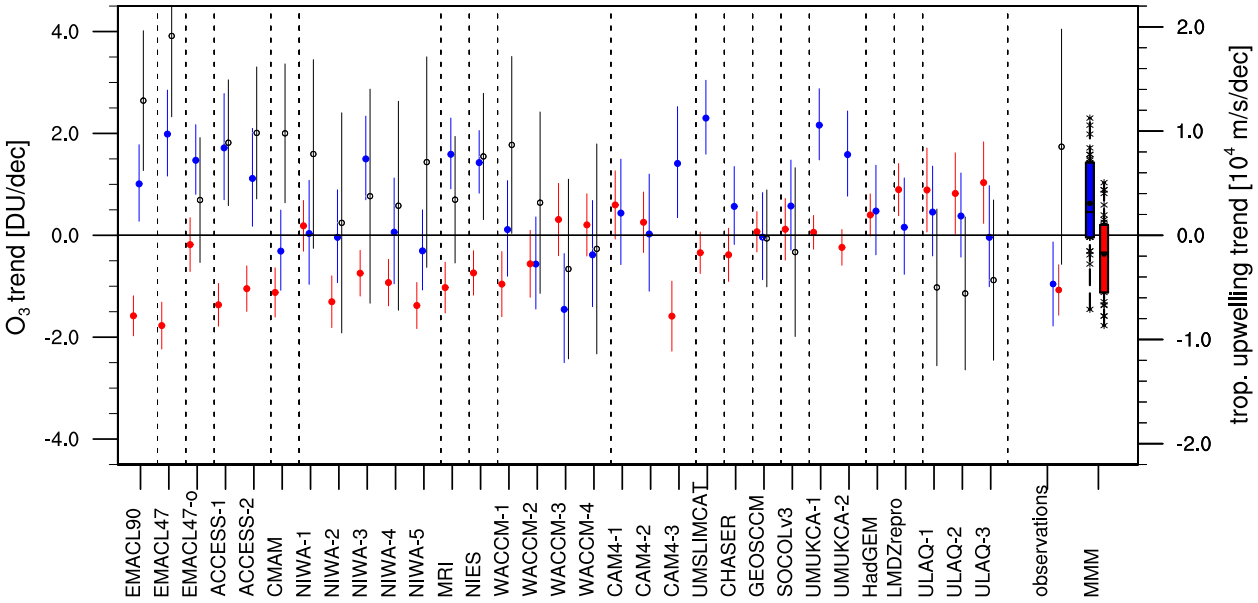

Figure 2. LS ozone trends and their uncertainties in the tropics (20°N–20°S, red dots) and northern mid-latitudes (30°N-50°N, blue dots) together with tropical upwelling trend (black circles, for all simulations providing TEM diagnostics) for the period 1998–2018 for all REF-C2 simulations. Dashed lines separate the individual models. Moreover, observational trends (1998–2018) and multi-model mean trends are given. Observational data for ozone are taken from BASIC$_{SG}$, and for tropical upwelling from ERA5 reanalysis. Error bars associated with each LS ozone trend represent the 90% confidence intervals. The multi-model mean trends are shown as boxplots: the black solid line in the box indicates the median, the black point the MMM and the colored box ranges from the 25th to the 75th percentile of the trends. Crosses denote trends of individual model simulations not lying within the box.

In the northern mid-latitudes less than half (40%) of the REF-C2 simulations show an increase in the LS ozone column, while the remaining 60% of the simulations show a non-significant change (either positive or negative). There is only one simulation (WACCM-3) that shows a significant decrease in the mid-latitude LS ozone column, ~~although it has a corresponding (non-significant) tropical trend~~and in this simulation the tropical ozone trend is positive (but not significant). The resulting MMM trend in the northern mid-latitudes is positive (+0.63 DU/dec) with a high inter-model spread: the 25th to 75th quantile of the distribution ranges from -0.04 to 1.42 DU/dec. Note here, that the observational trend (-0.96 DU/dec) lies ~~just~~ outside the 90% confidence interval of the MMM trend in the mid-latitudes ([-0.91 DU/dec; 2.16 DU/dec]).

Fig. 2 also reveals that over the years ~~1998–2018~~ 1998–2018 more than half of the model simulations have a dipole trend pattern in the LS ozone column, i.e. the sign of the tropical ozone trend is opposite to that in mid-latitudes. ~~A dipole~~ This trend

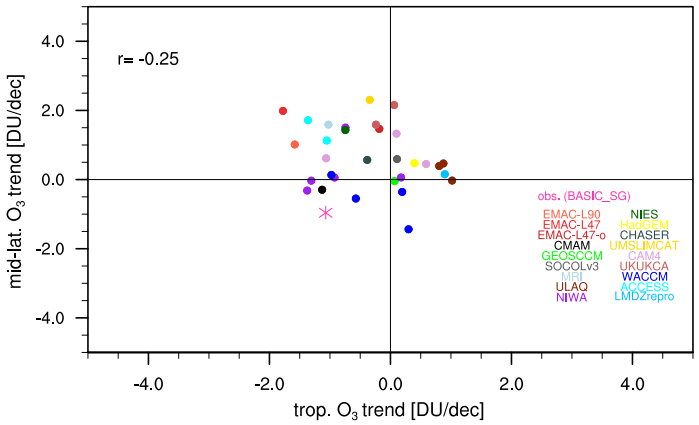

Figure 3. Inter-model correlation between tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trends, calculated over the period 1998–2018 for 31 CCMI REF-C2 simulations. All ensemble members of a particular model are shown in the same color. The observational ozone trends (BASIC$_{SG}$) are also included in here as a star.

.

pattern with negative LS ozone ~~trend~~ trends in the tropics and positive LS ozone ~~trend~~ trends in the northern mid-latitudes can be found for almost half the simulations (45%), and a ~~dipole~~ trend pattern with a positive ozone trend in the tropics and negative trend in the northern mid-latitudes is found in 13% of the simulations. The remaining simulations do not show this dipole, but either both have a positive trend in the tropics and the mid-latitudes (29%), or a negative trend in both tropics and mid-latitudes (13%, i.e. 3 simulations, namely NIWA-5, CMAM, WACCM-2). Only 3 out of 31 simulations simulate negative, but not significant, trends both in the tropics and northern extratropics, and thus they show a similar behavior to observations (see right of Fig. 2 and **?**). However, ~~the~~ their zonal trend patterns (see Fig. 1) reveal, that none of these three simulations reproduces the observed trend pattern with consistent negative trends from 50°S-50°N in the LS. Consequently it is important to keep in mind that the results of these (averaged) trends ~~are depending~~ depend on the choice of the latitude-pressure box, as the integration over a wider latitude band can lead to a cancellation of opposing trends. ~~To~~

Next, we analyze whether a systematic relationship between the LS tropical and mid-latitude ~~trend can be diagnosed from~~ trends exists in the CCM simulations. For this, the simulated northern mid-latitude LS ozone trends are plotted against the simulated tropical LS ozone trends over the time period ~~1998-2018~~ 1998–2018 for all 31 REF-C2 simulations and for the observed data-set BASIC$_{SG}$ in Fig. 3. As discussed above, in the LS the majority (45%) of the models have a negative ozone trend in the tropics and a positive trend in the northern mid-latitudes. ~~This~~ Moreover this illustration again highlights that the trends estimated from observational data are lying on the outer edge of the ~~distribution of the modeled trends~~ model trends distribution. The inter-model correlation between the tropical to mid-latitude trends is negative with a low ~~not significant~~ correlation coefficient (-0.25). Thus, for the chosen period the tropical ozone trends are ~~significantly, but~~ only weakly linked to mid-latitude

ozone trends in the models. ~~We expect a negative correlation here, because~~ However, we expected that the two trends are highly (negatively) correlated, as from our understanding increased tropical upwelling ~~, leading~~ leads to decreased tropical ozone, and this upwelling increase should be linked to an increased mid-latitude downwelling, ~~that~~ which would enhance ozone in the mid-latitudes. ~~The relationship of the trends to tropical upwelling will be further investigated in Sec.~~However Fig. 3 does not support this. Also slightly varying the period (i.e. looking at the periods 1999-2019, 2000-2020, 2001-2021) reveals very low negative or near zero correlations (not shown here). To get a better understanding of the processes leading to the given LS ozone trend patterns we will investigate the relationship of LS ozone trends to stratospheric transport trends in Section 3.4.

Overall we can conclude from the analysis of ozone trends in the suite of CCMI models (see Fig. 1, 2 and 3), that the LS ozone trends exhibit a considerably large spread across both the different models, but also across ~~individual ensemble members~~ ensemble members from a single model, in particular in the mid-latitudes. This indicates that ozone variability considerably influences the LS trends, in agreement with the recent studies by **?** and **?**. However, even when considering the high variability of possible trends in CCM simulations, the observational trends ~~emerges as a rather~~ emerge as an unlikely realization of the simulations over the period 1998–2018. In the next section, we will analyze the robustness of this finding by varying the period of the trend calculation, and providing an in-depth statistical analysis of the likelihood of the observed trend lying within the suite of modeled trends.

3.2 Robustness of lower stratospheric ozone trends

In the previous section we found that the observed LS negative ozone trend in the mid-latitudes together with a simultaneous negative trend in the tropics is unlikely ~~given~~ based upon the suite of CCM simulations. To further establish the robustness of this result, we here test whether this also holds for time periods that are slightly different to the ~~post-ODS~~ period 1998–2018 we considered before. Thus, in this section we first want to investigate how variability influences the ozone trends, and second we want to quantify the likelihood of the observed trend being a realization of the distribution of the modeled trends. To answer those questions, we calculate the LS ozone trends by varying the start and end years of the time period. In Fig. 4 (a) and (b), the observed tropical and mid-latitude ozone trend in the LS is shown for start years varying from 1995–2001 (y-axes) and end years from 2013–2019 (x-axes). Both tropical as well as mid-latitudes LS ozone trends are consistently negative for all chosen periods in the observations (top ~~raw~~row). This is in line with the results of **?**, who found that the observed negative sign of the tropical and mid-latitude trends remain insensitive to changing the end year. In the tropics, observational LS ozone trends are consistently negative with values between -0.64 and -1.24 DU/dec for all possible start end year combinations. In the mid-latitudes the trends are also negative for all shown time periods, but are more variable than in the tropics (values range between -0.11 and -1.22 DU/dec). In particular at mid-latitudes, the strongest negative trends are found for start years of 1996 to 1998, and a sudden decrease in the trend magnitude is found for the start year 1999 and 2000. Thus, the analysis in **??** and in the preceding section focused on a period with ~~particular~~ particularly strong negative mid-latitude ozone trends. Possible reasons for the sudden change in the trend, ~~like~~ such as the strong ENSO event in 1998, are discussed in ~~Sec.~~Section 4. Note that the trend magnitude ~~goes back up~~ increases again for the start year 2001, ~~so it again looks like~~ which again suggests that interannual variability influences the observational mid-latitude trends.

Fig. 4 (c) and (d) display the tropical and mid-latitude trends as a function of start and end year derived from the model simulations. To do so, a robust estimate of the trend probability distribution considering all model simulations was derived (see Section 2.3) and from this distribution the most likely trend is shown (see peak in the models' trend probability distributions of Fig. S1 and S2 in the supplement). In the tropics the ozone trends derived from the REF-C2 simulations are negative and ~~,~~ range from -0.74 to +0.02 DU/dec. In the mid-latitudes the trends are positive for all possible start/end year combinations, with values ranging from +0.4 to +1.48 DU/dec. In contrast to the sudden change in the mid-latitude observational trend for start years 1999 and 2000, in the REF-C2 simulations no such systematic change can be found. The estimated probability distributions of the trends from the REF-C2 simulations (see Figs. S1 and S2 in the supplement) are typically symmetric around their maximum value and show a single, central peak. The width of the distribution changes ~~for varying~~ when varying the start/end year combination, with narrower distributions for longer time periods. Moreover, ~~the~~ visual inspection of the distribution implies that the tropics (Fig. S1) ~~have generally~~ generally have Gaussian-like distributions, whereas the mid-latitudes (Fig. S2) often show a more peaked structure, i.e. with heavier ~~trails~~tails. Nevertheless, as an estimate of the width of the models trend distribution, we show in Fig. 4 (e) and (f) the standard deviation of the models distribution (in DU/dec) in the tropics and mid-latitudes, respectively. For longer time periods (values in lower right corner) the standard deviation of the models' trend is smaller, i.e. the distribution is narrower. This indicates that the influence of natural variability is less important for longer time periods, as should be expected.

Given the distributions representing the combined trends of the models, we can now quantify the disagreement between the observational trend estimate and the models' trend probability distributions for each start/end year combination. In Figs. 4 (g) and (h) the "probability of the disagreement" between observational and modeled LS ozone trends is given for the tropics and the mid-latitudes. The value of the "probability of disagreement" is calculated by the central interval of the models' probability distribution when taking the observed trend value as threshold of this interval. Thus, a probability value of 90% indicates that the observed trend falls within the inner 90% of the distribution, i.e., only 10% of the distribution is more extreme than the observed trend: the smaller the given "probability of disagreement" value, the higher is the probability that the observed trend lies within the models' distribution. In the tropics, the observed LS ozone trend falls within the 13% to 73% interval of the modeled probability distribution, i.e. the observed trends are generally likely representations of the models' trends. The agreement is best for short time periods (values in diagonal in Fig. 4 (g)), mostly because of the broader distribution (see Fig. 4 (e) and Fig. S1). Also for early start years (in particular 1995) and end years ranging from 2013 to 2018, the disagreement is small, because model trends are strongly negative for this period (see Fig. 4 (c)). In the mid-latitudes, the observed trend generally lies at more distant parts of the models' trends distribution (73% to 96%), i.e. the observed trend is a more extreme value in the models' distribution. The disagreement is smallest for both the earlier periods (lower left, start years 1995–1997 and end years 2013–2015) and the later periods (upper ~~left~~right, start years 1999–2001 and end years 2017–2019). This coincides with the generally smaller negative trends in those periods in observations (see Fig. 4 (b)) and rather constant trend distributions in the models (see Fig. 4 (d)). For the periods with the strongest negative observed trend (start years 1996–1998), the observed trend lies within the central 90% or higher of the models' distribution, i.e. is an unlikely representation from the modeled

trends. The sudden decrease in the observed trend magnitude for start years 1999 (Fig. 4 (b)) is reflected by a decrease of the central interval to about 75%. In general, one might have expected that longer periods lead to better agreement of the observed and modeled trend due to the smaller influence of variability (see Figs 4 (e) and (f))- as we do in the models- however, we do not find this to be true for either the tropics or the mid-latitudes.

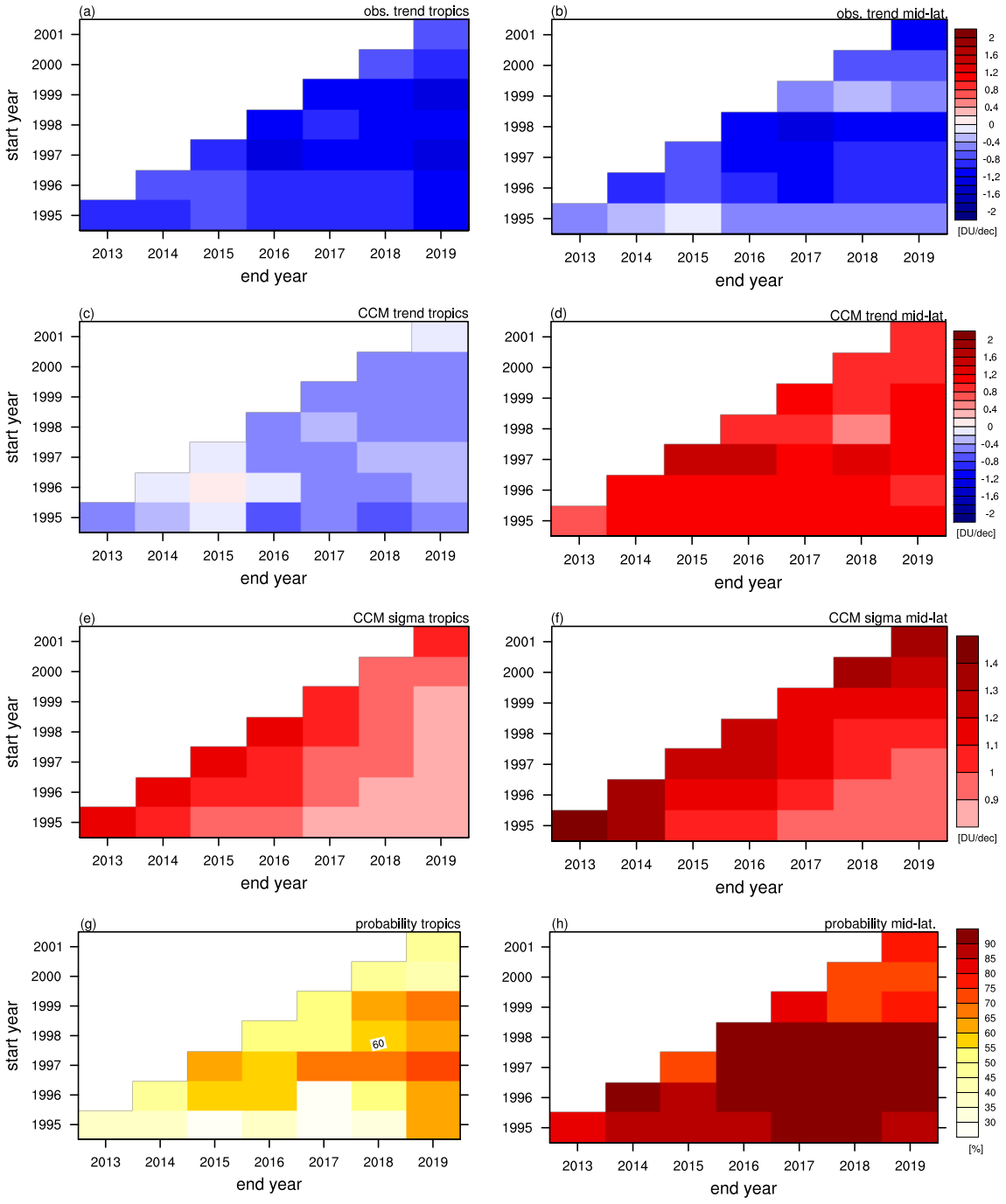

Figure 4. Tropical (right) and mid-latitude (left) LS ozone trends (in DU/dec) as function of different periods for the observational trend of BASIC~~SD~~ SG ((a)+(b)), the most likely trend of the modeled REF-C2 probability distribution ((c)+(d)) and the 1-σ standard deviation (in DU/dec) of the mean obtained from the probability distribution ((e)+(f)). The panels (g)+(h) show the "probability of disagreement" (in %) between observed trends and the REF-C2 trend probability distribution. In all panels the x-coordinate denotes the different end years (2013-2019) and the y-coordinate the different start years (1995-2001).

3.3 Convergence of future lower stratospheric ozone trends

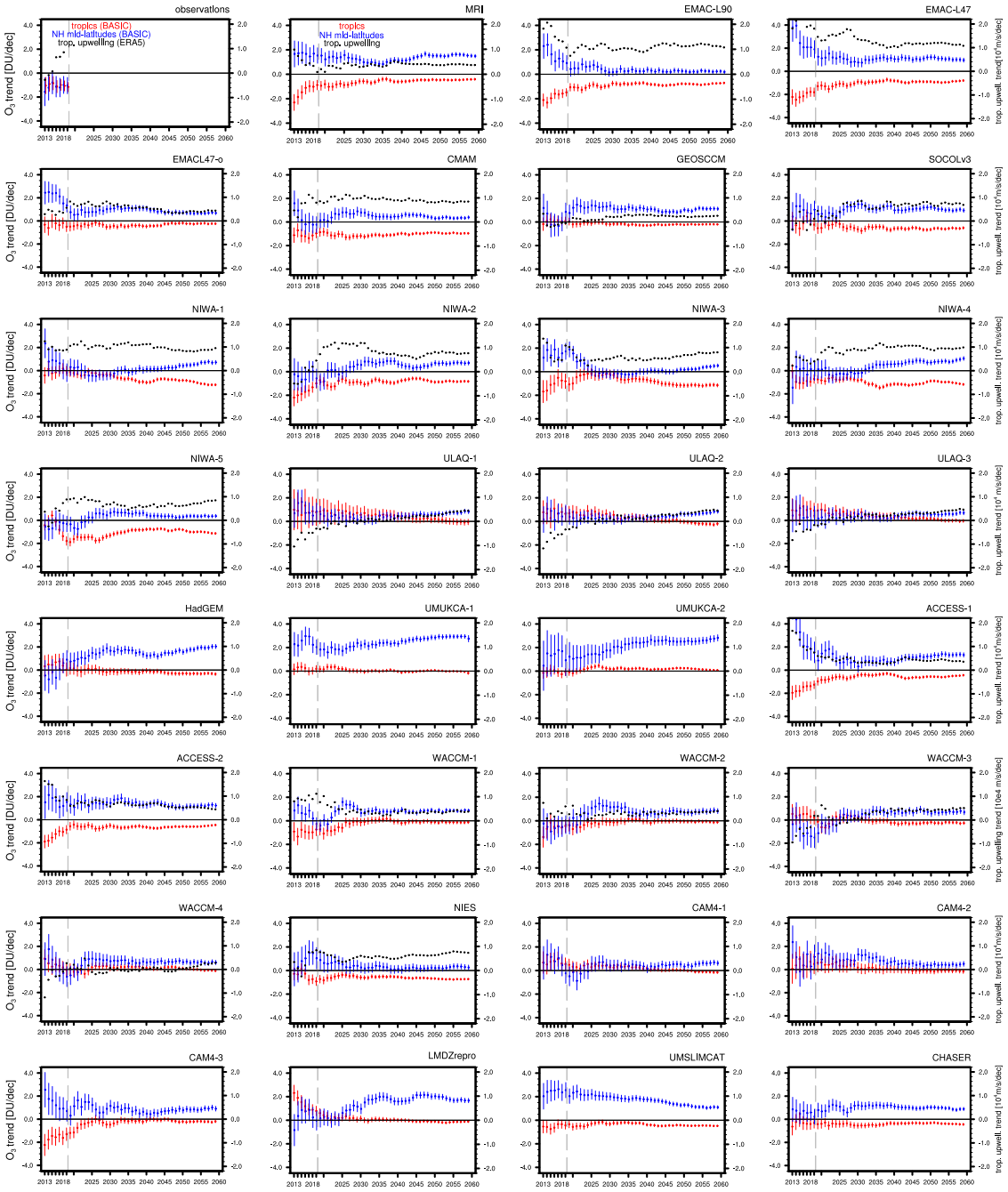

Figure 5. Tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trend and their uncertainties (in DU/dec) of observations (BASIC$_{SG}$) and ~~of~~ REF-C2 ~~model~~ simulations as a function of the end year (red and blue dots, respectively). Tropical upwelling ~~trend is~~ trends are included for all REF-C2 simulations, where TEM diagnostics was available (black dots); observational tropical upwelling is taken from ERA5 reanalysis. The end year varies from 2013 to 2019 for observational data and from 2013 to 2060 for REF-C2 simulations. Error bars associated with each trend represent the 90% confidence intervals.

~~In the last~~ In the previous section, the ~~robustness of ozone trend~~ ozone trend robustness was analyzed for time periods of up to 25 years. We will show in the following that, as the considered time periods ~~become longer.~~ are extended, the influence of natural variability decreases ~~more and more, and~~ and the trends converge to the trend forced by long-term ~~trends in~~ GHG and ODS ~~concentrations~~concentration changes. To analyze the timing and the values of the trends' convergence, we extend the period for the trend calculation into the future for all REF-C2 simulations.

Fig. 5 shows the tropical and northern mid-latitude LS ozone trends together with the tropical upwelling trend (black; if available) for periods with the fixed start year 1998, and the end year varying from 2013 up to 2060, by extending the time period by steps of one year. For reference, the observational trends of ozone (from BASIC$_{SG}$) and tropical upwelling (from ERA5) are shown in the upper left panel of Fig. 5, with the last available end point in the year 2019. As shown in the last section, the trends derived from observational data are consistently negative both in the tropics and in the northern mid-latitudes.

As discussed in Sec. 3.1, the ozone trends exhibit a strong inter-model spread for the observational time periods. Both tropical and mid-latitude ozone trends in the individual model simulations vary considerably for different end point years within the observational period (left of the vertical dashed gray lines)~~, with the~~ . The northern mid-latitude trend is generally more variable than the tropical trend. For longer time periods extending into the future, ~~for all simulations~~ the uncertainties in the LS ozone trends decline and ~~trends converge~~the trends converge in all simulations. All model simulations consistently ~~evolve a dipole trend pattern, with significant and persistent negative~~ simulate persistent negative or near-zero trends in the tropics and ~~persistent positive~~ positive or near-zero trends in the northern mid-latitudes~~(for a summary, see also Fig. 8). However~~. However, the timing of convergence of the trends to ~~the dipole~~ this trend pattern is rather different in the simulations, as can be inferred from Fig. 5, i.e. the convergence appears to be model dependent. For some models, the trends vary little for end years after 2020 (e.g. MRI in Fig. 5), while in other models, the trends still vary considerably until end years around 2030 to 2040 (e.g. the four WACCM ensemble members in Fig. 5). The timing of the convergence ~~basically~~ is controlled by the ratio of the year-to-year variability to the strength of the forced trends. The relative forcing by ODS versus GHG changes over time~~for the given periods~~, and thereby the forced ozone trends vary over the time periods as well, making it difficult to quantify an exact date of convergence. Still, the trend estimates for the entire period 1998 to 2060 do converge to stable values for almost all models, thus representing the forced trend for this time period. ~~For all model simulations, the long-term LS ozone trends over the entire period (1998–2060) are negative in the tropics and positive in the extratropics, but the trend magnitudes vary~~ The trend magnitudes over this long period vary strongly between the models~~. Thus model trends vary~~ , from -0.10 to -1.32 DU/dec in the tropics ~~,~~and from +0.39 to +2.00 DU/dec in the mid-latitudes. Comparing this to the model range of the ~~somewhat~~ shorter time period 1998–2040, we see that the tropical trend (+0.06 to -1.12 DU/dec) has not converged to the end point values of 2060, yet. The mid-latitude trend (+0.54 to +2.15 DU/dec) is however close to the 2060 values.

Overall, the mid-latitude trends converge to positive values in the majority of the model simulations (about 85%) by 2030. Thus, if both the year-to-year variability and the forced response of the models is simulated realistically, we should expect the emergence of positive mid-latitude trends from observational records within the next decade.

3.4 Influence of ~~tropical upwelling~~ transport processes on LS ozone trends ~~in CCMs~~

Reply to Anonymous Referee #2 (ACP-2020-947)

We thank referee #2 for the positive and constructive comments on our manuscript. Due to the suggestions of referee #1 we include some additional analysis to this paper. With that we gained additional insight to the possible processes determining the ozone trends in the LS. Due to the new results we also reorganized the structure of the manuscript in the last sections (see new section 3.4 and 3.5). Moreover, note that we do not consider interannual correlations any more, as we felt that not too much was learned here.

Below this reply, you find the revised manuscript considering all questions and comments. Additionally, we highlighted the changes of the manuscript for the pages 1-19, and attached them to the reply. From Section 3.4 on, the changes were rather comprehensive (e.g. replacement of a whole section), that it doesn't make sense to highlight them.

Minor issues:

1) There are a lot of minor editorial issues, especially missing hyphens (e.g. 'free running') and missing commas. I assume that the editorial office will sort these out if the authors don't find them.
-> We re-read the manuscript carefully and found some of these minor issues. If there are remaining ones, they will surely be corrected during the typesetting process.

2) Page 1. Line 10-11. It is not clear what the 'systematic change' relates to. The sentence mentions 'different analysis periods', but for the change to be systematic the period would have to be changing in a particular direction?
-> Rephrased.

3) Page 2. Line 4. Need to say '1987 Montreal Protocol and later Adjustments/Amendments'. Also, different ODSs started to decline at different times. Some (e.g. HCFCs) might still be growing. You mean the total halogen loading from ODSs.
-> Done.

4) Page 2. Line 8. There were still ODSs in atmosphere in 1980. Need to say 'pre-1980' or similar.
-> Done.

5) Page 3. Line 12. Put references in chronological order.
-> Done.

6) Page 4. Line 14. Ball et al (2018) also included some SD runs which showed very poor agreement with the observed trends. That could be mentioned here.
-> Done.

7) Page 5. Footnote. 'fourth'.
-> Done.

