

# *The Met Office HadGEM3-ES chemistry–climate model: evaluation of stratospheric dynamics and its impact on ozone*

Article

Published Version

Creative Commons: Attribution 3.0 (CC-BY)

Open Access

Hardiman, S. C., Butchart, N., O' Connor, F. M. and Rumbold, S. T. ORCID: <https://orcid.org/0000-0001-8138-4541> (2017) The Met Office HadGEM3-ES chemistry–climate model: evaluation of stratospheric dynamics and its impact on ozone. *Geoscientific Model Development*, 10 (3). pp. 1209-1232. ISSN 1991-9603 doi: 10.5194/gmd-10-1209-2017 Available at <https://centaur.reading.ac.uk/70162/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://dx.doi.org/10.5194/gmd-10-1209-2017>

To link to this article DOI: <http://dx.doi.org/10.5194/gmd-10-1209-2017>

Publisher: European Geosciences Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



# The Met Office HadGEM3-ES chemistry–climate model: evaluation of stratospheric dynamics and its impact on ozone

Steven C. Hardiman<sup>1</sup>, Neal Butchart<sup>1</sup>, Fiona M. O'Connor<sup>1</sup>, and Steven T. Rumbold<sup>1,a</sup>

<sup>1</sup>Met Office Hadley Centre, Met Office, Fitzroy Road, Exeter, Devon, EX1 3PB, UK

<sup>a</sup>now at: National Centre for Atmospheric Science, University of Reading, RG6 6BB, UK

*Correspondence to:* Steven C. Hardiman (steven.hardiman@metoffice.gov.uk)

Received: 1 November 2016 – Discussion started: 30 November 2016

Revised: 22 February 2017 – Accepted: 24 February 2017 – Published: 20 March 2017

**Abstract.** Free-running and nudged versions of a Met Office chemistry–climate model are evaluated and used to investigate the impact of dynamics versus transport and chemistry within the model on the simulated evolution of stratospheric ozone. Metrics of the dynamical processes relevant for simulating stratospheric ozone are calculated, and the free-running model is found to outperform the previous model version in 10 of the 14 metrics. In particular, large biases in stratospheric transport and tropical tropopause temperature, which existed in the previous model version, are substantially reduced, making the current model more suitable for the simulation of stratospheric ozone. The spatial structure of the ozone hole, the area of polar stratospheric clouds, and the increased ozone concentrations in the Northern Hemisphere winter stratosphere following sudden stratospheric warmings, were all found to be sensitive to the accuracy of the dynamics and were better simulated in the nudged model than in the free-running model. Whilst nudging can, in general, provide a useful tool for removing the influence of dynamical biases from the evolution of chemical fields, this study shows that issues can remain in the climatology of nudged models. Significant biases in stratospheric vertical velocities, age of air, water vapour, and total column ozone still exist in the Met Office nudged model. Further, these can lead to biases in the downward flux of ozone into the troposphere.

## 1 Introduction

Previous studies have identified numerous couplings between ozone, greenhouse gases, tropospheric ozone precursors and stratospheric ozone-depleting substances, and climate change. Increased carbon dioxide and near-surface ozone levels, for example, can impact vegetation and the strength of the land carbon sink (Sitch et al., 2007). Gas-phase constituents, such as tropospheric and stratospheric ozone, have contributed to historical climate forcing (Stevenson et al., 2013; Myhre et al., 2013) and the inclusion of interactive chemistry, at least in some models, could affect estimates of climate sensitivity (Nowack et al., 2015). Likewise, climate change can impact on atmospheric composition through changes in the strength of the Brewer–Dobson circulation (Butchart and Scaife, 2001; Butchart et al., 2006), changes in methane lifetime (Johnson et al., 2001; Voulgarakis et al., 2013), changes in background and peak surface ozone concentrations (Fiore et al., 2012), temperature-dependent chemical reaction rates (Waugh, 2009a), and the timescale for the stratospheric ozone layer to recover (WMO, 2011). Increasingly, there is also recognition of the extensive coupling between the troposphere and stratosphere, with stratospheric ozone recovery impacting on tropospheric composition through stratosphere–troposphere exchange (e.g. Zeng et al., 2010) and photolysis rates (e.g. Zhang et al., 2014) and also impacting on surface climate (Morgenstern et al., 2009).

As a result, coupled chemistry–climate models have evolved to encompass both stratospheric and tropospheric chemistry coupled to state-of-the-art atmosphere–ocean climate models, in order for such couplings to be stud-

ied and fully understood. Chemistry–climate models are also used to provide policy-relevant information, such as the assessment of strategies for mitigating and adapting to a changing climate with changing atmospheric composition (Eyring and Lamarque, 2012; Prinn, 2013). However, because of their inherent complexity, there is a strong need for comprehensive assessment and benchmarking of such models to sit alongside their development. In particular, the use of quantitative performance metrics (Waugh and Eyring, 2008) to both track the development of an individual model and/or to benchmark the performance of a multi-model ensemble (Eyring et al., 2008) is important. These performance metrics have traditionally been used to consider how well individual model processes are simulated. In the present study, we take this further, considering the impacts of model processes on each other.

Nudging the dynamics of chemistry–climate model simulations towards observations is a technique used both to look at the impact of specific physical processes on atmospheric composition, and/or to remove the influence of unrealistic model climatology from the evolution of chemical fields. Case studies covering just the length of a single observational campaign and simulations covering long-term trends over the historical period are both ways in which the use of nudged chemistry–climate models can enhance our understanding of the evolution of the chemical composition of the atmosphere. For example, Laat et al. (2001) consider the evolution of tropospheric ozone concentrations over the Indian Ocean during the spring of 1995 to evaluate the large-scale advection processes and associated tracer transport in their model. Dameris et al. (2005) consider the impact of various “forcings” (including sea surface temperatures, volcanoes, and the solar cycle) on chemical composition to investigate which processes are well/poorly represented in models. Akiyoshi et al. (2016) present a case study of the evolution of chemical species during the stratospheric sudden warming of winter 2010 using both a nudged model and observations to study the structure in the chemical fields. A more general overview of the impact of nudging on chemistry–climate models is given in Jöckel et al. (2006, 2015), Telford et al. (2013), and Tilmes et al. (2016).

In the present study, the stratospheric dynamics, transport, and simulated total column ozone (TCO) in free-running and nudged versions of the Met Office chemistry–climate model, HadGEM3-ES, are evaluated. The nudged simulations here make it possible to determine the ways in which biases in the model dynamical fields affect the accuracy of simulated TCO, and thereby help attribute the remaining biases in TCO to other components of the model (i.e. the transport and chemistry schemes).

This study is set out as follows. Section 2 describes the model setup and the simulations evaluated here. Section 3 presents the results and is split into sections focusing on model metrics and the dynamics and TCO of the tropics and extratropics. Conclusions and discussion are given in Sect. 4.

## 2 Model setup and simulations

The Met Office model configuration used in this study is the chemistry–climate model HadGEM3-ES. The underlying atmosphere model is the Global Atmosphere 4.0 (GA4.0) configuration of HadGEM3 (Walters et al., 2014) and is based on the Met Office’s Unified Model (MetUM). It has a horizontal resolution of  $1.875^\circ$  longitude  $\times$   $1.25^\circ$  latitude and 85 levels in the vertical, covering an altitude range of 0–85 km. This is coupled to the Global Land 4.0 (GL4.0) configuration of the JULES land surface model (Walters et al., 2014). For simulations requiring ocean and sea ice components, the Nucleus for European Modelling of the Ocean (NEMO v3.4; Madec, 2008) model, with a  $1^\circ$  resolution (ORCA-1) and 70 vertical levels, is used along with the Los Alamos sea ice model (CICE v4.1; Hunke and Lipscomb, 2008).

This configuration represents a significant improvement in the physical model since the Met Office’s contribution (Morgenstern et al., 2010) to the Chemistry–Climate Model Validation activity 2 (CCMVal-2; Eyring et al., 2008). For example, the horizontal and vertical resolutions have increased from  $3.75^\circ$  longitude  $\times$   $2.5^\circ$  latitude and 60 vertical levels (model lid at 84 km). There have also been improvements to the atmosphere model physics and the addition of new ocean and sea ice components, all of which are documented in detail in Hewitt et al. (2011), Walters et al. (2011), and Walters et al. (2014). A significant result of these model improvements is the much reduced temperature bias at the tropical tropopause layer, which in CCMVal-2 required the models based on MetUM to prescribe water vapour in this region. Water vapour is modelled interactively in the HadGEM3-ES simulations reported here.

This atmosphere-only or coupled atmosphere–ocean model HadGEM3 is, in turn, coupled to the gas-phase chemistry component of the United Kingdom Chemistry and Aerosol (UKCA) model (Morgenstern et al., 2009; O’Connor et al., 2014). The chemistry scheme is a combination of the stratospheric chemistry from Morgenstern et al. (2009) with the TropIsop tropospheric chemistry scheme from O’Connor et al. (2014). Photolysis rates are calculated interactively using the Fast-JX scheme (Telford et al., 2013), and interactive lightning emissions are scaled to give  $5 \text{ Tg N year}^{-1}$  (O’Connor et al., 2014). Details of the simulation of polar stratospheric clouds (PSCs) are given in Sect. 2 of Morgenstern et al. (2009) and Sect. 2 of Chipperfield and Pyle (1998). Above the nitric acid trihydrate (NAT) point (195 K), reactions occur on liquid sulfuric acid aerosols. Below this temperature the model forms solid NAT particles, and then below the ice point (188 K) the model forms ice particles. There is no representation of supercooled ternary solutions. The deposition schemes have been improved since the Met Office’s CCMVal-2 configuration, with interactive wet deposition now applied to a wider range of species, and the tabulated dry deposition scheme replaced by a resistance-in-series approach (O’Connor et al., 2014). The interactive

mass-based aerosol scheme (Bellouin et al., 2011) is unchanged from that used in CCMVal-2. Thus, the HadGEM3 model coupled to the UKCA chemistry scheme and the Coupled Large-scale Aerosol Simulator for Studies In Climate (CLASSIC) aerosol scheme (Bellouin et al., 2011) is referred to as HadGEM3-ES.

The results shown in this paper come from HadGEM3-ES simulations set up to follow the Chemistry–Climate Model Initiative (CCMI) reference simulations (Morgenstern et al., 2017). These include a single ensemble member for both the atmosphere-only historical simulation (REF-C1) and the coupled atmosphere–ocean historical and future simulation (REF-C2), which begin in 1960, as described in Eyring et al. (2013). The greenhouse gases (GHGs), ozone-depleting substances (ODSs), tropospheric ozone precursor emissions, aerosol and aerosol precursor emissions, sea surface temperatures (SSTs) and sea ice concentrations (for the atmosphere-only REF-C1 simulation), and the forcings from solar variability and stratospheric volcanic aerosol are all as described in Eyring et al. (2013).

The coupled (REF-C2) simulation is spun up to 1960 conditions as follows. A 400-year spin-up of the coupled atmosphere–ocean model to a perpetual pre-industrial state is followed by a transient spin-up of the coupled model, without interactive chemistry, to 1950 conditions. Chemistry is then included, and a 10-year spin-up to 1960 conditions is performed, as recommended by Eyring et al. (2013). For the atmosphere-only simulations, this 10-year spin-up from 1950 with chemistry included (Eyring et al., 2013) is all that is required for the atmosphere to equilibrate.

Alongside the free-running atmosphere-only historical simulations (REF-C1), simulations in which temperature and horizontal wind fields are nudged (Telford et al., 2008) towards the ERA-Interim reanalysis (Dee et al., 2011) are also run (REF-C1SD). Nudging is applied over the vertical range of 2.5–51 km and is smoothly increased/decreased over two model levels at the bottom/top of this vertical range. Surface pressure is not nudged, since HadGEM3-ES has a non-hydrostatic terrain-following dynamical core in which surface pressure is not a prognostic and, further, the difference in horizontal resolution between the model and the reanalysis data would lead to a mismatch in details of the orography. McLandress et al. (2014) found that discontinuities in the upper stratospheric temperatures exist in ERA-Interim, in 1985 and 1998, due to changes in the satellite radiance data used. These discontinuities led to erroneous jumps in ozone concentrations in the upper stratosphere in their model, and therefore, in the “smoothed” nudged simulations detailed in Table 1, they were removed here using the technique of McLandress et al. (2014). To avoid introducing spurious noise, Merryfield et al. (2013) found that the relaxation timescale must be longer than the time intervals between the reanalysis fields that are being nudged towards (6 h for ERA-Interim) and noted in particular that relaxation timescales of 24 and 48 h both gave good results (see

**Table 1.** Model simulations.

Name	Time period	Coupled ocean?	Nudging time-scale	Smoothing?
REF-C1	1960–2010	No	N/A	N/A
REF-C2	1960–2100	Yes	N/A	N/A
REF-C1SD-24 h	1980–2010	No	24 h	No
REF-C1SD-48 h	1980–2010	No	48 h	No
REF-C1SD-24 h, smoothed	1980–2010	No	24 h	Yes
REF-C1SD-48 h, smoothed	1980–2010	No	48 h	Yes
CCMVal-2 (UMUKCA-METO)	1960–2005	No	N/A	N/A

their Fig. 23). After some subjective trials, 24 and 48 h were also found to be appropriate timescales for HadGEM3-ES, at least for the fields of interest here, and results using both timescales are included below.

Details of these simulations are summarized in Table 1. Free-running simulations are run over the period 1960–2010 (REF-C1) and 1960–2100 (REF-C2), and nudged simulations are run over the period 1980–2010 (using initial conditions taken from REF-C1). As such, we analyse the period 1980–2010 in this study.

### 3 Results

#### 3.1 Metrics

Metrics for evaluating the processes in chemistry–climate models relevant for the simulation of stratospheric ozone were developed as part of the CCMVal-2 project (Eyring et al., 2008). The metrics for dynamical processes are listed in Butchart et al. (2010, 2011). These dynamical metrics include one for the polar vortex final warming time but, for reasons explained later in this section, we choose to evaluate final warming using the method of Hardiman et al. (2011), and thus this metric is not directly comparable and not included here. Table 2 lists the metrics used in this study.

