Dear Editor,

please find below the revised manuscript with tracked changes, addressing the reviewers' comments and suggestions. The most important changes to the manuscript include:

- Following a comment by Reviewer #1, the argument of different SST/SIC impacting on total column ozone has been removed, as we found the differences between the simulations to be not significant to a 95% confidence level (according to the t-test in Appendix A3).
- Following a comment by both reviewers, we added a table (Table 4) showing methane and methylchloroform lifetimes as a measure of tropospheric oxidation capacity.
- In addition, we added an appendix (Appendix B) to discuss a recently reported bug in the EMAC output, which affects the two wind components which we used in our analysis (ua and va variables). At this stage it is clearly not possible to repeat all the simulations, therefore we applied a 10% corrections to the wind variables to account for this problem, as suggested by a test simulation performed after the bug was detected. Note that the bug affects *only the output*, and not the internal consistency of the model dynamics, which is correctly calculated. The relevant Figures (5-9 and S3-S5) have been updated accordingly: they show some minor differences with respect to the previous version, but the conclusions of this paper do not change.

The ESMValTool has been further developed and updated in the last months, including several minor bug fixes:

- Preprocessing of the model output for EVAL2 and TS2000 revised.
- Fixed calculation of the standard deviations in the annual cycle plots, to account for missing values in the normalization of the weights.
- Resolution of the hybrid-to-pressure levels increased in the lower troposphere.

All the plots in the paper have been redone accordingly. The differences with respect to the previous versions are minor and do not affect any of the conclusions. Nevertheless, for completeness we report all the differences in the following table.

Fig.	**Content**	**Changes**
1	ta annual cycle	No differences
2	ta zonal mean	Minor differences at lower levels (p>850 hPa)
3	h2o annual cycle	Minor difference in the standard deviation of HALOEdata
4	h2o annual cycle	Minor difference in the standard deviation of HALOEdata
5	ECVs rmsd metrics	Differences in ua and va; minor differences in zg
6	ECVs bias metrics	Differences in ua and va; minor differences in zg
7	ECVs Taylor diagrams	Differences in ua (bottom panels)
8	ua annual cycle	10% difference due to time-filter correction, see Appendix B
9	ua zonal mean (DJF)	10% difference due to time-filter correction, see Appendix B; minor differences at lower levels (p>850 hPa)
10	toz zonal cycle	No differences
11	Ozone metrics	No differences
12	Ozone Taylor diagrams	No differences
13	tropoz lat-lon	No differences
14	tropoz zonal cycle	No differences
15	Ozone annual cycle	No differences
16	Ozone vertical profiles	No differences
17	NO_x vertical profiles	Differences in EVAL2 (in particular for the Thule profile)
18	CO annual cycle	No differences
19	CO vertical profiles	No differences

Manuscript prepared for Geosci. Model Dev. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 21 January 2015

Quantitative evaluation of ozone and selected climate parameters in a set of EMAC simulations

M. Righi[1], V. Eyring[1], K.-D. Gottschaldt[1], C. Klinger[1,*], F. Frank[1], P. Jöckel[1], and I. Cionni[2]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile (ENEA), Rome, Italy
[*]now at: Ludwig-Maximilians-Universität München, Fakultät für Physik, Meteorologisches Institut, Munich, Germany

Correspondence to: M. Righi (mattia.righi@dlr.de)