8) Page 8. Line 22. Please state which model was used by Stone et al. Is that one of the CCMI models?
-> Done.

9) Page 9. Caption (and elsewhere). Post 1998 is not the 'post ODS' period. ODSs are still present and different ones have different trends. Total chlorine and bromine are declining, which is not the same thing. You should find another description.

-> Thank you for this helpful comment, we now do not use the expression post-ODS any more, which was indeed a min-nomer.

10) Page 12. Line 3. ‘depend’.
-> Done.

11) Page 12. Line 11. ‘low not significant’. This reads strangely. Maybe it is a lack of a comma, but could also be better to say ‘non-significant’.
-> Done.

Analysis of recent lower stratospheric ozone trends in chemistry climate models

Simone Dietmüller[1], Hella Garny[1,2], Roland Eichinger[2,1,3], and William T. Ball[4,5,6]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Ludwig Maximilians Universität, Faculty of Physics, Institute for Meteorology, Munich, Germany
[3]Charles University, Department of Atmospheric Physics, Faculty of Mathematics and Physics, Prague, Czech Republic
[4]Institute for Atmospheric and Climate Science, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland
[5]Physikalisch-Meteorologisches Observatorium Davos World Radiation Centre, Dorfstrasse 33, 7260 Davos Dorf, Switzerland
[6]Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, TU Delft, Stevinweg 1, 2628 CN Delft, the Netherlands

Correspondence to: S. Dietmüller (simone.dietmueller@dlr.de)

Abstract. Recent observations show a significant decrease of lower stratospheric (LS) ozone concentrations in tropical and mid-latitude regions since 1998. By analyzing 31 chemistry climate model (CCM) simulations performed for the Chemistry Climate Model Initiative (CCMI, Morgenstern et al. (2017)), we find a large spread in the 1998–2018 trend patterns between different CCMs and between different realizations performed with the same CCM. The latter, in particular, indicates that natural variability strongly influences LS ozone trends. However none of the model simulations reproduces the observed ozone trend structure of coherent negative trends in the LS. In contrast to the observations, most models show a LS trend pattern with negative trends in the tropics (20°S-20°N) and positive trends in the northern mid-latitudes (30°-50°N) or vice versa. To investigate the influence of natural variability on historical LS ozone trends we analyze the sensitivity of observational trends and the models' trend probability distributions for varying periods with start dates between 1995 to 2001 and end dates between 2013 to 2019. Generally, modeled and observed LS trends remain robust for these different periods, however observational data show a change towards weaker mid-latitude trends for certain periods, likely forced by natural variability. Moreover we show that in the tropics the observed trends agree well with the models' trend distribution, whereas in the mid-latitudes the observational trend is typically an extreme value of the models' distribution. We further investigate the LS ozone trends for extended periods reaching into the future and find that all models develop a positive ozone trend at mid-latitudes and the trends converge to constant values by the period that spans 1998–2060. Intermodel correlations between ozone trends and measures of transport circulation trends confirm the dominant role of greenhouse gas (GHG)-driven tropical upwelling enhancement on the tropical LS ozone decrease. Mid-latitude ozone, on the other hand, appears to be influenced by multiple competing factors: an enhancement in the shallow branch decreases ozone, while an enhancement in the deep branch increases ozone and, furthermore, mixing plays a role here too. Sensitivity simulations with fixed forcing of GHGs or ozone depleting substances (ODSs) reveal that the GHG-driven increase in circulation strength do not lead to a net trend in LS mid-latitude column ozone. Rather, the positive ozone trends simulated consistently in the models in this region emerge from the decline of ODSs, i.e. the ozone recovery. Therefore, we hypothesize that next to the influence of natural variability, the disagreement of

modeled and observed LS mid-latitude ozone trends could indicate a mismatch in the relative role of the response of ozone to ODS versus GHG-forcing in the models.

1 Introduction

Stratospheric ozone is essential for protecting the Earth's surface from ultra violet radiation, which is harmful for plants, animals and humans. Human-made ozone depleting substance (ODS) emissions significantly reduced ozone concentrations for some decades after 1960. After controlling the use of ODSs by the 1987 Montreal protocol and later adjustments, however, ODS concentrations started to decline in the mid-to-late 1990s (e.g. Newman et al., 2007; Chipperfield et al., 2017). As a consequence, total stratospheric ozone is expected to recover in the future. Dhomse et al. (2018) have analyzed the recovery of stratospheric ozone mixing ratios of the CCMI-1 (Chemistry Climate Model Intercomparison project part 1) climate projection simulations. They found that the ozone layer is simulated to return to a pre-1980 ODS level between 2030 and 2060, depending on the region. However, they discovered a large spread among the individual models, which shows that there are many uncertainties in these projections. The evolution of stratospheric ozone in the 21st century does not only result from a decrease in ODS concentrations but also from an interplay between changes in both the atmospheric composition and the circulation (World Meteorological Organization (WMO) 2014). Increasing anthropogenic greenhouse gas (GHG) emissions (CO_2, CH_4, N_2O) leads to enhanced tropical upwelling and thereby to an acceleration of tracer transport along the stratospheric overturning circulation (e.g. Butchart, 2014; Eichinger et al., 2019). On the other hand, increasing GHGs also slows down ozone depletion through GHG-induced stratospheric cooling (e.g. Jonsson et al., 2004; Oman et al., 2010; Bekki et al., 2013; Dietmüller et al., 2014; Marsh et al., 2016) and emissions of CH_4 and N_2O additionally impact ozone through chemical processes (e.g. Ravishankara et al., 2009; Kirner et al., 2015; Revell et al., 2012; Winterstein et al., 2019).

In the recent years, a number of studies have analyzed observational records to identify ozone trends in the stratosphere (e.g. Harris et al., 2015; Steinbrecht et al., 2017; Weber et al., 2018). These studies consistently report an ozone recovery in the upper stratosphere after the turnaround of the ODS concentrations around the year 1998. In the lower stratosphere (LS), however, most observed ozone trends are not statistically significant for such a relatively short period due to large internal variability and instrumental difficulties (e.g. Steinbrecht et al., 2017). Subsequently, Ball et al. (2018) analyzed LS ozone trends from satellite data between 1998 and 2016 in detail making use of a dynamical (multiple) linear regression analysis. They identified a statistically significant decline of LS ozone between 60°S and 60°N in that period, of approximately 2 DU in the LS below 24 km of altitude. The implication was that the stratospheric ozone column was continuing to decline, because the LS ozone reduction more than offsets the positive trend in the upper stratosphere. Shortly afterwards Wargan et al. (2018) studied ozone trends in the reanalysis products MERRA-2 and GEOS-RPIT. In the tropics they detected a positive ozone trend in a 5 km layer above the tropopause and a negative trend at 7-15 km above the tropopause. Nevertheless, in the northern and southern mid-latitude LS they detected a negative ozone trend. As such, there are some similarities to the findings of Ball et al. (2018), but there are also quantitative differences, for example the positive trend in the 5 km layer or a missing overall statistically significant decrease in the column integrated ozone. Wargan et al. (2018) suggested that the negative mid-latitude trend might

be explained by enhanced isentropic transport between the tropical and mid-latitude LS. However, the recent study of Orbe et al. (2020) explicitly demonstrated that in the NH this mid-latitude ozone decrease is primarily associated with large scale advection. Furthermore, they showed that the observed changes in advection and in ozone are well within the range of model variability (gauged from one CCM). By means of using a chemistry transport model (CTM) and extending the analysis period to the year 2017, Chipperfield et al. (2018) suggested that the negative LS ozone trends are only a result of large natural variability. They showed that there was a strong positive ozone anomaly in 2017 which is driven by short term dynamical transport of ozone, and concluded that this points to large year-to-year variability rather than to an ongoing downward trend. However, an update of the data set which was used in Ball et al. (2018) showed that the large interannual variability alone cannot explain the entire trend in Chipperfield et al. (2018) (see Ball et al., 2019): the larger year-to-year variability in the SH was implicated to result from a non-linear interaction between the quasi-biennal-oscillation (QBO) and seasonal variability and despite this large variability the observed negative LS ozone trend remains.

To improve confidence in future projections of the ozone layer it is important to evaluate the skill of chemistry climate models (CCMs) in simulating the observed ozone trends over recent decades. A direct comparison between the CCM multi-model-mean (MMM) values and observational data showed that the ozone trend profiles of modeled MMM data agree well with observations, except in the lowermost mid-latitude stratosphere (SPARC CCMVal, 2010; WMO, 2018). The most recent study of Ball et al. (2020) investigated LS ozone trends of the 1998–2016 period in merged satellite data and compared them to the ozone trends in CCMs using the climate projection simulations of the CCMVal2 project. Similar to the observations, the CCMs showed a decline in LS ozone in the tropics, likely due to enhanced tropical upwelling, following from an increase in greenhouse gases (see e.g. Randel et al., 2008). In contrast to the observations, however, models do not show a decrease, but rather an increase in LS mid-latitude ozone. Ball et al. (2020) argue that these discrepancies in the LS between models and observations can possibly be explained by differences in the horizontal two-way mixing between the tropics and mid-latitudes, though they did not provide explicit evidence from the models (see also Wargan et al., 2018). The study suggested that the negative mid-latitude observational trend is caused by an intensification of two-way mixing (by analyzing effective diffusivity in reanalysis data). On the other hand enhanced downwelling of ozone-rich air to the mid-latitudes could consequently lead to a positive trend in the mid-latitudes. Apparently, the processes that determine mid-latitude LS ozone in models and observations are not understood so far.

In the present study, we seek to quantify whether the observed LS ozone trends lie within the suite of modeled trends. If yes, this would imply that the observed trend is just one realization of possible trends given within the large year-to-year variability. If not, this would imply that either models do not represent year-to-year variability correctly, or that there is a forced trend in the real world that is not adequately represented in the models. In contrast to the study of Ball et al. (2020) we are using the simulation data of a more recent inter-model comparison project (namely the Chemistry Climate Model Initiative, phase 1, CCMI-1) and analyze the ozone trends for a wider range of updated current state-of-the-art CCMs including all their ensemble simulations.

A brief description of the model simulations, of the observational data sets and of the methods used is presented in Section 2. In Section 3 we show our results. We provide a detailed comparison of ozone trends over the years 1998–2018 in different

CCM simulations and observations (Section 3.1). Here we focus on LS ozone trends, and we investigate how natural variability influences these LS ozone trends (Section 3.2 and 3.3). We link LS ozone trends with stratospheric transport trends (Section 3.4) and we investigate how ozone trends are forced by GHG and ODS emissions (Section 3.5). A discussion of the reasons for the disagreement in the LS mid-latitude ozone trends between models and observations and the conclusions follow in Sections 4 and 5, respectively.

2 Data and Methods

2.1 Models and Simulations

In the present study, we analyze the model output from 18 state-of-the-art CCMs from the Chemistry Climate Model Initiative phase 1 (CCMI-1, Morgenstern et al. (2017)). Tab. 1 lists all these CCMs together with their references, the forcing that underlies the sea surface temperatures (SSTs), and the simulation type considered. A detailed overview of all models that participated in CCMI-1 can be found in Morgenstern et al. (2017). We mainly evaluate the long term 'free running' simulations of CCMI-1 (REF-C2), as they span the time period 1998–2018. We do not use REF-C1 'free running' simulations of the recent past or the specified dynamics simulations (REF-C1SD), as they only span the period from 1998 to 2010. Moreover we want to point out that the specified dynamics simulations performed for CCMI do not represent stratospheric circulation better than the 'free running' simulations: Chrysanthou et al. (2019) compared stratospheric residual circulation among specified dynamic (SD) simulations and found that the spread in these simulations is even larger than in REF-C2. Furthermore Ball et al. (2018) showed poor agreement with the observed ozone trend for some selected SD simulations of CCMI. For the REF-C2 model simulations used in our study, all available ensemble members of the individual models are taken into account. The ensemble size of a certain simulation (if ensemble simulations were performed) is also given in Tab. 1 (brackets after simulations). Thus for the REF-C2 simulations, 18 models performed a total of 31 realizations (six models performed multiple ensemble members simulations). The REF-C2 simulations include hindcast and forecast periods spanning 1960–2100. They are all 'free running' simulations, thus each model simulation has its own internal variability. Note that REF-C2 simulations use a variety of different SSTs and SICs (sea ice concentrations), either prescribed climate model SST fields from offline model simulations (of the same or of a different model) or they are coupled to an interactive ocean and sea ice module. Moreover the representation of the QBO is different across the CCMs, with models having an internally generated QBO (e.g. MRI, EMAC-L90), nudged QBO (e.g. NIES, WACCM, SOCOLv3, EMAC-L47, EMAC-L47-o) or no QBO (e.g. CMAM, LMDZ). REF-C2 reference simulations follow the WMO (2011) A1 scenario for ODSs and the RCP 6.0 scenario (Meinshausen et al., 2011) for other greenhouse gases, tropospheric ozone precursors, and aerosol and aerosol precursor emissions. For anthropogenic emissions, the CCMI recommendation was to use MACCity (Granier et al., 2011) until 2000, followed by RCP 6.0 emissions. Besides the REF-C2 simulations we also consider the 11 sensitivity simulations with fixed greenhouse gases (fGHG) and with fixed ODSs (fODS) in our analysis. These sensitivity scenarios are both based on the REF-C2 simulation. However in case of the fGHG simulations CO_2, CH_4, N_2O, and other non-ozone depleting GHGs are held at their 1960 value, and so we are able to study the impact due to ODS concentration changes only (i.e. in the absence of GHG-induced climate change). In the case of the fODS simulations

the ODS concentrations are fixed to the 1960 level throughout the simulation. All models providing both of these sensitivity simulations are given in Tab 1.

Table 1. Overview of the CCMI simulations, analyzed for the present study. For the individual CCMs their reference(s), their SSTs and their available simulations (REF-C2, fGHG, fODS) are given. The numbers in brackets behind the simulations indicate the number of realizations of each REF-C2, fGHG or fODS simulation. Detailed information about the models' SSTs and the models' representation of the QBO are given in the supplement of Morgenstern et al. (2017).

CCMI Model	Reference(s)	SSTs	Analyzed Simulation
CMAM	Jonsson et al. (2004)	prescribed	REF-C2(1), fGHG(1), fODS(1)
	Scinocca et al. (2008)		
CESM1-WACCM	Solomon et al. (2015); Garcia et al. (2017)	interactive	REF-C2(4)*, fGHG(3), fODS(3)
	Marsh et al. (2013)		
EMAC-L90	Jöckel et al. (2010, 2016)	prescribed	REF-C2(1)
EMAC-L47	Jöckel et al. (2010, 2016)	prescribed	REF-C2(1)
EMAC-L47-o	Jöckel et al. (2010, 2016)	interactive	REF-C2(1)**
GEOSCCM	Molod et al. (2012, 2015)	prescribed	REF-C2(1)
	Oman et al. (2011, 2013)		
MRI	Deushi and Shibata (2011)	interactive	REF-C2(1)
	Yukimoto et al. (2011, 2012)		
SOCOLv3	Stenke et al. (2013); Revell et al. (2015)	prescribed	REF-C2(1)
NIWA-UKCA	Morgenstern et al. (2009, 2013)	interactive	REF-C2(5), fGHG(2), fODS(2)
	Stone et al. (2015)		
ULAQ	Pitari et al. (2014)	prescribed	REF-C2(3), fGHG(1), fODS(1)
HadGEM	Walters et al. (2014); Madec et al. (2015)	interactive	REF-C2(1)
	Hunke et al. (2010); Morgenstern et al. (2009)		
	O'Connor et al. (2014); Hardiman et al. (2017)		
UMUKCA	Morgenstern et al. (2009); Bednarz et al. (2016)	prescribed	REF-C2(2)
ACCESS-CCM	Morgenstern et al. (2009, 2013)	prescribed	REF-C2(3), fGHG(1), fODS(1)
	Stone et al. (2015)		
NIES	Imai et al. (2013); Akiyoshi et al. (2016)	prescribed	REF-C2(1), fGHG(1), fODS(1)
UMSLIMCAT	Tian and Chipperfield (2005)	prescribed	REF-C2(1), fGHG(1), fODS(1)
CHASER	Sudo and Akimoto (2007)	interactive	REF-C2(1), fGHG(1), fODS(1)
LMDz-REPROBUS	Marchand et al. (2012); Szopa et al. (2013)	interactive	REF-C2(1)
	Dufresne et al. (2013)		
CESM1-CAM4-Chem	Tilmes et al. (2016)	interactive	REF-C2 (3)

* The fourth ensemble of WACCM (WACCM-4) was provided by M. Abalos; ** EMAC-L47 simulations are not ensembles, as one simulation is with prescribed SSTs and one with interactive ocean

2.2 Observational data

For observations, we make use of the BAyeSian Integrated and Consolidated (BASIC) ozone composite that merges SWOOSH (Davis et al., 2016) and GOZCARDS (Froidevaux et al., 2015) through the BASIC method of Ball et al. (2017). The method was developed to account for artefacts in composite datasets that are a consequence of merging observations from different instruments that each have unique spatial and temporal observing characteristics. As a result, these artefacts can alias in regression analysis and bias, e.g., trend estimates (see examples in Ball et al. (2017)). BASIC composites aim to account for and reduce artefacts using an empirically driven Bayesian inference methodology, but it relies on the availability of already developed ozone composites. Here, BASIC$_{SG}$ has been extended to the end of 2019 using the latest versions of GOZCARDS, v2.20, and SWOOSH, v2.6. As such BASIC$_{SG}$ covers 1985-2019 as monthly mean zonal means on a 10° latitude grid from 60°S–60°N and over a pressure range of 147–1 hPa (∼13–48 km). BASIC$_{SG}$ was presented in Ball et al. (2018), and a sensitivity analysis of trends was applied to it in Ball et al. (2019), with examples of data artefacts that it addresses in the accompanying appendix and supplementary materials, respectively.

To obtain an observationally constrained estimate of tropical upwelling and extratropical downwelling mass fluxes, we use ECMWF's fifth generation of atmospheric reanalysis data, ERA5 (Hersbach et al., 2020). The mass fluxes are calculated from 6-hourly data on the reduced set of pressure levels.

2.3 Statistical Methods

In some parts of our analysis, and to make a robust comparison between multiple models and a single 'real-world' realization, i.e. observations, we form probability distributions to estimate the combined probability of the ozone trends from all REF-C2 models. To do so, we calculate the linear trend and the associated uncertainty using a least squares method for every simulation. Then, to build the trend probability distribution of the models, first one of the 18 CCMI models is randomly selected, assuming that the models are randomly uniformly distributed. In case the selected CCM provided ensemble member simulations, in a second step one of these members is randomly chosen, thus taking into account that ensemble members are treated differently than individual models. In the next step, the trend estimate ($t^{M_{i,k}}$) of the specific randomly selected CCMI model M_i with ensemble member k is calculated by randomly choosing an ozone trend value from the trends associated and assumed normal distribution $\mathcal{N}$, which is based on the mean $\mu_{M_{i,k}}$ and standard deviation $\sigma_{M_{i,k}}$ of the simulations linear trend. Thus we can write the trend estimate of the selected model simulation as: $t^{M_{i,k}} = \mathcal{N}(\mu_{M_{i,k}}; \sigma_{M_{i,k}})$. In order to take into account the uncertainty of the single observational dataset (σ_{obs}), we also add to the calculated model trend estimate a random estimate of the observational noise by taking the observational standard deviation of the linear regression coefficient. We repeat the above described procedure 50 000 times. With that we have a large sample of model trends and can build up a robust probability density function (PDF) of the REF-C2 ozone trends. From these estimated PDFs we can then estimate the probability of a given trend relative to the models. We derive a "probability of disagreement" between the observational and the modeled trend distribution by taking the central interval of the models' trend distribution with the observed trend value as threshold of this interval. To calculate this central interval we order the the 50 000 values from the REF-C2 trend distribution according to

their probability values and then sum up the ordered probability values until the value of the observed trend is reached. This probability value indicates our estimate of whether the observations agree with the models, i.e. high probability values indicate that a disagreement between models and observations is less likely due to chance.

2.4 Analysis Methods

We here provide a short description of our methodology to analyze transport processes, which follows the studies of Dietmüller et al. (2018) and Eichinger et al. (2019). Stratospheric mean AoA is defined as the mean residence time of an air parcel in the stratosphere (Hall and Plumb, 1994; Waugh and Hall, 2002). In the CCMs, the AoA tracer is implemented as an inert tracer with a mixing ratio that linearly increases over time as lower boundary condition. AoA is then calculated as the time lag between the local mixing ratio at a certain grid point and the current mixing ratio at a reference point.

The residual circulation transit time (RCTT) is the hypothetical age that air would have if it only followed the residual circulation, thus without processes such as eddy mixing or diffusion. RCTTs are calculated by backward trajectories on the basis of the Transformed Eulerian Mean (TEM) meridional and vertical velocities (referred to as residual velocities) with a standard fourth-order Runge-Kutta integration (Birner and Bönisch, 2011). The RCTT is then the time that these backward trajectories require to reach the tropopause from their respective starting point in the stratosphere. The RCTT differs from AoA because of resolved and unresolved mixing. In the stratosphere, this is due to the mixing of air between branches and the in-mixing of air from the mid-latitudes into the tropical pipe, which leads to recirculation of old air around the BDC branches. In global model studies, this effect has been named aging by mixing (AbM) and is interpreted as the difference between AoA and RCTT (e.g. Garny et al., 2014).

3 Results

3.1 Ozone trends over the period 1998–2018 in CCM simulations and observations

In this section we analyze the ozone trends of all 'free running' CCMI-1 simulations (REF-C2), including all ensemble realizations of each model, for the period 1998–2018 together with the observational data, BASIC$_{SG}$. We chose the period 1998–2018 to be consistent with the observational trend estimate in the ozone recovering phase as presented by Ball et al. (2018). Note that ODSs are declining in this period as a result of the Montreal Protocol and its amendments. By using the REF-C2 simulations we include a wide spectrum of SST variability in the different CCMs, as they use either an interactive ocean or prescribed SSTs from a coupled ocean-atmosphere model simulation (see Tab.1). Ozone trends are calculated by simple linear regression, using the monthly deseasonalized ozone time series. We refrain from excluding sources of variability such as QBO, ENSO (El Nino Southern Oscillation), solar cycle or volcanic eruptions in the regression analysis to capture the full range of variability of ozone trends over the given period. Hence our trend estimates have to be interpreted as resulting from both forced trends (e.g., via GHG increases and ODS decreases) as well as from natural and internal climate variability. In the following we compare the calculated ozone trend from the observational data to the trends presented in Ball et al. (2018, 2019, 2020) that used a

dynamical linear modeling (DLM) approach, which attempts to take natural sources of variability into account. In a nutshell, DLM has many similarities with ordinary least squares multiple linear regression (MLR), using predictor variables to account for some of the variability in the timeseries (e.g. solar variability, the QBO). Where DLM primarily differs from MLR is in allowing for a non-linear trend to be estimated and for the seasonal cycle to evolve with time, and therefore the shape of these terms is not predefined. For more details, see Laine et al. (2014) and Ball et al. (2018).

The panels of Fig. 1 show a latitude-pressure cross-section of the ozone trend for observations (first panel of Fig. 1) and all free running CCMI model simulations. Generally, the linear trend fit we perform on the BASIC$_{SG}$ data yields similar spatial patterns and magnitudes to those estimated in Ball et al. (2018) with the DLM approach (see their Fig. 1f). There are a few small differences, e.g., our linear trend fit results in larger positive trends in the upper stratosphere over the southern tropics of $\sim 1\%$, a slightly less negative trend in the northern hemisphere middle stratosphere ($<1\%$), and consistently large and negative trends close to 100 hPa in the tropics as opposed to a smaller and insignificant trend at around 10°S and over 100-80 hPa in the DLM estimate as shown by Ball et al. (2019). Most notably, linear trend calculations result in small positive trends (up to $\sim$3%) in the southern mid-latitude lower stratosphere, as opposed to overall negative but insignificant trends reported by Ball et al. (2019) in that region. However, the comparison reveals that the overall magnitude and trend pattern is also captured by the simple linear regression, i.e. it is not dependent on the exact method used to calculate the trends. Therefore, we proceed with using a linear fitting approach for the comparison between observations and CCMs, though the above caveats should be kept in mind when comparing with a full regression analysis using DLM (Ball et al., 2019).

Overall, large inter-model variability of the trends derived from the individual REF-C2 simulations (including all ensemble members) is revealed in Fig. 1. Nevertheless, a number of features can be identified that are consistent over most models and all their ensemble members. In the upper stratosphere (1-10 hPa) nearly all simulations consistently show an overall positive ozone trend. This ozone increase can be explained by the decrease of ODSs (see e.g. WMO, 2018) and by a slow down in ozone destruction rates as the stratosphere cools from GHG increases (see e.g. Portmann and Solomon, 2007), as will be further discussed in Sec. 3.5. This upper stratospheric ozone trend has been found for climate model simulations and for observational data in several studies before (e.g. SPARC CCMVal, 2010; Harris et al., 2015; Steinbrecht et al., 2017; Ball et al., 2018; WMO, 2018; Ball et al., 2020). However, in the lower stratosphere (30-100/150 hPa) we find a wide spread in the ozone trends among the CCM simulations over recent decades. Many REF-C2 simulations exhibit negative trends in the tropical LS, and they are comparable to the observational trend in magnitude and structure. In agreement with earlier studies (e.g. WMO, 2018; Orbe et al., 2020), we will show in Section 3.4 that this tropical ozone decrease is related to enhanced tropical upwelling in a warmer climate. However, there are also simulations showing a positive LS ozone trend in the tropics (i.e., GEOSCCM, SOCOLv3, NIWA-1, WACCM-3/4, CAM4-1/2, LMDZrepro, HadGEM; note that the number of the ensemble run is denoted with -1, -2 and so on). At northern and southern mid- and high-latitudes most simulations exhibit a positive trend, but with a pronounced intermodel spread. Only a few simulations show negative trends in either northern or southern mid-latitudes (e.g. GEOSCCM, WACCM-3, WACCM-4), but it is important to point out here that none of the 31 simulations reproduces the observed negative ozone trend pattern with an ozone decrease covering the tropical belt and extending to the mid-latitude (50°S-50°N), as shown in the upper-left panel and previously in Ball et al. (2018, 2019). This discrepancy in the

LS ozone trend between observations and models has been reported before (e.g. ozone trends, based on CCMI simulations, (WMO, 2018; Orbe et al., 2020), and in comparison to CCMVal-2 simulations (Ball et al., 2020)). For CCMs that provide multiple ensemble members (WACCM, NIWA, ULAQ, ACCESS, CAM4 and UMUKCA), we also identify a large ensemble spread in the simulated LS ozone trends. For example in WACCM two ensemble members simulate positive tropical ozone trends, while the two other members simulate negative tropical ozone trends. In WACCM (as well as in NIWA and CAM4), the coupled ocean allows for differences in the SST variability between the ensemble members, possibly explaining the large spread in tropical ozone trends. However, as is also the case for models with prescribed SSTs (ACCESS, ULAQ, UMUKCA) that exhibit a large spread between the simulations, the SST variability is not the only reason for the different trend pattern, as was similarly reported and discussed by Ball et al. (2020) for CCMVal-2 models. The large spread in LS ozone trends between ensemble members is further in agreement with the study of Stone et al. (2018). They used a nine member ensemble of a 'free running' CCM simulation (CESM1-WACCM) and showed that LS ozone trends over the years 1998–2016 are characterized by large internal variability, with e.g. the LS ozone trend ranging from +6% to -6% per decade. But note, again, that none of these ensemble members showed the coherent decrease in ozone in the tropics and extratropics as found in observations (Ball et al., 2020).

Following this qualitative discussion on the spread in the ozone trend pattern between the CCM simulations, we now turn to the LS ozone trends with a more quantitative comparison of the apparent inconsistencies between observations and CCMs. We calculate the trends of the deseasonalized LS ozone columns for the period 1998–2018 in two regions: the inner tropics (20°N-20°S) and in the northern mid-latitudes (30°N-50°N). We choose the northern mid-latitude band 30°N-50°N for direct comparability with the study of Ball et al. (2020). The pressure range of the lower stratosphere was taken to be 30–100 hPa for the tropics and 30–150 hPa for the mid-latitudes, to take into account the differences in latitudinal tropopause heights. Trends and their uncertainties (represented by the 90% confidence interval of the linear slope) are shown for each of the 31 available REF-C2 simulations of 18 different CCMs in Fig. 2. We decided to focus on the northern mid-latitudes here, because the SH mid-latitude trends are likely more strongly influenced by the large chemical depletion of ozone within the polar vortex. We will come back to the LS ozone trends of the southern mid-latitudes in Section 3.5.