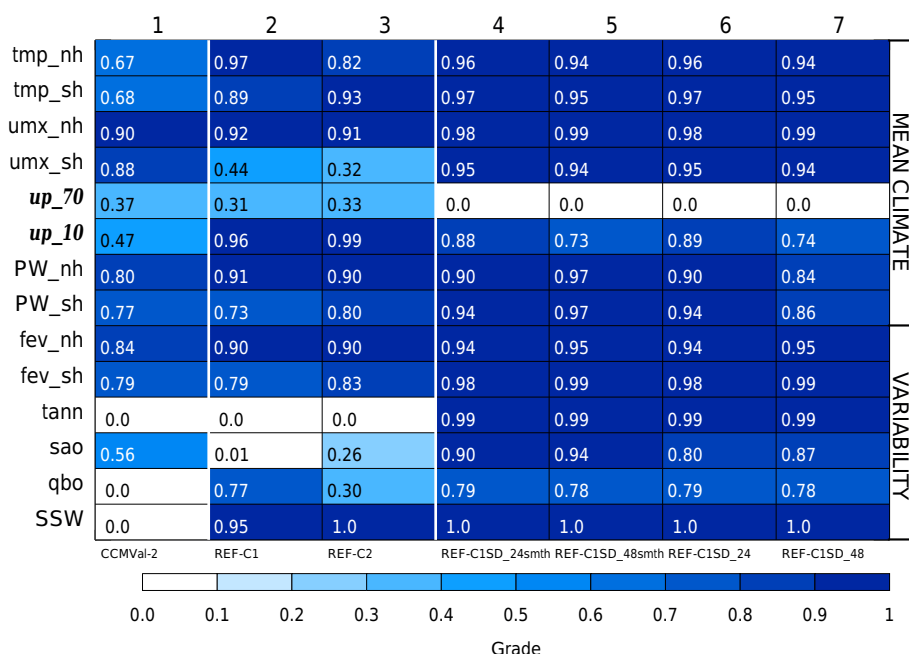
Following the method of Waugh and Eyring (2008), “grades” are associated with each metric to measure how accurately it is simulated, and these are calculated as follows:

$$g = 1 - \frac{1}{3} \frac{|\mu_{\text{model}} - \mu_{\text{obs}}|}{\sigma_{\text{obs}}}, \quad (1)$$

where  $g$  is the grade assigned to the metric (and is set to 0 if calculated to have a negative value),  $\mu_{\text{model}}$  and  $\mu_{\text{obs}}$  are the model and observational mean values of the metric, and  $\sigma_{\text{obs}}$  is the interannual standard deviation of the observations (a proxy for observational uncertainty). Thus, a value of 1 represents the model having an identical mean value to reanalysis (the “observations”), and a value of 0 represents the model mean value deviating by more than 3 standard deviations from the reanalysis. Here, we recalculate these metrics

**Table 2.** Metrics.

Name	Description
Mean climate	
tmp_nh	60–90° N December–January–February temperatures at 50 hPa
tmp_sh	60–90° S September–October–November temperatures at 50 hPa
umx_nh	Maximum Northern Hemisphere eastward wind in December–January–February at 10 hPa
umx_sh	Maximum Southern Hemisphere eastward wind in June–July–August at 10 hPa
up_70	Tropical upwelling mass flux at 70 hPa
up_10	Tropical upwelling mass flux at 10 hPa
PW_nh	Slope of the regression of the February and March 50 hPa temperatures 60–90° N on the 100 hPa January and February heat flux 40–80° N
PW_sh	Slope of the regression of the August and September 50 hPa temperatures 60–90° S on the 100 hPa July and August heat flux 40–80° N
Variability	
fev_nh	Amplitude of the leading mode of variability (EOF) of the 50 hPa zonal-mean zonal wind for the Northern Hemisphere, poleward of 45° (EOFs are scaled to have the same standard deviation as the original data)
fev_sh	Amplitude of the leading mode of variability (EOF) of the 50 hPa zonal-mean zonal wind for the Southern Hemisphere, poleward of 45° (EOFs are scaled to have the same standard deviation as the original data)
tann	Amplitude of the annual cycle at 2 hPa in the zonal-mean zonal wind, 10° S–10° N
SAO	Amplitude of the semi-annual oscillation at 1 hPa in the zonal-mean zonal wind, 10° S–10° N
QBO	Amplitude of the quasi-biennial oscillation at 20 hPa in the zonal-mean zonal wind, 10° S–10° N
SSW	Frequency per year of major sudden stratospheric warmings, defined using reversal of the zonal-mean zonal wind at 10 hPa, 60° N



**Figure 1.** Metrics of dynamical fields and processes (see Table 2). Bold italic font indicates metrics which are not directly constrained in the nudged simulations. Column numbers are printed above each column, and the model simulation is printed below each column. For details of model simulations, see Table 1 (where “24smth” corresponds to “24 h, smoothed”, etc.).

for the Met Office model used in CCMVal-2 (UMUKCA-METO, REF-B1 simulation) using the years 1980–2010 of the ERA-Interim reanalyses (Dee et al., 2011) instead of the years 1980–2000 of the ERA40 reanalysis. These recalculated CCMVal-2 metrics can then be directly compared to those for all the free-running and nudged CCMI simulations. Figure 1 displays these metrics in the same style as Butchart et al. (2010).

It is interesting to note that the UMUKCA-METO values for some of these metrics show a significant degradation compared to those given in Butchart et al. (2010) for the same simulation. Reasons for this are that

- the reanalysis dataset used here as the benchmark is ERA-Interim as opposed to ERA-40 and
- the analysis here is over the period 1980–2010 as opposed to 1980–2000 as used in CCMVal-2.

In particular, using a different period can substantially alter the values of some metrics. For example, the PW<sub>sh</sub> diagnostic considers the variability in the heat flux and polar vortex temperatures in the Southern Hemisphere high-latitude winter. The sudden warming observed in 2002 (the only Southern Hemisphere sudden warming on record) significantly increases the overall variability in both these quantities. The semi-annual oscillation (measured by the SAO metric) increases in amplitude for the years 2000–2010, such that its mean amplitude for the period 1980–2000 is  $15 \text{ m s}^{-1}$  and this increases to  $17 \text{ m s}^{-1}$  for the period 1980–2010. This increase is not captured in the free-running simulations. The trend in mass upwelling in the tropical lower stratosphere (measured by the up<sub>70</sub> diagnostic) is, for ERA-Interim, almost steady over the period 1980–1995, but shows a strong downward trend over the period 1995–2010, which is again not captured in the free-running simulations. This sensitivity shows that a need to analyse over the full 30 years is common to all simulations for calculation of the most reliable metric scores.

Since reanalysis datasets and the period analysed will continue to be updated, there are issues with referring back to the values of metrics in previous reports (see also Austin et al., 2003). These issues could be minimized by

- using information from multiple reanalyses datasets as the metric “observations” and
- ensuring that the period analysed is of sufficient length to reduce the impact of interannual variability,

where the “interannual variability” in this case is the interannual standard deviation of the observations, as noted above in Eq. (1). Of course, if possible, recalculating metrics from older simulations and reports, using identical benchmark datasets and time periods for consistency, would allow for the cleanest comparison to the latest simulations. In any case, metrics continue to provide an invaluable and concise

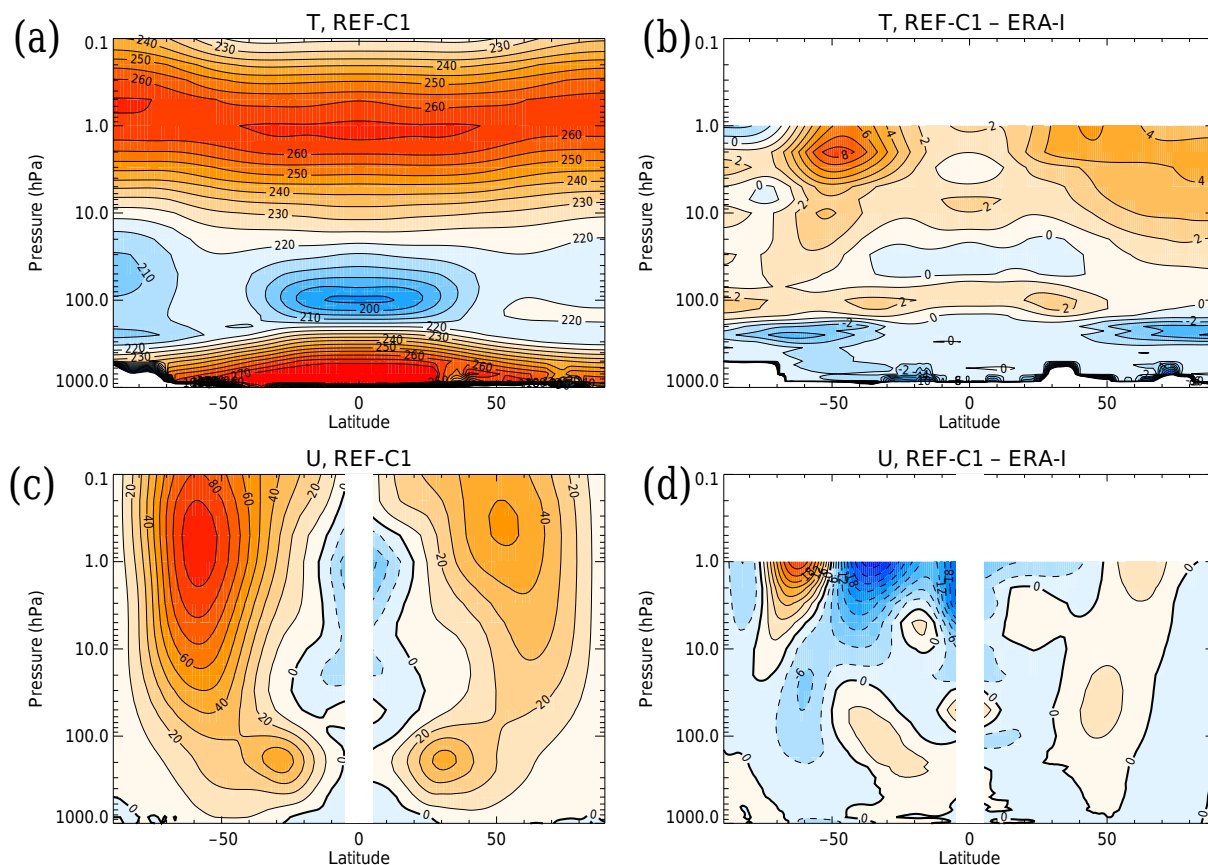
indication of current model performance, indicating diagnostics where models are performing well and those where improvement is required.

Comparing column 1 with columns 2 and 3 of Fig. 1, the free-running version of HadGEM3-ES is shown to perform better than UMUKCA-METO in 10 of the 14 metrics (with umx<sub>sh</sub> and SAO significantly better in UMUKCA-METO, and up<sub>70</sub> and PW<sub>sh</sub> better in UMUKCA-METO but not significantly so). Further, as noted above, the SAO metric is particularly sensitive to the period analysed, so the differences in this metric between UMUKCA-METO and the CCMI simulations cannot be considered reliable (i.e. robust across different periods). Thus, apart from the strength of the Southern Hemisphere polar night jet, the dynamics of HadGEM3-ES show improvements over (or no difference to) the version of HadGEM used for CCMVal-2 (documented in Morgenstern et al., 2010).

As denoted in Fig. 1 and Table 2, the metrics are divided into those that measure the mean climate of model simulations and those that measure their variability. This division follows that in Butchart et al. (2010, 2011). Figure 1 demonstrates quite clearly that, whilst the nudged simulations (columns 4–7) are graded similarly to the free-running simulations (columns 2–3) in terms of mean climate metrics (an aspect in which the free-running model is already very good, though again with the exception of the Southern Hemisphere polar night jet strength), the nudged simulations outperform the free-running simulations in terms of variability.

The nudged simulations that use the discontinuity corrected ERA-Interim dataset (McLandress et al., 2014; columns 4 and 5 of Fig. 1) show a better performance in the semi-annual oscillation metric than those without this correction (columns 6 and 7 of Fig. 1), although given that the evaluation is against the unmodified ERA-Interim dataset it is unclear why this should be the case. Certainly, it is expected that the only differences in performance between the nudged simulations with and without the discontinuities removed would be in the upper stratosphere (where the correction is applied) – a region assessed here only by the SAO metric.

The nudged simulations perform very well ( $g > 0.9$ ) in almost all metrics, with the exceptions of tropical upwelling (up<sub>70</sub> and up<sub>10</sub>) and the quasi-biennial oscillation (QBO). Surprisingly, at both 70 and 10 hPa the tropical upwelling in the free-running model is closer to the reanalysis than in the nudged model. Note, however, that due to the inherent noise and uncertainty in vertical velocities in reanalyses, vertical velocity is not nudged; only horizontal velocities,  $u$  and  $v$ , are nudged. If the nudged  $u$  and  $v$  winds do not have zero horizontal divergence then they will force spurious gravity and acoustic modes that will be reflected in spurious vertical velocities. Furthermore, if  $u$  and  $v$  are not in geostrophic balance then the nudging will introduce ageostrophic motions. Also note that upwelling (or, more particularly, the residual circulation) may not be entirely due to dynamics, as previ-



**Figure 2.** (a) Zonal-mean annual mean temperature for the REF-C1 simulation; panel (b) is the same as (a) but for differences between the REF-C1 simulation and ERA-Interim; (c) zonal-mean zonal wind, for December–January–February (Northern Hemisphere) and June–July–August (Southern Hemisphere), for the REF-C1 simulation; panel (d) is the same as (c) but for differences between the REF-C1 simulation and ERA-Interim. The years 1980–2010 are used.

ously thought, but perhaps also influenced by diabatic heating (Ming et al., 2016a, b), something that is not constrained in any of the simulations (except indirectly, by nudging the temperature field). Indeed, some transport calculations (e.g. for descent in the polar stratosphere; Tegtmeier et al., 2008) use the diabatic rather than the kinematic vertical velocity (see Butchart, 2014). Thus, even though they use the same numerical advection schemes, the stratospheric transport in nudged simulations need not be more accurate than in free-running models, as discussed in more detail below. Note also that in both the free-running and nudged simulations the tropical upwelling at 10 hPa is significantly closer to the reanalysis than upwelling at 70 hPa. This may be due to the model simulating a different structure of meridional circulation relative to that of the reanalysis (i.e. differences in shallow versus deep circulations; Birner and Bönisch, 2011).

The grading of the QBO metric below 0.8 for the nudged simulations is somewhat more surprising. Although the QBO is internally generated in the free-running REF-C1 and REF-C2 simulations, the QBO metric depends only on zonal wind which is directly nudged in the REF-C1SD simulations.

In fact, the nudged model accurately simulates the quasi-biennial oscillation in the zonal-mean winds at 20 hPa used in this metric, matching the reanalysis winds closely, except not quite reaching the peak values of the oscillation and thus underestimating the amplitude of the relevant Fourier harmonics by 4 % (not shown). However, since the power-spectrum approach inherent in this metric does not give a measure of uncertainty, this is calculated differently (by subsampling the data; Butchart et al., 2010). This produces an estimate of uncertainty that is small in magnitude and leads to this metric being very sensitive and thus lower than might be expected in the nudged simulations. Caution is therefore needed when interpreting this metric for any model. Indeed, the sensitivity of this metric is only apparent due to the use of nudged simulations, thus demonstrating the importance of the nudged simulations for testing the robustness and reliability of metrics involving quantities that are directly nudged.

Figure 1 shows that, whilst there are small differences between the nudged simulations with 24 and 48 h relaxation timescales, there are (with the exception of the SAO and heat flux metrics) no significant differences between the simula-



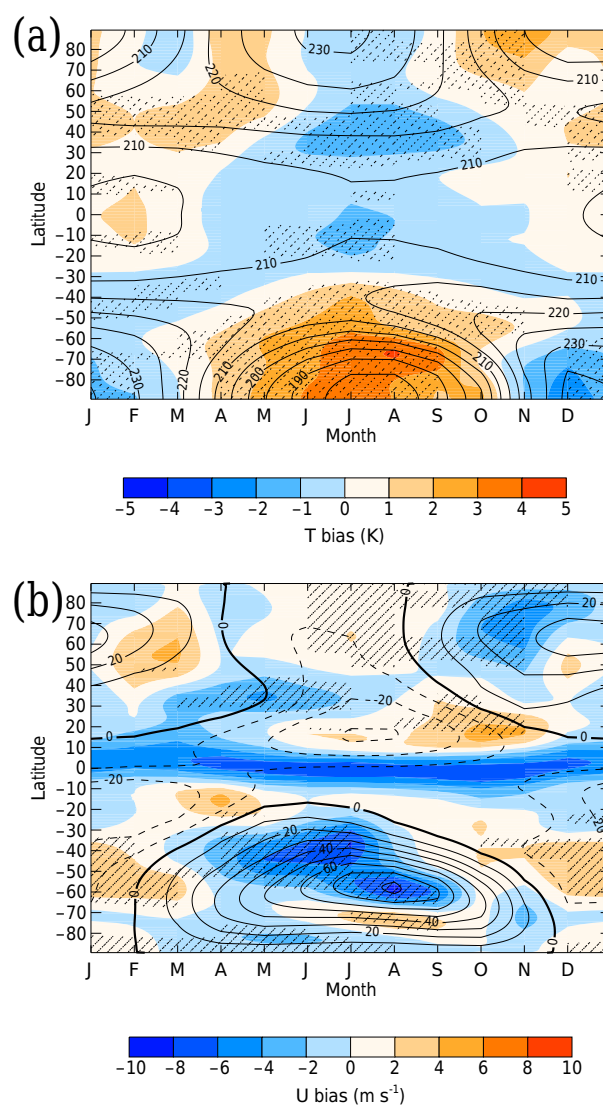
tions using smoothed and unsmoothed datasets. From this point on, we will just consider the simulations using the smoothed dataset, with a particular focus on the 24 h relaxation timescale integration (REF-C1SD-24 h, smoothed).