In the tropics about half (42%) of the REF-C2 simulations show a significant decrease, about the same (42%) show a non-significant change, and about 15% a significant increase in the integrated tropical LS ozone column. Note that significance is defined as the non-overlap of the error bars (90% confidence interval) with the zero trend. The resulting MMM ozone trend (see red bar on right of Fig. 2) is negative (-0.37 DU/dec), but it is insignificant due to the considerable spread among the different models. The 25th-75th quantile of the distribution ranges from -1.12 to 0.20 DU/dec (see edges of box on the right of Fig. 2). Note that for the calculation of the MMM trend, we choose to weight each of the 31 simulations equally (i.e., not taking into account that some models have multiple ensemble members) because the trend variations among ensemble members are as large as among the different models over this period.

The observed tropical LS ozone trend of -1.07 DU/dec is statistically significant at the 90% level. Thus the observed tropical trend is more strongly negative than the MMM trend, but lies within the 90% confidence interval of the MMM trend ([-1.76 DU/dec; 1.03 DU/dec]).

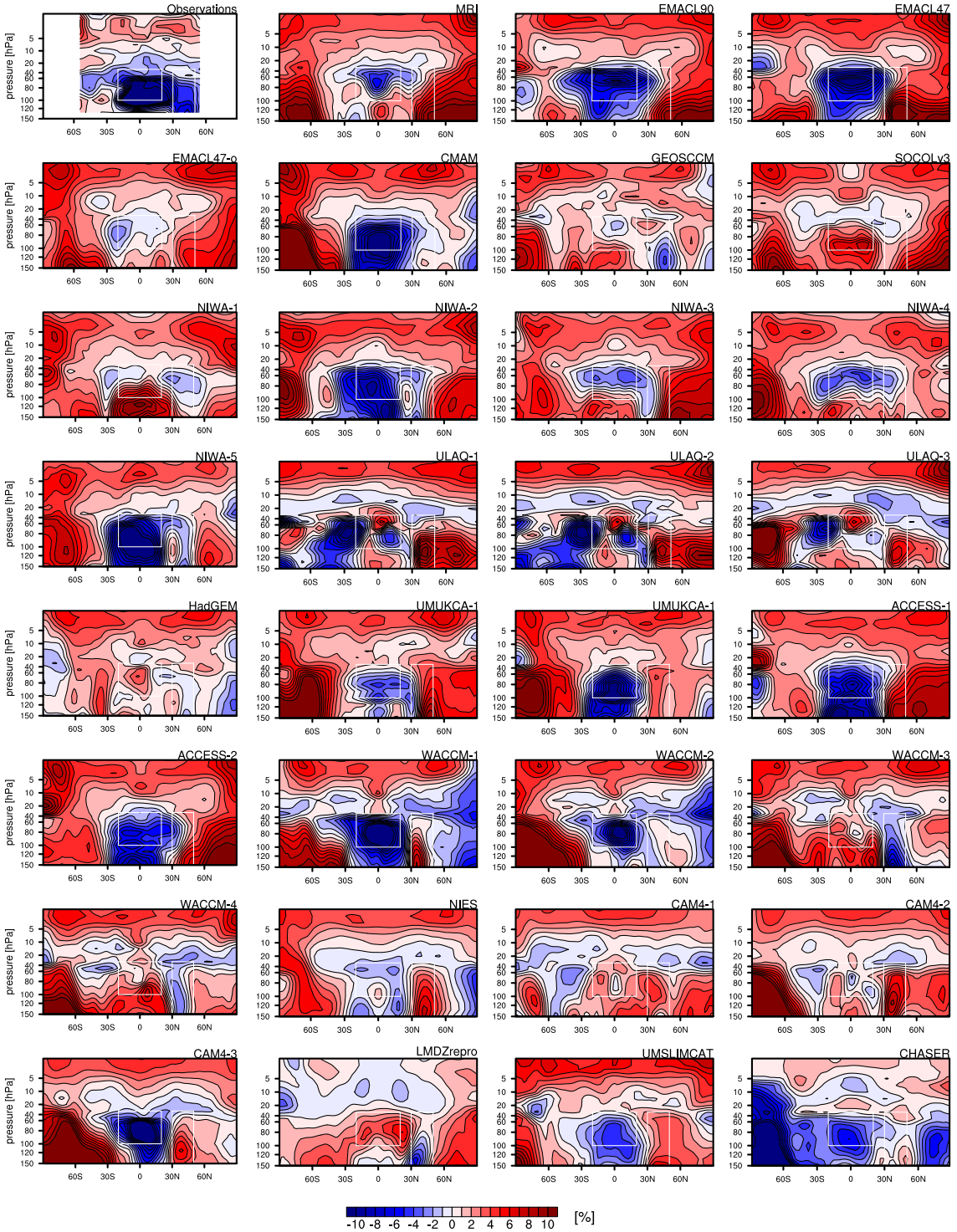

Figure 1. Latitude-pressure cross section of the ozone trend over the period 1998–2018 for the observational data set BASIC$_{SG}$ and for all CCMI REF-C2 simulations. Trends are given as relative ozone changes over the whole time period. Boxes illustrate the regions selected to integrate ozone in the LS for trend comparisons later in this study, i.e. in the tropics (20°N-20°S, 30–100hPa) and in the northern mid-latitudes (30°N-50°N, 30–150 hPa).

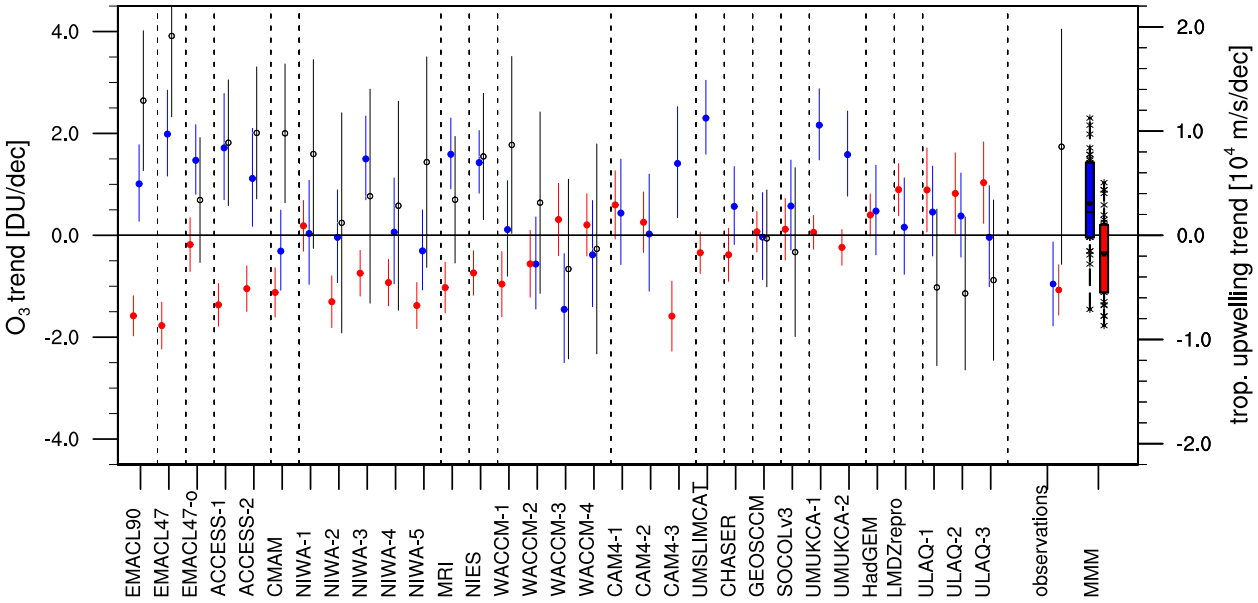

Figure 2. LS ozone trends and their uncertainties in the tropics (20°N–20°S, red dots) and northern mid-latitudes (30°N-50°N, blue dots) together with tropical upwelling trend (black circles, for all simulations providing TEM diagnostics) for the period 1998–2018 for all REF-C2 simulations. Dashed lines separate the individual models. Moreover, observational trends (1998–2018) and multi-model mean trends are given. Observational data for ozone are taken from BASIC$_{SG}$, and for tropical upwelling from ERA5 reanalysis. Error bars associated with each LS ozone trend represent the 90% confidence intervals. The multi-model mean trends are shown as boxplots: the black solid line in the box indicates the median, the black point the MMM and the colored box ranges from the 25th to the 75th percentile of the trends. Crosses denote trends of individual model simulations not lying within the box.

In the northern mid-latitudes less than half (40%) of the REF-C2 simulations show an increase in the LS ozone column, while the remaining 60% of the simulations show a non-significant change (either positive or negative). There is only one simulation (WACCM-3) that shows a significant decrease in the mid-latitude LS ozone column, and in this simulation the tropical ozone trend is positive (but not significant). The resulting MMM trend in the northern mid-latitudes is positive (+0.63 DU/dec) with a high inter-model spread: the 25th to 75th quantile of the distribution ranges from -0.04 to 1.42 DU/dec. Note here, that the observational trend (-0.96 DU/dec) lies outside the 90% confidence interval of the MMM trend in the mid-latitudes ([-0.91 DU/dec; 2.16 DU/dec]).

Fig. 2 also reveals that over the years 1998–2018 more than half of the model simulations have a dipole trend pattern in the LS ozone column, i.e. the sign of the tropical ozone trend is opposite to that in mid-latitudes. This trend pattern with

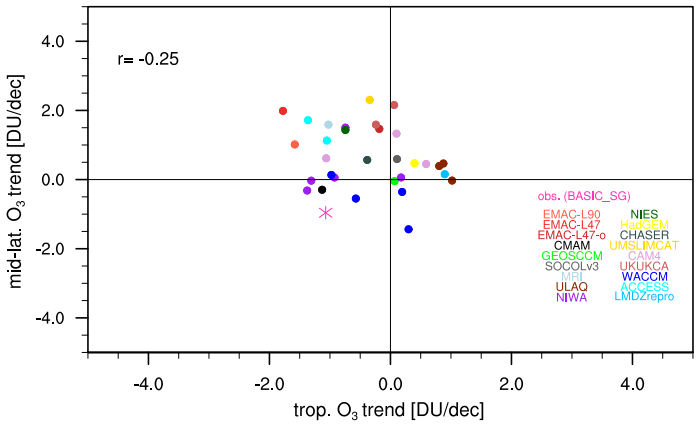

Figure 3. Inter-model correlation between tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trends, calculated over the period 1998–2018 for 31 CCMI REF-C2 simulations. All ensemble members of a particular model are shown in the same color. The observational ozone trends (BASIC$_{SG}$) are also included in here as a star.

.

negative LS ozone trends in the tropics and positive LS ozone trends in the northern mid-latitudes can be found for almost half the simulations (45%), and a trend pattern with a positive ozone trend in the tropics and negative trend in the northern mid-latitudes is found in 13% of the simulations. The remaining simulations do not show this dipole, but either both have a positive trend in the tropics and the mid-latitudes (29%), or a negative trend in both tropics and mid-latitudes (13%, i.e. 3 simulations, namely NIWA-5, CMAM, WACCM-2). Only 3 out of 31 simulations simulate negative, but not significant, trends both in the tropics and northern extratropics, and thus they show a similar behavior to observations (see right of Fig. 2 and Ball et al. (2019)). However, their zonal trend patterns (see Fig. 1) reveal, that none of these three simulations reproduces the observed trend pattern with consistent negative trends from 50°S-50°N in the LS. Consequently it is important to keep in mind that the results of these (averaged) trends depend on the choice of the latitude-pressure box, as the integration over a wider latitude band can lead to a cancellation of opposing trends.

Next, we analyze whether a systematic relationship between the LS tropical and mid-latitude trends exists in the CCM simulations. For this, the simulated northern mid-latitude LS ozone trends are plotted against the simulated tropical LS ozone trends over the time period 1998–2018 for all 31 REF-C2 simulations and for the observed data-set BASIC$_{SG}$ in Fig. 3. As discussed above, in the LS the majority (45%) of the models have a negative ozone trend in the tropics and a positive trend in the northern mid-latitudes. Moreover this illustration again highlights that the trends estimated from observational data are lying on the outer edge of the model trends distribution. The inter-model correlation between the tropical to mid-latitude trends is negative with a low correlation coefficient (-0.25). Thus, for the chosen period the tropical ozone trends are only weakly linked to mid-latitude ozone trends in the models. However, we expected that the two trends are highly (negatively) correlated,

as from our understanding increased tropical upwelling leads to decreased tropical ozone, and this upwelling increase should be linked to an increased mid-latitude downwelling, which would enhance ozone in the mid-latitudes. However Fig. 3 does not support this. Also slightly varying the period (i.e. looking at the periods 1999-2019, 2000-2020, 2001-2021) reveals very low negative or near zero correlations (not shown here). To get a better understanding of the processes leading to the given LS ozone trend patterns we will investigate the relationship of LS ozone trends to stratospheric transport trends in Section 3.4.

Overall we can conclude from the analysis of ozone trends in the suite of CCMI models (see Fig. 1, 2 and 3), that the LS ozone trends exhibit a considerably large spread across both the different models, but also across ensemble members from a single model, in particular in the mid-latitudes. This indicates that ozone variability considerably influences the LS trends, in agreement with the recent studies by Chipperfield et al. (2018) and Stone et al. (2018). However, even when considering the high variability of possible trends in CCM simulations, the observational trends emerge as an unlikely realization of the simulations over the period 1998–2018. In the next section, we will analyze the robustness of this finding by varying the period of the trend calculation, and providing an in-depth statistical analysis of the likelihood of the observed trend lying within the suite of modeled trends.

3.2 Robustness of lower stratospheric ozone trends

In the previous section we found that the observed LS negative ozone trend in the mid-latitudes together with a simultaneous negative trend in the tropics is unlikely based upon the suite of CCM simulations. To further establish the robustness of this result, we here test whether this also holds for time periods that are slightly different to the period 1998–2018 we considered before. Thus, in this section we first want to investigate how variability influences the ozone trends, and second we want to quantify the likelihood of the observed trend being a realization of the distribution of the modeled trends. To answer those questions, we calculate the LS ozone trends by varying the start and end years of the time period. In Fig. 4 (a) and (b), the observed tropical and mid-latitude ozone trend in the LS is shown for start years varying from 1995–2001 (y-axes) and end years from 2013–2019 (x-axes). Both tropical as well as mid-latitudes LS ozone trends are consistently negative for all chosen periods in the observations (top row). This is in line with the results of Ball et al. (2019), who found that the observed negative sign of the tropical and mid-latitude trends remain insensitive to changing the end year. In the tropics, observational LS ozone trends are consistently negative with values between -0.64 and -1.24 DU/dec for all possible start end year combinations. In the mid-latitudes the trends are also negative for all shown time periods, but are more variable than in the tropics (values range between -0.11 and -1.22 DU/dec). In particular at mid-latitudes, the strongest negative trends are found for start years of 1996 to 1998, and a sudden decrease in the trend magnitude is found for the start year 1999 and 2000. Thus, the analysis in Ball et al. (2018, 2019) and in the preceding section focused on a period with particularly strong negative mid-latitude ozone trends. Possible reasons for the sudden change in the trend, such as the strong ENSO event in 1998, are discussed in Section 4. Note that the trend magnitude increases again for the start year 2001, which again suggests that interannual variability influences the observational mid-latitude trends.

Fig. 4 (c) and (d) display the tropical and mid-latitude trends as a function of start and end year derived from the model simulations. To do so, a robust estimate of the trend probability distribution considering all model simulations was derived (see

Section 2.3) and from this distribution the most likely trend is shown (see peak in the models' trend probability distributions of Fig. S1 and S2 in the supplement). In the tropics the ozone trends derived from the REF-C2 simulations are negative and range from -0.74 to +0.02 DU/dec. In the mid-latitudes the trends are positive for all possible start/end year combinations, with values ranging from +0.4 to +1.48 DU/dec. In contrast to the sudden change in the mid-latitude observational trend for start years 1999 and 2000, in the REF-C2 simulations no such systematic change can be found. The estimated probability distributions of the trends from the REF-C2 simulations (see Figs. S1 and S2 in the supplement) are typically symmetric around their maximum value and show a single, central peak. The width of the distribution changes when varying the start/end year combination, with narrower distributions for longer time periods. Moreover, visual inspection of the distribution implies that the tropics (Fig. S1) generally have Gaussian-like distributions, whereas the mid-latitudes (Fig. S2) often show a more peaked structure, i.e. with heavier tails. Nevertheless, as an estimate of the width of the models trend distribution, we show in Fig. 4 (e) and (f) the standard deviation of the models distribution (in DU/dec) in the tropics and mid-latitudes, respectively. For longer time periods (values in lower right corner) the standard deviation of the models' trend is smaller, i.e. the distribution is narrower. This indicates that the influence of natural variability is less important for longer time periods, as should be expected.

Given the distributions representing the combined trends of the models, we can now quantify the disagreement between the observational trend estimate and the models' trend probability distributions for each start/end year combination. In Figs. 4 (g) and (h) the "probability of the disagreement" between observational and modeled LS ozone trends is given for the tropics and the mid-latitudes. The value of the "probability of disagreement" is calculated by the central interval of the models' probability distribution when taking the observed trend value as threshold of this interval. Thus, a probability value of 90% indicates that the observed trend falls within the inner 90% of the distribution, i.e., only 10% of the distribution is more extreme than the observed trend: the smaller the given "probability of disagreement" value, the higher is the probability that the observed trend lies within the models' distribution. In the tropics, the observed LS ozone trend falls within the 13% to 73% interval of the modeled probability distribution, i.e. the observed trends are generally likely representations of the models' trends. The agreement is best for short time periods (values in diagonal in Fig. 4 (g)), mostly because of the broader distribution (see Fig. 4 (e) and Fig. S1). Also for early start years (in particular 1995) and end years ranging from 2013 to 2018, the disagreement is small, because model trends are strongly negative for this period (see Fig. 4 (c)). In the mid-latitudes, the observed trend generally lies at more distant parts of the models' trends distribution (73% to 96%), i.e. the observed trend is a more extreme value in the models' distribution. The disagreement is smallest for both the earlier periods (lower left, start years 1995–1997 and end years 2013–2015) and the later periods (upper right, start years 1999–2001 and end years 2017–2019). This coincides with the generally smaller negative trends in those periods in observations (see Fig. 4 (b)) and rather constant trend distributions in the models (see Fig. 4 (d)). For the periods with the strongest negative observed trend (start years 1996–1998), the observed trend lies within the central 90% or higher of the models' distribution, i.e. is an unlikely representation from the modeled trends. The sudden decrease in the observed trend magnitude for start years 1999 (Fig. 4 (b)) is reflected by a decrease of the central interval to about 75%. In general, one might have expected that longer periods lead to better agreement of the observed and modeled trend due to the smaller influence of variability (see Figs 4 (e) and (f))- as we do in the models- however, we do not find this to be true for either the tropics or the mid-latitudes.

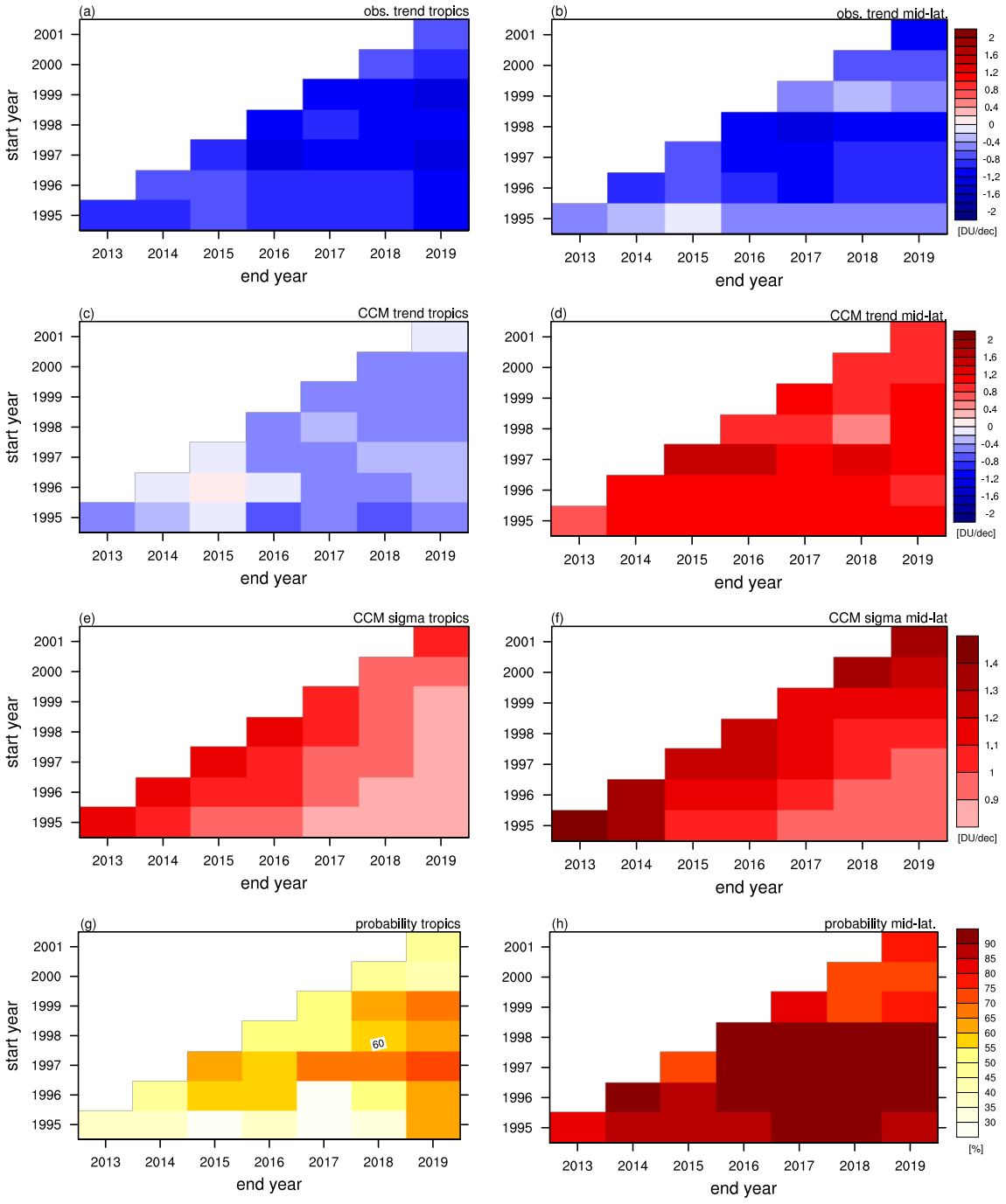

Figure 4. Tropical (right) and mid-latitude (left) LS ozone trends (in DU/dec) as function of different periods for the observational trend of BASIC_{SG}((a)+(b)), the most likely trend of the modeled REF-C2 probability distribution ((c)+(d)) and the 1-σ standard deviation (in DU/dec) of the mean obtained from the probability distribution ((e)+(f)). The panels (g)+(h) show the "probability of disagreement" (in %) between observed trends and the REF-C2 trend probability distribution. In all panels the x-coordinate denotes the different end years (2013-2019) and the y-coordinate the different start years (1995-2001).

3.3 Convergence of future lower stratospheric ozone trends

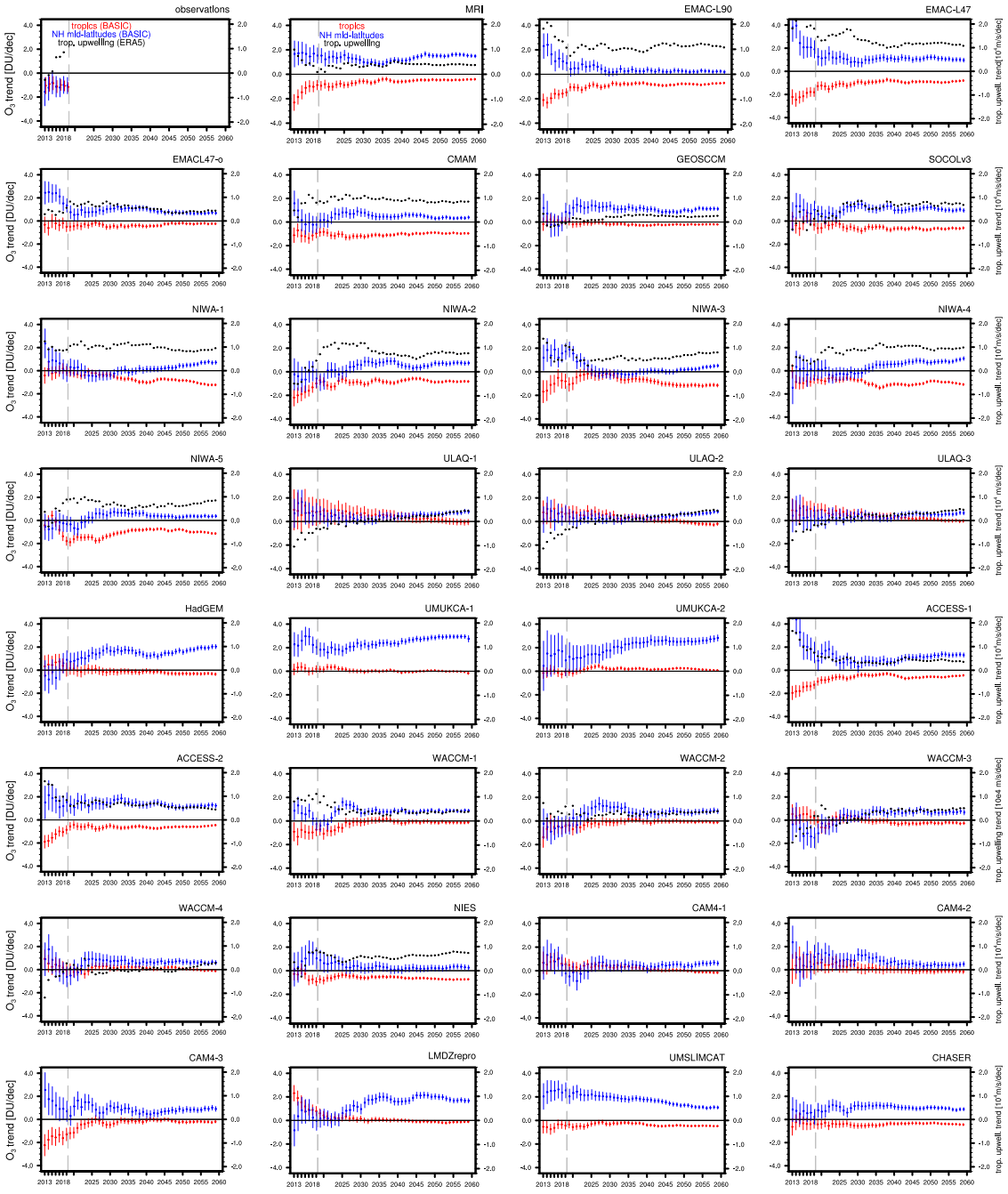

Figure 5. Tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trend and their uncertainties (in DU/dec) of observations (BASIC$_{SG}$) and REF-C2 simulations as a function of the end year (red and blue dots, respectively). Tropical upwelling trends are included for all REF-C2 simulations, where TEM diagnostics was available (black dots); observational tropical upwelling is taken from ERA5 reanalysis. The end year varies from 2013 to 2019 for observational data and from 2013 to 2060 for REF-C2 simulations. Error bars associated with each trend represent the 90% confidence intervals.

In the previous section, the ozone trend robustness was analyzed for time periods of up to 25 years. We will show in the following that, as the considered time periods are extended, the influence of natural variability decreases and the trends converge to the trend forced by long-term GHG and ODS concentration changes. To analyze the timing and the values of the trends' convergence, we extend the period for the trend calculation into the future for all REF-C2 simulations.

Fig. 5 shows the tropical and northern mid-latitude LS ozone trends together with the tropical upwelling trend (black; if available) for periods with the fixed start year 1998, and the end year varying from 2013 up to 2060, by extending the time period by steps of one year. For reference, the observational trends of ozone (from BASIC$_{SG}$) and tropical upwelling (from ERA5) are shown in the upper left panel of Fig. 5, with the last available end point in the year 2019. As shown in the last section, the trends derived from observational data are consistently negative both in the tropics and in the northern mid-latitudes.

As discussed in Sec. 3.1, the ozone trends exhibit a strong inter-model spread for the observational time periods. Both tropical and mid-latitude ozone trends in the individual model simulations vary considerably for different end point years within the observational period (left of the vertical dashed gray lines). The northern mid-latitude trend is generally more variable than the tropical trend. For longer time periods extending into the future, the uncertainties in the LS ozone trends decline and the trends converge in all simulations. All model simulations consistently simulate persistent negative or near-zero trends in the tropics and positive or near-zero trends in the northern mid-latitudes. However, the timing of convergence of the trends to this trend pattern is rather different in the simulations, as can be inferred from Fig. 5, i.e. the convergence appears to be model dependent. For some models, the trends vary little for end years after 2020 (e.g. MRI in Fig. 5), while in other models, the trends still vary considerably until end years around 2030 to 2040 (e.g. the four WACCM ensemble members in Fig. 5). The timing of the convergence is controlled by the ratio of the year-to-year variability to the strength of the forced trends. The relative forcing by ODS versus GHG changes over time, and thereby the forced ozone trends vary over the time periods as well, making it difficult to quantify an exact date of convergence. Still, the trend estimates for the entire period 1998 to 2060 do converge to stable values for almost all models, thus representing the forced trend for this time period. The trend magnitudes over this long period vary strongly between the models, from -0.10 to -1.32 DU/dec in the tropics and from +0.39 to +2.00 DU/dec in the mid-latitudes. Comparing this to the model range of the shorter time period 1998–2040, we see that the tropical trend (+0.06 to -1.12 DU/dec) has not converged to the end point values of 2060, yet. The mid-latitude trend (+0.54 to +2.15 DU/dec) is however close to the 2060 values.

Overall, the mid-latitude trends converge to positive values in the majority of the model simulations (about 85%) by 2030. Thus, if both the year-to-year variability and the forced response of the models is simulated realistically, we should expect the emergence of positive mid-latitude trends from observational records within the next decade.

3.4 Influence of transport processes on LS ozone trends

In this section we aim at improving our understanding of how transport processes control the LS ozone trends in the models. As is well known from earlier studies, tropical upwelling significantly influences stratospheric ozone in the tropics (e.g. Oman et al., 2010). Enhanced tropical upwelling leads to more transport of tropospheric ozone-poor air into the tropical LS. Moreover, a faster removal of ozone in the tropical pipe reduces the residence time in the LS. To analyze how tropical and mid-latitude LS

ozone trends are influenced by transport processes, we show in Fig. 2 the tropical upwelling trends (20°N-20°S, 70 hPa) for all simulations providing TEM diagnostics. This shows that models with strong positive tropical upwelling trends also have large negative tropical ozone trends. However, for the mid-latitude trend it is difficult to visually detect a clear relation with tropical upwelling trends.

Therefore we analyze the relation of tropical upwelling and extratropical downwelling trends to LS ozone trends in terms of a correlation analysis. Fig. 6a shows the inter-model correlation between the tropical upwelling mass flux trends at different stratospheric levels and tropical LS ozone column trends over a sub-set of 20 REF-C2 simulations. Additionally the correlation of the northern mid-latitude downwelling mass flux trends at different levels and LS ozone column trends is provided in Fig. 6b. As above we calculate the trends over the period 1998-2018 and tropical ozone trends are averaged over 20°N-20°S and mid-latitude ozone trends over 30°-50°N.

The correlation profiles between tropical ozone column trends and tropical upwelling trends (red line in Fig. 6a) show significant high negative correlations (r≈-0.8) at all levels between 30 and 100 hPa. Thus, as expected, changes in tropical upwelling at all levels below 30 hPa highly influence LS tropical ozone. This is in line with previous studies (e.g. Oman et al., 2010; SPARC CCMVal, 2010). Between 10 and 30 hPa, the correlation decreases with altitude and becomes insignificant. The correlation values of tropical ozone trends to downwelling trends are positive and also rather high (Fig. 6b). This is clear, as upwelling is directly linked to downwelling, however the negative sign of downwelling causes a sign reversal of the correlation coefficients.

For ozone trends in the northern mid-latitudes (30°-50°N), the correlation of LS ozone to tropical upwelling trends varies in altitude from about -0.2 to +0.4 (solid blue lines in Fig. 6a): It is weakly negative up to 100 hPa; above, the correlation turns to positive values (r≈ 0.4 at 70 hPa). Compared to the relation of upwelling trends to tropical ozone trends, these correlations are quite low and not significant at the 95% level - moreover these correlations are not robust when slightly varying in the period (not shown). The same is true for correlations between mid-latitude ozone trends and downwelling trends (see solid blue lines in Fig. 6b). A possible reason for the non-robust and not significant correlations might be the choice of the mid-latitude averaging region from 30°-50°N. This region can partly include regions of upwelling at some pressure levels, and the location of the turnaround latitude is model dependent. Not accounting for a dynamically consistent averaging region might obscure the correlation analysis. Therefore, we additionally define a dynamically more consistent mid-latitude region by averaging the LS ozone column from the turnaround latitudes of the BDC to 50°N. For each month the averages were taken by calculating the position of the residual streamfunction maximum at each level, and then averaging the LS ozone column from this turnaround latitude to 50°N. It was further ensured that tropospheric air is not included in the averages (which could happen at levels below the tropical tropopause) by using only the region above the tropopause.

The ozone trends in this dynamically defined box are slightly higher compared to the fixed latitudinal region between 30° and 50°N, but given the large spread in trends this difference is not significantly different (see Table 2; the same is true for the longer period 1998–2040, not shown). The correlation profiles for LS ozone trends within this dynamically defined mid-latitude box is included in Figs. 6a and b (see dashed blue line): Due to the dynamical consistency of mid-latitude ozone and the downwelling region, the correlations increase in absolute number compared to the correlations with ozone trends in the fixed

boxes, and the correlations are more robust across different periods (not shown). In particular, the correlation of ozone trends in the dynamically defined averaging box to downwelling peaks at 100hPa with a significant correlation coefficient (r≈0.5). Up- and downwelling at around 100 hPa reflects the shallow branch of the BDC (see e.g. **?**Dietmüller et al., 2018). Thus, the significant positive correlation of downwelling trends around this level to mid-latitude ozone trends suggests that an enhanced shallow branch leads to a decrease in ozone in this region. This would be consistent with enhanced horizontal advection via the shallow branch that transports tropical ozone poor air to the mid-latitudes. The fact that correlations decrease to insignificant correlation values above (and correlations to tropical upwelling even change sign) likely reflects the relation of mid-latitude ozone trends to downward transport of ozone via the deep branch. Thus, overall the correlation analysis suggests that the two competing transport processes of shallow horizontal versus deep vertical advection influence ozone in the mid-latitude LS.

In general, the weaker correlations of mid-latitude ozone to up/downwelling compared to tropical ozone suggests that mid-latitude ozone changes are controlled by a variety of processes, possibly also including two-way mixing. Furthermore, not only changes in the transport strength, but also in the background ozone gradients, can lead to changes in the transport of ozone. For example, the increase in upper stratospheric ozone mixing ratios could lead to enhanced downward transport of ozone despite an unchanged downwelling strength.

To better elucidate the role of different transport processes in the different regions, we additionally analyze the local correlation of AoA trends to the ozone trends for a subset of 9 REF-C2 simulations that provide the necessary diagnostics (namely EMAC-L90, EMAC-L47-1, ACCESS-1, WACCM-1, CMAM, GEOS, SOCOL, MRI, NIWA-1). As shown in Fig. 7a, in the middle stratosphere the correlation coefficients are relatively weak, consistent with the expectation that chemical processes play an important role there. In the LS, we find very high correlations (larger than 0.8) between ozone and AoA trends in the tropics and extending to about 40°N. Thus, inter-model differences in ozone trends are highly controlled by differences in transport trends in this region. Negative correlation values can be found in the LS mid-latitudes north of about 40°N and above 80 to 60 hPa. Interestingly, in the SH correlations are positive throughout the LS. To analyze the role of different transport processes, we separate AoA into the components of residual transport (RCTT) and aging by mixing (AbM; for details see section 2.4). The inter-model correlations between ozone trends and RCTT and AbM trends, respectively, are shown in Fig. 7 b and c. In the LS, RCTT trends are highly positively correlated to ozone trends between 40°S-40°N, whereas for latitudes poleward of 40° the correlation coefficients turn to negative values. AbM trends and ozone trends correlate strongly (and positively) in the LS for latitudes poleward of 30°. This again underlines that in the tropical LS residual transport changes largely control the ozone trends: negative RCTT trends (indicating faster upwelling) are associated with negative ozone trends. This is also in line with the findings of Fig. 6a. In the LS mid-latitudes, on the other hand, both changes in residual transport (RCTTs) and in mixing (AbM) have an impact on ozone trends, leading to the non-homogeneous correlation structure with AoA trends (Fig. 7a). In the region of our interest, i.e. 30°N-50°N, the different transport processes of residual transport with its deep and shallow branch and of two-way mixing appear to influence ozone trends: the RCTT correlations (Fig. 7b) suggest that an enhancement of the meridional component of the residual circulation (shallow branch) leads to an ozone decrease up to 40°N by enhanced transport of tropical ozone-poor air to mid-latitudes. This is in line with the significant positive correlation of models' LS ozone and downwelling trends, that we presented in Fig. 6b. The negative correlations between RCTT and ozone

trends north of 40°N indicate that ozone trends are driven by vertical downwelling (from the deep branch) here: enhanced downwelling (lower transit time) is associated with transport of ozone-rich air from above. Moreover mixing processes play a role in the mid-latitude region. The correlation of AbM trends with ozone trends is positive (r≈ 0.6) north of 30°N in the LS, indicating that mixing is strongly influencing ozone trends in this region as well. Overall Fig. 7 reveals that transport processes in the LS mid-latitudes are complex, as this region is influenced by many competing transport processes. We will discuss this issue further in Section 4.

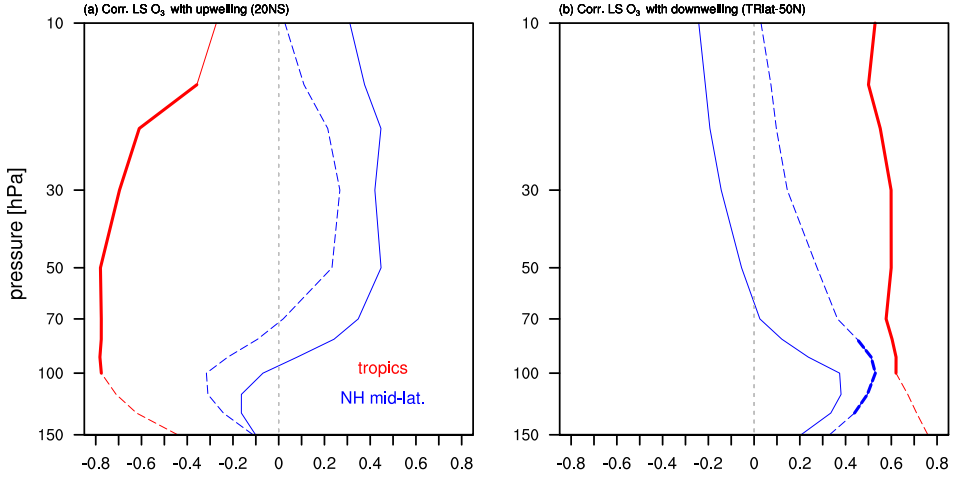

Figure 6. Vertical profile of the inter-model correlation coefficients for (a) tropical upwelling (20°N-20°S) trends (kg/s/dec) to tropical (red line) and northern mid-latitude (blue line) LS ozone column trends and for (b) downwelling mass flux (between the turnaround latitudes and 50°N) trends (kg/s/dec) to tropical and northern mid-latitude LS ozone column trends. Correlations are calculated for upwelling and downwelling trends between 10 and 150 hPa. Mid-latitudes ozone trends are averaged over the latitude band of 30°N-50°N (solid blue line) and also over the dynamical defined latitude band between the turnaround latitudes to 50° (dashed blue line). Trends are calculated over the period 1998–2018 for a subset of 20 REF-C2 simulations. Correlation coefficients which are significant on the 95% level are highlighted in bold.

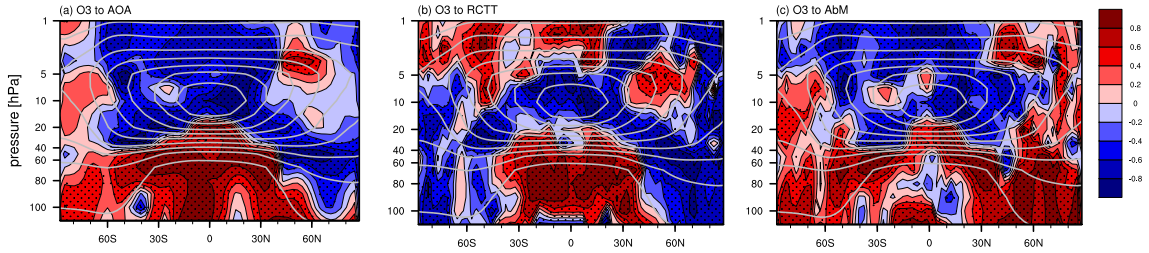

Figure 7. Inter-model correlation coefficients between local ozone trends and (a) local AoA trends, (b) local RCTT trends and (c) local AbM trends. Trends are calculated over the period 1998–2018 for a subset of 9 REF-C2 model simulations. White contours show the MMM ozone climatology, and the stippled regions mark where correlation coefficients are significant on the 95% level.

Table 2. MMM and observational ozone trends, calculated over the period 1998–2018 for tropical upwelling at 70 and 100 hPa, for extra-tropical downwelling at 70 and 100 hPa, for the LS tropical ozone column and for the northern mid-latitude ozone column. Note that LS mid-latitude ozone trends are averaged over the fixed latitude band of 30°N-50°N and also over the dynamical defined latitude band between the turnaround latitudes to 50°. MMM trends and their standard deviation are given over a subset of 20 REF-C2 simulations. Observational based data for up- and downwelling are taken from ERA5 reanalysis and observational data for ozone from BASIC$_{SG}$.

	MMM	**observations**
trop. upwelling trend (70 hPa) [kg/sec/dec]	0.78 10E+7 $\pm$ 1.92 10E+7	1.53 10E+7
trop. upwelling trend (100 hPa) [kg/sec/dec]	1.62 10E+7 $\pm$ 2.21 10E+7	3.14 10E+7
downwelling trend (70 hPa) [kg/sec/dec]	-0.22 10E+7 $\pm$ 1.19 10E+7	-0.35 10E+7
downwelling trend (100 hPa) [kg/sec/dec]	-0.69 10E+7 $\pm$ 2.12 10E+7	-0.15 10E+7
trop. ozone trend [DU/dec]	-0.53 $\pm$ 0.91	-1.07
mid-lat. (fixed) ozone trend [DU/dec]	0.47 $\pm$ 0.87	-0.96
mid-lat. (dyn) ozone trend [DU/dec]	0.78 $\pm$ 0.91	–

3.5 Forced ozone trends in models

In the previous sections we analyzed the ozone trends of the recent 20 year period in detail and found that modeled and observed ozone trends disagree, especially in the northern mid-latitude LS. Assuming the observational data are correct, the question that arises from our results is whether the disagreement stems from the influence of natural variability, or whether the forced response to GHG or ODS concentrations is not captured correctly in the models. Thus in the following, we will investigate the relative role of GHG versus ODS forcing on the ozone trends in the models for the observational period and periods extending into the future. Figs. 8a and 8b show upper and lower stratosphere MMM ozone trends in the tropics (20°N-20°S), in the northern mid-latitudes (30°-50°N) and in the southern mid-latitudes (30°-50°S) for the REF-C2 simulations as well as for the sensitivity simulations with fixed ODS (fODS) and with fixed GHG (fGHG) concentrations (for a detailed description of these sensitivity simulations see Section 2.1). These MMM ozone trends are calculated for the recent time period (1998–2018), for a time period, which extends into the future (1998–2040) and for a future time period (2050–2100). We also include the respective observational trends for 1998–2018. Note that for the calculation of the MMM trends only 10 model simulations are taken into account, as the fODS and fGHG simulations have a smaller sample size than the REF-C2 simulations (see Tab. 1). Moreover we exclude ULAQ for the MMM calculation, as its values are clear outliers compared to other models, such that it would shift the MMM to lower absolute values. Note further that the MMM ozone trends are calculated as the average of the ensemble-means from each model. This ensures that models are weighted equally regardless their ensemble size, which is desirable here as we aim to extract the forced trends, in particular for the longer time periods. Next to the trends averaged over the tropics and mid-latitudes, Fig. 9 shows the latitudinal distribution of the ozone column trends in the upper and lower stratosphere over the period 1998-2040 for the REF-C2, fODS and fGHG simulations. Note that we show the trend over the period 1998–2040 here, as we expect the forced signal to emerge more clearly for this period compared to the shorter observational period.

In the upper stratosphere, the MMM ozone trends over the periods 1998–2018 and 1998–2040 are positive and of the same magnitude in tropical and mid-latitude regions (Fig. 8a). The 1998–2018 MMM trends are more than twice as strong as the observed trends (dots in Fig. 8a), with only one model simulation having lower trend values (in the tropics and NH). Even for the short period of 20 years, the ozone trends are consistently positive for both the models and the observations, indicating that the upper stratosphere MMM trend is robust to inter-annual variability. Therefore, this likely is the forced signal driven by GHG and ODS changes. The analysis of the models' latitudinal distribution in upper stratospheric ozone column trends shows no considerable latitudinal variation (see Fig. 9a). The positive upper stratospheric MMM trend can be explained by the combined effect of still decreasing ODS concentrations at the beginning of the trend periods 1998-2018 and 1998–2040 and by rising GHG concentrations causing stratospheric cooling. The contribution of these two effects is quantified by comparing fGHG, fODS and REF-C2 simulations. In fGHG, the GHG-driven increase of the stratospheric circulation (resulting mostly from the increase in SSTs), as well as GHG induced stratospheric cooling is excluded. In fODS, the chemical ozone destruction via ODS concentrations is excluded. Upper stratospheric ozone trends in fGHG and fODS are positive, but considerably lower than in REF-C2, with trends in fODS having the lowest values. This is in particular true for the extended period 1998–2040,

where we expect clearly forced trends. The weaker upper stratospheric ozone trend in the fGHG simulations can be explained by the missing additional ozone increase due to GHG-induced stratospheric cooling, as ozone is photochemically controlled in these upper regions. The weaker trend in the fODS simulations can be explained by the missing additional increase via the recovery from ODS destruction. The comparison of fODS and fGHG trends over the period 1998–2040 reveals that about 2/3 of the REF-C2 upper stratospheric trend is due to the ODS forced trend. The upper stratospheric trends over the second half of the century (2050–2100) reveal that the ceasing influence of ODS forcing manifests in decreasing ozone trends in the fGHG simulations. However, the ODS forcing still contributes to the ozone increase by about as much as the GHG forcing.

For the LS, Fig. 8b highlights that ozone trends are highly variable in particular for the shorter period of about 20 years and that the MMM ozone trends over the period 1998–2018 and 1998–2040 are negative in the tropics and positive in the mid-latitudes in the REF-C2 simulations. In general, the mid-latitude ozone trends are very variable both in the northern and southern mid-latitudes, but the southern mid-latitude trends are somewhat lower (and negative in some models) for the shorter period. Also in observations, the SH mid-latitude trend is more uncertain and variable (compare observational estimates in Fig. 8b, and see Ball et al., 2019).

In order to attribute modeled LS ozone trends to GHG and ODS changes, we compare the ozone trends of the REF-C2 to fGHG and fODS simulations in Fig. 9b (see also MMM trends in Table S1 of the supplement). For the short time period of about 20 years we find that the MMM mid-latitude ozone trends are positive and overall similar between the fGHG and the REF-C2 simulations. The fODS simulations, in contrast, show a negative MMM mid-latitude trend, but with a very high intermodel spread. Compared to the REF-C2 simulations, the tropical LS trends are less negative in the fGHG simulations and more negative in the fODS simulations. This it what we expect from the missing influence of the GHG concentration rise on tropical upwelling. But note that trends of fODS and fGHG are not significantly different from the REF-C2 simulation. The small, mostly not significant differences (not shown) with its high intermodel spread in the fGHG, fODS and REF-C2 trends over the quite short observational period (1998–2018) again underlines the conclusion that variability strongly impacts LS ozone trends.

For the longer time period (1998–2040), the MMM fGHG trend in the tropical LS is near zero (see Fig. 8b and Table S1). In contrast to the trends over the short time period it can be clearly distinguished from the negative REF-C2 trend and also from the negative MMM fODS trend which is comparable to the REF-C2 trend. This can be explained by the absence of GHG-induced enhancement of tropical upwelling, which strongly influences tropical LS ozone trends. The latitudinal distribution in Fig. 9b shows in more detail the tropical LS ozone column trends in the individual fGHG simulations (red thin lines): most models show trends near zero in the tropical region. The slightly negative ozone trends in the tropics in two models are a bit surprising. However, they probably can be explained by the fact that the upper stratospheric ozone increase can reduce the UV radiation reaching the LS, and thus less ozone is produced there chemically (see e.g. Meul et al., 2014). In the mid-latitudes, the MMM trend in the fGHG simulations is positive and only slightly smaller than the REF-C2 trend, whereas the fODS MMM trend is near zero (see Fig. 8b). This indicates that enhanced downwelling associated with the strengthened circulation plays a minor role in this selected region and is consequently not responsible for the positive trend found in REF-C2. This weak influence of downwelling trends on mid-latitude ozone trends is consistent with the results presented in Section 3.4. There, we found that

downwelling mass flux via the deep branch and mid-latitude ozone trends are only weakly related in REF-C2. Moreover, the near zero ozone trend in the fODS simulations underlines that the mid-latitude ozone trends are strongly influenced by ODS recovery. This might be through decreased local ozone destruction (as ODSs are still decreasing), or through ozone transport from upper or polar regions, where ozone is increasing strongly because of the ”closure of the ozone hole“. Thus, ozone increases in the mid-latitudes, even without an enhanced transport circulation.

To better understand the fact that the mid-latitude fODS trend is near zero, although we expect transport-induced changes in the LS, we show in Fig. 9 (blue thick line) the latitudinal distribution of the MMM fODS LS ozone partial column trend. Here we see that the LS mid-latitude band between 30°N-50°N lies just within a region where ozone trends are shifting from negative to positive values. The MMM trend is negative between 30°N-40°N and positive between 40°N-50°N, explaining the near zero mid-latitude trend over the total latitude band. We suppose that the negative trend 30°-40°N can be explained by enhanced advection through the shallow branch and/or two-way mixing and the positive trend between 40°N-50°N by enhanced downwelling, as suggested by the correlations with RCTTs (Fig. 7b). However, the individual models show quite noisy behaviour in the latitudinal distribution of LS mid-latitude ozone trends, mainly in the NH (thin blue lines in Fig. 9b), indicating that the relative role of trends in the different transport processes might differ in models. The trends in the fGHG simulations are near zero in the inner tropics and positive at all other latitudes, indicating that the recovery from ODSs leads to an increase in ozone almost everywhere throughout the LS. The latitudinal distributions thus indicate that the GHG-driven circulation changes would induce a decrease in ozone from the tropics up to 40°N/S (leading to a near zero trend in the region 30°N-50°N), but due to the recovery of ozone from ODSs, the trend is essentially shifted to positive values, so that the average trend over 30°N-50°N is positive.

The LS ozone trends calculated over the period 2050–2100 confirm the role of ODSs in influencing the mid-latitude ozone trends: despite a strong increase in tropical upwelling in this period (not shown), which drives the strong decrease of tropical ozone in the REF-C2 and likewise the fODS simulations, mid-latitude MMM ozone trends are essentially zero (or slightly negative in the NH) in the fGHG simulation. The effects of an ODS recovery on mid-latitude ozone are smaller in this period due to the declining influence of ODSs, but in the SH mid-latitudes this still leads to a robust positive ozone trend.

Overall our analysis of the fODS and fGHG simulations suggests that the recovery from ODSs is a dominant player for LS mid-latitude ozone trends. GHG-induced circulation strengthening also impacts LS mid-latitude ozone trends, but the competing transport effects via shallow and deep branches lead only to small transport induced trends when averaged over the region from 30°N-50°N.

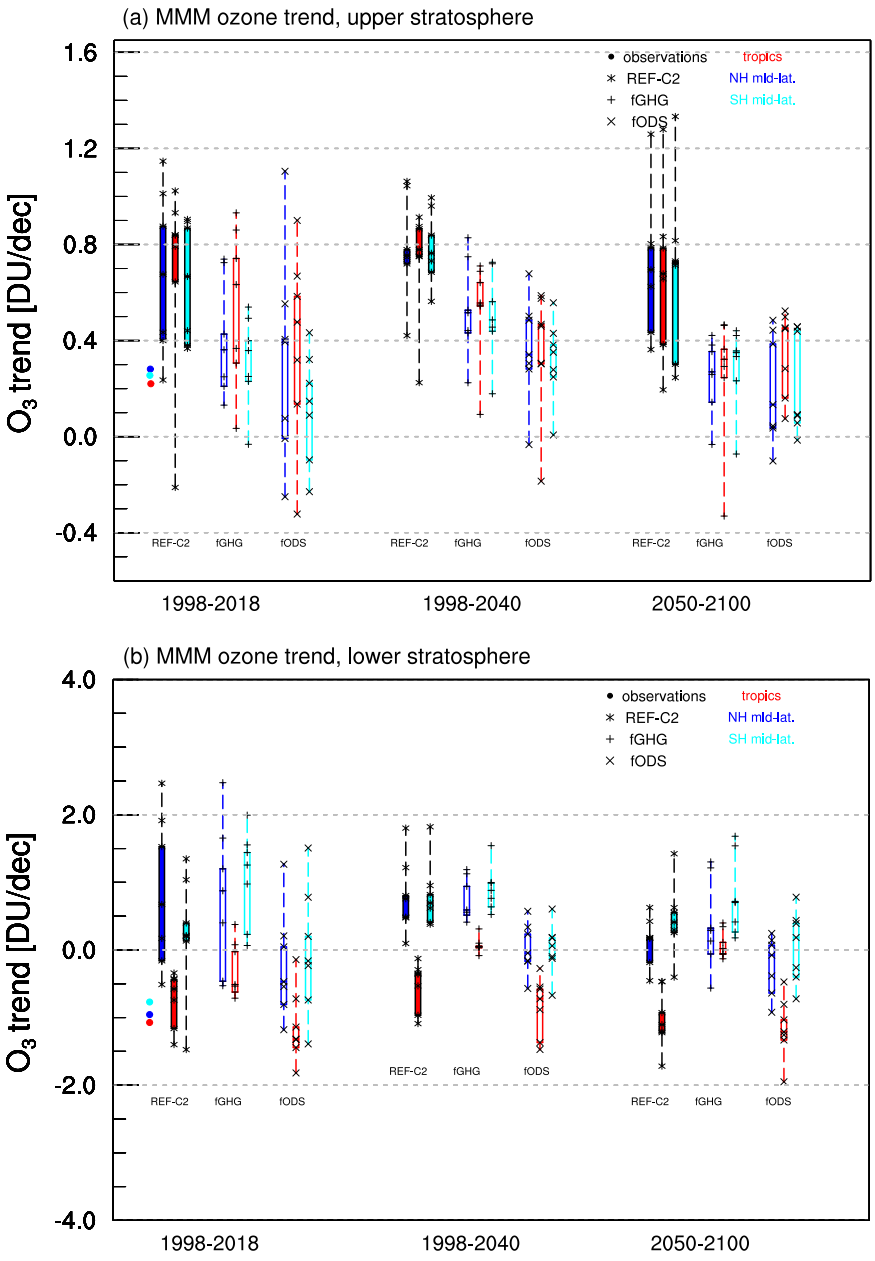

Figure 8. MMM ozone column trends in the tropics (red, 20°N-20°S), in the northern mid-latitudes (blue, 30°-50°N) and in the southern mid-latitudes (cyan, 30°-50°S) for thee different periods (i.e. 1998–2018, 1998–2040, 2050–2100) for (a) the upper stratosphere (1-10 hPa) and (b) the LS (30-100/150 hPa). The boxes extend from the lower to upper quartile of the data with a line for the median and with whiskers to show the minimum and maximum values of the LS MMM ozone trends. MMM trends are given for REF-C2 simulations (filled boxes) as well as for fGHG and fODS simulations (not-filled boxes). Note here that for the estimate of MMM trends only 10 model simulations are taken into account, as fixed GHG simulations have this smaller sample size, and we want to ensure that all three simulation types include the same models for the MMM trend estimate. Individual model trends are denoted by black stars for REF-C2, by black pluses for fGHG and by black crosses for fODS. Observational data are included for the trends over the period 1998–2018 (red, blue, cyan points, respectively).