Despite the issues caused by changing the reanalysis dataset and analysing over a different period, it is worth noting that, if a “direct” comparison is made, then values for the free-running CCMI simulations (REF-C1 and REF-C2) are above the CCMVal-2 multi-model mean (Butchart et al., 2010) for 10 of the 14 metrics. The exceptions are the Southern Hemisphere jet maximum (umx\_sh), tropical mean upwelling at 70 hPa (up\_70), and the tropical annual cycle (tann) and semi-annual oscillation (SAO). Note also that, since the differences in the reanalysis dataset and period analysed cause the metric grades of the Met Office CCMVal-2 model (UMUKCA-METO) to get worse (as already noted above), this adds confidence that the CCMI model shows improvement over the CCMVal-2 model in terms of these metrics (assuming the differences when recalculating the grades of UMUKCA-METO can be considered representative of the CCMVal-2 multi-model mean).

### 3.2 Dynamics

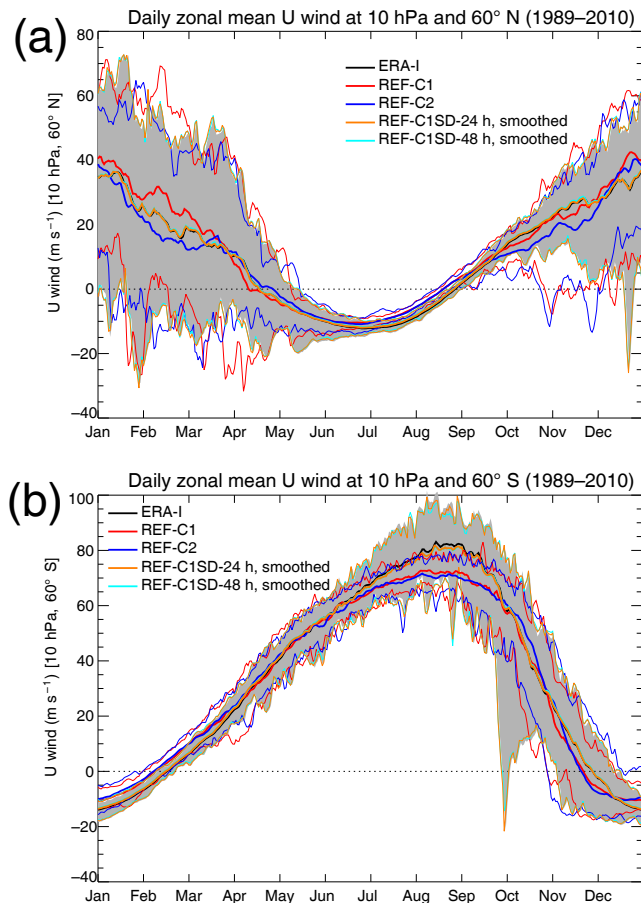
Figure 2 shows climatologies of the annual mean zonal-mean temperature and zonal wind in the REF-C1 simulation and biases in this simulation relative to ERA-Interim. A cold bias in the troposphere and a warm bias at the tropical tropopause, which have existed in all the Met Office HadGEM models (Hardiman et al., 2015), exist also in the REF-C1 simulation, but these biases are small ( $< 1$  K cold bias in the tropical troposphere, and a 1–2 K warm bias at the tropical tropopause; Fig. 2b). Also, as demonstrated in the metrics tmp\_nh and tmp\_sh in Fig. 1, the biases in extratropical temperature at 50 hPa are small ( $\sim 0.5$  K in the Northern Hemisphere and  $\sim 1$  K in the Southern Hemisphere). Temperature biases of up to 8 K do exist in the upper stratosphere, but these are less important than biases at the tropical tropopause (which influence stratospheric water vapour) and the extratropical lower stratosphere (which affect polar stratospheric cloud formation) and thus will not significantly affect model performance. Figure 2d shows that the strong eastward jet bias seen at around 1 hPa in the Southern Hemisphere (related to the poorly graded umx\_sh in Fig. 1) is accompanied by a westward bias just equatorward of the jet. This dipole structure to the bias is indicative of the jet being too strong because it is located too far poleward (possibly an issue with the way in which non-orographic gravity waves are attenuated in the upper stratosphere; Scaife et al., 2002). These biases in temperature and zonal wind are, as expected, largely removed in the nudged simulations (Fig. 1).

Figure 3 considers the seasonal cycle in temperature at 50 hPa (relevant to polar stratospheric cloud formation during winter and spring) and zonal wind at 10 hPa (a measure of polar vortex variability). Figure 3a shows that there are biases



**Figure 3.** Biases in the climatological seasonal cycle of the REF-C1 simulation, relative to ERA-Interim, for zonal-mean (a) temperature (50 hPa) and (b) zonal wind (10 hPa). Black contours show ERA-I values, with contour intervals of 5 K and  $10 \text{ m s}^{-1}$ , respectively, and coloured shading shows the bias (REF-C1 minus ERA-I), with contour intervals 1 K and  $2 \text{ m s}^{-1}$ , respectively. Stippling shows regions where the bias is statistically significant at the 95 % level as calculated using a  $t$  test. Tick marks indicate the middle of each month.

in the 50 hPa temperature in both the Northern Hemisphere and Southern Hemisphere high latitudes. The seasonal cycle in temperature is too weak in both hemispheres, but this signal is more pronounced in the Southern Hemisphere, with up to a 4 K warm bias seen in August. In both hemispheres, a warm bias of 1–2 K is seen in polar spring. In the nudged version of the model, temperature biases are largely removed, with biases at 50 hPa ranging from  $-0.88$  to  $+0.10$  K (not shown).



**Figure 4.** Polar vortex variability for the (a) Northern Hemisphere and (b) Southern Hemisphere. Thick solid lines show mean values, and maximum and minimum values are shown by thin solid lines for the model simulations and shading for ERA-I over the years 1989–2010. Tick marks indicate the middle of each month.

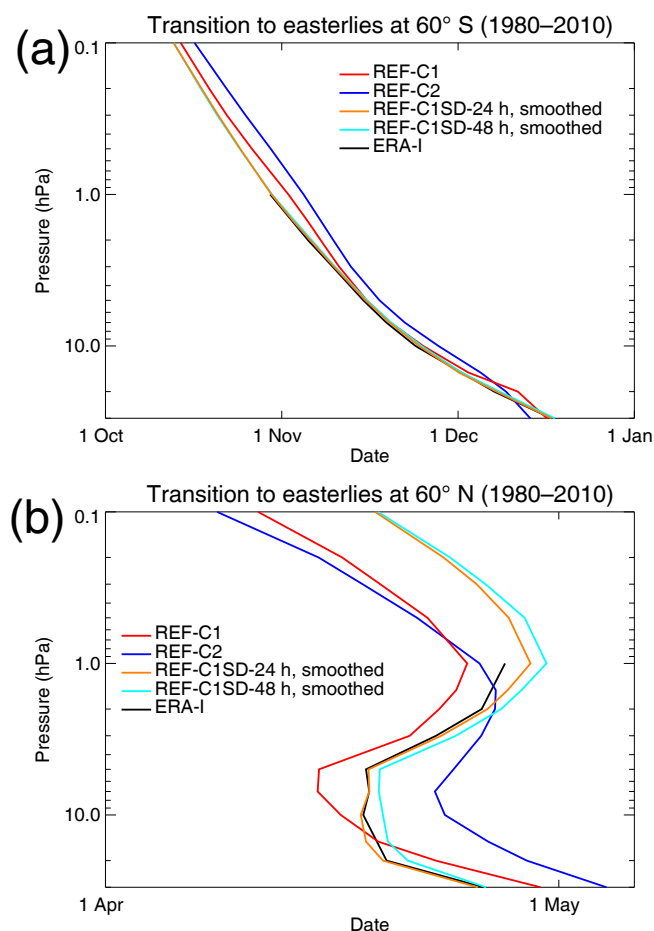
Figure 3b shows that the winter polar vortex (at 10 hPa) in both hemispheres is biased weak relative to the ERA-Interim reanalysis, consistent with the warm biases in the polar vortex shown in Fig. 3a. The weak bias is most significant in the Southern Hemisphere winter, with a negative bias of up to  $6 \text{ m s}^{-1}$  in magnitude seen there. Again, this bias is removed in the nudged model, with biases in zonal-mean wind at 10 hPa showing magnitudes between  $-0.92$  and  $+0.66 \text{ m s}^{-1}$ . For both 50 hPa temperature and 10 hPa zonal winds, the biases in the REF-C2 simulation resemble those found in REF-C1 and hence are not shown. However, the magnitude of warm biases in the extratropical Northern Hemisphere is greater in REF-C2, as discussed further below (see Fig. 6).

### 3.2.1 High latitudes

A detailed look at the strength and variability of the zonal-mean wind at 10 hPa in both hemispheres (Fig. 4) demonstrates that this is well simulated in the northern high latitudes in all seasons, with the free-running models showing a small negative bias and slightly too much interannual variability in October and November. However, the vortex strength and variability in Southern Hemisphere winter and early spring are too weak in the free-running models. Despite this, the time of the vortex breakup, determined as the time when the zonal wind transitions from eastward to westward, is shown to be very accurately simulated in both hemispheres. Since the polar vortex acts as a barrier to transport, this vortex breakup allows transport of ozone into and out of the polar region, impacting springtime TCO in the high latitudes. Accurate simulation of the vortex breakup time is also important since the dynamical impact of the Southern Hemisphere extratropical stratosphere on the troposphere is shown to be greatest during the time of the vortex breakup (Kidston et al., 2015).

Figure 5 shows this polar vortex breakup time at all altitudes for both hemispheres. This is accurately simulated in all simulations. The largest bias is seen in the Northern Hemisphere lower stratosphere for REF-C2 where the vortex breakup is around 10 days late, although even this is well within the 95 % confidence interval for vortex breakup times calculated using ERA-Interim (Hardiman et al., 2011). As mentioned above, we do not include this metric in Fig. 1 since we take a different approach to that in Butchart et al. (2010), using instead an approach used in previous multi-model studies (Eyring et al., 2006). Hardiman et al. (2011) demonstrated that the time of the “final warming” of the polar vortex can be adequately calculated using monthly mean data in both hemispheres, and can be accurately calculated using monthly mean data in the Southern Hemisphere where the vertical profile of the final warming time is far simpler than in the Northern Hemisphere. In multi-model studies (the primary use of metrics), this has the advantage of requiring lower volumes of model data, and it also removes the noise associated with daily data (something which is done in a less physically intuitive way, by using a low-pass filter, for the metric used in Butchart et al., 2010).

Of course, another important factor in determining the simulated heterogeneous ozone depletion is the area of the PSCs. In this study, the size of the area in which temperature at 50 hPa falls below 195 K is used as a proxy for the PSC area. Figure 6a shows that the average October daily PSC area in the Southern Hemisphere high latitudes is too low in the free-running model, consistent with the warm biases in the Southern Hemisphere high-latitude temperatures at 50 hPa shown in Fig. 3a. The average daily October PSC area across all years (1980–2010), in units of  $10^6 \text{ km}^2$ , is 1.0 in REF-C1, 1.6 in REF-C2, and 4.0 in both nudged simulations. The nudged simulations, as expected, show excel-



**Figure 5.** Polar vortex final warming times, as defined by the final transition from eastward to westward of the zonal-mean zonal wind at 60°, for (a) the Southern Hemisphere and (b) the Northern Hemisphere. Climatologies for the years 1980–2010 are shown.

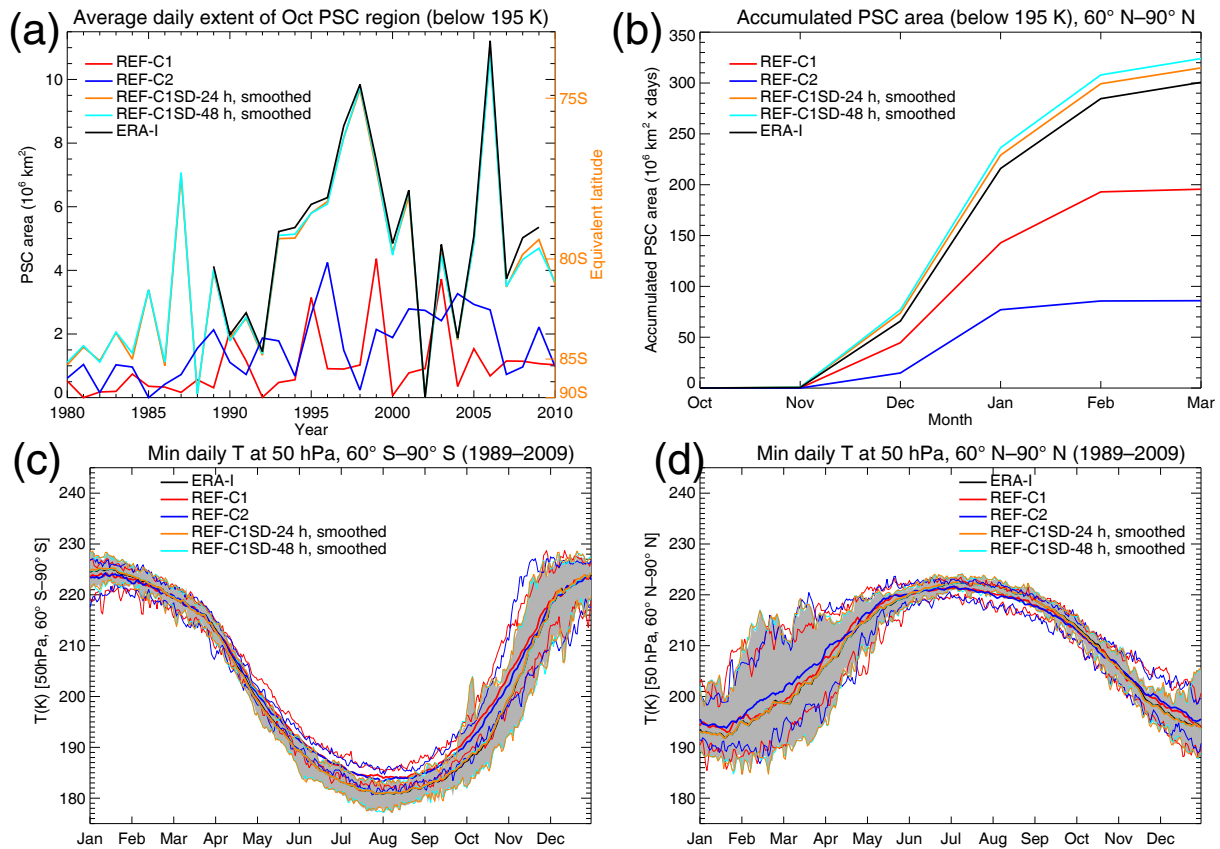
lent agreement with ERA-Interim in this diagnostic. Thus, PSC area in the free-running models is around one-third of the value as calculated from ERA-Interim temperatures, and this is likely to have implications for heterogeneous ozone depletion. Figure 6b shows that, similarly in the northern high latitudes, the accumulated PSC area throughout Northern Hemisphere winter in the free-running models is, on average, around half the value it should be (according to ERA-Interim). There is substantial variability in the accumulated PSC area found in earlier REF-C1 and REF-C2 simulations (not shown or documented here) such that the large differences in accumulated PSC area between the REF-C1 and REF-C2 simulations shown here lie within the expected variability. On average, the CCMval models were found to underestimate PSC area as compared to ERA40 (Butchart et al., 2011), and so this problem is not unique to HadGEM3-ES. Again, the nudged simulations show an accumulated PSC area that is in good agreement with ERA-Interim. Figure 6c and d show minimum daily temperatures at 50 hPa in the

southern and northern high latitudes, respectively, and show more clearly that the warm biases in the free-running simulations are somewhat larger in the Southern Hemisphere winter than in the Northern Hemisphere winter, with warm biases of up to 4 K seen in the Southern Hemisphere (consistent with Fig. 3a). The variability in these minimum daily temperatures is shown to be too large in October and November in the Southern Hemisphere of the free-running simulations, but to be in good agreement with the reanalysis in the Northern Hemisphere in all simulations.