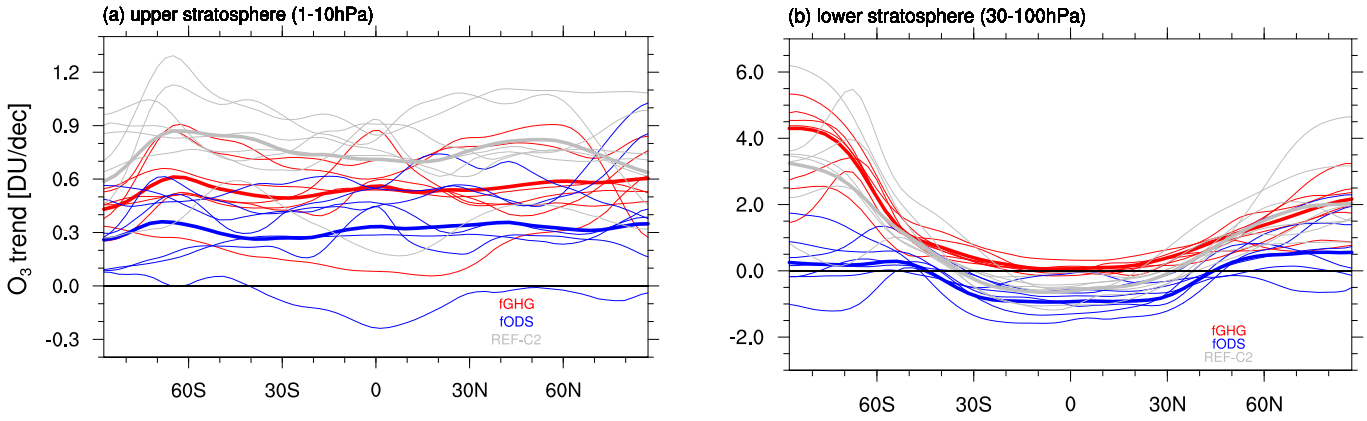

Figure 9. Latitudinal distribution of ozone column trends over the period 1998–2040 for all REF-C2 (gray lines), fGHG (red lines), fODS simulations (blue lines) for (a) the upper stratosphere (1-10 hPa) and (b) the LS (30-100 hPa). Thick lines indicate the MMM ozone trends.

4 Discussion

In the previous sections we analyzed ozone trends over periods spanning the past two decades (i.e., 1998–2018) in detail. We found that modeled and observational ozone trends agree well in the tropical lower stratosphere, but in the northern mid-latitude LS the observed ozone trend represents an extreme value in the distribution of model trends.

In the following, possible reasons for the discrepancy between the mid-latitude ozone trends in the model simulations and the observations will be discussed. One possible reason for the disagreement between modeled and observed LS ozone trends could be issues with the satellite records. For example, instrument biases and drifts can lead to large uncertainties in the observations, particularly in the lower stratosphere. The effect can manifest as steps in the data when instruments, which have different vertical resolutions, are added that can influence trend estimates (REFS as here - I think Gaudel is not relevant here). For a thorough discussion on this topic, see Harris et al. (2015); Ball et al. (2017); Petropavlovskikh et al. (2019). However, for the sake of this discussion we will assume that the observational data record is correct. Hence, the question that arises from our results is whether the disagreement stems from the influence of natural variability, or whether it is related to the forced trend, or more specifically:

- The mean value of the modeled trend distributions might be incorrect. In other words, the forced trend might not be captured correctly by the models.

– If we assume that modeled trend distributions are correct, the observed ozone trend as an unlikely representation might emerge due to very anomalous conditions during the considered periods. This may be caused by extrema in natural variability in the beginning of the time series (late 1990s), and/or in the end of the time series (late 2010s).

– The modeled trend distribution constructed from the REF-C2 simulations might be biased because natural variability (e.g. QBO and ENSO) is not represented adequately in the models. This could lead to an overly narrow trend distribution, and thus would make the observed trend seem more unlikely than it is.

While it is not easily possible to test which of the above explanations is correct, in the following we will discuss their possible contributions to the diagnosed disagreement in the light of our results and what is known from literature.

Representation of forced trends

Based on the CCMI-1 data, we confirmed previous studies in that the decrease in tropical LS ozone is strongly related to the GHG-driven increase of tropical upwelling. The tropical upwelling trend derived from reanalysis (ERA5) lies in the range of the upwelling trends simulated by the models, but on the upper end of the range. This is consistent with tropical ozone trends, which are on the stronger (more negative) end of the trend range simulated by the models as well. Circulation trends derived from reanalysis bear considerable uncertainty (e.g. Abalos et al., 2015), however reanalyses tend to agree better in the recent decades (T. Birner, personal communication, S-RIP report). Therefore, the upwelling trend derived over the period 1998–2018 from ERA5 is likely better constrained compared to earlier periods.

In the mid-latitudes (30°N-50°N), we find that the GHG-driven circulation changes do not lead to a net trend in ozone. This is evident from the fODS simulations (see Section 3.5) and from the vanishing mid-latitude LS ozone trends over the period 2050–2100, when the influence of ODSs cease. The correlation analysis in Sec. 3.4 revealed that competing processes influence ozone trends in this region: an enhanced shallow branch in the LS can decrease ozone due to enhanced horizontal advection, while enhanced downwelling in the deep branch increases ozone (see correlation to RCTTs, Fig. 7b). In the fODS simulations, those competing influences lead to negative LS ozone trends equatorward 40°N/S, and to positive ozone trends poleward 40°N/S (see Fig.9). Thus, this leads to nearly vanishing ozone trends in the mid-latitude region defined as 30°-50°N. The consistent simulation of positive ozone trends in the mid-latitude LS in the REF-C2 MMM for the recent past and the coming decades is thus a result of the ODS concentration decline rather than of GHG-driven circulation changes. The effects of declining ODS concentrations on LS mid-latitude ozone can either be related to the chemical recovery of ozone, leading to local increases in ozone, or maybe more importantly to enhanced ozone transport into this region. Another effect can be induced by the circulation changes due to ODS-driven ozone changes, that have been shown to have had a strong impact on AoA trends in the past (Polvani et al., 2019; Abalos et al., 2019). However, future circulation changes due to this effect are shown to be weak (Polvani et al., 2019). Furthermore, ozone-induced circulation changes are stronger in the SH, not consistent with approximately symmetric ozone trends in the mid-latitudes of both hemispheres.

Given that the positive mid-latitude ozone trends in models are driven by ODSs rather than by GHG changes, the discrepancy to the observed trend could indicate a mismatch in the relative role of the response of ozone to ODS versus GHG forcing. This

means that either the GHG-driven circulation change in the models could be underestimated or differ in structure, or the ODS-driven ozone increase in the mid-latitude LS could be overestimated in the models. As for the latter, we showed that upper stratospheric ozone increases more strongly in the models than in the observational data (see Fig. 8a). Thus, one hypothesis would be that the ODS-driven recovery of stratospheric ozone in the period since the late 1990s is generally overestimated in the models, which would then make negative ozone trends in the mid-latitude LS unlikely in the models. As for the effects of the GHG-driven circulation changes, we mentioned earlier that the MMM tropical upwelling trend is weaker compared to the estimate from ERA5 reanalysis (see Tab. 2). However, the generally consistent tropical ozone trends between models and observations rule out a vast underestimation of tropical upwelling changes. Rather, structural circulation trend differences could contribute to the disagreement in the mid-latitudes. An indication of which is the lower mid-latitude downwelling trend diagnosed from ERA5 as opposed to the models (see Tab. 2). This is also consistent with the finding of poleward shifted turnaround latitudes by Orbe et al. (2020), as discussed below. While it is a likely explanation that structural circulation trends or anomalies contribute to the observed ozone trends, it is not easily possible to separate the role of natural variability in forming those structural circulation trends (see discussion on natural variability below).

In general, since LS mid-latitude ozone trends are driven by competing transport processes (see Section 3.4), the mis-match of trends in this region between models and observations might also indicate a mis-representation of transport processes in the models. We show that ozone trends in the LS correlate well with trends in the passive AoA tracer, indicating that the differences in ozone trends between models are transport-driven. While there is a long-standing discrepancy of AoA trends derived from observations and models in the mid-stratosphere, AoA trends in the mid-latitude LS tend to agree well between models and observations (see e.g. Chapter 5 of WMO, 2018). On the other hand, climatological mean AoA in the suite of CCMI models used in this study varies considerably between models, and it was shown that this is due to differences in mixing effects on AoA (Dietmüller et al., 2018).

The studies of Wargan et al. (2018) and Ball et al. (2020) argue that the LS mid-latitude ozone decrease in observational data is possibly linked to enhanced two way mixing. Ball et al. (2020) used effective diffusivity (Haynes and Shuckburgh, 2000) as diagnostic for horizontal mixing and found that in reanalysis data (JRA-55, ERA-Interim) mixing is enhanced in the 1998–2018 period. In an earlier study, also Ray et al. (2010) showed a substantial increase in effective diffusivity under a changing climate for CCMs and reanalyses data (JRA-25, ERA-40). Recently, Orbe et al. (2020) used the TEM budget analysis of an idealized short lived tracer (that covaries with ozone on interannual and decadal time scales) in 10 free-running ensemble member simulations with the GEOSCCM model, in order to identify the mechanism that is driving the negative LS ozone trends. In contrast to the studies of Ball et al. (2019) and Wargan et al. (2018), the study by Orbe et al. (2020) showed that the mixing effect is not as important for the LS mid-latitude ozone trend. Rather they found a poleward expansion of the residual circulation in the LS with weaker downwelling in the subtropics, and stronger downwelling over the mid-latitudes, leading to negative LS trends in the NH. However, as discussed in Orbe et al. (2020), mixing must be considered in the context of the specific tracer that is analyzed (i.e. short lived tracers are less sensitive to mixing). As such, the analysis of the TEM budget for the tracer ozone could be a focus in further investigations.

Overall, the LS ozone trends are strongly affected by variability over the short period, making it difficult to infer whether the forced trends in models and observations agree. For the models, we extended the time period into the future to investigate for which period length the trends converge. We find that the inter-model spread of the ozone trends substantially diminishes for the longer time period (1998–2040), but to a different extent for different regions (see Fig. 8). In the upper stratosphere, MMM trends are significantly positive already for the shorter period 1998–2018. In the LS, the MMM ozone trends consistently show positive trends in the mid-latitudes for the period 1998–2040, with a comparably low inter-model spread. Thus the question arises as to whether we can expect observational data to also show a positive ozone trend in the mid-latitudes in the future. If the forced model trends are assumed to be correct, we should expect this positive trend to emerge by about 2030 to 2040 (compare Fig. 5).

Influence of natural variability on the observed trend

Sources of natural variability that strongly influence LS ozone are volcanic eruptions, the QBO and ENSO. No major volcanic eruption occurred during the analyzed period, so we will disregard this source of variability. The influence of the QBO and ENSO on the hemispheric mean mid-latitude ozone is of the same magnitude, and thus they can both impact LS ozone trends, as shown by the study of Olsen et al. (2019).

We know from earlier studies that the QBO has a strong dynamical effect on the sub-tropical and mid-latitude LS ozone (e.g. Randel and Wu, 2007). Moreover it was recently shown that ozone trends in the mid-latitudes are directly linked to the QBO, as the QBO induces a secondary circulation (see e.g. Ball et al. (2019) and A. Stenke personal communication, EGU 2020). In 2016, the typical QBO phasing was disrupted, and this has been shown to be associated with negative LS ozone anomalies in the tropics (Kusuma et al., 2019). These negative anomalies at the end of the time period would lead to a strengthened negative ozone trend, and our analysis indeed shows slightly stronger negative tropical ozone trends for the end year 2016 compared to 2015 (see Fig. 4a). The mid-latitude ozone trend is also stronger for the end year 2016, which however does not fit expectations (QBO-induced anomalies are of a different sign in tropics and extratropics, see e.g. Randel and Wu, 2007). Another way, in which the QBO could lead to decadal scale variability in ozone, and thus influence the trends, was recently reported (J. Neu personal communication, AGU 2018): since the QBO's influence on tropical upwelling depends on the season, the timing of the QBO phases is crucial for its influence on trace gas concentrations. Similarly, Ball et al. (2019) pointed out that non-linear attribution may be required to capture the QBO's impact.

One of the strongest warm ENSO events on record occurred in late 1997 (Jensen et al., 1998). By using CCM (WACCM) simulations with prescribed SSTs from observations, Calvo et al. (2010) showed that this strong ENSO event was associated with low ozone values in the tropics and high values in the mid-latitudes. This is in line with observational results by Randel et al. (2009). Consequently, mid-latitude ozone trends should be more negative when beginning the time period with this warm ENSO year. This is consistent with the strong mid-latitude trends in the BASIC$_{SG}$ data-set for the start years 1998 (and less so for 1996-1997, see Fig.4 (b)). However, as the tropical trend is not associated with weaker negative trends for the start year 1998, this explanation is again not fully consistent.

As stated earlier, we have refrained from applying a multiple linear regression (MLR), which potentially would take at least part of the named sources of variability into account. If the trend strengths and patterns are strongly influenced by anomalous natural variability events, one might argue that removing this variability via an MLR method would have a large impact on the trends. However, the trend estimates by Ball et al. (2018), that take ENSO and QBO variability into account differ only in details from our trend estimates. Note that an MLR method might not fully account for the induced signals by QBO or ENSO, because, as mentioned above, their influence is likely non-linearly dependent on the signal strength and the signal timing. Thus, an MLR analysis cannot conclusively clarify the role of natural variability for the observed trends.

Overall, the sudden systematic change in the magnitude of the mid-latitude observational trend (Fig. 4 (b)) indicates that natural variability (in particular the strong ENSO event in 1997) influenced the observed trends over the analyzed periods, and contributed to the particularly strong disagreement of observed and modeled mid-latitude trends for the relevant time periods. However, the expected effects of QBO and ENSO events on the trends are not entirely consistent between tropics and mid-latitudes. Possibly an exceptional combination of different factors led up to the particular observed trend pattern, causing the mid-latitude trends to be more anomalous than the tropical trends in comparison to the trend distribution derived from the models.

Representation of natural variability in models

Above, we argued that natural variability likely influenced the observed ozone trends, and that might partly explain that trends over the observed period disagree with the trends in model simulations. However, how large this disagreement is, depends on the underlying trend distribution derived from the models. For example, if the influence of natural variability is underestimated in the models, the trend distribution is too narrow.

The QBO is represented differently in the individual CCMs: some models generate a QBO internally, some models nudge winds towards a given QBO, and in some models, the representation of the QBO is missing entirely (for more details see Morgenstern et al., 2017). Thus, over the whole suite of models, this could cause an underestimation of ozone variability in the models and therewith consequently a too narrow trend distribution. Moreover, as the QBO signal is treated differently across the REF-C2 model setups, we can also expect that the inter-model differences in the QBO representation contributes to the spread in ozone trends over recent decades.

The analyzed 'free running' REF-C2 simulations either use an interactive ocean model, or use SSTs from other model simulations that are coupled to an ocean model. However, these coupled models still have biases with respect to the simulation of ENSO (Bellenger et al., 2014), thus ENSO-related variability in LS ozone might also be underrepresented.

Further, even if the QBO and ENSO are represented with the correct signal strength (e.g., by nudging the QBO, and prescribing observed SSTs), the induced circulation anomalies might not be captured entirely by the models. Hence, even if hindcast simulations with prescribed observed SSTs are used, it is not guaranteed that the effects of natural variability on ozone trends are fully captured. It would be interesting to compare the modeled trend distributions from the REF-C2 simulations to such hindcast simulations (REF-C1), however, in CCMI-1 the data of those hindcast simulations are only available until 2010. The

assessment of the representation of natural variability and its effects on ozone would require a more in-depth analysis, which we leave for future studies.

5 Conclusions

In the present study, we analyzed in detail lower stratospheric ozone trends for the recent period 1998–2018, and variations of this period, using a total of 31 simulations of different state-of-the-art chemistry climate models and compared them to the observation based dataset BASIC$_{SG}$. Moreover, we linked the ozone trends to stratospheric circulation trends and discussed the reasons for the differences in the LS ozone trends between models and observations. The main findings of our study are summarized in the following.

1) LS ozone trends over the period 1998–2018 vary strongly across different models and among different ensemble members of the same model. Therefore, internal variability strongly influences the LS ozone trends over this short time period. But even if this high variability is taken into account, none of the model simulations reproduces the pattern of observational ozone trends with negative values extending from the southern to the northern mid-latitudes. Thus the observed LS ozone trend pattern is a rather unlikely realization in state-of-the-art CCM simulations.

2) The models' LS ozone trend (given as the most likely values of the models' trend probability distribution) remains negative in the tropics and positive in the mid-latitudes for variations in the time period between 1995 and 2019. Although there is quite a large spread in the magnitude of model trends, the trends do not show a systematic change for the different periods. For observations, LS trends remain negative in both the tropics and the mid-latitudes for all these periods. In contrast to the models' consistent trend we find a systematic shift in the trend magnitude towards less negative mid-latitude trends for the start years 1999 and 2000, which is likely associated with natural variability.

3)In the tropics, the observed trends are a likely representation by the models' trend distribution. However in the mid-latitudes the observational trends represent an extreme value of the models' probability distribution.

4) Tropical LS ozone trends are linked to the GHG-driven increase of tropical upwelling, confirming previous studies. The robust positive mid-latitude LS ozone trends simulated in the models, on the other hand, are found to be driven by changes in ODS- rather than GHG-driven circulation changes. The effects of the latter average to about zero ozone trends between 30° and 50°N, because of competing processes of advection along the shallow versus deep circulation branch, and of two-way mixing.

5) In all models, negative trends in the tropics and positive trends in the mid-latitudes emerge for periods extending into the future (2040), but the models differ in the timing by which trends stabilise. If ozone variability and forced trends would be realistically simulated in the models, we should expect positive mid-latitude ozone trends to emerge in the next 1-2 decades from observational records, too.

Finally we discussed the question as to whether the apparent discrepancy between model and observational trends is due to the misrepresentation of certain processes in the models (e.g., mixing strength, residual circulation strength) or due to inadequate representation of natural variability (ENSO/QBO). Or additionally, the observational trend could just be an extreme (but plausible) realization of the models' trend distribution. Another hypothesis that could emerge from our results is that the discrepancy of mid-latitude ozone trends might stem from an overestimation of ODS-induced ozone recovery in the recent decades in models compared to observations. This effect would be consistent with the weaker upper-stratospheric ozone trends

in the observations compared to models. However, this hypothesis needs further investigation, as does the role of different transport processes for LS ozone trends.

Author contributions. SD performed the data analysis and produced the figures. SD, RE and HG made substantial contributions to conception and design, analysis and interpretation of the data. WTB provided the observational ozone trends and contributed to the interpretation of the results. Moreover all authors participated in drafting the article.

DATA: All data of CCMVal-2 and CCMI-1 used in this study can be obtained through the British Atmospheric Data Centre (BADC) archive (ftp://ftp.ceda.ac.uk). CESM1-WACCM data have been downloaded from http://www.earthsystemgrid.org. For instructions for access to both archives see http://blogs.reading.ac. uk/ccmi/badc-data-access

The authors declare that they have no conflict of interest.

Acknowledgements. This study was funded by the Helmholtz Association under grant VH-NG-1014 (Helmholtz-Hochschul- Nachwuchsforschergruppe MACClim). We acknowledge the modeling groups for making their simulations available for this analysis, the joint WCRP SPARC/IGAC Chemistry-Climate Model Initiative (CCMI) for organizing and coordinating the model data analysis activity, and the British Atmospheric Data Centre (BADC) for collecting and archiving the CCMI model output. We thank Marta Abalos for providing us an additional (forth) ensemble member of WACCM. Moreover, we want to acknowledge that the EMAC simulations were done within the project ESCiMo (Earth System Chemistry integrated Modelling), a national (German) contribution to the ChemistryClimate Model Initiative, and have been performed at the German Climate Computing Centre DKRZ through support from the Bundesministerium fuer Bildung und Forschung (BMBF). W.T.B. was funded by the SNSF project 200020_182239 (POLE)". RE acknowledges support by GA CR under grant nos. 16-01562J and 18-01625S. BASIC$_{SG}$ for 1985-2019 will be available for download from https://data.mendeley.com/datasets/2mgx2xzzpk/4 following review of this paper. GOZCARDS ozone data contributions from L. Froidevaux, R. Wang, J. Anderson, and R. A. Fuller at the Jet Propulsion Laboratory are gratefully acknowledged. Last, we also thank the reviewers for their constructive comments on our manuscript.

References

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.

Abalos, M., Polvani, L., Calvo, N., Kinnison, D., Ploeger, F., Randel, W., and Solomon, S.: New insights on the impact of ozone-depleting substances on the Brewer-Dobson circulation, Journal of Geophysical Research: Atmospheres, 124, 2435–2451, 2019.

Akiyoshi, H., Nakamura, T., Miyasaka, T., Shiotani, M., and Suzuki, M.: A nudged chemistry-climate model simulation of chemical constituent distribution at northern high-latitude stratosphere observed by SMILES and MLS during the 2009/2010 stratospheric sudden warming, Journal of Geophysical Research: Atmospheres, 121, 1361–1380, 2016.

Ball, W. T., Alsing, J., Mortlock, D. J., Rozanov, E. V., Tummon, F., and Haigh, J. D.: Reconciling differences in stratospheric ozone composites, Atmospheric Chemistry and Physics, 17, 12 269–12 302, 2017.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., et al.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmospheric Chemistry and Physics, 18, 1379–1394, 2018.

Ball, W. T., Alsing, J., Staehelin, J., Davis, S. M., Froidevaux, L., and Peter, T.: Stratospheric ozone trends for 1985–2018: sensitivity to recent large variability, Atmospheric Chemistry and Physics Discussions, 2019, 1–27, doi:10.5194/acp-2019-243, https://www.atmos-chem-phys-discuss.net/acp-2019-243/, 2019.

Ball, W. T., Chiodo, G., Abalos, M., Alsing, J., and Stenke, A.: Inconsistencies between chemistry–climate models and observed lower stratospheric ozone trends since 1998, Atmospheric chemistry and physics, 20, 9737–9752, 2020.

Bednarz, E. M., Maycock, A. C., Abraham, N. L., Braesicke, P., Dessens, O., and Pyle, J. A.: Future Arctic ozone recovery: the importance of chemistry and dynamics, Atmospheric Chemistry and Physics, 16, 12 159–12 176, 2016.

Bekki, S., Rap, A., Poulain, V., Dhomse, S., Marchand, M., Lefèvre, F., Forster, P. M., Szopa, S., and Chipperfield, M. P.: Climate impact of stratospheric ozone recovery, Geophysical Research Letters, 40, 2796–2800, 2013.

Bellenger, H., Guilyardi, É., Leloup, J., Lengaigne, M., and Vialard, J.: ENSO representation in climate models: From CMIP3 to CMIP5, Climate Dynamics, 42, 1999–2018, 2014.

Birner, T. and Bönisch, H.: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere, Atmospheric Chemistry and Physics, 11, 817–827, 2011.

Butchart, N.: The Brewer-Dobson circulation, Reviews of Geophysics, 52, 157–184, doi:10.1002/2013RG000448, http://dx.doi.org/10.1002/2013RG000448, 2014.

Calvo, N., Garcia, R., Randel, W., and Marsh, D.: Dynamical mechanism for the increase in tropical upwelling in the lowermost tropical stratosphere during warm ENSO events, Journal of the Atmospheric Sciences, 67, 2331–2340, 2010.

Chipperfield, M. P., Bekki, S., Dhomse, S., Harris, N. R., Hassler, B., Hossaini, R., Steinbrecht, W., Thiéblemont, R., and Weber, M.: Detecting recovery of the stratospheric ozone layer, Nature, 549, 211–218, 2017.

Chipperfield, M. P., Dhomse, S., Hossaini, R., Feng, W., Santee, M. L., Weber, M., Burrows, J. P., Wild, J. D., Loyola, D., and Coldewey-Egbers, M.: On the Cause of Recent Variations in Lower Stratospheric Ozone, Geophysical Research Letters, 2018.

Chrysanthou, A., Maycock, A. C., Chipperfield, M. P., Dhomse, S., Garny, H., Kinnison, D., Akiyoshi, H., Deushi, M., Garcia, R. R., Jöckel, P., et al.: The effect of atmospheric nudging on the stratospheric residual circulation in chemistry–climate models, Atmospheric Chemistry and Physics, 19, 11 559–11 586, 2019.

Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H., Fujiwara, M., and Damadeo, R.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: A long-term database for climate studies, Earth system science data, 8, 461, 2016.

Deushi, M. and Shibata, K.: Development of a Meteorological Research Institute chemistry-climate model version 2 for the study of tropospheric and stratospheric chemistry, Pap. Meteorol. Geophys., 62, 1–46, 2011.

Dhomse, S. S., Kinnison, D., Chipperfield, M. P., Salawitch, R. J., Cionni, I., Hegglin, M. I., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bednarz, E. M., et al.: Estimates of ozone return dates from Chemistry-Climate Model Initiative simulations, Atmospheric Chemistry and Physics, 18, 8409–8438, 2018.

Dietmüller, S., Ponater, M., and Sausen, R.: Interactive ozone induces a negative feedback in CO2-driven climate change simulations, Journal of Geophysical Research: Atmospheres, 119, 1796–1805, 2014.

Dietmüller, S., Eichinger, R., Garny, H., Birner, T., Boenisch, H., Pitari, G., Mancini, E., Visioni, D., Stenke, A., Revell, L., et al.: Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models, Atmospheric Chemistry and Physics, 18, 6699–6720, 2018.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., et al.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, 2013.

Eichinger, R., Dietmüller, S., Garny, H., Šácha, P., Birner, T., Bönisch, H., Pitari, G., Visioni, D., Stenke, A., Rozanov, E., et al.: The influence of mixing on the stratospheric age of air changes in the 21st century, Atmospheric Chemistry and Physics, 19, 921–940, 2019.

Froidevaux, L., Anderson, J., Wang, H.-J., Fuller, R., Schwartz, M., Santee, M., Livesey, N., Pumphrey, H., Bernath, P., Russell III, J., et al.: Global OZone Chemistry And Related Datasets for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H 2 O, and O 3., Atmospheric Chemistry & Physics Discussions, 15, 2015.

Garcia, R. R., Smith, A. K., Kinnison, D. E., Cámara, Á. d. l., and Murphy, D. J.: Modification of the Gravity Wave Parameterization in the Whole Atmosphere Community Climate Model: Motivation and Results, Journal of the Atmospheric Sciences, 74, 275–291, 2017.

Garny, H., Birner, T., Bönisch, H., and Bunzel, F.: The effects of mixing on age of air, Journal of Geophysical Research: Atmospheres, 119, 7015–7034, 2014.

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, 2011.

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

Hardiman, S. C., Butchart, N., O'Connor, F. M., and Rumbold, S. T.: The Met Office HadGEM3-ES chemistry–climate model: evaluation of stratospheric dynamics and its impact on ozone, Geoscientific Model Development, 10, 1209–1232, 2017.

Harris, N. R., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., Steinbrecht, W., Anderson, J., Bhartia, P. K., Boone, C., et al.: Past changes in the vertical distribution of ozone–Part 3: Analysis and interpretation of trends, Atmospheric Chemistry and Physics, 15, 9965–9982, 2015.

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

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., and Elliott, S.: CICE: the Los Alamos Sea Ice Model Documentation and Software User?s Manual Version 4.1 LA-CC-06-012, T-3 Fluid Dynamics Group, Los Alamos National Laboratory, 675, 2010.

Imai, K., Manago, N., Mitsuda, C., Naito, Y., Nishimoto, E., Sakazaki, T., Fujiwara, M., Froidevaux, L., von Clarmann, T., Stiller, G. P., et al.: Validation of ozone data from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES), Journal of Geophysical Research: Atmospheres, 118, 5750–5769, 2013.

Jensen, M. P., Mather, J. H., and Ackerman, T. P.: Observations of the 1997–98 warm ENSO event at the Manus Island ARM site, Geophysical research letters, 25, 4517–4520, 1998.

Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), Geosci. Model Dev., **3**, 717–752, doi:10.5194/gmd-3-717-2010, 2010.