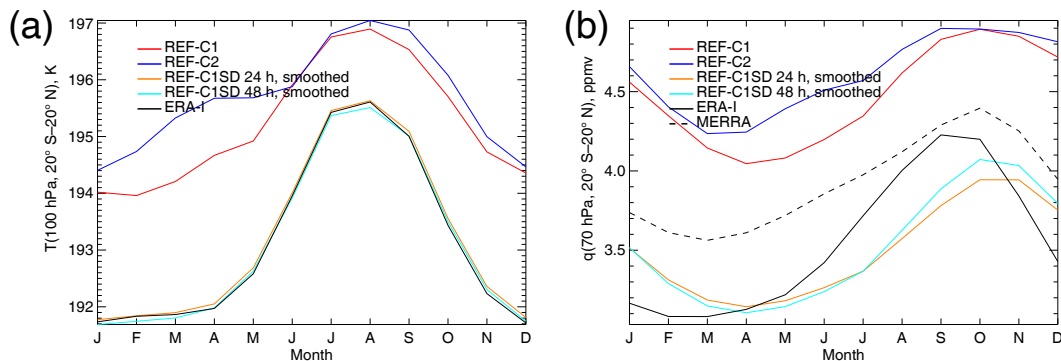
### 3.2.2 Tropics

Traditionally, the Met Office climate model has suffered from a warm bias in the tropical tropopause region (Hardiman et al., 2015) leading to very high stratospheric water vapour concentrations. In HadGEM3-ES, however, this bias is relatively small (around 1–2 K; see Fig. 7a), leading to concentrations of water vapour (Fig. 7b) that are only around 0.6 ppmv too high in the stratosphere relative to MERRA (Rienecker et al., 2011)<sup>1</sup>. The remaining 1–2 K bias in temperature is caused, in part, by simulated ozone concentrations that are too high (see Fig. 17 below and also O'Connor et al., 2009; Hardiman et al., 2015). The difference in 100 hPa tropical temperature between REF-C1 and REF-C2 in January–May (Fig. 7a) is localized to heights of around 150–50 hPa. Since this difference does not extend throughout the troposphere, it is thought unlikely to be due to differences in sea surface temperatures per se (Hardiman et al., 2007). The same difference as that seen in 100 hPa temperature is also seen in 70 hPa water vapour concentrations (Fig. 7b), though it is delayed by 2 months, consistent with the time taken for air parcels to rise from 100 to 70 hPa in the tropics. In all months, tropical tropopause temperature and water vapour concentrations in REF-C1 are closer to the observations than those in REF-C2 (Fig. 7). This may be expected, since REF-C1 is an atmosphere-only simulation, and thus forcing from sea surface temperatures will be in line with observations, whereas REF-C2 is a coupled atmosphere–ocean simulation. Temperatures in the nudged model are in line with observations (Fig. 7a) leading to lower water vapour concentrations (Fig. 7b). However, note that just nudging the temperatures and horizontal winds is not enough to remove any bias in water vapour concentrations (see also Hardiman et al., 2015). These are too low relative to the MERRA reanalysis by around 0.5 ppmv (Fig. 7b), although Fig. 7 of Hardiman et al. (2015) suggests that improvements to the ice microphysics scheme in more recent versions of HadGEM may account for a significant fraction of this bias. They also have an offset seasonal cycle, indicative of tropical upwelling that is too weak in the model (see Figs. 9 and 10 below).

<sup>1</sup>MERRA is used in Fig. 7b as it is shown in Hardiman et al. (2015) to more accurately reproduce water vapour concentrations than ERA-Interim, as compared against the SWOOSH dataset.



**Figure 6.** (a) Average daily October nitric acid trihydrate (NAT) PSC area, at 50 hPa, in the Southern Hemisphere, defined as the area poleward of 60° S with daily mean temperatures below 195 K. (b) Accumulated daily PSC area, at 50 hPa, in the Northern Hemisphere, defined as the area poleward of 60° N with daily mean temperature below 195 K. (c) Minimum 50 hPa daily mean temperature in the region 60–90° S. (d) Minimum 50 hPa daily mean temperature in the region 60–90° N. Thick and thin lines, and shading, in panels (c) and (d) are as in Fig. 4. All panels are averaged over the years 1989–2009. Note that the temperature is used as a proxy for PSC area here, and thus these are estimates of the PSC area seen by the interactive chemistry.

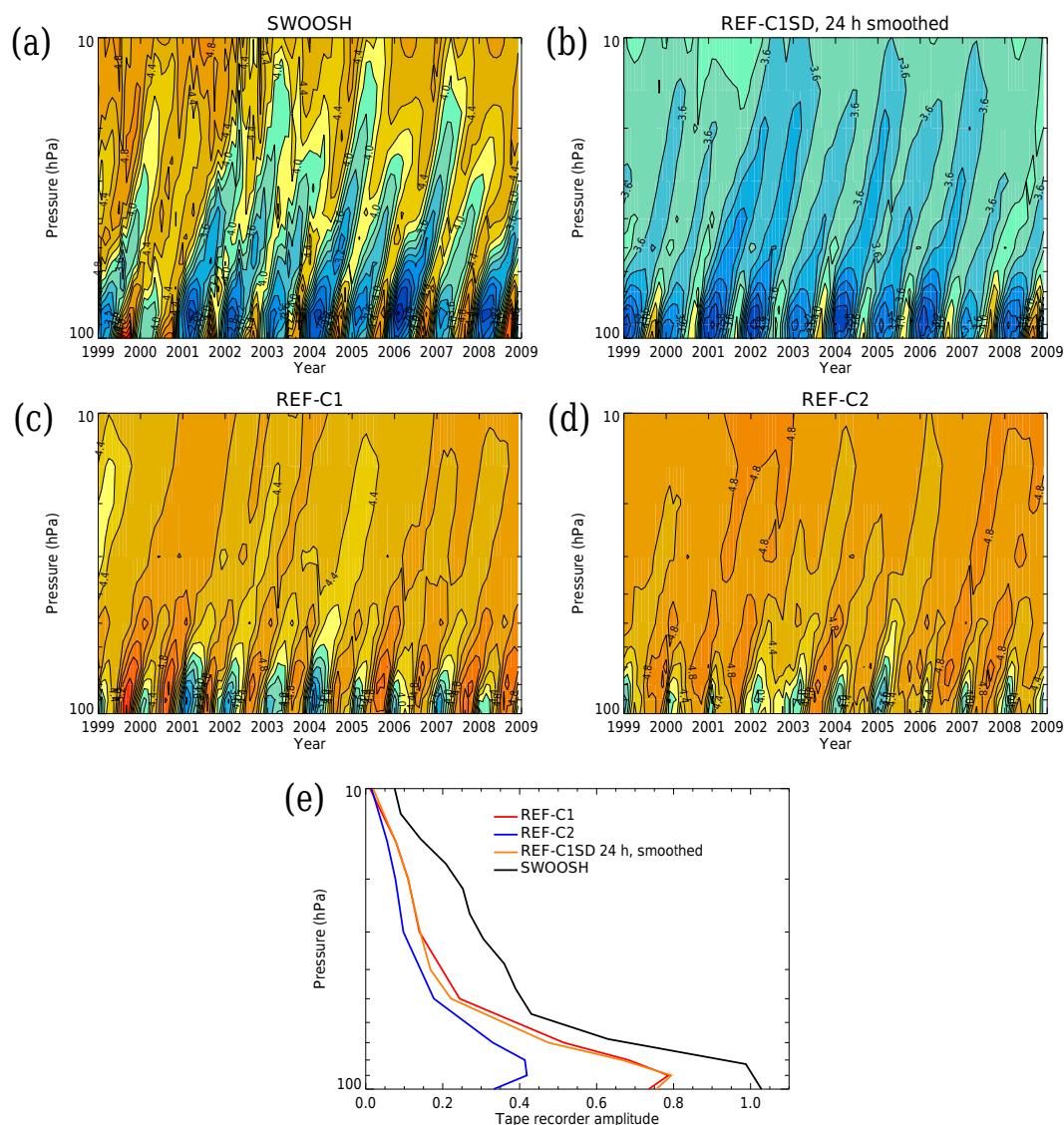


**Figure 7.** Tropical (20° S–20° N) seasonal cycle in (a) temperature ( $T$ ) and (b) water vapour ( $q$ ), averaged over the years 1980–1999, as compared to ERA-Interim reanalysis (for  $T$ ), and ERA-I and MERRA reanalyses (for  $q$ ). Tick marks indicate the middle of each month.

Accurate water vapour concentrations are very important for correctly simulating chemical species in the stratosphere, including ozone. Water vapour, although not constrained in the nudged model, is strongly influenced by the cold-point

temperature at the tropical tropopause. The annual cycle in cold-point temperature causes an equivalent annual cycle in water vapour concentrations entering the stratosphere in the tropics, and the upward transport of water vapour in the trop-

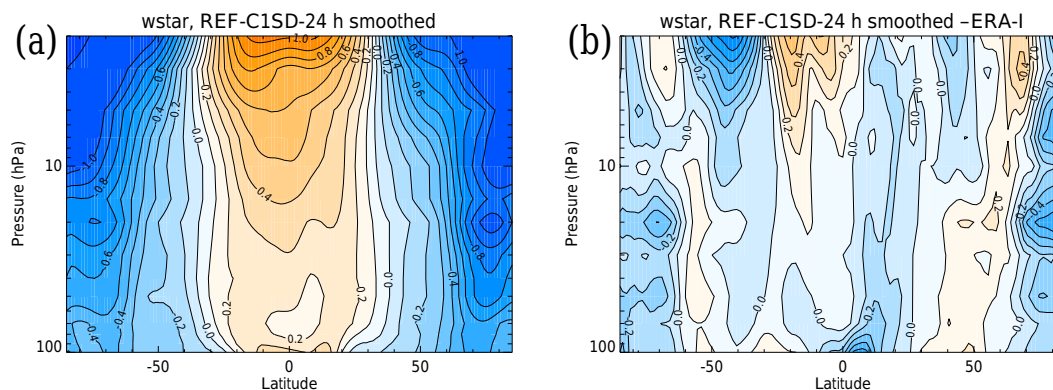




**Figure 8.** Tropical tape recorder signal,  $q$  (ppmv) averaged 10° S–10° N, for (a) SWOOSH data, and the (b) REF-C1SD 24 h smoothed, (c) REF-C1, and (d) REF-C2 simulations. (e) Amplitude of tape recorder calculated, at each height, as the amplitude of the Fourier harmonic corresponding to the annual cycle.

ics gives rise to the so-called “tape recorder” signal, shown in Fig. 8. Due to an 8 K warm bias in tropical tropopause temperature in the UMUKCA-METO CCMVal-2 simulation (Morgenstern et al., 2010), stratospheric water vapour had to be prescribed in that model and the tape recorder signal was therefore not simulated (Morgenstern et al., 2009). A significant improvement in the tropical tropopause temperature bias in HadGEM3-ES means that the tape recorder is simulated in this model. The tape recorder in the nudged (Fig. 8b) and free-running models (Fig. 8c–d) is compared against the Stratospheric Water and Ozone Satellite Homogenized dataset (SWOOSH – <http://www.esrl.noaa.gov/csd/groups/csd8/swoosh/>; Fig. 8a). The tape recorder signal appears more coherent much higher into the stratosphere in

the nudged simulation. However, Fig. 8e shows that this is not due to the amplitude of the annual cycle harmonic (the seasonal cycle in the tape recorder signal) being greater in the nudged simulation than in the free-running simulations. A reduced amplitude in some of the sub-annual harmonics in the nudged simulation (not shown) may explain the increased coherence. Whilst water vapour concentrations are slightly low in the mid-stratosphere in the nudged simulation (by 0.53 ppmv at 30 hPa), they are closer to observations in the lower stratosphere than in the free-running model. Water vapour concentrations are slightly high in the free-running model (by 0.42 ppmv in REF-C1 and 0.57 ppmv in REF-C2 at 30 hPa). However, sensitivity experiments in a different version of the HadGEM3 model have shown changes in wa-



**Figure 9.** Zonal-mean annual mean climatologies in residual vertical velocity for (a) REF-C1SD (nudged simulation) and (b) differences between the REF-C1SD simulation and ERA-Interim. The years 1989–2009 are used. Unlike temperature and zonal wind, the biases in residual vertical velocity are not negligible for the nudged simulations (see text for details).

ter vapour  $< 0.75$  ppmv to have no significant impact on the simulated stratospheric chemistry (not shown).

Whilst temperatures and horizontal winds are forced close to the ERA-Interim reanalysis in the nudged model, vertical winds are notoriously difficult to simulate accurately and are therefore not nudged. Figure 9 demonstrates that, as shown in Fig. 1, nudging temperature and horizontal wind fields does not imply that the simulated vertical wind field will also be close to the reanalysis (and, further, there is reasonable agreement in the average magnitude of the vertical wind field across different reanalyses; Butchart et al., 2011; Abalos et al., 2015). At some locations, the biases in residual vertical velocity in the nudged simulations (Fig. 9b) are of the same magnitude as the absolute values (Fig. 9a).

Although the HadGEM3-ES simulations do capture the double-peaked nature of the 70 hPa residual vertical velocity in the tropics (Fig. 10a), like other models the peaks are too hemispherically symmetric (Butchart et al., 2010) and are biased low in both hemispheres. As a consequence, the upwelling mass flux from troposphere to stratosphere (Fig. 10b) is too weak, particularly in the nudged simulations. Figure 10a and b show values of vertical velocity and upwelling, respectively, to be around 20 % lower in REF-C1SD-24 h than in the free-running simulations. This weak bias is much greater in the Northern Hemisphere winter (Fig. 10c) than in the Southern Hemisphere winter (Fig. 10d). Thus, Figs. 9 and 10 show that the stratospheric circulation is very difficult to simulate accurately, even in nudged simulations.