Jöckel, P., Holger, T. O., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C.A, M., Brinkop, S., Cai, D., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neumaier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth System Chemistry Integrated Modelling (ESCiMO) with the Modular Earth Submodel System (MESSy, version 2.51), Geosci. Model Dev., **9**, 1153–1200, doi:doi:10.5194/gmd-9-1153-2016, 2016.

Jonsson, A., De Grandpre, J., Fomichev, V., McConnell, J., and Beagley, S.: Doubled CO2-induced cooling in the middle atmosphere: Photochemical analysis of the ozone radiative feedback, Journal of Geophysical Research: Atmospheres, 109, 2004.

Kirner, O., Ruhnke, R., and Sinnhuber, B.-M.: Chemistry–Climate Interactions of Stratospheric and Mesospheric Ozone in EMAC Long-Term Simulations with Different Boundary Conditions for CO2, CH4, N2O, and ODS, Atmosphere-Ocean, 53, 140–152, 2015.

Kusuma, L., Lubis, S. W., and Setiawan, S.: Unprecedented Quasi-Biennial Oscillation (QBO) disruption in 2015-2016: Implications for tropical waves and ozone, in: IOP Conference Series: Earth and Environmental Science, vol. 284, p. 012016, IOP Publishing, 2019.

Laine, M., Latva-Pukkila, N., and Kyrölä, E.: Analysing time-varying trends in stratospheric ozone time series using the state space approach, Atmospheric Chemistry and Physics, 14, 9707–9725, 2014.

Madec, G. et al.: NEMO ocean engine, 2015.

Marchand, M., Keckhut, P., Lefebvre, S., Claud, C., Cugnet, D., Hauchecorne, A., Lefèvre, F., Lefebvre, M.-P., Jumelet, J., Lott, F., et al.: Dynamical amplification of the stratospheric solar response simulated with the Chemistry-Climate model LMDz-Reprobus, Journal of Atmospheric and Solar-Terrestrial Physics, 75, 147–160, 2012.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate change from 1850 to 2005 simulated in CESM1 (WACCM), Journal of climate, 26, 7372–7391, 2013.

Marsh, D. R., Lamarque, J.-F., Conley, A. J., and Polvani, L. M.: Stratospheric ozone chemistry feedbacks are not critical for the determination of climate sensitivity in CESM1 (WACCM), Geophysical Research Letters, 43, 3928–3934, 2016.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M., Lamarque, J.-F., 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, 2011.

Meul, S., Langematz, U., Oberländer, S., Garny, H., and Jöckel, P.: Chemical contribution to future tropical ozone change in the lower stratosphere, Atmospheric Chemistry and Physics, 14, 2959–2971, 2014.

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, 2012.

Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, Geoscientific Model Development, 8, 1339–1356, 2015.

Morgenstern, O., Braesicke, P., O'Connor, F., Bushell, A., Johnson, C., Osprey, S., and Pyle, J.: Evaluation of the new UKCA climate-composition model–Part 1: The stratosphere, Geoscientific Model Development, 2, 43–57, 2009.

Morgenstern, O., Zeng, G., Luke Abraham, N., Telford, P. J., Braesicke, P., Pyle, J. A., Hardiman, S. C., O'Connor, F. M., and Johnson, C. E.: Impacts of climate change, ozone recovery, and increasing methane on surface ozone and the tropospheric oxidizing capacity, Journal of Geophysical Research: Atmospheres, 118, 1028–1041, 2013.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abrams, N. L., Akkyoshi, H., Archibald, A. T., Bekki, S., Butchart, N., Chipperfield, M. P., Deushi, M., Dhomse, S. S., Garcia, R. R., Hardiman, S. C., Horowitz, L. W., Jöckel, P., Josse, B., Kinnison, D., Lin, M. Y., Mancini, E., Manyin, M. E., Marchand, M., Marecal, V., Michou, M., Pitari, L. D. O. N. G. ans Plummer, D. A., Revell, L. E., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T. Y., Tilmes, S., Yamashita, Y., Yoshida, K., and Zeng, G.: Review of the global models used within the Chemistry-Climate Model Initiative (CCMI), Geosci. Model Dev., 10, 639–671, doi:10.5194/gmd-10-639-2017, 2017.

Newman, P., Daniel, J., Waugh, D., and Nash, E.: A new formulation of equivalent effective stratospheric chlorine (EESC), 2007.

O'Connor, F., Johnson, C., Morgenstern, O., Abraham, N., Braesicke, P., Dalvi, M., Folberth, G., Sanderson, M., Telford, P., Voulgarakis, A., et al.: Evaluation of the new UKCA climate-composition model–Part 2: The Troposphere, Geoscientific Model Development, 7, 41–91, 2014.

Olsen, M. A., Manney, G. L., and Liu, J.: The ENSO and QBO Impact on Ozone Variability and Stratosphere-Troposphere Exchange Relative to the Subtropical Jets, Journal of Geophysical Research: Atmospheres, 124, 7379–7392, 2019.

Oman, L., Plummer, D., Waugh, D., Austin, J., Scinocca, J., Douglass, A., Salawitch, R., Canty, T., Akiyoshi, H., Bekki, S., et al.: Multimodel assessment of the factors driving stratospheric ozone evolution over the 21st century, Journal of Geophysical Research: Atmospheres, 115, 2010.

Oman, L., Ziemke, J., Douglass, A., Waugh, D., Lang, C., Rodriguez, J., and Nielsen, J.: The response of tropical tropospheric ozone to ENSO, Geophysical Research Letters, 38, 2011.

Oman, L. D., Douglass, A. R., Ziemke, J. R., Rodriguez, J. M., Waugh, D. W., and Nielsen, J. E.: The ozone response to ENSO in Aura satellite measurements and a chemistry-climate simulation, Journal of Geophysical Research: Atmospheres, 118, 965–976, 2013.

Orbe, C., Wargan, K., Pawson, S., and Oman, L. D.: Mechanisms Linked to Recent Ozone Decreases in the Northern Hemisphere Lower Stratosphere, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 631, 2020.

Petropavlovskikh, I., Godin-Beekmann, S., Hubert, D., Damadeo, R., Hassler, B., and Sofieva, V.: SPARC/IO3C/GAW report on Long-term Ozone Trends and Uncertainties in the Stratosphere, 2019.

Pitari, G., Aquila, V., Kravitz, B., Robock, A., Watanabe, S., Cionni, I., Luca, N. D., Genova, G. D., Mancini, E., and Tilmes, S.: Stratospheric ozone response to sulfate geoengineering: Results from the Geoengineering Model Intercomparison Project (GeoMIP), Journal of Geophysical Research: Atmospheres, 119, 2629–2653, 2014.

Polvani, L. M., Wang, L., Abalos, M., Butchart, N., Chipperfield, M., Dameris, M., Deushi, M., Dhomse, S., Jöckel, P., Kinnison, D., et al.: Large Impacts, Past and Future, of Ozone-Depleting Substances on Brewer-Dobson Circulation Trends: A Multimodel Assessment, Journal of Geophysical Research: Atmospheres, 124, 6669–6680, 2019.

Portmann, R. W. and Solomon, S.: Indirect radiative forcing of the ozone layer during the 21st century, Geophysical research letters, 34, 2007.

Randel, W. J. and Wu, F.: A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data, Journal of Geophysical Research: Atmospheres, 112, 2007.

Randel, W. J., Garcia, R., and Wu, F.: Dynamical balances and tropical stratospheric upwelling, Journal of the atmospheric sciences, 65, 3584–3595, 2008.

Randel, W. J., Garcia, R. R., Calvo, N., and Marsh, D.: ENSO influence on zonal mean temperature and ozone in the tropical lower stratosphere, Geophysical Research Letters, 36, 2009.

Ravishankara, A., Daniel, J. S., and Portmann, R. W.: Nitrous oxide (N2O): the dominant ozone-depleting substance emitted in the 21st century, science, 326, 123–125, 2009.

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, 115, 2010.

Revell, L., Tummon, F., Stenke, A., Sukhodolov, T., Coulon, A., Rozanov, E., Garny, H., Grewe, V., and Peter, T.: Drivers of the tropospheric ozone budget throughout the 21st century under the medium-high climate scenario RCP 6.0, Atmospheric Chemistry and Physics, 15, 5887–5902, 2015.

Revell, L. E., Bodeker, G. E., Smale, D., Lehmann, R., Huck, P. E., Williamson, B. E., Rozanov, E., and Struthers, H.: The effectiveness of N2O in depleting stratospheric ozone, Geophysical research letters, 39, 2012.

Scinocca, J., McFarlane, N., Lazare, M., Li, J., and Plummer, D.: Technical Note: The CCCma third generation AGCM and its extension into the middle atmosphere, Atmospheric Chemistry and Physics, 8, 2008.

Solomon, S., Kinnison, D., Bandoro, J., and Garcia, R.: Simulation of polar ozone depletion: An update, Journal of Geophysical Research: Atmospheres, 120, 7958–7974, 2015.

SPARC CCMVal: SPARC Report No 5 (2010) Chemistry-Climate Model Validation, WCRP-132, WMO/TD-No. 1526, 2010.

Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., et al.: An update on ozone profile trends for the period 2000 to 2016, Atmospheric chemistry and physics, 17, 10 675–10 690, 2017.

Stenke, A., Schraner, M., Rozanov, E., Egorova, T., Luo, B., and Peter, T.: The SOCOL version 3.0 chemistry–climate model: description, evaluation, and implications from an advanced transport algorithm, Geoscientific Model Development, 6, 1407–1427, 2013.

Stone, K., Morgenstern, O., Karoly, D., Klekociuk, A., French, W., Abraham, N., and Schofield, R.: Evaluation of the Australian Community Climate and Earth-System Simulator Chemistry-Climate Model., Atmospheric Chemistry & Physics Discussions, 15, 2015.

Stone, K. A., Solomon, S., and Kinnison, D. E.: On The Identification of Ozone Recovery, Geophysical Research Letters, 2018.

Sudo, K. and Akimoto, H.: Global source attribution of tropospheric ozone: Long-range transport from various source regions, Journal of Geophysical Research: Atmospheres, 112, 2007.

Szopa, S., Balkanski, Y., Schulz, M., Bekki, S., Cugnet, D., Fortems-Cheiney, A., Turquety, S., Cozic, A., Déandreis, C., Hauglustaine, D., et al.: Aerosol and ozone changes as forcing for climate evolution between 1850 and 2100, Climate dynamics, 40, 2223–2250, 2013.

Tian, W. and Chipperfield, M. P.: A new coupled chemistry–climate model for the stratosphere: The importance of coupling for future O3-climate predictions, Quarterly Journal of the Royal Meteorological Society, 131, 281–303, 2005.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., Smith, A. K., Neely, R. R., Conley, A., Vitt, F., et al.: Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), Geoscientific Model Development Discussions, 2016.

Walters, D., Williams, K., Boutle, I., Bushell, A., Edwards, J., Field, P., Lock, A., Morcrette, C., Stratton, R., Wilkinson, J., et al.: The Met Office Unified Model global atmosphere 4.0 and JULES global land 4.0 configurations, Geoscientific Model Development, 7, 361–386, 2014.

Wargan, K., Orbe, C., Pawson, S., Ziemke, J. R., Oman, L. D., Olsen, M. A., Coy, L., and Knowland, K. E.: Recent decline in extratropical lower stratospheric ozone attributed to circulation changes, Geophysical Research Letters, 2018.

Waugh, D. and Hall, T.: Age of stratospheric air: Theory, observations, and models, Reviews of Geophysics, 40, doi:10.1029/2000RG000101, 2002.

Weber, M., Coldewey-Egbers, M., Fioletov, V. E., Frith, S. M., Wild, J. D., Burrows, J. P., Long, C. S., and Loyola, D.: Total ozone trends from 1979 to 2016 derived from five merged observational datasets–the emergence into ozone recovery, Atmospheric Chemistry and Physics, 18, 2097–2117, 2018.

Winterstein, F., Tanalski, F., Jöckel, P., Dameris, M., and Ponater, M.: Implication of strongly increased atmospheric methane concentrations for chemistry–climate connections, Atmospheric Chemistry and Physics, 19, 7151–7163, 2019.

WMO: Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring ProjectReport, World Meteorological Organization, pp. Report No. 58, 588 pp, 2018.

Yukimoto, S., Yoshimura, H., Hosaka, M., Sakami, T., Tsujino, H., Hirabara, M., Tanaka, T., Deushi, M., Obata, A., Nakano, H., et al.: Development of the Meteorological Research Institute-Earth System Model (MRIESM1)?model description, Tech. rep., Technical reports of the Meteorological Research Institute, 2011.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., et al.: A new global climate model of the Meteorological Research Institute: MRI-CGCM3 model description and basic performance, Journal of the Meteorological Society of Japan. Ser. II, 90, 23–64, 2012.

Supplement of "Analysis of recent lower stratospheric ozone trends in chemistry climate models"

Dietmueller et al.

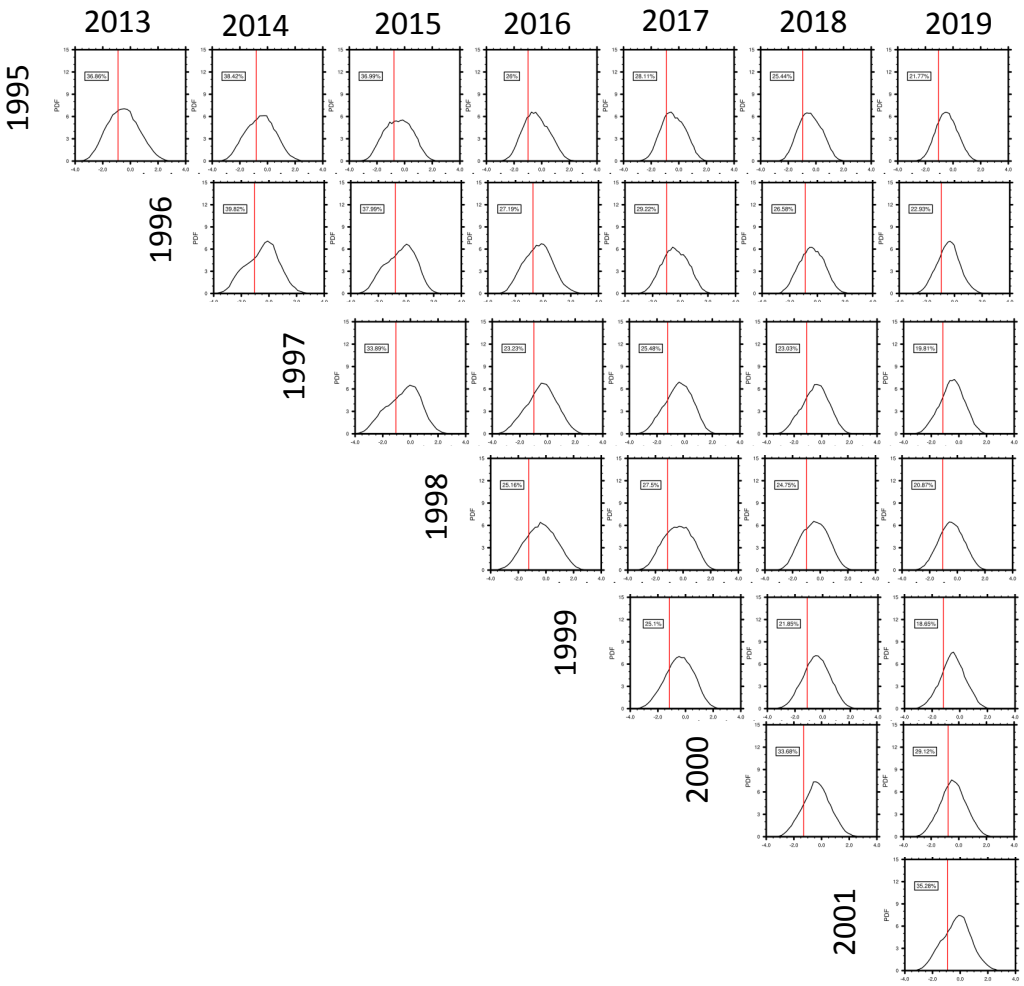

Figure S1. Probability distribution function (PDF) of the models' LS tropical ozone trend (20°S-20°N, 30-100 hPa) as function of different periods. In all panels the x-coordinate denotes the different end years (2013-2019) and the y-coordinate the different start years (1995-2001). The red line indicates the respective observational ozone trend value. Moreover the probability (in %) of the observational trend lying within the models' distribution is given within each panel.

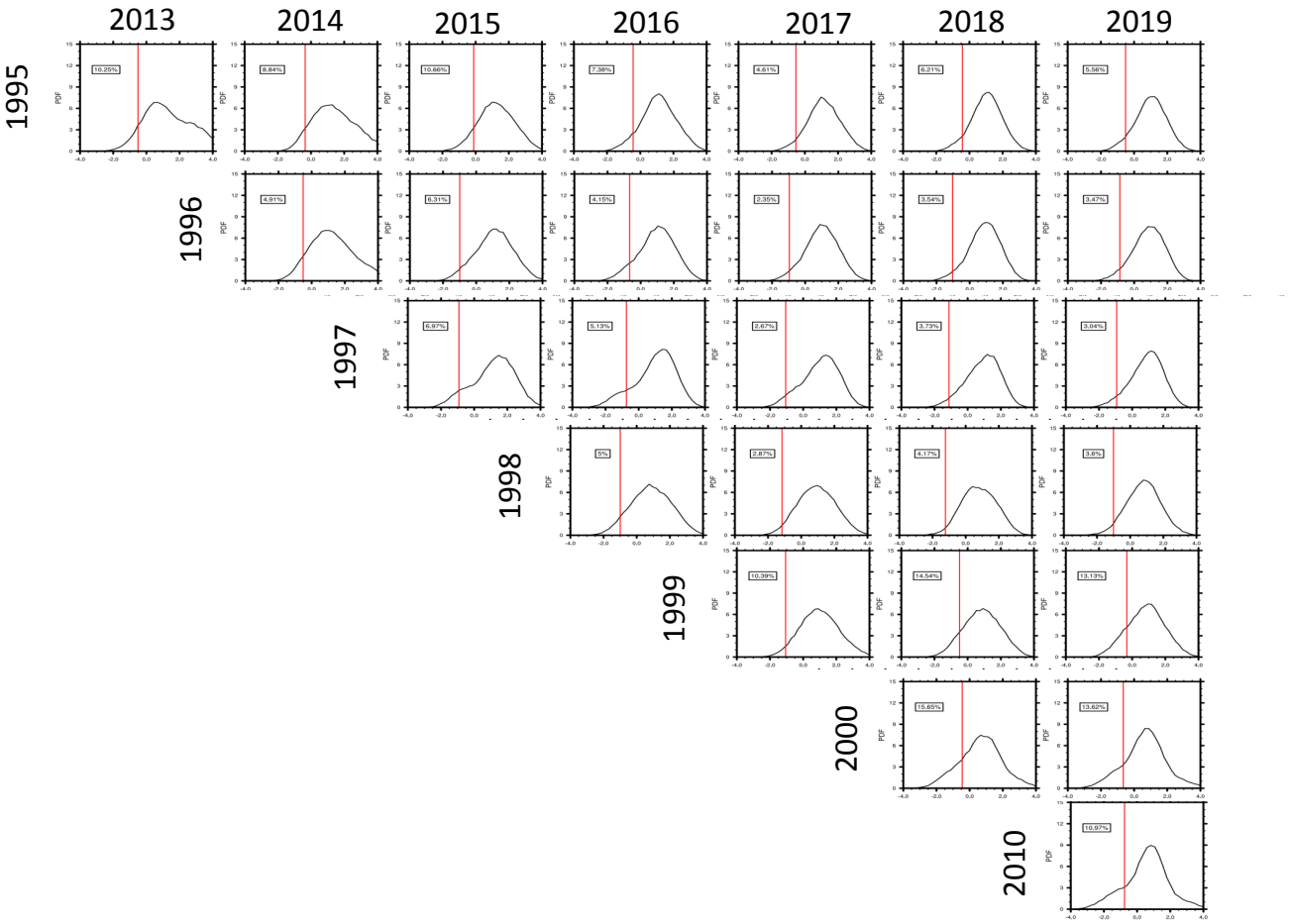

Figure S2. Same as Fig. S1, but PDFs for the models' LS northern mid-latiudes ozone trend (30°N-50°N, 30-150 hPa).

Table S1. MMM ozone trends and their 1-sigma standard deviations [DU/dec] in the lower stratosphere (LS, 30-100/150 hPa) and in the upper stratosphere (US, 1-10 hPa) for the tropical and northern mid-latitudes. Trends are given for the REF-C2 simulation, as well as for the sensitivity simulations with fixed GHG and fixed ODS concentrations (fGHG and fODS). Trends are calculated over the period 1998-2018 and 1998-2040.

MMM trop. O3 trend	**LS**	**US**
REF-C2 (1998-2018)	-0.71±0.40	0.68±0.41
REF-C2 (1998-2040)	-0.55±0.38	0.71±0.25
fGHG (1998-2018)	-0.35±0.41	0.64±0.26
fGHG (1998-2040)	0.03±0.06	0.61±0.07
fODS (1998-2018)	-1.09±0.59	0.51±0.27
fODS (1998-2040)	-0.85±0.48	0.44±0.12
MMM NH mid-lat. O3 trend	**LS**	**US**
REF-C2 (1998-2018)	0.76±1.12	0.64±0.34
REF-C2 (1998-2040)	0.79±0.60	0.75±0.20
fGHG (1998-2018)	0.62±1.11	0.43±0.25
fGHG (1998-2040)	0.68±0.22	0.58±0.16
fODS (1998-2018)	-0.18±0.84	0.41±0.39
fODS (1998-2040)	0.04±0.40	0.43±0.15

Analysis of recent lower stratospheric ozone trends in chemistry climate models

Simone Dietmüller[1], Hella Garny[1,2], Roland Eichinger[2,1,3], and William T. Ball[4,5,6]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Ludwig Maximilians Universität, Faculty of Physics, Institute for Meteorology, Munich, Germany
[3]Charles University, Department of Atmospheric Physics, Faculty of Mathematics and Physics, Prague, Czech Republic
[4]Institute for Atmospheric and Climate Science, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland
[5]Physikalisch-Meteorologisches Observatorium Davos World Radiation Centre, Dorfstrasse 33, 7260 Davos Dorf, Switzerland
[6]Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, TU Delft, Stevinweg 1, 2628 CN Delft, the Netherlands

Correspondence to: S. Dietmüller (simone.dietmueller@dlr.de)

Abstract. Recent observations show a significant decrease of lower stratospheric (LS) ozone concentrations in tropical and mid-latitude regions since 1998. By analyzing 31 chemistry climate model (CCM) simulations performed for the Chemistry Climate Model Initiative (CCMI, **?**), we find a large spread in the 1998–2018 trend patterns between different CCMs and between different realizations performed with the same CCM. The latter, in particular, indicates that natural variability strongly influences LS ozone trends. However none of the model simulations reproduces the observed ozone trend structure of coherent negative trends in the LS. In contrast to the observations, most models show a ~~dipole trend pattern in the LS~~ LS trend pattern with negative trends in the tropics (20°S-20°N) and positive trends in the northern mid-latitudes (30°-50°N) or vice versa. To investigate the influence of natural variability on ~~the~~ historical LS ozone trends we analyze the sensitivity of observational trends and the models' trend probability distributions for ~~slightly varied post-ODS (ozone depleting substances) periods .~~ varying periods with start dates between 1995 to 2001 and end dates between 2013 to 2019. Generally, modeled and observed LS trends remain robust for ~~different post-ODS~~ these different periods, however observational data show a ~~systematic~~ change towards weaker mid-latitude trends for certain periods, likely forced by natural variability~~for certain periods~~. Moreover we ~~can~~ show that in the tropics the observed trends agree ~~quite~~ well with the models' trend distribution, whereas in the mid-latitudes the observational trend is ~~a rather~~ typically an extreme value of the models' distribution. We further investigate the LS ozone trends for extended periods reaching into the future and find that all models develop a ~~dipole trend pattern in the future, i.e. in almost all models the~~ positive ozone trend at mid-latitudes and the trends converge to constant values ~~for the entire period~~ by the period that spans 1998–2060. ~~An investigation of interannual ozone variability also reveals a clear dipole pattern in ozonevariability in all CCMs and in observational data, however it is more pronounced in the models. Thus, although the LS~~ Intermodel correlations between ozone trends and measures of transport circulation trends confirm the dominant role of greenhouse gas (GHG)-driven tropical upwelling enhancement on the tropical LS ozone decrease. Mid-latitude ozone, on the other hand, appears to be influenced by multiple competing factors: an enhancement in the shallow branch decreases ozone, while an enhancement in the deep branch increases ozone and, furthermore, mixing plays a role here too. Sensitivity simulations

with fixed forcing of GHGs or ozone depleting substances (ODSs) reveal that the GHG-driven increase in circulation strength do not lead to a net trend in LS mid-latitude column ozone. Rather, the positive ozone trends simulated consistently in the models in this region emerge from the decline of ODSs, i.e. the ozone ~~variability pattern is similar, the probability of overall negative LS ozone trends simultaneously in the tropics and mid-latitudes is higher in observations. To access the dynamical influence on LS ozone trends, the models' tropical upwelling trends are correlated against their LS ozone trends. For the period 1998–2018 tropical ozone trends are negatively correlated (-0.83) and~~ recovery. Therefore, we hypothesize that next to the influence of natural variability, the disagreement of modeled and observed LS mid-latitude ~~trends are positively correlated (+0.49). However, the correlation in the mid-latitudes is rather weak and not robust for slightly varied time periods, which indicates that other processes like two-way mixing play an important role here, too~~ozone trends could indicate a mismatch in the relative role of the response of ozone to ODS versus GHG-forcing in the models.

1 Introduction

Stratospheric ozone is essential for protecting the Earth's surface from ultra violet radiation, which is harmful for plants, animals and humans. Human-made ozone depleting substance (ODS) emissions significantly reduced ozone concentrations for some decades after 1960. After controlling the use of ODSs by the 1987 Montreal protocol and later adjustments, however, ~~ODSs~~ ODS concentrations started to decline in the mid-to-late 1990s (e.g. **??**). As a consequence, total stratospheric ozone is expected to recover in the future. **?** have analyzed the recovery of stratospheric ozone mixing ratios of the CCMI-1 (Chemistry Climate Model Intercomparison project part 1) climate projection simulations. They found that the ozone layer is simulated to return to ~~pre-ODS~~ a pre-1980 ODS level between 2030 and 2060, depending on the region. However, they discovered a large spread among the individual models, which shows that there are many uncertainties in these projections. The evolution of stratospheric ozone in the 21st century does not only result from a decrease in ~~ODSs~~ ODS concentrations but also from an interplay between changes in both the atmospheric composition and the circulation (World Meteorological Organization (WMO) 2014). Increasing anthropogenic greenhouse gas (GHG) emissions (CO_2, CH_4, N_2O) ~~lead~~ leads to enhanced tropical upwelling and thereby to an acceleration of tracer transport along the stratospheric overturning circulation (e.g. **??**). On the other hand, increasing GHGs also ~~slow~~ slows down ozone depletion through GHG-induced stratospheric cooling (e.g. **?????**) and emissions of CH_4 and N_2O additionally impact ozone ~~by~~ through chemical processes (e.g. **????**).