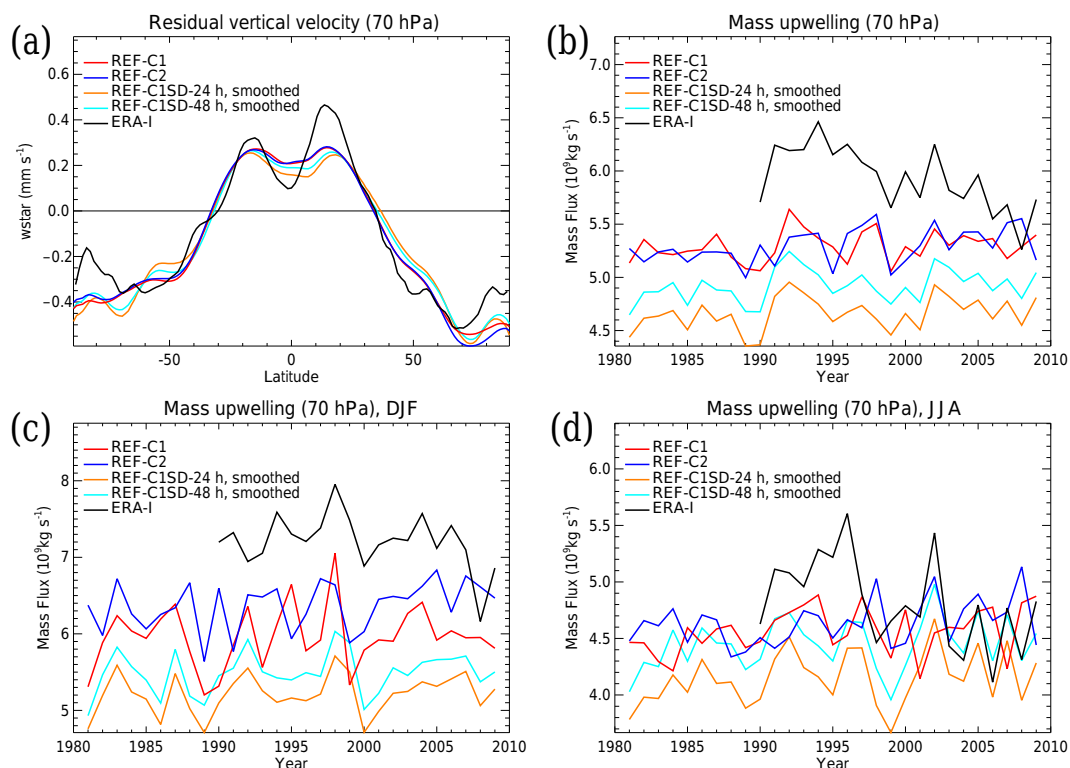
An alternative diagnostic of the strength of stratospheric transport is the so-called age of air (Fig. 11). The mean age of stratospheric air (Waugh, 2009b) denotes the time since that parcel of air was last in contact with the troposphere and thus gives an indication of the rate of transport to different regions within the stratosphere. Figure 11a shows that age of air is too old in the lower stratosphere in the tropics (by up to 0.5 years compared to age inferred from CO<sub>2</sub> observations) – consistent with too little upwelling shown in

Fig. 10b. However, age of air is too young throughout much of the stratosphere (Fig. 11b), which cannot be explained by biases in upwelling from the troposphere to the stratosphere alone (Birner and Bönisch, 2011). Nonetheless, the age simulated by HadGEM3-ES represents a significant improvement on that seen in the Met Office UMUKCA-METO CCMVal-2 simulation (Morgenstern et al., 2010), in which stratospheric air was 1–2 years too old. Moreover, the age simulated by HadGEM3-ES is in much better agreement with observations (Fig. 11). Furthermore, Linz et al. (2016) argue that it is the latitudinal gradient in age of air, and not age itself, that best diagnoses the strength of the meridional mass circulation and that this gradient, at any height, is independent of the circulation above. This latitudinal gradient is much improved in the HadGEM3-ES model as compared to UMUKCA-METO. For example, at 21 km, the latitudinal gradient (35–45° N to 10° S–10° N) in HadGEM3-ES is 1.7 years, in line with the observations, whereas it is 3.2 years in UMUKCA-METO.

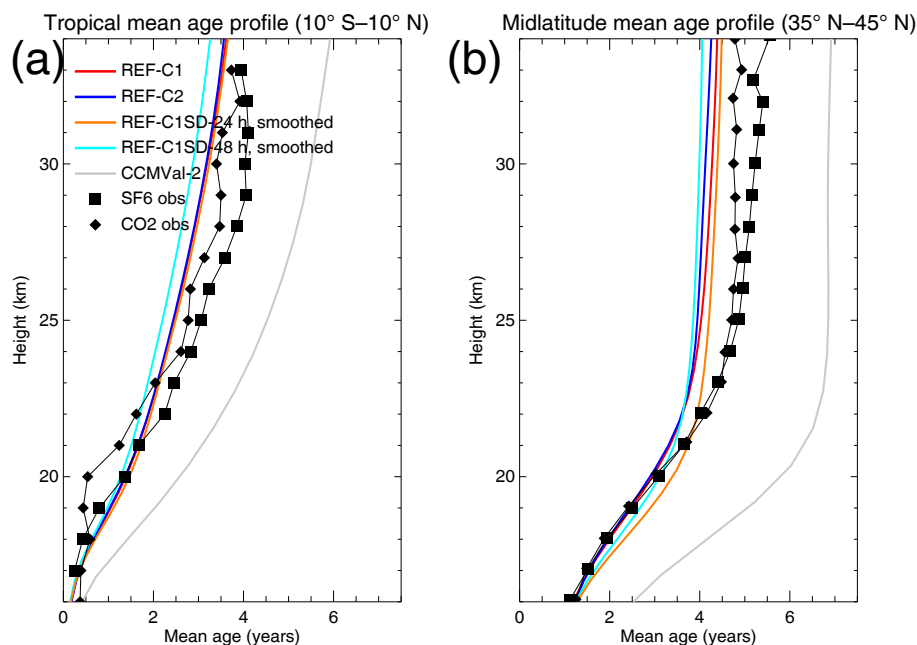
### 3.3 Ozone

Figure 12 shows time series of TCO as simulated in the free-running and nudged models, compared to the Total Ozone Mapping Spectrometer (TOMS) satellite data (McPeters et al., 1998). Near-global (60° S–60° N) annual mean ozone (Fig. 12a) is biased high relative to observations by around 10 Dobson units (DUs). Near-global ozone loss is slightly stronger in the nudged model than in the free-running model, such that near-global TCO in the nudged model agrees well with the TOMS data after around 1990.

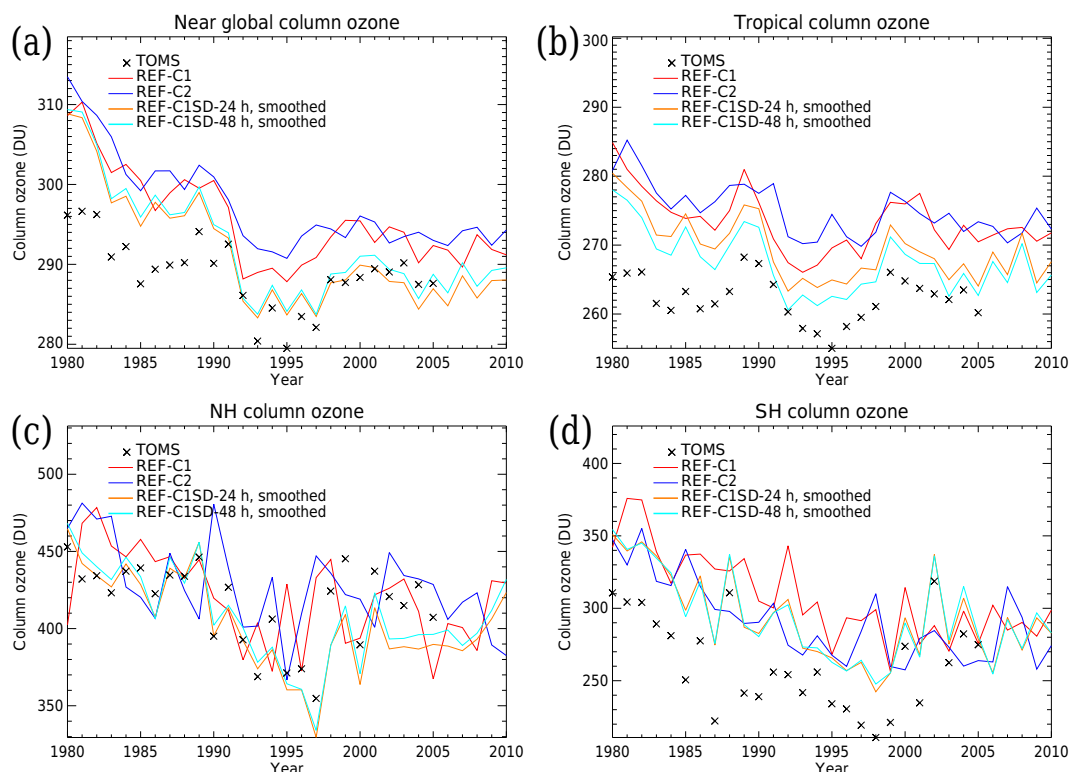
Figure 13a shows the global net annual mean stratosphere–troposphere exchange (STE) of ozone (i.e. the net mass flux of ozone across the tropopause; see caption of Fig. 13 for details). Consistent with Fig. 10b, which showed the tropical mass upwelling from the troposphere to the stratosphere to be biased weak, the STE ozone flux in the model simulations is found to be too low as compared to ERA-Interim. Currently,



**Figure 10.** (a) Residual vertical velocity at 70 hPa (1989–2009) and tropical mass upwelling through 70 hPa for (b) annual mean, (c) December–January–February, and (d) June–July–August, as calculated for free-running simulations, nudged simulations, and ERA-Interim. Mass upwelling in (b) is calculated using seasonal means as in Butchart et al. (2010), such that the annual means plotted above the tick marks refer to December–November means.



**Figure 11.** Stratospheric age of air (1990–2010) in the (a) tropics (observations from Andrews et al., 2001) and (b) Northern Hemisphere midlatitudes (observations from Engel et al., 2009). The period 1990–2010 was chosen for CCM1 model simulations to allow for age of air to adjust during the first 10 years of the nudged simulations. The period 1980–2000 was used for the CCMVal-2 model simulation (historical period only). The exact period chosen makes very little difference to the diagnosed age of air (not shown).



**Figure 12.** Total column ozone (TCO): (a) annual mean near global ( $60^{\circ}$  S– $60^{\circ}$  N), (b) annual mean tropics ( $20^{\circ}$  S– $20^{\circ}$  N), (c) Northern Hemisphere March ( $60$ – $90^{\circ}$  N), and (d) Southern Hemisphere October ( $60$ – $90^{\circ}$  S).

the best estimate of STE ozone flux inferred from observations is  $550 \pm 140 \text{ Tg O}_3 \text{ year}^{-1}$  (Olsen et al., 2001); thus, even the ERA-Interim estimate of STE ozone flux is around  $250 \text{ Tg O}_3 \text{ year}^{-1}$  too low. Figure 13b and c show that, consistent with Fig. 10c and d, the bias in STE ozone flux (as compared to ERA-Interim) is more prominent in the Northern Hemisphere winter than in the Southern Hemisphere winter. The similarity between Figs. 10 and 13 demonstrates the influence of the stratospheric meridional circulation on the STE ozone flux. A bias in STE ozone flux will have implications for extratropical tropospheric climate (see Sect. 7.3 of Butchart, 2014), surface ozone concentrations (e.g. Zhang et al., 2014), and the global tropospheric ozone budget (Wild, 2007; Young et al., 2013).

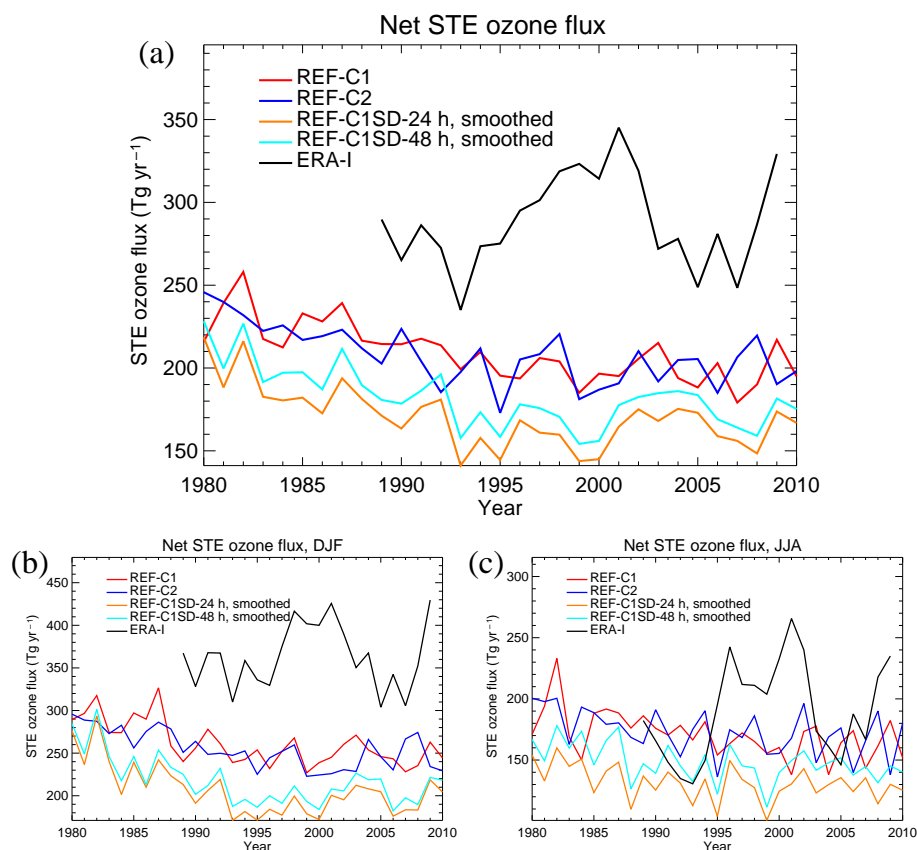
### 3.3.1 High latitudes

The change in TCO in the high latitudes, during the period 1980–2010, is similar in all simulations (Fig. 12c, d) and agrees well with the TOMS observations. However, TCO that is too high is indicative of an ozone hole that is too small in area. Further, we have seen 50 hPa temperatures biased high in the free-running model (Fig. 3a), PSC areas biased too low (Fig. 6), and negative biases in the Southern Hemisphere polar vortex strength (Fig. 4b). Figure 14 shows TCO over the South Pole in October, averaged over the years 1997–

2002, as compared against the 220 DU contour from the TOMS satellite data averaged over the same 6 years. Southern Hemisphere high-latitude TCO is biased high, by around 40 DU, in all versions of the model (Fig. 12d). Figure 3-11c from Chap. 3 of WMO (2011) shows this bias to be within the 95 % prediction interval of the CCMVal-2 model simulations. Nevertheless, this bias leads to a simulated ozone hole (area with TCO values below 220 DU that is too small. Hence, an accurate simulation of PSC areas (Fig. 6a) is insufficient to eliminate errors in the areal extent of the ozone hole in HadGEM3-ES, at least when the nudging is to ERA-Interim temperatures. On the other hand, the nudging does remove errors in the orientation of the ozone hole which is slightly displaced from the pole (Fig. 14). The phase of the “croissant” shape in maximum ozone around  $60^{\circ}$  S is also more accurately simulated in the nudged model, with a minimum value around  $50^{\circ}$  W, in line with TOMS. In the free-running simulations, the location of the minimum varies from around  $60^{\circ}$  W to around  $110^{\circ}$  W. Whilst REF-C1 simulates a more accurate phase than REF-C2, errors are most pronounced from  $60^{\circ}$  E to  $30^{\circ}$  W, where TCO is too high at  $60^{\circ}$  S.

Northern high-latitude zonal-mean TCO is very well simulated (Fig. 12c). In terms of azonal ozone structure, conclusions for the Northern Hemisphere (Fig. 15) are the same as for the Southern Hemisphere. The amplitudes of the two



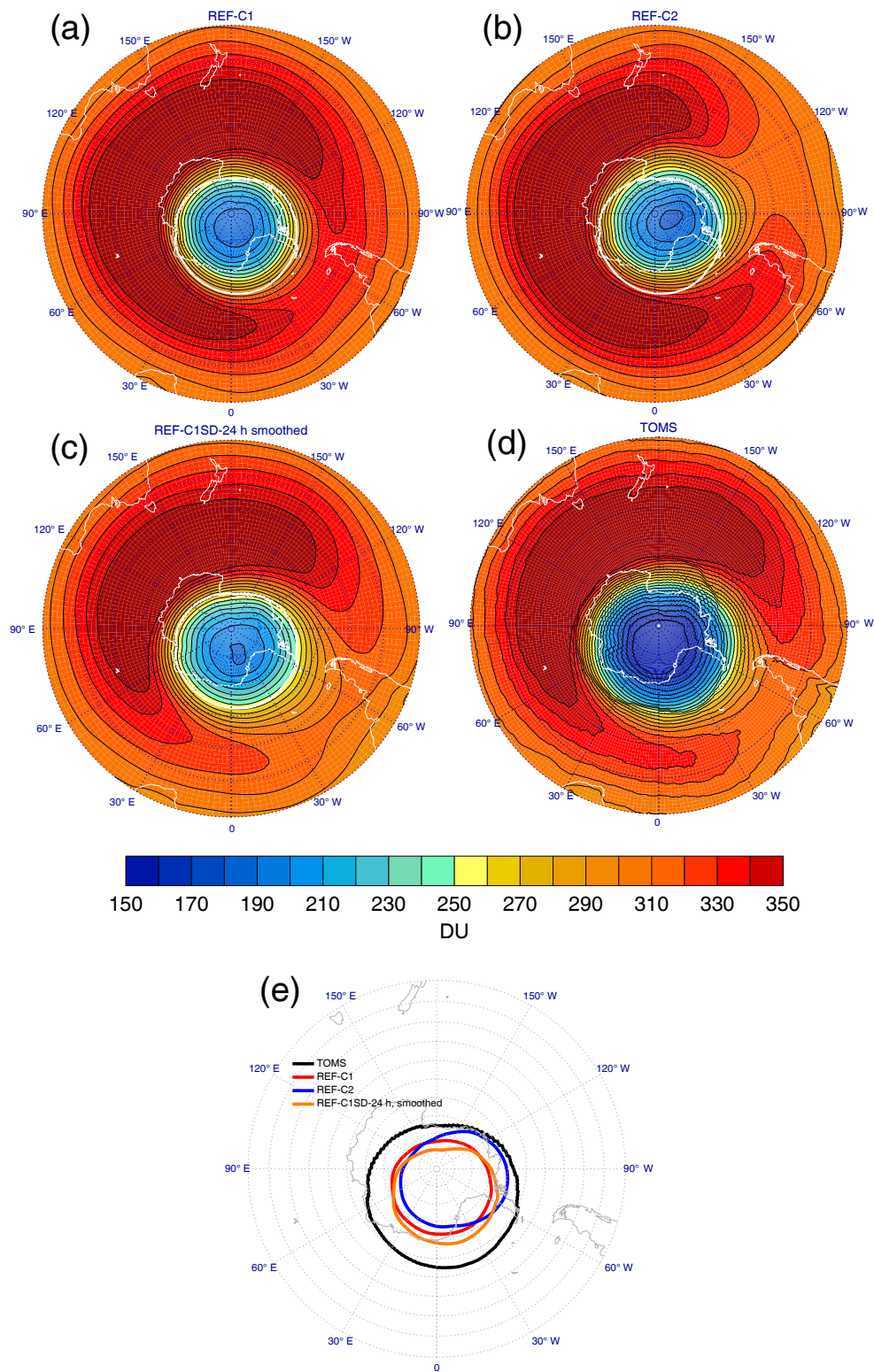


**Figure 13.** Stratosphere–troposphere exchange of ozone for (a) annual mean, (b) December–January–February, and (c) June–July–August. This flux of ozone across the tropopause is calculated using monthly mean residual vertical velocity and ozone mass mixing ratio, following Hegglin and Shepherd (2009). The tropopause is here defined as the 100 hPa surface equatorward of 50° and the 200 hPa surface poleward of 50°.

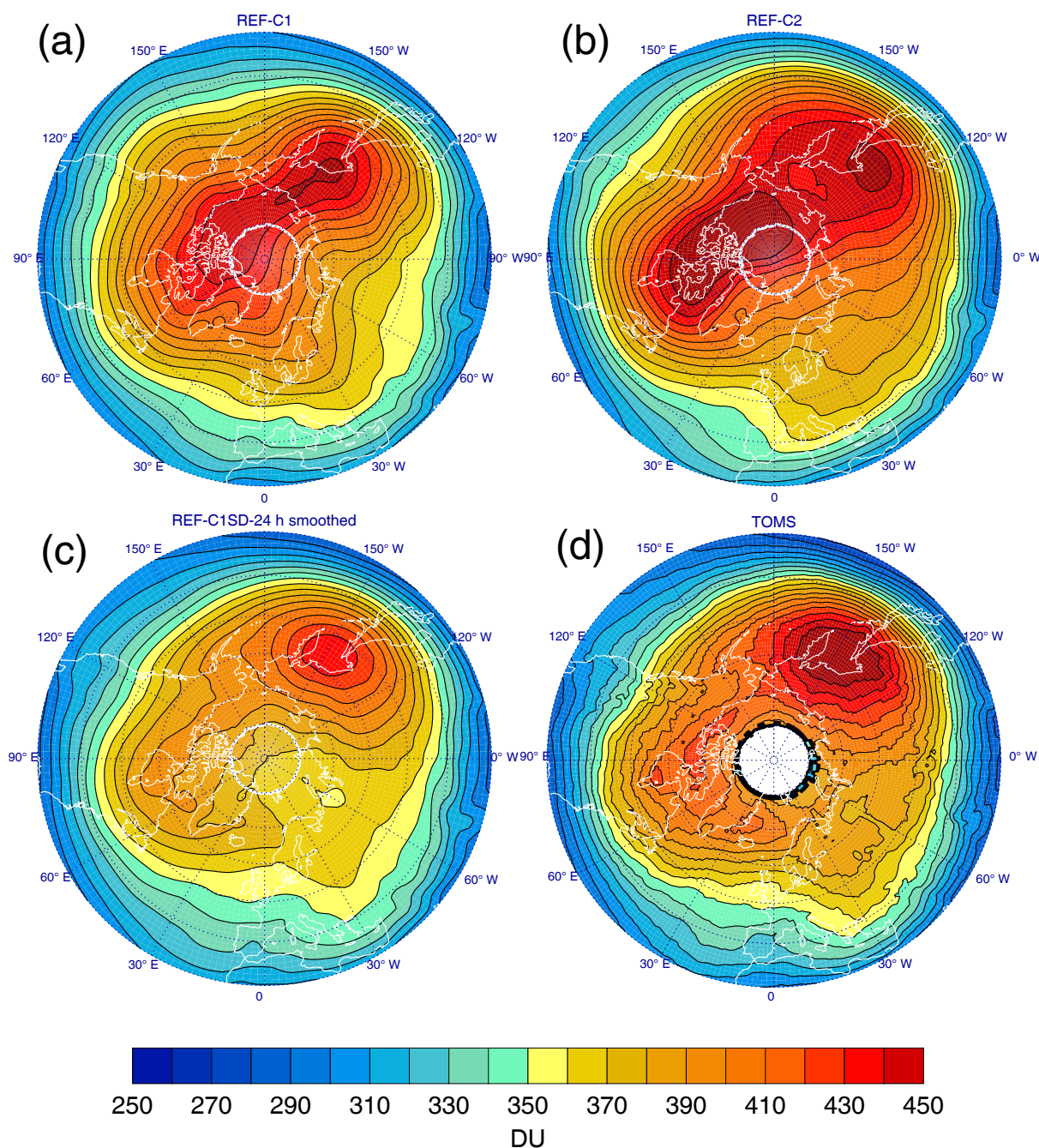
ozone maxima simulated around 120° E and 140° W are similar in the free-running model (especially in REF-C2). In the nudged simulation, however, the amplitude of the 150° W maximum is far greater than that of the 120° E maximum, in closer agreement with TOMS. Biases in the zonal asymmetry of ozone (i.e. the croissant shape in the Southern Hemisphere and larger maximum around 150° W in the Northern Hemisphere) arise due to corresponding biases in the amplitude and phase of the planetary stationary waves in the stratosphere which, again, are eliminated by the nudging. The fact that free-running models in general are unable to reproduce the correct phase (and amplitude) for the stationary waves (see Figs. 8 and 9 of Butchart et al., 2011) makes it rather difficult to determine what phase to include when prescribing zonally asymmetric ozone forcings in models without interactive chemistry. In the absence of improvement to the simulated phase of stationary waves, the results here show that prescribing zonally asymmetric ozone will almost always lead to different TCO from that obtained by the same model using self-determined ozone.

A further way in which dynamics influence ozone concentrations is through the enhanced poleward transport that fol-

lows sudden stratospheric warmings (SSWs; Akiyoshi et al., 2016). Figure 16 shows the average positive ozone anomaly following a SSW, which increases ozone concentrations by around 15 % compared to their climatological values. In the middle stratosphere, where ozone is dynamically controlled, the anomalies in the nudged simulation agree well with ERA-Interim but at higher levels, where chemistry starts to dominate, the anomalies are too large (cf. Fig. 16b, e and a, d). Equally, without nudging, the model simulates a realistic adiabatic temperature increase, associated with the SSWs (cf. Fig. 16i and g), and consequently realistic ozone anomalies in the month following the SSWs (cf. Fig. 16c, f and a, d) but, interestingly, the structure of these temperature and ozone anomalies in the upper stratosphere is less accurate than in the nudged simulation. As well as SSWs influencing ozone, it is also the case that zonally asymmetric ozone can increase the frequency of simulated SSWs (Albers et al., 2013), thus creating the possibility for a feedback in models with interactive chemistry.



**Figure 14.** Climatological TCO during October in the Southern Hemisphere for (a) REF-C1, (b) REF-C2, (c) REF-C1SD-24 h (smoothed), and (d) TOMS. Panel (e) indicates the ozone hole, defined as the 220 DU contour. White contour in (a), (b), and (c) shows TOMS 220 DU contour. TCO in REF-C1SD is still biased high, but the ozone hole has the correct shape. The years 1997–2002 are used in all cases.



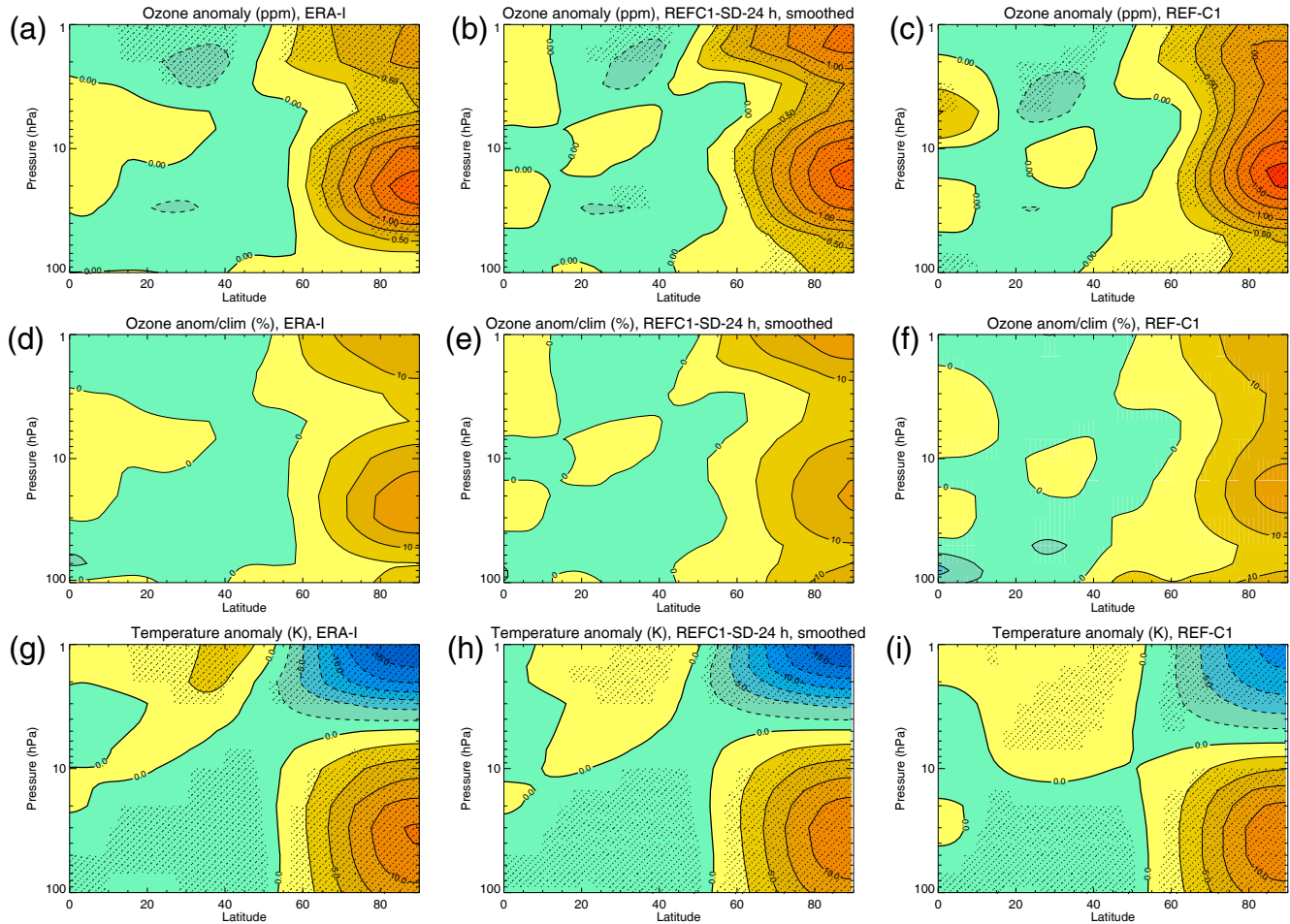
**Figure 15.** The same as in Fig. 14, but for climatological TCO in Northern Hemisphere March.

### 3.3.2 Tropics

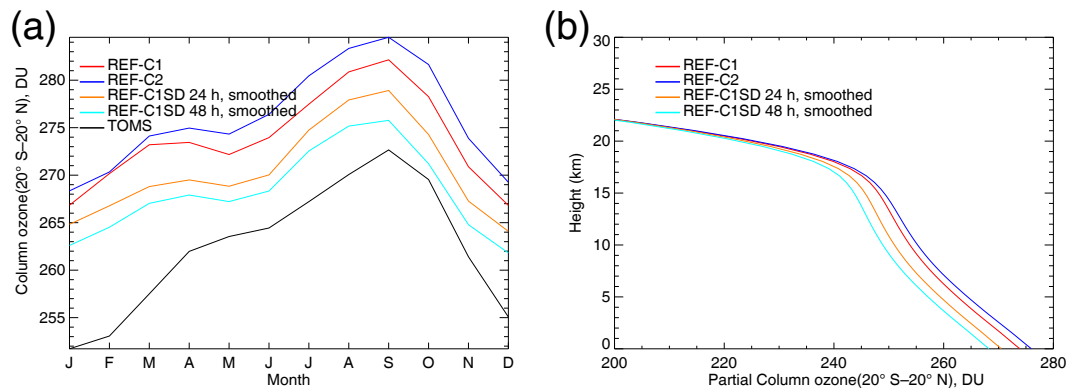
The simulated interannual variability in tropical TCO (Fig. 12b), in both free-running and nudged simulations, agrees well with the observations. However, all simulations show a  $\sim 6$  DU reduction in TCO over the period 1980–1995 which is much larger than the observed reduction of  $\sim 2$  DU (consistent with Fig. 3-6a from Chap. 3 of WMO, 2011). Furthermore, TCO is again biased high, with average biases of

12.6 DU in the free-running model and 7.0 DU in the nudged model (Fig. 12b). The largest biases, relative to TOMS, occur in December–January–February (Fig. 17a). As noted in Fig. 7, the bias in REF-C1 is smaller than that in REF-C2. Whilst tropical temperature and water vapour concentrations can influence TCO, they are clearly not the only influences on simulated tropical ozone. Cold-point temperature is constrained to reanalyses in the nudged model and water vapour concentrations in the nudged model are too low relative to





**Figure 16.** Anomalies, averaged over the 30 days following a stratospheric sudden warming, in (a)–(c) ozone volume mixing ratio (ppmv), (d)–(f) ozone as a percentage of climatological values, and (g)–(i) temperature (K) for ERA-Interim, the 24 h nudged simulation, and the free-running REF-C1 simulation. Stippling shows regions where the anomalies are statistically significantly different from zero, with 95 % confidence, as calculated using a  $t$  test.