In the recent years, a number of studies have analyzed observational records to identify ozone trends in the stratosphere (e.g. **???**). These studies consistently report an ozone recovery in the upper stratosphere after the turnaround of the ODS concentrations ~~in the year 1998 (start of the post-ODS period).~~ around the year 1998. In the lower stratosphere (LS), however, most observed ozone trends are ~~statistically not significant for that~~ not statistically significant for such a relatively short period due to large internal variability and instrumental difficulties (e.g. **?**). Subsequently, **?** analyzed LS ozone trends from satellite data between 1998 and 2016 in detail making use of a dynamical (multiple) linear regression analysis. ~~This way, they~~ They identified a statistically significant decline of LS ozone between 60°S and 60°N in that period, ~~by~~ of approximately 2 DU in the LS below 24 km of altitude. The implication was that the stratospheric ozone column was continuing to decline, because

the LS ozone reduction more than offsets the positive trend in the upper stratosphere. Shortly afterwards **?** studied ozone trends in the reanalysis products MERRA-2 and GEOS-RPIT. In the tropics they detected a positive ozone trend in a 5 km layer above the tropopause and a negative trend at 7-15 km above the tropopause. Nevertheless, in the northern and southern mid-latitude LS they detected a negative ozone trend. As such, there are some similarities to the findings of **?**, but there are also ~~major~~ quantitative differences, for example the positive trend in the 5 km layer or a missing overall ~~negative trend in the LS~~statistically significant decrease in the column integrated ozone. **?** suggested that the negative mid-latitude trend might be explained by enhanced isentropic transport between the tropical and mid-latitude LS. However, the recent study of **?** explicitly demonstrated that in the NH this mid-latitude ozone decrease is primarily associated with large scale advection. Furthermore, they showed that the observed changes in advection and in ozone are well within the range of model variability (gauged from one CCM). By means of using a chemistry transport model (CTM) and extending the analysis period to the year 2017, **?** suggested that the negative LS ozone trends are only a result of large natural variability. They showed that there ~~is~~ was a strong positive ozone anomaly in 2017 which is driven by short term dynamical transport of ozone, and concluded that this points to large year-to-year variability rather than to an ongoing downward trend. However, an update of the data set which was used in **?** showed that the large interannual variability alone cannot explain the entire trend in **?** (see **?**): the larger year-to-year variability in the SH was implicated to ~~results~~ result from a non-linear interaction between the quasi-biennal-oscillation (QBO) and seasonal variability and despite this large variability the observed negative LS ozone trend remains. ~~To get more~~

To improve confidence in future projections of the ozone layer it is important to ~~know how well~~ evaluate the skill of chemistry climate models (CCMs) ~~simulate the observed post-ODS peak ozone trends~~ in simulating the observed ozone trends over recent decades. A direct comparison between the CCM multi-model-mean (MMM) values and observational data showed that the ozone trend profiles of modeled MMM data agree well with observations, except in the lowermost mid-latitude stratosphere ~~(??)~~(**??**). The most recent study of **?** investigated LS ozone trends of the 1998–2016 period in merged satellite data and compared them to the ozone trends in CCMs using the climate projection simulations of the CCMVal2 project. Similar to the observations, the CCMs showed a decline in LS ozone in the tropics, likely due to enhanced tropical upwelling, following from an increase in greenhouse gases (see e.g. **?**). In contrast to the observations, however, models do not show a decrease, but rather an increase in LS mid-latitude ozone. **?** argue that these discrepancies in the LS between models and observations can possibly be explained by differences in the horizontal two-way mixing between the tropics and mid-latitudes, though they did not provide explicit evidence from the models ~~, only observations (see also **?**). This~~ (see also **?**). The study suggested that the negative mid-latitude observational trend is caused by an intensification of two-way mixing (by analyzing effective diffusivity in reanalysis data). On the other hand enhanced downwelling of ozone-rich air to the mid-latitudes could consequently lead to a positive trend in the mid-latitudes. Apparently, the processes that determine mid-latitude LS ozone in models and observations are not understood so far.

In the present study, we seek to quantify whether the observed LS ozone trends lie within the suite of modeled trends. If yes, this would imply that the observed trend is just one realization of possible trends given within the large year-to-year variability. If not, this would imply that either models do not represent year-to-year variability correctly, or that there is a forced trend in the real world that is not adequately represented in the models. In contrast to the study of **?** we are using the simulation data of

a more recent inter-model comparison project (namely the Chemistry Climate Model Initiative, phase 1, CCMI-1) and analyze the ozone trends for a wider range ~~and for improved~~ of updated current state-of-the-art CCMs including all their ensemble simulations.

A brief description of the model simulations, of the observational data ~~set~~ sets and of the ~~used statistical methods~~ methods used is presented in Section ~~2.1~~2. In Section 3 we show our results~~: in section 3.1 we~~. We provide a detailed comparison of ozone trends over the years 1998–2018 in different CCM simulations and observations ~~, with a~~ (Section 3.1). Here we focus on LS ozone trends, and we investigate how natural variability influences these LS ozone trends (Section 3.2 and 3.3). ~~Moreover we~~ We link LS ozone trends with stratospheric ~~upwelling~~ transport trends (Section 3.4) ~~, and compare the year-to-year variability in observations and in all CCM simulations~~ and we investigate how ozone trends are forced by GHG and ODS emissions (Section ~~??~~3.5). A discussion of the reasons for the disagreement in the LS mid-latitude ozone ~~trend~~ trends between models and observations and the conclusions follow in Sections 4 and 5, respectively.

2 Data and Methods

2.1 Models and Simulations

In the present study, we analyze the model output from ~~overall~~ 18 state-of-the-art CCMs from the Chemistry Climate Model Initiative phase 1 (CCMI-1, **?**). Tab. 1 lists all these CCMs together with their references, ~~their underlying~~ the forcing that underlies the sea surface temperatures (SSTs), and the simulation type considered. A detailed overview of all models that participated in CCMI-1 can be found in **?**. We mainly evaluate the long term ~~free running~~ 'free running' simulations of CCMI-1 (REF-C2), as they span the time period ~~we are interested in (namely~~ 1998–2018~~)~~. We do not use ~~the the free running~~REF-C1 'free running' simulations of the recent past or the specified dynamics simulations (REF-C1SD), as they only span the period from 1998 to 2010. Moreover we want to point out that the specified dynamics simulations performed for CCMI do not represent stratospheric circulation better than the ~~free running~~ 'free running' simulations: **?** compared stratospheric residual circulation among specified dynamic (SD) simulations and found that the spread in these simulations is even larger than in REF-C2. Furthermore **?** showed poor agreement with the observed ozone trend for some selected SD simulations of CCMI. For the REF-C2 model simulations used in our study, all available ensemble members of the individual models are taken into account. The ensemble size of a certain simulation (if ensemble simulations were performed) is also given in Tab. 1 (brackets after simulations). Thus for the REF-C2 ~~simulation~~ simulations, 18 models performed a total of 31 realizations (~~6~~ six models performed multiple ensemble members simulations). The REF-C2 simulations include hindcast and forecast periods spanning 1960–2100. They are all ~~free running~~ 'free running' simulations, thus each model simulation has its own internal variability. Note that REF-C2 simulations use a variety of different SSTs and SICs (sea ice concentrations), either prescribed climate model SST fields from offline model simulations (of the same or of a different model) or they are coupled to an interactive ocean and sea ice module. Moreover the representation of the QBO is different across the CCMs, with models having an internally generated QBO (e.g. MRI, EMAC-L90), nudged QBO (e.g. NIES, WACCM, SOCOLv3, EMAC-L47, EMAC-L47-o) or no QBO (e.g. CMAM, LMDZ). REF-C2 reference simulations follow the WMO (2011) A1 scenario for ~~ozone-depleting~~

~~substances~~ ODSs and the RCP 6.0 scenario (**?**) for other greenhouse gases, tropospheric ozone precursors, and aerosol and aerosol precursor emissions. For anthropogenic emissions, the CCMI recommendation was to use MACCity (**?**) until 2000, followed by RCP 6.0 emissions. Besides the REF-C2 simulations we also consider the 11 sensitivity simulations with fixed greenhouse gases (fGHG) and with fixed ODSs (fODS) in our analysis. ~~This sensitivity scenario is~~ These sensitivity scenarios are both based on the REF-C2 simulation~~, but the GHGs~~. However in case of the fGHG simulations CO_2, CH_4, N_2O, and other non-ozone depleting GHGs are held at their 1960 value, ~~thus we can~~ and so we are able to study the impact due to ODS concentration changes only (i.e. in the absence of ~~GHG induced~~ GHG-induced climate change). In the case of the fODS simulations the ODS concentrations are fixed to the 1960 level throughout the simulation. All models providing ~~this~~ both of these sensitivity simulations are given in Tab 1.

Table 1. Overview of the CCMI simulations, analyzed for the present study. For the individual CCMs their reference(s), their SSTs and their available simulations (REF-C2~~and ,~~ fGHG, fODS) are given. The numbers in brackets behind the simulations indicate the number of realizations of each REF-C2~~and ,~~ fGHG or fODS simulation. Detailed information about the models' SSTs and the models' representation of the QBO are given in the supplement of **?**.

CCMI Model	Reference(s)	SSTs	Analyzed Simulation
CMAM	**?**	prescribed	REF-C2(1), fGHG(1), fODS(1)
	?		
CESM1-WACCM	**??**	interactive	REF-C2(4)*, fGHG(~~1~~)3), fODS(3)
	?		
EMAC-L90	**??**	prescribed	REF-C2(1) ~~, fGHG(1)~~
EMAC-L47	**??**	prescribed	REF-C2(1)
EMAC-L47-o	**??**	interactive	REF-C2(1)**
GEOSCCM	**??**	prescribed	REF-C2(1)
	??		
MRI	**?**	interactive	REF-C2(1)
	??		
SOCOLv3	**??**	prescribed	REF-C2(1)
NIWA-UKCA	**??**	interactive	REF-C2(5), fGHG(~~3~~)2), fODS(2)
	?		
ULAQ	**?**	prescribed	REF-C2(3), fGHG(1), fODS(1)
HadGEM	**??**	interactive	REF-C2(1)
	??		
	??		
UMUKCA	**??**	prescribed	REF-C2(2)
ACCESS-CCM	**??**	prescribed	REF-C2(3), fGHG(1), fODS(1)
	?		
NIES	**??**	prescribed	REF-C2(1), fGHG(1), fODS(1)
UMSLIMCAT	**?**	prescribed	REF-C2(1), fGHG(1), fODS(1)
CHASER	**?**	interactive	REF-C2(1), fGHG(1), fODS(1)
LMDz-REPROBUS	**??**	interactive	REF-C2(1)
	?		
CESM1-CAM4-Chem	**?**	interactive	REF-C2 (3)

* The fourth ensemble of WACCM (WACCM-4) was provided by M. Abalos; ** EMAC-L47 simulations are not ensembles, as one simulation is with prescribed SSTs and one with interactive ocean

2.2 Observational data

For observations, we make use of the BAyeSian Integrated and Consolidated (BASIC) ozone composite that merges SWOOSH (**?**) and GOZCARDS (**?**) through the BASIC method of **?**. The method was developed to account for artefacts in composite datasets that are a consequence of merging observations from different instruments that each have unique spatial and temporal observing characteristics. As a result, these artefacts can alias in regression analysis and bias, e.g., trend estimates (see examples in **?**). BASIC composites aim to account for and reduce artefacts using an empirically driven Bayesian inference methodology, but it relies on the availability of already developed ozone composites. Here, BASIC$_{SG}$ has been extended to the end of 2019 using the latest versions of GOZCARDS, v2.20, and SWOOSH, v2.6. As such BASIC$_{SG}$ covers 1985-2019 as monthly mean zonal means on a 10° latitude grid from 60°S–60°N and over a pressure range of 147–1 hPa (∼13–48 km). BASIC$_{SG}$ was presented in **?**, and a sensitivity analysis of trends was applied to it in **?**, with examples of data artefacts that it addresses in the accompanying appendix and supplementary materials, respectively.

To obtain an observationally constrained estimate of tropical upwelling and extratropical downwelling mass fluxes, we use ECMWF's fifth generation of atmospheric reanalysis data, ERA5 (**?**). The mass fluxes are calculated from 6-hourly data on the reduced set of pressure levels.

2.3 Statistical Methods

In some parts of our analysis, and to make a robust comparison between multiple models and a single 'real-world' realization, i.e. observations, we form probability distributions to estimate the combined probability of the ozone trends from all REF-C2 models. To do so, we calculate the linear trend and the associated uncertainty using a least squares method for every simulation. Then, to build the trend probability distribution of the models~~trend~~, first one of the 18 CCMI models is randomly selected, assuming that the models are randomly ~~uniform~~ uniformly distributed. In case the selected CCM ~~has ensemble members,~~ provided ensemble member simulations, in a second step one of these members is ~~then also randomly chosen~~randomly chosen, thus taking into account that ensemble members are treated differently than individual models. In the next step, the trend estimate ($t^{M_{i,k}}$) of the specific randomly selected CCMI model M_i with ensemble member k is calculated by randomly choosing ~~a~~ an ozone trend value from the trends associated and assumed normal distribution $\mathcal{N}$, which is based on the mean $\mu_{M_{i,k}}$ and standard deviation $\sigma_{M_{i,k}}$ of the simulations linear trend. Thus we can write the trend estimate of the selected model simulation as: $t^{M_{i,k}} = \mathcal{N}(\mu_{M_{i,k}}; \sigma_{M_{i,k}})$. In order to take into account the uncertainty of the single observational dataset (σ_{obs}), we also add to the calculated model trend estimate a random estimate of the observational noise by taking the observational standard deviation of the linear regression coefficient. We repeat the above described procedure 50 000 times. With that we have a large ~~amount~~ sample of model trends and can build up a robust probability density function (PDF) of the REF-C2 ozone trends. From these estimated PDFs we can then estimate the probability of a given trend relative to the models~~ in order to compare the model simulations~~. We derive a "probability of disagreement" between the observational and the modeled trend distribution by taking the central interval of the models' trend distribution with the observed trend value as threshold of this interval. To calculate this central interval we order the the 50 000 values from the REF-C2 trend distribution according to

their probability values and then sum up the ordered probability values until the value of the observed trend is reached. This probability value indicates our estimate of whether the observations agree with the models, i.e. high probability values indicate that a disagreement between models and observations is less likely due to chance.

2.4 Analysis Methods

We here provide a short description of our methodology to analyze transport processes, which follows the studies of **?** and **?**. Stratospheric mean AoA is defined as the mean residence time of an air parcel in the stratosphere (**??**). In the CCMs, the AoA tracer is implemented as an inert tracer with a mixing ratio that linearly increases over time as lower boundary condition. AoA is then calculated as the time lag between the local mixing ratio at a certain grid point and the current mixing ratio at a reference point.

The residual circulation transit time (RCTT) is the hypothetical age that air would have if it only followed the residual circulation, thus without processes such as eddy mixing or diffusion. RCTTs are calculated by backward trajectories on the basis of the Transformed Eulerian Mean (TEM) meridional and vertical velocities (referred to as residual velocities) with a standard fourth-order Runge-Kutta integration (**?**). The RCTT is then the time that these backward trajectories require to reach the tropopause from their respective starting point in the stratosphere. The RCTT differs from AoA because of resolved and unresolved mixing. In the stratosphere, this is due to the mixing of air between branches and the in-mixing of air from the mid-latitudes into the tropical pipe, which leads to recirculation of old air around the BDC branches. In global model studies, this effect has been named aging by mixing (AbM) and is interpreted as the difference between AoA and RCTT (e.g. **?**).

3 Results

3.1 Ozone trends over the period 1998–2018 in CCM simulations and observations

In this section we analyze the ozone trends of all ~~free running~~ 'free running' CCMI-1 simulations (REF-C2), including all ensemble realizations of each model~~for the post-ODS~~ , for the period 1998–2018 together with the observational data, BASIC$_{SG}$. We chose the period 1998–2018 to be consistent with the observational trend estimate in the ozone recovering phase as presented by **?**. Note that ODSs are declining in this period as a result of the Montreal Protocol and its amendments. By using the REF-C2 simulations we include a wide spectrum of SST variability in the different CCMs, as they use either an interactive ocean or prescribed SSTs from a coupled ocean-atmosphere model simulation (see Tab.1). Ozone trends are calculated by simple linear regression~~(see section 2.3)~~, using the monthly deseasonalized ozone time series. We refrain from excluding sources of variability such as QBO, ENSO (El Nino Southern Oscillation), solar cycle or volcanic eruptions in the regression analysis to capture the full range of variability of ozone trends over the given period. Hence our trend estimates have to be interpreted as resulting from both forced trends (e.g., via GHG increases and ODS decreases) as well as from natural and internal climate variability. In the following we compare the ~~so~~ calculated ozone trend from the observational data to the ~~trend estimate by using~~ trends presented in **???** that used a dynamical linear modeling (DLM) approach, ~~that does attempt~~ which attempts to take

natural sources of variability into account~~, as presented in ??~~. In a nutshell, DLM has many similarities with ordinary least squares multiple linear regression (MLR), using predictor variables to account for some of the variability in the timeseries (e.g. solar variability, the QBO). Where DLM primarily differs from MLR is in allowing for a non-linear trend to be estimated and for the seasonal cycle to evolve with time, and therefore the shape of these terms is not predefined. For more details, see **?** and **?**.

The panels of Fig. 1 show a latitude-pressure cross-section of the ~~post-ODS~~ ozone trend for observations (first panel of Fig. 1) and all free running CCMI model simulations. Generally, the linear trend fit we perform on the BASIC~~SD~~ ~~SG~~ data yields similar spatial patterns and magnitudes to those estimated in **?** with the DLM approach (see their Fig. 1f). There are a few small differences, e.g., our linear trend fit results in larger positive trends in the upper stratosphere over the southern tropics of $\sim 1\%$, a slightly less negative trend in the northern hemisphere middle stratosphere ($<1\%$), and consistently large and negative trends close to 100 hPa in the tropics as opposed to a smaller and insignificant trend at around 10°S and over 100-80 hPa in the DLM estimate as shown by **?**. Most notably, linear trend calculations result in small positive trends (up to $\sim$3%) in the southern mid-latitude lower stratosphere, as opposed to overall negative but insignificant trends reported by **?** in that region. However, the comparison reveals that the overall magnitude and trend pattern is also captured by the simple linear regression, i.e. it is not dependent on the exact method used to calculate the trends. Therefore, we proceed with using a linear fitting approach for the comparison between observations and CCMs, though the above caveats should be kept in mind when comparing with a full regression analysis using DLM (**?**).

Overall, large inter-model variability of the trends derived from the individual REF-C2 simulations (including all ensemble members) is revealed in Fig. 1. Nevertheless, a number of features can be identified that are consistent over most models and all their ensemble members. In the upper stratosphere (1-10 hPa) nearly all simulations consistently show an overall positive ~~trend in ozone~~ ozone trend. This ozone increase can be explained by the decrease of ODSs (see e.g. **?**) and by a slow down in ozone destruction rates as the stratosphere cools from GHG increases (see e.g. **?**)~~.~~, as will be further discussed in Sec. 3.5. This upper stratospheric ozone trend has been found for climate model simulations and for observational data in several studies before (e.g. **??????**). However, in the lower stratosphere (30-100/150 hPa) we find a wide spread in the ozone trends among the CCM simulations over recent decades. Many REF-C2 simulations exhibit negative trends in the tropical LS, and they are comparable to the observational trend in magnitude and structure. In agreement with earlier studies ~~(e.g. ?)~~(e.g. **??**), we will show in Section 3.4 that this tropical ozone decrease is related to enhanced tropical upwelling in a warmer climate. However, there are also simulations showing a positive LS ozone trend in the tropics (~~e.g.~~ i.e., GEOSCCM, SOCOLv3, NIWA-1, WACCM-3/4, CAM4-1~~, LMDZrepro~~/2, LMDZrepro, HadGEM; note that the number of the ensemble run is denoted with -1, -2 and so on). At northern and southern mid- and high-latitudes most ~~CCMs~~ simulations exhibit a positive trend, but with a pronounced ~~spread between model simulations~~intermodel spread. Only a few simulations show negative trends in either northern or southern mid-latitudes (e.g. GEOSCCM, WACCM-3, WACCM-4), but it is important to point out here that none of the 31 simulations reproduces the observed negative ozone trend pattern with an ozone decrease covering the tropical belt and extending to the mid-latitude (50°S-50°N), as shown in the upper-left panel and previously in **??**. This discrepancy in the LS ozone trend between observations and models has been reported before (e.g. ozone ~~profile~~ trends, based on CCMI

simulations, ~~(?)~~(??), and in comparison to CCMVal-2 simulations (**?**)). For CCMs that provide multiple ensemble members (WACCM, NIWA, ULAQ, ACCESS, CAM4 and UMUKCA), we also identify a large ensemble spread in the simulated LS ozone trends. For example in WACCM two ensemble members simulate positive tropical ozone trends, while the two other members simulate negative tropical ozone trends. In WACCM (as well as in NIWA and CAM4), the coupled ocean allows for differences in the SST variability between the ensemble members, possibly explaining the large spread in tropical ozone trends. However, as is also the case for models with prescribed SSTs (ACCESS, ULAQ, UMUKCA) ~~,~~that exhibit a large spread between the simulations, the SST variability is not the only reason for the different trend pattern, as was similarly reported and discussed by **?** for CCMVal-2 models. The large spread in LS ozone trends between ensemble members is further in agreement with the study of **?**. They used a nine member ensemble of a ~~free running CCM simulation~~ 'free running' CCM simulation (CESM1-WACCM) and showed that LS ozone trends over the years 1998–2016 are characterized by large internal variability, with e.g. the LS ozone trend ranging from +6% to -6% per decade. But note, again, that none of these ensemble members showed the coherent decrease in ozone in the tropics and extratropics as found in observations (**?**).

Following this qualitative discussion on the spread in the ozone trend pattern between the CCM simulations, we now turn to the LS ozone trends with a more quantitative comparison of the apparent inconsistencies between observations and CCMs. We calculate the trends of the deseasonalized LS ozone columns for the ~~post-ODS~~ period 1998–2018 in two regions: the inner tropics (20°N-20°S) and in the northern mid-latitudes (30°N-50°N). We choose the northern mid-latitude band 30°N-50°N ~~in order to be comparable to~~ for direct comparability with the study of **?**. The pressure range of the lower stratosphere was taken to be 30–100 hPa for the tropics and 30–150 hPa for the mid-latitudes, to take into account the differences in ~~the~~ latitudinal tropopause heights. Trends and their uncertainties (represented by the 90% confidence interval of the linear slope) are shown for each of the 31 available REF-C2 simulations of 18 different CCMs in Fig. 2. We decided to focus on the northern mid-latitudes here, because the SH mid-latitude trends are likely more strongly influenced by the large chemical depletion of ozone within the polar vortex. ~~However, we~~ We will come back to the LS ozone trends of the southern mid-latitudes ~~later on (see section 4)~~in Section 3.5.

In the tropics about half (42%) of the REF-C2 simulations show a significant decrease, about the same (42%) show a non-significant change, and about 15% a significant increase in the integrated tropical LS ozone column. Note that significance is defined as the non-overlap of the error bars (90% confidence interval) with the zero trend. The resulting MMM ozone trend (see red bar on right of Fig. 2) is negative (-0.37 DU/dec), but it is insignificant due to the considerable spread among the different models~~, with the~~ . The 25th-75th quantile of the distribution ~~ranging~~ ranges from -1.12 to 0.20 DU/dec (see edges of box on the right of Fig. 2). Note that for the ~~MMM trend~~ calculation of the MMM trend, we choose to weight each of the 31 simulations ~~is weighted equally ,~~ equally (i.e., not taking into account that some models have multiple ensemble members~~,~~) because the trend ~~variation~~ variations among ensemble members are as large as among the different models ~~.~~over this period.

The observed tropical LS ozone trend of -1.07 DU/dec is statistically significant at the 90% level. Thus the observed tropical trend is more strongly negative than the MMM trend, but lies within the 90% confidence interval of the MMM trend (~~being~~ [-1.76 DU/dec; 1.03 DU/dec]).

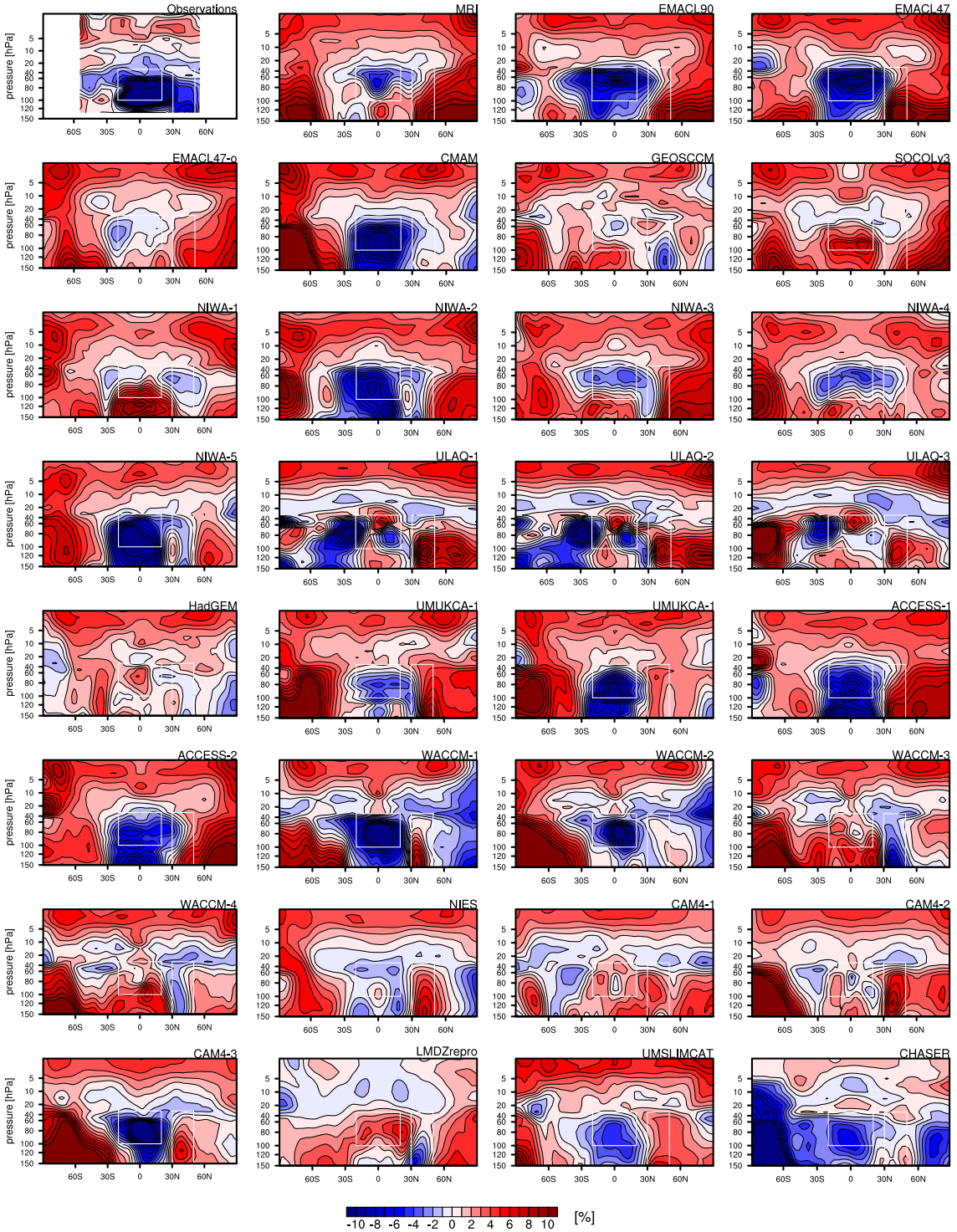

Figure 1. Latitude-pressure cross section of the ozone trend over the ~~post-ODS~~ period 1998–2018 for the observational data set BASIC$_{SG}$ and for all CCMI REF-C2 simulations. Trends are given as relative ozone changes over the whole time period. Boxes illustrate the regions selected to integrate ozone in the LS for trend comparisons later in this study, i.e. in the tropics (20°N-20°S, 30–100hPa) and in the northern mid-latitudes (30°N-50°N, 30–150 hPa).

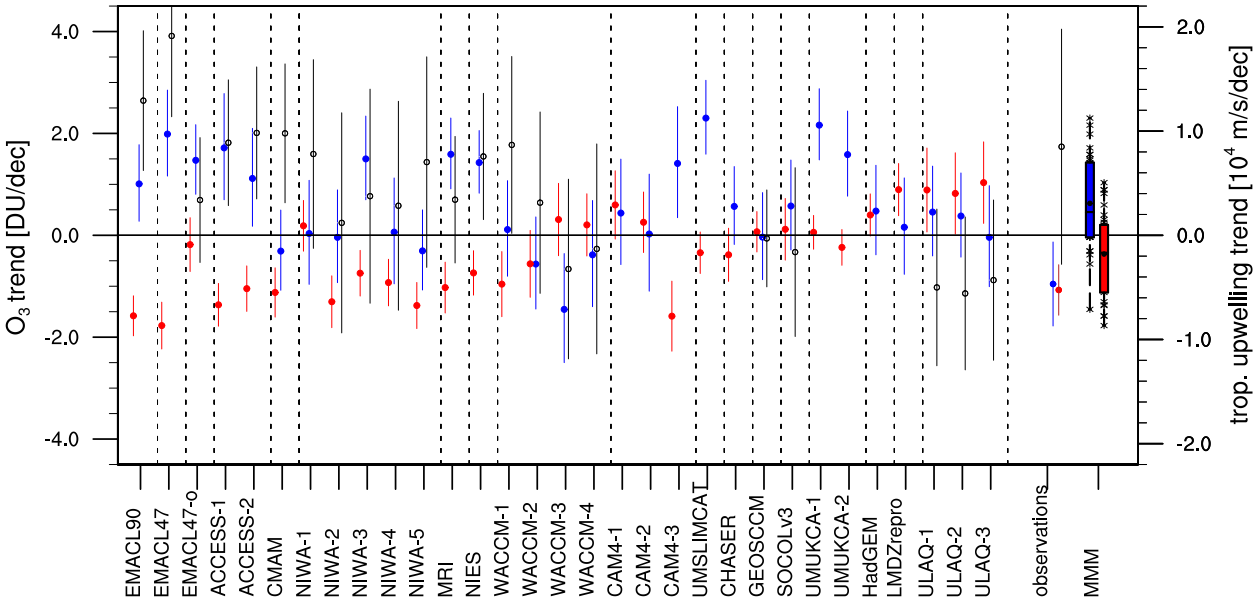

Figure 2. LS ozone trends and their uncertainties in the tropics (20°N–20°S, red dots) and northern mid-latitudes (30°N-50°N, blue dots) together with tropical upwelling trend (black circles, for all simulations providing TEM diagnostics) for the period 1998–2018 for all REF-C2 simulations. Dashed lines separate the individual models. Moreover, observational trends (1998–2018) and multi-model mean trends are given. Observational data for ozone are taken from BASIC$_{SG}$, and for tropical upwelling from ERA5 reanalysis. Error bars associated with each LS ozone trend represent the 90% confidence intervals. The multi-model mean trends are shown as boxplots: the black solid line in the box indicates the median, the black point the MMM and the colored box ranges from the 25th to the 75th percentile of the trends. Crosses denote trends of individual model simulations not lying within the box.

In the northern mid-latitudes less than half (40%) of the REF-C2 simulations show an increase in the LS ozone column, while the remaining 60% of the simulations show a non-significant change (either positive or negative). There is only one simulation (WACCM-3) that shows a significant decrease in the mid-latitude LS ozone column, ~~although it has a corresponding (non-significant) tropical trend~~and in this simulation the tropical ozone trend is positive (but not significant). The resulting MMM trend in the northern mid-latitudes is positive (+0.63 DU/dec) with a high inter-model spread: the 25th to 75th quantile of the distribution ranges from -0.04 to 1.42 DU/dec. Note here, that the observational trend (-0.96 DU/dec) lies ~~just~~ outside the 90% confidence interval of the MMM trend in the mid-latitudes ([-0.91 DU/dec; 2.16 DU/dec]).

Fig. 2 also reveals that over the years ~~1998–2018~~ 1998–2018 more than half of the model simulations have a dipole trend pattern in the LS ozone column, i.e. the sign of the tropical ozone trend is opposite to that in mid-latitudes. ~~A dipole~~ This trend

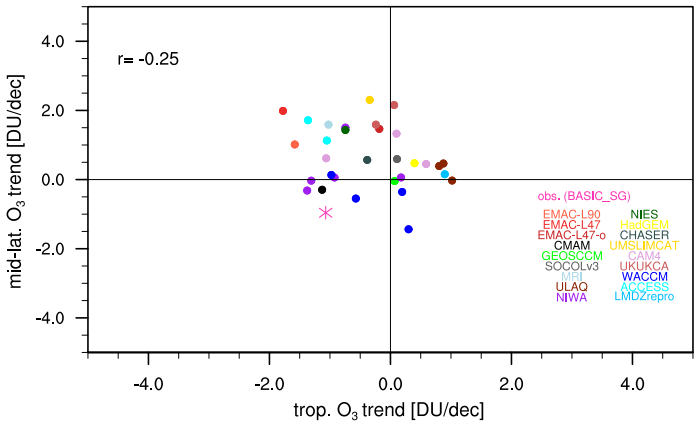

Figure 3. Inter-model correlation between tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trends, calculated over the period 1998–2018 for 31 CCMI REF-C2 simulations. All ensemble members of a particular model are shown in the same color. The observational ozone trends (BASIC$_{SG}$) are also included in here as a star.

.

pattern with negative LS ozone ~~trend~~ trends in the tropics and positive LS ozone ~~trend~~ trends in the northern mid-latitudes can be found for almost half the simulations (45%), and a ~~dipole~~ trend pattern with a positive ozone trend in the tropics and negative trend in the northern mid-latitudes is found in 13% of the simulations. The remaining simulations do not show this dipole, but either both have a positive trend in the tropics and the mid-latitudes (29%), or a negative trend in both tropics and mid-latitudes (13%, i.e. 3 simulations, namely NIWA-5, CMAM, WACCM-2). Only 3 out of 31 simulations simulate negative, but not significant, trends both in the tropics and northern extratropics, and thus they show a similar behavior to observations (see right of Fig. 2 and **?**). However, ~~the~~ their zonal trend patterns (see Fig. 1) reveal, that none of these three simulations reproduces the observed trend pattern with consistent negative trends from 50°S-50°N in the LS. Consequently it is important to keep in mind that the results of these (averaged) trends ~~are depending~~ depend on the choice of the latitude-pressure box, as the integration over a wider latitude band can lead to a cancellation of opposing trends. ~~To~~

Next, we analyze whether a systematic relationship between the LS tropical and mid-latitude ~~trend can be diagnosed from~~ trends exists in the CCM simulations. For this, the simulated northern mid-latitude LS ozone trends are plotted against the simulated tropical LS ozone trends over the time period ~~1998-2018~~ 1998–2018 for all 31 REF-C2 simulations and for the observed data-set BASIC$_{SG}$ in Fig. 3. As discussed above, in the LS the majority (45%) of the models have a negative ozone trend in the tropics and a positive trend in the northern mid-latitudes. ~~This~~ Moreover this illustration again highlights that the trends estimated from observational data are lying on the outer edge of the ~~distribution of the modeled trends~~ model trends distribution. The inter-model correlation between the tropical to mid-latitude trends is negative with a low ~~not significant~~ correlation coefficient (-0.25). Thus, for the chosen period the tropical ozone trends are ~~significantly, but~~ only weakly linked to mid-latitude

ozone trends in the models. ~~We expect a negative correlation here, because~~ However, we expected that the two trends are highly (negatively) correlated, as from our understanding increased tropical upwelling ~~, leading~~ leads to decreased tropical ozone, and this upwelling increase should be linked to an increased mid-latitude downwelling, ~~that~~ which would enhance ozone in the mid-latitudes. ~~The relationship of the trends to tropical upwelling will be further investigated in Sec.~~However Fig. 3 does not support this. Also slightly varying the period (i.e. looking at the periods 1999-2019, 2000-2020, 2001-2021) reveals very low negative or near zero correlations (not shown here). To get a better understanding of the processes leading to the given LS ozone trend patterns we will investigate the relationship of LS ozone trends to stratospheric transport trends in Section 3.4.

Overall we can conclude from the analysis of ozone trends in the suite of CCMI models (see Fig. 1, 2 and 3), that the LS ozone trends exhibit a considerably large spread across both the different models, but also across ~~individual ensemble members~~ ensemble members from a single model, in particular in the mid-latitudes. This indicates that ozone variability considerably influences the LS trends, in agreement with the recent studies by **?** and **?**. However, even when considering the high variability of possible trends in CCM simulations, the observational trends ~~emerges as a rather~~ emerge as an unlikely realization of the simulations over the period 1998–2018. In the next section, we will analyze the robustness of this finding by varying the period of the trend calculation, and providing an in-depth statistical analysis of the likelihood of the observed trend lying within the suite of modeled trends.

3.2 Robustness of lower stratospheric ozone trends

In the previous section we found that the observed LS negative ozone trend in the mid-latitudes together with a simultaneous negative trend in the tropics is unlikely ~~given~~ based upon the suite of CCM simulations. To further establish the robustness of this result, we here test whether this also holds for time periods that are slightly different to the ~~post-ODS~~ period 1998–2018 we considered before. Thus, in this section we first want to investigate how variability influences the ozone trends, and second we want to quantify the likelihood of the observed trend being a realization of the distribution of the modeled trends. To answer those questions, we calculate the LS ozone trends by varying the start and end years of the time period. In Fig. 4 (a) and (b), the observed tropical and mid-latitude ozone trend in the LS is shown for start years varying from 1995–2001 (y-axes) and end years from 2013–2019 (x-axes). Both tropical as well as mid-latitudes LS ozone trends are consistently negative for all chosen periods in the observations (top ~~raw~~row). This is in line with the results of **?**, who found that the observed negative sign of the tropical and mid-latitude trends remain insensitive to changing the end year. In the tropics, observational LS ozone trends are consistently negative with values between -0.64 and -1.24 DU/dec for all possible start end year combinations. In the mid-latitudes the trends are also negative for all shown time periods, but are more variable than in the tropics (values range between -0.11 and -1.22 DU/dec). In particular at mid-latitudes, the strongest negative trends are found for start years of 1996 to 1998, and a sudden decrease in the trend magnitude is found for the start year 1999 and 2000. Thus, the analysis in **??** and in the preceding section focused on a period with ~~particular~~ particularly strong negative mid-latitude ozone trends. Possible reasons for the sudden change in the trend, ~~like~~ such as the strong ENSO event in 1998, are discussed in ~~Sec.~~Section 4. Note that the trend magnitude ~~goes back up~~ increases again for the start year 2001, ~~so it again looks like~~ which again suggests that interannual variability influences the observational mid-latitude trends.

Fig. 4 (c) and (d) display the tropical and mid-latitude trends as a function of start and end year derived from the model simulations. To do so, a robust estimate of the trend probability distribution considering all model simulations was derived (see Section 2.3) and from this distribution the most likely trend is shown (see peak in the models' trend probability distributions of Fig. S1 and S2 in the supplement). In the tropics the ozone trends derived from the REF-C2 simulations are negative and ~~,~~ range from -0.74 to +0.02 DU/dec. In the mid-latitudes the trends are positive for all possible start/end year combinations, with values ranging from +0.4 to +1.48 DU/dec. In contrast to the sudden change in the mid-latitude observational trend for start years 1999 and 2000, in the REF-C2 simulations no such systematic change can be found. The estimated probability distributions of the trends from the REF-C2 simulations (see Figs. S1 and S2 in the supplement) are typically symmetric around their maximum value and show a single, central peak. The width of the distribution changes ~~for varying~~ when varying the start/end year combination, with narrower distributions for longer time periods. Moreover, ~~the~~ visual inspection of the distribution implies that the tropics (Fig. S1) ~~have generally~~ generally have Gaussian-like distributions, whereas the mid-latitudes (Fig. S2) often show a more peaked structure, i.e. with heavier ~~trails~~tails. Nevertheless, as an estimate of the width of the models trend distribution, we show in Fig. 4 (e) and (f) the standard deviation of the models distribution (in DU/dec) in the tropics and mid-latitudes, respectively. For longer time periods (values in lower right corner) the standard deviation of the models' trend is smaller, i.e. the distribution is narrower. This indicates that the influence of natural variability is less important for longer time periods, as should be expected.

Given the distributions representing the combined trends of the models, we can now quantify the disagreement between the observational trend estimate and the models' trend probability distributions for each start/end year combination. In Figs. 4 (g) and (h) the "probability of the disagreement" between observational and modeled LS ozone trends is given for the tropics and the mid-latitudes. The value of the "probability of disagreement" is calculated by the central interval of the models' probability distribution when taking the observed trend value as threshold of this interval. Thus, a probability value of 90% indicates that the observed trend falls within the inner 90% of the distribution, i.e., only 10% of the distribution is more extreme than the observed trend: the smaller the given "probability of disagreement" value, the higher is the probability that the observed trend lies within the models' distribution. In the tropics, the observed LS ozone trend falls within the 13% to 73% interval of the modeled probability distribution, i.e. the observed trends are generally likely representations of the models' trends. The agreement is best for short time periods (values in diagonal in Fig. 4 (g)), mostly because of the broader distribution (see Fig. 4 (e) and Fig. S1). Also for early start years (in particular 1995) and end years ranging from 2013 to 2018, the disagreement is small, because model trends are strongly negative for this period (see Fig. 4 (c)). In the mid-latitudes, the observed trend generally lies at more distant parts of the models' trends distribution (73% to 96%), i.e. the observed trend is a more extreme value in the models' distribution. The disagreement is smallest for both the earlier periods (lower left, start years 1995–1997 and end years 2013–2015) and the later periods (upper ~~left~~right, start years 1999–2001 and end years 2017–2019). This coincides with the generally smaller negative trends in those periods in observations (see Fig. 4 (b)) and rather constant trend distributions in the models (see Fig. 4 (d)). For the periods with the strongest negative observed trend (start years 1996–1998), the observed trend lies within the central 90% or higher of the models' distribution, i.e. is an unlikely representation from the modeled

trends. The sudden decrease in the observed trend magnitude for start years 1999 (Fig. 4 (b)) is reflected by a decrease of the central interval to about 75%. In general, one might have expected that longer periods lead to better agreement of the observed and modeled trend due to the smaller influence of variability (see Figs 4 (e) and (f))- as we do in the models- however, we do not find this to be true for either the tropics or the mid-latitudes.

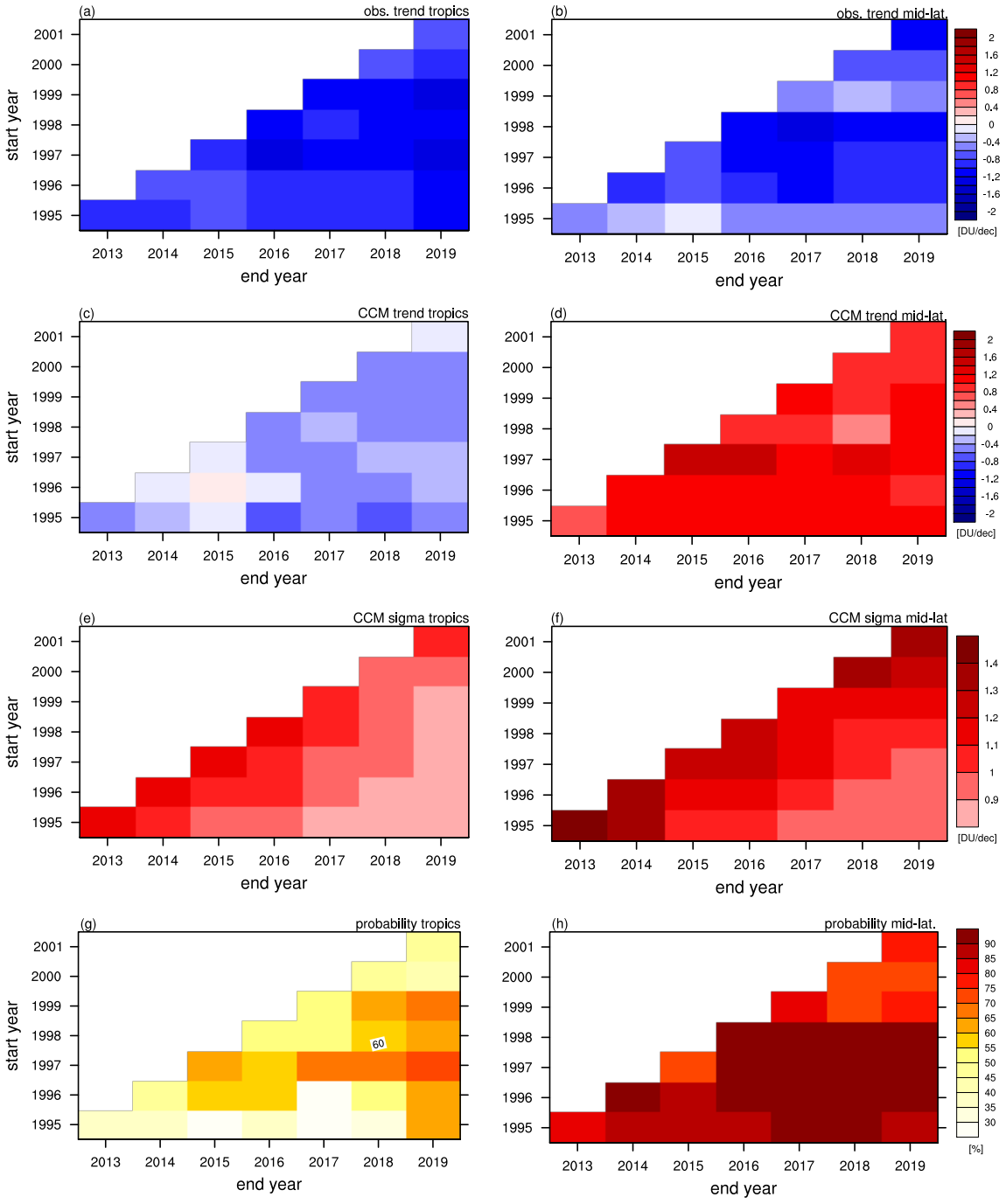

Figure 4. Tropical (right) and mid-latitude (left) LS ozone trends (in DU/dec) as function of different periods for the observational trend of BASIC~~SD~~ SG ((a)+(b)), the most likely trend of the modeled REF-C2 probability distribution ((c)+(d)) and the 1-σ standard deviation (in DU/dec) of the mean obtained from the probability distribution ((e)+(f)). The panels (g)+(h) show the "probability of disagreement" (in %) between observed trends and the REF-C2 trend probability distribution. In all panels the x-coordinate denotes the different end years (2013-2019) and the y-coordinate the different start years (1995-2001).

3.3 Convergence of future lower stratospheric ozone trends

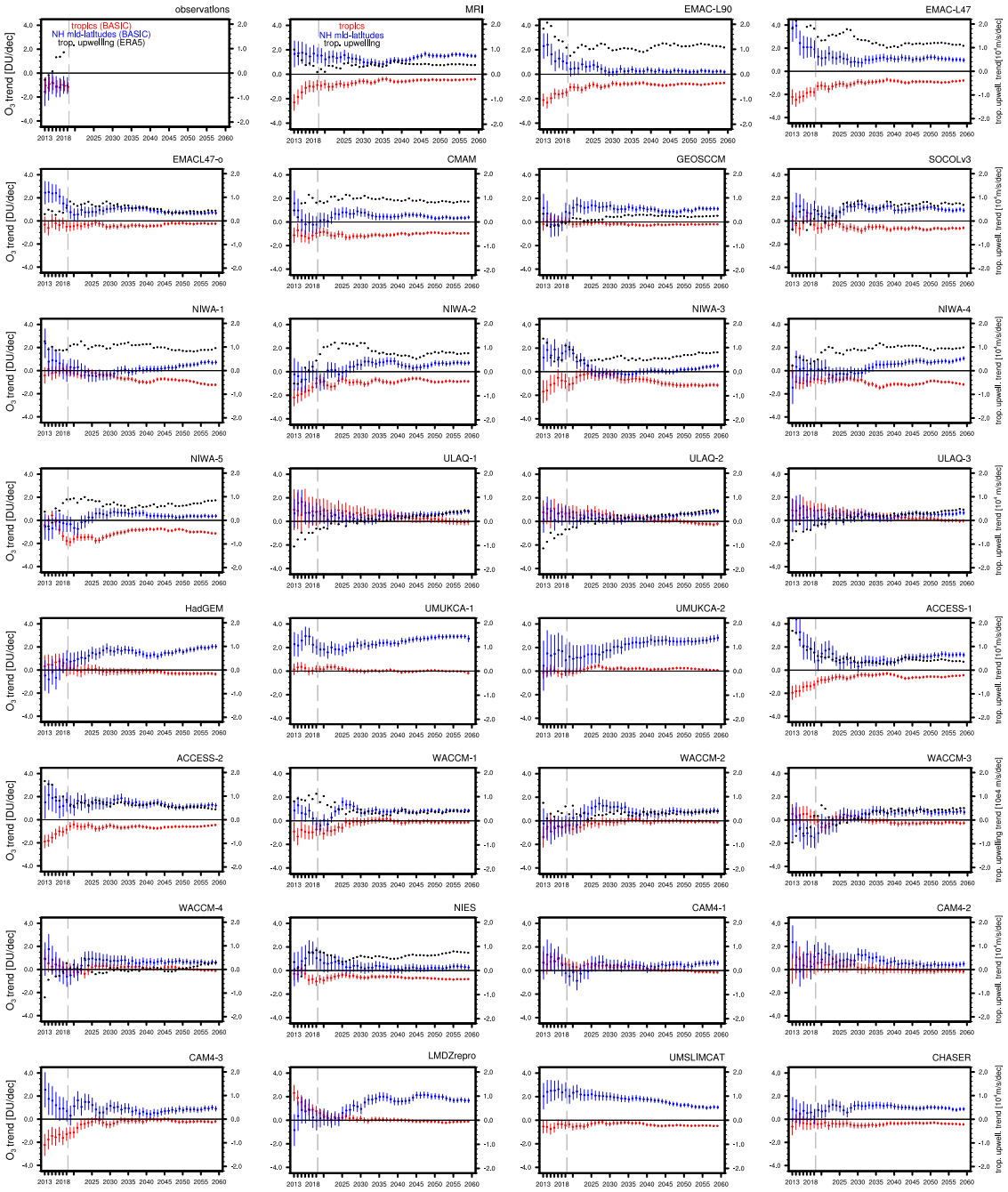

Figure 5. Tropical (20°S-20°N) and northern mid-latitude (30°N-50°N) LS ozone column trend and their uncertainties (in DU/dec) of observations (BASIC$_{SG}$) and ~~of~~ REF-C2 ~~model~~ simulations as a function of the end year (red and blue dots, respectively). Tropical upwelling ~~trend is~~ trends are included for all REF-C2 simulations, where TEM diagnostics was available (black dots); observational tropical upwelling is taken from ERA5 reanalysis. The end year varies from 2013 to 2019 for observational data and from 2013 to 2060 for REF-C2 simulations. Error bars associated with each trend represent the 90% confidence intervals.

~~In the last~~ In the previous section, the ~~robustness of ozone trend~~ ozone trend robustness was analyzed for time periods of up to 25 years. We will show in the following that, as the considered time periods ~~become longer.~~ are extended, the influence of natural variability decreases ~~more and more, and~~ and the trends converge to the trend forced by long-term ~~trends in~~ GHG and ODS ~~concentrations~~concentration changes. To analyze the timing and the values of the trends' convergence, we extend the period for the trend calculation into the future for all REF-C2 simulations.

Fig. 5 shows the tropical and northern mid-latitude LS ozone trends together with the tropical upwelling trend (black; if available) for periods with the fixed start year 1998, and the end year varying from 2013 up to 2060, by extending the time period by steps of one year. For reference, the observational trends of ozone (from BASIC$_{SG}$) and tropical upwelling (from ERA5) are shown in the upper left panel of Fig. 5, with the last available end point in the year 2019. As shown in the last section, the trends derived from observational data are consistently negative both in the tropics and in the northern mid-latitudes.

As discussed in Sec. 3.1, the ozone trends exhibit a strong inter-model spread for the observational time periods. Both tropical and mid-latitude ozone trends in the individual model simulations vary considerably for different end point years within the observational period (left of the vertical dashed gray lines)~~, with the~~ . The northern mid-latitude trend is generally more variable than the tropical trend. For longer time periods extending into the future, ~~for all simulations~~ the uncertainties in the LS ozone trends decline and ~~trends converge~~the trends converge in all simulations. All model simulations consistently ~~evolve a dipole trend pattern, with significant and persistent negative~~ simulate persistent negative or near-zero trends in the tropics and ~~persistent positive~~ positive or near-zero trends in the northern mid-latitudes~~(for a summary, see also Fig. 8). However~~. However, the timing of convergence of the trends to ~~the dipole~~ this trend pattern is rather different in the simulations, as can be inferred from Fig. 5, i.e. the convergence appears to be model dependent. For some models, the trends vary little for end years after 2020 (e.g. MRI in Fig. 5), while in other models, the trends still vary considerably until end years around 2030 to 2040 (e.g. the four WACCM ensemble members in Fig. 5). The timing of the convergence ~~basically~~ is controlled by the ratio of the year-to-year variability to the strength of the forced trends. The relative forcing by ODS versus GHG changes over time~~for the given periods~~, and thereby the forced ozone trends vary over the time periods as well, making it difficult to quantify an exact date of convergence. Still, the trend estimates for the entire period 1998 to 2060 do converge to stable values for almost all models, thus representing the forced trend for this time period. ~~For all model simulations, the long-term LS ozone trends over the entire period (1998–2060) are negative in the tropics and positive in the extratropics, but the trend magnitudes vary~~ The trend magnitudes over this long period vary strongly between the models~~. Thus model trends vary~~ , from -0.10 to -1.32 DU/dec in the tropics ~~,~~and from +0.39 to +2.00 DU/dec in the mid-latitudes. Comparing this to the model range of the ~~somewhat~~ shorter time period 1998–2040, we see that the tropical trend (+0.06 to -1.12 DU/dec) has not converged to the end point values of 2060, yet. The mid-latitude trend (+0.54 to +2.15 DU/dec) is however close to the 2060 values.

Overall, the mid-latitude trends converge to positive values in the majority of the model simulations (about 85%) by 2030. Thus, if both the year-to-year variability and the forced response of the models is simulated realistically, we should expect the emergence of positive mid-latitude trends from observational records within the next decade.

3.4 Influence of ~~tropical upwelling~~ transport processes on LS ozone trends ~~in CCMs~~

Table 2. MMM and observational ozone trends, calculated over the period 1998–2018 for tropical upwelling at 70 and 100 hPa, for extratropical downwelling at 70 and 100 hPa, for the LS tropical ozone column and for the northern mid-latitude ozone column. Note that LS mid-latitude ozone trends are averaged over the fixed latitude band of 30° N-50° N and also over the dynamical defined latitude band between the turnaround latitudes to 50° . MMM trends and their standard deviation are given over a subset of 20 REF-C2 simulations. Observational based data for up- and downwelling are taken from ERA5 reanalysis and observational data for ozone from

~~EMAC-L90 -0.71~~ heighttrop. upwelling trend (70 hPa)	[~~-0.82; -0.54~~kg/sec/dec]
~~EMAC-L47 -0.60~~ trop. upwelling trend (100 hPa)	[~~-0.82; -0.54~~kg/sec/dec]
~~EMAC-L47-o -0.76~~ downwelling trend (70 hPa)	[~~-0.86; -0.61~~kg/sec/dec]
~~NIES -0.78~~ downwelling trend (100 hPa)	[~~-0.87; -0.64~~kg/sec/dec]
~~ACCESS -2 -0.54~~ heighttrop. ozone trend	[~~-0.71; -0.32~~DU/dec]
~~NIWA -2 -0.37~~ mid-lat. (fixed) ozone trend	[~~-0.58; -0.11~~DU/dec]
~~NIWA -3 -0.29~~ mid-lat. (dyn) ozone trend	[~~-0.52; -0.02~~DU/dec]
~~NIWA -4 -0.33 -0.55; -0.080.35 0.09; 0.56-0.64 -0.78; -0.45NIWA -5 -0.11 [-0.36; 0.17]0.08 [-0.19; 0.35]-0.37 -0.58; -0.11ULAQ -10.34~~	

3.5 Forced ozone trends in models

4 ~~Discussion~~

In the previous sections we analyzed the ozone trends of ~~recent post-ODS periods (i.e. 1998-2018 and variations on this period) the recent 20 year period~~ in detail and found that modeled and ~~observational~~ observed ozone trends disagree, especially in the northern mid-latitude LS. ~~One possible reason for the disagreement between modeled and observed LS ozone trends could be problems with the satellite records, as e.g. instrument biases and drifts can lead to large uncertainties in the observations (see e.g. ???). However, assuming the~~ Assuming the observational data are correct, the question that arises from our results is whether the disagreement stems from the influence of natural variability, or whether ~~it is related to the forced trend. In~~ the forced response to GHG or ODS concentrations is not captured correctly in the models. Thus in the following, we will ~~discuss this question using Fig. 8, which summarizes our results: it shows the observational~~ investigate the relative role of GHG versus ODS forcing on the ozone trends in the models for the observational period and periods extending into the future. Figs. 8a and 8b show upper and lower stratosphere MMM ozone trends in the tropics (20°N-20°S), in the northern mid-latitudes (30°-50°N) and in the southern mid-latitudes (30°-50°S) for the REF-C2 simulations as well as ~~the REF-C2~~ for the sensitivity simulations with fixed ODS (fODS) and with fixed GHG (fGHG) concentrations (for a detailed description of these sensitivity simulations see Section 2.1). These MMM ozone trends ~~and its spread for periods of different length, for the tropics, the northern mid-latitudes and the southern mid-latitudes in the upper and lower stratosphere~~are calculated for the recent time period (1998–2018), for a time period, which extends into the future (1998–2040) and for a future time period (2050–2100).