**Figure 17.** (a) Seasonal cycle in tropical TCO, averaged over the years 1980–1999, as compared to TOMS satellite data. Tick marks indicate the middle of each month. (b) Vertical profile of partial column ozone, integrated downwards from the top of the model.

MERRA (Fig. 7), yet TCO, although improved, is still too high even in the nudged model (Fig. 17a). Figure 17b shows that this high bias primarily occurs in the tropical tropopause region (as shown also for the Met Office CCMVal-2 model by Fig. 7 of Gettelman et al., 2010), and thus the bias exists throughout the troposphere.

#### 4 Conclusions

This study analyses the historical period (1980–2010) of free-running and nudged simulations using HadGEM3-ES, the Met Office chemistry–climate model as configured for inclusion in the Chemistry–Climate Model Initiative. In the nudged model configuration, the relaxation timescale of the applied nudging was found to be important (Merryfield et al., 2013) although it was not the case that a single timescale could be found in which all metrics were improved. In the present study, 24 and 48 h nudging timescales were both found to give good results overall for the stratospheric fields considered.

Metrics of dynamical processes relevant for the simulation of stratospheric ozone were calculated for all model configurations. These were compared against the metrics as recalculated over the period 1980–2010 for the previous model configuration, UMUKCA-METO, used in CCMVal-2 (Morgenstern et al., 2010). The free-running model configuration is shown to have significantly improved since the UMUKCA-METO configuration, performing better in 10 of the 14 metrics considered here. The grades associated with some metrics were found to be sensitive to the reanalysis period used, implying that the period used should be of a sufficient length to reduce the impact of interannual variability. As such, a direct backward comparison of the metric grades in this paper to those of the CCMVal-2 model simulations (Butchart et al., 2010) is not possible. However, assuming that the change in the grades awarded to the UMUKCA-METO simulation (as recalculated using the period 1980–2010) is representative of that for other chemistry–climate models, it is likely that the HadGEM3-ES free-running model performs better than the CCMVal-2 multi-model mean in 10 of the 14 metrics.

Particularly significant improvements to the free-running model are that HadGEM3-ES no longer suffers from the large positive bias in stratospheric age of air or large warm bias in tropical tropopause temperature that were present in UMUKCA-METO (Morgenstern et al., 2009). More realistic stratospheric water vapour concentrations make HadGEM3-ES more suitable for accurately simulating stratospheric ozone concentrations (Hardiman et al., 2015). Issues do remain with the free-running model climatology, however. The seasonal cycle in extratropical winds and temperatures is found to be slightly weak in the model. This is most noticeable in the Southern Hemisphere polar vortex, which is too weak (by up to  $6 \text{ m s}^{-1}$ ) and therefore too warm (by up to

4 K). There are also ongoing moderate biases in temperature, water vapour, ozone, and upwelling mass flux in the tropics.

Metrics are split into those assessing mean climate and those assessing variability. The mean climate was found to be well simulated in both free-running and nudged versions of HadGEM3-ES with the notable exception of stratospheric transport, as diagnosed by the upwelling mass flux in the tropics. Vertical velocities are very noisy in reanalysis data (Butchart, 2014) and therefore cannot be nudged towards. As such, the diabatic component of stratospheric transport is difficult to constrain, even in nudged simulations. However, the variability in the nudged simulations was found to be significantly closer to the reanalysis than the variability in the free-running simulations. The nudged simulations showed grades above 0.9 for all variability metrics, except that diagnosing the accuracy of the quasi-biennial oscillation. In this case, the measure of variability used for the quasi-biennial oscillation was found to make the metric too sensitive in general, demonstrating the use of nudged simulations for ensuring the robustness and reliability of metrics involving quantities that are directly nudged.

Comparison of the free-running model climatology to that of the nudged version shows that accurately simulated dynamics, specifically temperature and horizontal wind fields, do play a role in the spatial structure of the ozone hole. This structure is correct in both hemispheres in the nudged model. However, the high ozone biases that exist in the tropics and southern high latitudes of the free-running model persist also in the nudged model, and these are therefore not solely attributable to biases in the dynamical fields. Thus, despite the fact that the area of Southern Hemisphere polar stratospheric clouds is correctly simulated in the nudged model, the ozone hole area, defined as the area over which TCO drops to below 220 DU, is too small in both free-running and nudged models (an issue which is not unique to HadGEM3-ES, as shown by Fig. 1 of Austin et al., 2010).

Tropical TCO is improved in the nudged simulations over that seen in the free-running model, but is still biased high relative to observations, with these biases occurring in the tropical tropopause region. It is worth noting that both water vapour and TCO are not perfect in the nudged simulation, and significant biases in the simulated transport and chemistry still exist in this model.

The fact that tropical upwelling and the stratospheric meridional circulation are found difficult to constrain and, indeed, are found to be worse in the nudged simulations than in the free-running simulations, means that ozone fluxes, in particular from the stratosphere to the troposphere, are not well constrained in the nudged model either, with obvious implications for the simulated extratropical tropospheric ozone budget. Again, this issue is not unique to HadGEM3-ES – even the ERA-Interim reanalysis shows ozone fluxes from the stratosphere to the troposphere with only around half the value inferred from observations.

In summary, biases in transport and ozone remain in the nudged simulations, demonstrating that these biases are not solely due to the model dynamics. Nevertheless, HadGEM3-ES is found to have good climatology and variability in basic meteorological fields, and a realistic simulation of stratospheric ozone loss. HadGEM3-ES represents a significant improvement over its predecessor, UMUKCA-METO.

**Code and data availability.** Due to intellectual property right restrictions, we cannot provide either the source code or documentation papers for the Unified Model (UM). The Met Office Unified Model is available for use under licence. A number of research organizations and national meteorological services use the UM in collaboration with the Met Office to undertake basic atmospheric process research, produce forecasts, develop the UM code and build and evaluate Earth system models. For further information on how to apply for a licence, see <http://www.metoffice.gov.uk/research/modelling-systems/unified-model>. JULES is available under licence free of charge. For further information on how to gain permission to use JULES for research purposes, see <https://jules.jchmr.org/software-and-documentation>. The model code for NEMO v3.4 is available from the NEMO website ([www.nemo-ocean.eu](http://www.nemo-ocean.eu)). Upon registering, individuals can access the code using the open-source subversion software (<http://subversion.apache.org/>). The revision number of the base NEMO code used for this paper is 3309. The model code for CICE is freely available from the United States Los Alamos National Laboratory (<http://oceans11.lanl.gov/trac/CICE/wiki/SourceCode>), again using subversion. The revision number for the version used for this paper is 430. The data will be submitted to the British Atmospheric Data Centre (BADC) database for the CCM1 project.

**Author contributions.** Steven C. Hardiman wrote Sects. 1, 3, and 4 of the paper and produced the figures. Fiona M. O'Connor wrote Sect. 2 of the paper. Steven C. Hardiman, Neal Butchart, and Fiona M. O'Connor contributed to running model integrations and to discussion on the structure and content of the paper. Steven T. Rumbold processed the chemistry and aerosol emissions datasets used in model integrations.

**Acknowledgements.** The work of Steven C. Hardiman, Neal Butchart, Fiona M. O'Connor, and Steven T. Rumbold was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). Steven C. Hardiman and Neal Butchart were also supported by the European Commission's 7th Framework Programme, under grant agreement no. 603557, StratoClim project. Fiona M. O'Connor also acknowledges additional funding received from the Horizon 2020 European Union's Framework Programme for Research and Innovation "Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach (CRESCENDO)" project under grant agreement no. 641816. The authors would like to acknowledge Jeff Knight for processing the sea surface temperature and sea ice datasets used in the model integrations. ERA-Interim data used in

this study were provided by ECMWF. Met Office CCMVal-2 data used in this study are stored at the British Atmospheric Data Centre (BADC). We acknowledge use of the MONSooN high-performance computing system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, a strategic partnership between the Met Office and the Natural Environment Research Council.

Edited by: J. Williams

Reviewed by: two anonymous referees

## 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, *J. Geophys. Res.-Atmos.*, 120, 7534–7554, doi:10.1002/2015JD023182, 2015.
- 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, *J. Geophys. Res.-Atmos.*, 121, 1361–1380, doi:10.1002/2015JD023334, 2016.
- Albers, J. R., McCormack, J. P., and Nathan, T. R.: Stratospheric ozone and the morphology of the northern hemisphere planetary waveguide, *J. Geophys. Res.*, 118, 563–576, doi:10.1029/2012JD017937, 2013.
- Andrews, A. E., Boering, K. A., Daube, B. C., Wofsy, S. C., Loewenstein, M., Jost, H., Podolske, J. R., Webster, C. R., Herman, R. L., Scott, D. C., Flesch, G. J., Moyer, E. J., Elkins, J. W., Dutton, G. S., Hurst, D. F., Moore, F. L., Ray, E. A., Romashkin, P. A., and Strahan, S. E.: 2001: Mean ages of stratospheric air derived from in situ observations of CO<sub>2</sub>, CH<sub>4</sub>, and N<sub>2</sub>O, *J. Geophys. Res.*, 106, 32295–32314, doi:10.1029/2001JD000465, 2001.
- Austin, J., Shindell, D., Beagley, S. R., Brühl, C., Dameris, M., Manzini, E., Nagashima, T., Newman, P., Pawson, S., Pitari, G., Rozanov, E., Schnadt, C., and Shepherd, T. G.: Uncertainties and assessments of chemistry-climate models of the stratosphere, *Atmos. Chem. Phys.*, 3, 1–27, doi:10.5194/acp-3-1-2003, 2003.
- Austin, J., Struthers, H., Scinocca, J., Plummer, D. A., Akiyoshi, H., Baumgaertner, A. J. G., Bekki, S., Bodeker, G. E., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M. P., Cugnet, D., Dameris, M., Dhomse, S., Frith, S., Garny, H., Gettelman, A., Hardiman, S. C., Jöckel, P., Kinnison, D., Kubin, A., Lamarque, J. F., Langematz, U., Mancini, E., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Nielsen, J. E., Pitari, G., Pyle, J., Rozanov, E., Shepherd, T. G., Shibata, K., Smale, D., Teyssède, H., and Yamashita, Y.: Chemistry-climate model simulations of spring Antarctic ozone, *J. Geophys. Res.*, 115, D00M11, doi:10.1029/2009JD013577, 2010.
- Bellouin, N., Rae, J., Jones, A., Johnson, C., Haywood, J., and Boucher, O.: Aerosol forcing in the Climate Model Intercomparison Project (CMIP5) simulations by HadGEM2-ES and the role of ammonium nitrate, *J. Geophys. Res.*, 116, D20206, doi:10.1029/2011JD016074, 2011.
- Birner, T. and Bönisch, H.: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere, *At-*

- mos. Chem. Phys., 11, 817–827, doi:10.5194/acp-11-817-2011, 2011.
- Butchart, N. and Scaife, A. A.: Removal of chlorofluorocarbons by increased mass exchange between stratosphere and troposphere in a changing climate, *Nature*, 410, 799–802, doi:10.1038/35071047, 2001.
- Butchart, N., Scaife, A. A., Bourqui, M., de Grandpré, J., Hare, S. H. E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D., and Sigmond, M.: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation, *Clim. Dynam.*, 27, 727–741, doi:10.1007/s00382-006-0162-4, 2006.
- Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Krüger, K., Kushner, P., Newman, P., Osprey, S. M., Perlwitz, J., Sassi, F., Sigmond, M., and Wang L.: Stratospheric dynamics, in: *Evaluation of Chemistry-Climate Models*, Rep. 5, edited by: Eyring, V., Shepherd, T. G., and Waugh, D. W., WCRP-132, WCRP, Geneva, Switzerland, 109–148, 2010.
- Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Haynes, P. H., Krüger, K., Kushner, P. J., Newman, P. A., Osprey, S. M., Perlwitz, J., Sigmond, M., Wang, L., Akiyoshi, H., Austin, J., Bekki, S., Baumgaertner, A., Braesicke, P., Brühl, C., Chipperfield, M., Dameris, M., Dhomse, S., Eyring, V., Garcia, R., Garny, H., Jockel, P., Lamarque, J.-F., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Pawson, S., Plummer, D., Pyle, J., Rozanov, E., Scinocca, J., Shepherd, T. G., Shibata, K., Smale, D., Teyssède, H., Tian, W., Waugh, D., and Yamashita, Y.: Multi-model climate and variability of the stratosphere, *J. Geophys. Res.*, 116, D05102, doi:10.1029/2010JD014995, 2011.
- Butchart, N.: The Brewer-Dobson circulation, *Rev. Geophys.*, 52, 157–184, doi:10.1002/2013RG000448, 2014.
- Chipperfield, M. P. and Pyle, J. A.: Model sensitivity studies of Arctic ozone depletion, *J. Geophys. Res.*, 103D21, 28389–28403, doi:10.1029/98JD01960, 1998.
- Dameris, M., Grewe, V., Ponater, M., Deckert, R., Eyring, V., Mager, F., Matthes, S., Schnadt, C., Stenke, A., Steil, B., Brühl, C., and Giorgetta, M. A.: Long-term changes and variability in a transient simulation with a chemistry-climate model employing realistic forcing, *Atmos. Chem. Phys.*, 5, 2121–2145, doi:10.5194/acp-5-2121-2005, 2005.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, I., Biblot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Greer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, I., Kallberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteorol. Soc.*, 137, 553–597, 2011.
- Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., Moore, F., Hurst, D., Elkins, J., Schauffler, S., Andrews, A., and Boering, K.: Age of stratospheric air unchanged within uncertainties over the past 30 years, *Nat. Geosci.*, 2, 28–31, doi:10.1038/NGEO388, 2009.
- Eyring, V., Butchart, N., Waugh, D. W., Akiyoshi, H., Austin, J., Bekki, S., Bodeker, G. E., Boville, B. A., Brühl, C., Chipperfield, M. P., Cordero, E., Dameris, M., Deushi, M., Fioletov, V. E., Frith, S. M., Garcia, R. R., Giorgetta, A., Giorgetta, M. A., Grewe, V., Jourdain, L., Kinnison, D. E., Mancini, E., Manzini, E., Marchand, M., Marsh, D. R., Nagashima, T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Shepherd, T. G., Shibata, K., Stolarski, R. S., Struthers, H., Tian, W., and Yoshiki M.: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, *J. Geophys. Res.*, 111, D22308, doi:10.1029/2006JD007327, 2006.
- Eyring, V., Chipperfield, M. P., Giorgetta, M. A., Kinnison, D. E., Manzini, E., Matthes, K., Newman, P. A., Pawson, S., Shepherd, T. G., and Waugh, D. W.: Overview of the new CCM-Val reference and sensitivity simulations in support of upcoming ozone and climate assessments and the planned SPARC CCM-Val, *SPARC Newsletter*, 30, 20–26, 2008.
- Eyring, V. and Lamarque, J.-F.: Global chemistry-climate modeling and evaluation, *Eos Trans. AGU*, 93, 539, doi:10.1029/2012EO510012, 2012.
- Eyring, V., Lamarque, J.-F., Hess, P., et al.: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) Community Simulations in Support of Upcoming Ozone and Climate Assessments, *SPARC Newsletter*, 40, 48–66, 2013.
- Fiore, A. M., Naik, V., Spracklen, D., Steiner, A., Unger, N., Prather, M., Bergmann, D., Cameron-Smith, P. J., Cionni, I., Collins, W., Dalsøren, S., Eyring, V., Folberth, G., Ginoux, P., Horowitz, L. W., Josse, B., Lamarque, J.-F., MacKenzie, I. A., Nagashima, T., O'Connor, F. M., Righi, M., Rumbold, S., Shindell, D. T., Skeie, R. B., Sudo, K., Szopa, S., Takemura, T., and Zeng, G.: Global Air Quality and Climate, *Chem. Soc. Rev.*, 41, 6663–6683, doi:10.1039/C2CS35095E, 2012.
- Giorgetta, M. A., Hegglin, M. I., Son, S.-W., Kim, J., Fujiwara, M., Birner, T., Kremser, S., Rex, M., Añel, J. A., Akiyoshi, H., Austin, J., Bekki, S., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M., Dameris, M., Dhomse, S., Garny, H., Hardiman, S. C., Jockel, P., Kinnison, D. E., Lamarque, J. F., Mancini, E., Marchand, M., Michou, M., Morgenstern, O., Pawson, S., Pitari, G., Plummer, D., Pyle, J. A., Rozanov, E., Scinocca, J., Shepherd, T. G., Shibata, K., Smale, D., Teyssède, H., and Tian, W.: Multimodel assessment of the upper troposphere and lower stratosphere: Tropics and global trends, *J. Geophys. Res.*, 115, D00M08, doi:10.1029/2009JD013638, 2010.
- Hardiman, S. C., Butchart, N., Haynes, P. H., and Hare, S. H. E.: A note on forced versus internal variability of the stratosphere, *Geophys. Res. Lett.*, 34, L12803, doi:10.1029/2007GL029726, 2007.
- Hardiman, S. C., Butchart, N., Charlton-Perez, A. J., Shaw, T. A., Akiyoshi, H., Baumgaertner, A., Bekki, S., Braesicke, P., Chipperfield, M., Dameris, M., Garcia, R. R., Michou, M., Pawson, S., Rozanov, E., and Shibata, K.: Improved predictability of the troposphere using stratospheric final warmings, *J. Geophys. Res.*, 116, D18113, doi:10.1029/2011JD015914, 2011.
- Hardiman, S. C., Boutle, I. A., Bushell, A. C., Butchart, N., Cullen, M. J. P., Field, P. R., Furtado, K., Manners, J. C., Milton, S. F., Morcrette, C., O'Connor, F. M., Shipway, B. J., Smith, C., Walters, D. N., Willett, M. R., Williams, K. D., Wood, N., Abraham, N. L., Keeble, J., Maycock, A. C., Thuburn, J., and Woodhouse, M. T.: Processes controlling tropical tropopause tempera-

- ture and stratospheric water vapor in climate models, *J. Climate*, 28, 6516–6535, doi:10.1175/JCLI-D-15-0075.1, 2015.
- Hegglin, M. I. and Shepherd, T. G.: Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux, *Nat. Geosci.*, 2, 687–691, doi:10.1038/NGEO604, 2009.
- Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., McLaren, A. J., and Hunke, E. C.: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, *Geosci. Model Dev.*, 4, 223–253, doi:10.5194/gmd-4-223-2011, 2011.
- Hunke, E. C. and Lipscomb, W. H.: CICE: the Los Alamos sea ice model, Documentation and software users manual, Version 4.0, Fluid Dynamics Group, Los Alamos National Laboratory, Los Alamos, 2008.
- Jöckel, P., Tost, H., Pozzer, A., Brühl, C., Buchholz, J., Ganzeveld, L., Hoor, P., Kerkweg, A., Lawrence, M. G., Sander, R., Steil, B., Stiller, G., Tanarhte, M., Taraborrelli, D., van Aardenne, J., and Lelieveld, J.: The atmospheric chemistry general circulation model ECHAM5/MESSy1: consistent simulation of ozone from the surface to the mesosphere, *Atmos. Chem. Phys.*, 6, 5067–5104, doi:10.5194/acp-6-5067-2006, 2006.
- Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., 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., Neu-maier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth System Chemistry integrated Modelling (ESCI-Mo) with the Modular Earth Submodel System (MESSy) version 2.51, *Geosci. Model Dev.*, 9, 1153–1200, doi:10.5194/gmd-9-1153-2016, 2016.
- Johnson, C. E., Stevenson, D. S., Collins, W. J., and Derwent, R. G.: Role of climate feedback on methane and ozone studied with a Coupled Ocean-Atmosphere-Chemistry Model, *Geophys. Res. Lett.*, 28, 1723–1726, doi:10.1029/2000GL011996, 2001.
- Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., and Gray, L. J.: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather, *Nat. Geosci.*, 8, 433–440, doi:10.1038/NGEO2424, 2015.
- Laat, A. T. J. de, Zachariasse, M., Roelofs, G. J. H., Velthoven, P. van, Dickerson, R. R., Rhoads, K. P., Oltmans, S. J., and Lelieveld, J.: Tropospheric O<sub>3</sub> distribution over the Indian Ocean during spring 1995 evaluated with a chemistry-climate model, *J. Geophys. Res.*, 104, 13881–13893, 2001.
- Linz, M., Plumb, R. A., Gerber, E. P., and Sheshadri, A.: The relationship between age of air and the diabatic circulation of the stratosphere, *J. Atmos. Sci.*, 73, 4507–4518, doi:10.1175/JAS-D-16-0125.1, 2016.
- Madec G.: “NEMO ocean engine”, Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No. 27, ISSN No. 1288–1619, 2008.
- McLandress, C., Plummer, D. A., and Shepherd, T. G.: Technical Note: A simple procedure for removing temporal discontinuities in ERA-Interim upper stratospheric temperatures for use in nudged chemistry-climate model simulations, *Atmos. Chem. Phys.*, 14, 1547–1555, doi:10.5194/acp-14-1547-2014, 2014.
- McPeters, R. D., Krueger, A. J., Bhartia, P. K., Herman, J. R., Welle-meyer, C. G., Seftor, C. J., Jaross, G., Torres, O., Moy, L., Labow, G., Byrly, W., Taylor, S. L., Swissler, T., and Cebula, R. P.: Earth Probe Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA Reference Publication 1998-206895, 1998.
- Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., Ajayamohan, R. S., and Fyfe, J. C.: The Canadian Seasonal to Interannual Prediction System, Part I: Models and initialization, *Mon. Weather Rev.*, 141, 2910–2945, doi:10.1175/MWR-D-12-00216.1, 2013.
- Ming, A., Hitchcock, P., and Haynes, P.: The Double Peak in Upwelling and Heating in the Tropical Lower Stratosphere, *J. Atmos. Sci.*, 73, 1889–1901, doi:10.1175/JAS-D-15-0293.1, 2016a.
- Ming, A., Hitchcock, P., and Haynes, P.: The Response of the Lower Stratosphere to Zonally Symmetric Thermal and Mechanical Forcing, *J. Atmos. Sci.*, 73, 1903–1922, doi:10.1175/JAS-D-15-0294.1, 2016b.
- Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 1: The stratosphere, *Geosci. Model Dev.*, 2, 43–57, doi:10.5194/gmd-2-43-2009, 2009.
- Morgenstern, O., Giorgetta, M. A., Shibata, K., Eyring, V., Waugh, D. W., Shepherd, T. G., Akiyoshi, H., Austin, J., Baumgaertner, A. J. G., Bekki, S., Braesicke, P., Brühl, C., Chipperfield, M. P., Cugnet, D., Dameris, M., Dhomse, S., Frith, S. M., Garny, H., Gettelman, A., Hardiman, S. C., Hegglin, M. I., Jöckel, P., Kinnison, D. E., Lamarque, J.-F., Mancini, E., Manzini, E., Marchand, M., Michou, M., Nakamura, T., Nielsen, J. E., Olivié, D., Pitari, G., Plummer, D. A., Rozanov, E., Scinocca, J. F., Smale, D., Teyssède, H., Toohey, M., Tian, W., and Yamashita, Y.: Review of the formulation of present-generation stratospheric chemistry-climate models and associated external forcings, *J. Geophys. Res.*, 115, D00M02, doi:10.1029/2009JD013728, 2010.
- Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, 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., Mancini, E., Manyin, M. E., Marchand, M., Marécal, V., Michou, M., Oman, L. D., Pitari, G., 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 phase 1 of the Chemistry–Climate Model Initiative (CCMI), *Geosci. Model Dev.*, 10, 639–671, doi:10.5194/gmd-10-639-2017, 2017.
- Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Bernsten, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J.-H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, *Atmos. Chem. Phys.*, 13, 1853–1877, doi:10.5194/acp-13-1853-2013, 2013.
- Nowack, P. J., Abraham, N. L., Maycock, A. C., Braesicke, P., Gregory, J. M., Joshi, M. M., Osprey, A., and Pyle, J. A.: A large ozone-circulation feedback and its implications for



- global warming assessments, *Nat. Clim. Change*, 5, 41–45, doi:10.1038/nclimate2451, 2015.
- O'Connor, F. M., Johnson, C. E., Morgenstern, O., and Collins, W. J.: Interactions between tropospheric chemistry and climate model temperature and humidity biases, *Geophys. Res. Lett.*, 36, L16801, doi:10.1029/2009GL039152, 2009.
- O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., Folberth, G. A., Sanderson, M. G., Telford, P. J., Voulgarakis, A., Young, P. J., Zeng, G., Collins, W. J., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 2: The Troposphere, *Geosci. Model Dev.*, 7, 41–91, doi:10.5194/gmd-7-41-2014, 2014.
- Olsen, S. C., McLinden, C. A., and Prather, M. J.: The stratospheric N<sub>2</sub>O–NO<sub>y</sub> system: Testing uncertainties in a 3-D framework, *J. Geophys. Res.*, 106, 28771–28784, doi:10.1029/2001JD000559, 2001.
- Prinn, R. G.: Development and application of earth system models, *P. Natl. Acad. Sci. USA*, 110, 3673–3680, doi:10.1073/pnas.1107470109, 2013.
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *J. Climate*, 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.
- Scaife, A. A., Butchart, N., Warner, C. D., and Swinbank, R.: Impact of a spectral gravity wave parametrization on the stratosphere in the Met Office Unified Model, *J. Atmos. Sci.*, 59, 1473–1489, 2002.
- Sitch, S., Cox, P. M., Collins, W. J., and Huntingford, C.: Indirect radiative forcing of climate change through ozone effects on the land-carbon sink, *Nature*, 448, 791–794, doi:10.1038/nature06059, 2007.
- Stevenson, D. S., Young, P. J., Naik, V., Lamarque, J.-F., Shindell, D. T., Voulgarakis, A., Skeie, R. B., Dalsoren, S. B., Myhre, G., Bernsten, T. K., Folberth, G. A., Rumbold, S. T., Collins, W. J., MacKenzie, I. A., Doherty, R. M., Zeng, G., van Noije, T. P. C., Strunk, A., Bergmann, D., Cameron-Smith, P., Plummer, D. A., Strode, S. A., Horowitz, L., Lee, Y. H., Szopa, S., Sudo, K., Nagashima, T., Josse, B., Cionni, I., Righi, M., Eyring, V., Conley, A., Bowman, K. W., Wild, O., and Archibald, A.: Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), *Atmos. Chem. Phys.*, 13, 3063–3085, doi:10.5194/acp-13-3063-2013, 2013.
- Tegtmeier, S., Krüger, K., Wohltmann, I., Schoellhammer, K., and Rex, M.: Variations of the residual circulation in the Northern Hemispheric winter, *J. Geophys. Res.*, 113, D16109, doi:10.1029/2007JD009518, 2008.
- Telford, P. J., Braesicke, P., Morgenstern, O., and Pyle, J. A.: Technical Note: Description and assessment of a nudged version of the new dynamics Unified Model, *Atmos. Chem. Phys.*, 8, 1701–1712, doi:10.5194/acp-8-1701-2008, 2008.
- Telford, P. J., Abraham, N. L., Archibald, A. T., Braesicke, P., Dalvi, M., Morgenstern, O., O'Connor, F. M., Richards, N. A. D., and Pyle, J. A.: Implementation of the Fast-JX Photolysis scheme (v6.4) into the UKCA component of the MetUM chemistry-climate model (v7.3), *Geosci. Model Dev.*, 6, 161–177, doi:10.5194/gmd-6-161-2013, 2013.
- 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., Val Martin, M., Tanimoto, H., Simpson, I., Blake, D. R., and Blake, N.: Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI), *Geosci. Model Dev.*, 9, 1853–1890, doi:10.5194/gmd-9-1853-2016, 2016.
- Voulgarakis, A., Naik, V., Lamarque, J.-F., Shindell, D. T., Young, P. J., Prather, M. J., Wild, O., Field, R. D., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Folberth, G. A., Horowitz, L. W., Josse, B., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Stevenson, D. S., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, *Atmos. Chem. Phys.*, 13, 2563–2587, doi:10.5194/acp-13-2563-2013, 2013.
- Walters, D. N., Best, M. J., Bushell, A. C., Copsey, D., Edwards, J. M., Falloon, P. D., Harris, C. M., Lock, A. P., Manners, J. C., Morcrette, C. J., Roberts, M. J., Stratton, R. A., Webster, S., Wilkinson, J. M., Willett, M. R., Boutle, I. A., Earnshaw, P. D., Hill, P. G., MacLachlan, C., Martin, G. M., Moufouma-Okia, W., Palmer, M. D., Petch, J. C., Rooney, G. G., Scaife, A. A., and Williams, K. D.: The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations, *Geosci. Model Dev.*, 4, 919–941, doi:10.5194/gmd-4-919-2011, 2011.
- Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., Lock, A. P., Morcrette, C. J., Stratton, R. A., Wilkinson, J. M., Willett, M. R., Bellouin, N., Bodas-Salcedo, A., Brooks, M. E., Copsey, D., Earnshaw, P. D., Hardiman, S. C., Harris, C. M., Levine, R. C., MacLachlan, C., Manners, J. C., Martin, G. M., Milton, S. F., Palmer, M. D., Roberts, M. J., Rodríguez, J. M., Tennant, W. J., and Vidale, P. L.: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations, *Geosci. Model Dev.*, 7, 361–386, doi:10.5194/gmd-7-361-2014, 2014.
- Waugh, D. W. and Eyring, V.: Quantitative performance metrics for stratospheric-resolving chemistry-climate models, *Atmos. Chem. Phys.*, 8, 5699–5713, doi:10.5194/acp-8-5699-2008, 2008.
- Waugh, D. W., Oman, L., Kawa, S. R., Stolarski, R. S., Pawson, S., Douglass, A. R., Newman, P. A., and Nielsen, J. E.: Impacts of climate change on stratospheric ozone recovery, *Geophys. Res. Lett.*, 36, L03805, doi:10.1029/2008GL036223, 2009.
- Waugh, D.: Atmospheric dynamics: The age of stratospheric air, *Nat. Geosci.*, 2, 14–16, doi:10.1038/ngeo397, 2009.
- Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, *Atmos. Chem. Phys.*, 7, 2643–2660, doi:10.5194/acp-7-2643-2007, 2007.
- WMO (World Meteorological Organization), Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project-Report No. 52, 516 pp., Geneva, Switzerland, 2011.
- Young, P. J., Archibald, A. T., Bowman, K. W., Lamarque, J.-F., Naik, V., Stevenson, D. S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dal-

- søren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Horowitz, L. W., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R. B., Shindell, D. T., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), *Atmos. Chem. Phys.*, 13, 2063–2090, doi:10.5194/acp-13-2063-2013, 2013.
- Zeng, G., Morgenstern, O., Braesicke, P., and Pyle, J. A.: Impact of stratospheric ozone recovery on tropospheric ozone and its budget, *Geophys. Res. Lett.*, 37, L09805, doi:10.1029/2010GL042812, 2010.
- Zhang, L., Jacob, D. J., Yue, X., Downey, N. V., Wood, D. A., and Blewitt, D.: Sources contributing to background surface ozone in the US Intermountain West, *Atmos. Chem. Phys.*, 14, 5295–5309, doi:10.5194/acp-14-5295-2014, 2014.