Abstract

Four simulations with the ECHAM/MESSy Atmospheric Chemistry (EMAC) model have been evaluated with the Earth System Model Validation Tool (ESMValTool) to identify differences in simulated ozone and selected climate parameters that resulted from (i) different setups of the EMAC model (nudged vs. free-running) and (ii) different boundary conditions (emissions, sea surface temperatures (SSTs) and sea-ice concentrations (SICs)). To assess the relative performance of the simulations, quantitative performance metrics are calculated consistently for the climate parameters and ozone. This is important for the interpretation of the evaluation results since biases in climate can impact on biases in chemistry and vice versa. The observational datasets used for the evaluation include ozonesonde and aircraft data, meteorological reanalyses and satellite measurements. The results from a previous EMAC evaluation of a model simulation with ~~weak~~ nudging towards realistic meteorology in the troposphere have been compared to new simulations with different model setups and updated emission datasets in free-running timeslice and nudged Quasi Chemistry-Transport Model (QCTM) mode. The latter two configurations are particularly important for chemistry-climate projections and for the quantification of individual sources (e.g., the transport sector) that lead to small chemical perturbations of the climate system, respectively. With the exception of some specific features which are detailed in this study, no large differences that could be related to the different setups ~~of the EMAC simulations~~ (nudged vs. free-running) of the EMAC simulations were found, which offers the possibility to evaluate and improve the overall model with the help of shorter nudged simulations. The main differences between the two setups is a better representation of the tropospheric and stratospheric temperature in the nudged simulations, which also better reproduce stratospheric water vapour concentrations, due to the improved simulation of the temperature in the tropical tropopause layer. Ozone and ozone precursor concentrations, on the other hand, are very similar in the different model setups, if similar boundary conditions are used. Different boundary conditions however lead to relevant differences in the four simulations. ~~SSTs and SICs, which are prescribed in all simulations , play a key role in the representation~~

Biases which are common to all simulations are the underestimation of the ozone hole ~~, which is significantly underestimated in some experiments. A bias that is present in all simulations is an~~ and the overestimation of tropospheric column ozone, ~~which is~~ the latter being significantly reduced when lower lightning emissions of nitrogen oxides are used. To further investigate possible other reasons for such bias, two sensitivity simulations with an updated scavenging routine and the addition of a newly proposed HNO_3-forming channel of the $HO_2 + NO$ reaction were performed. The update in the scavenging routine resulted in a slightly better representation of ozone compared to the reference simulation. The introduction of the new HNO_3-forming channel significantly reduces ~~this bias~~the overestimation of tropospheric ozone. Therefore, including the new reaction rate could potentially be important for a realistic simulation of tropospheric ozone, although laboratory experiments and other ~~models~~ model studies need to confirm this hypothesis and some modifications to the rate, which has a strong dependence on water vapour, might also still be needed.

1 Introduction

A correct representation of tropospheric and stratospheric ozone is crucial for reproducing past trends in climate variables (e.g., temperature) as well as for providing reliable projections of the chemistry-climate system in the 21st century. Tropospheric ozone burden has increased by around 30 % between 1850 and 2010 to a level of $\sim 340\,\mathrm{Tg}$ (Young et al., 2013), leading to a global mean radiative forcing (RF) of $\sim 0.4\,\mathrm{W\,m^{-2}}$ (Stevenson et al., 2013). This increase is particularly strong in the Northern Hemisphere (NH) mid-latitudes, due to the increased anthropogenic emissions. In the future, tropospheric ozone is projected to change, depending on the emission scenario and in particular the evolution of the ozone precursors nitrogen oxides ($NO_x = NO + NO_2$), carbon monoxide (CO), methane (CH_4), and non-methane hydrocarbons (NMHCs). For example, Cionni et al. (2011) found that trends in tropospheric column ozone contribute substantially to total column ozone trends in the 21st century in the four Representative Concentration Pathways (RCP; Moss et al., 2010), mainly because of the difference in methane concentrations and stratospheric in-

put of ozone, which result in a 10 DU (~ 109 Tg) increase compared to 2000 in RCP8.5 (Eyring et al., 2013a). On the other hand, stratospheric ozone has been subject to a major perturbation since the late 1970s due to anthropogenic emissions of ozone-depleting substances (ODSs), now successfully controlled under the Montreal Protocol and its Amendments and Adjustments (WMO, 2011). The ozone hole has been identified as the primary driver of changes in Southern Hemisphere (SH) summertime high-latitude surface climate over the past few decades (Thompson and Solomon, 2002, 2005; Thompson et al., 2005). Due to the projected disappearance of the ozone hole during the 21st century, a deceleration of the poleward side of the jet (a decrease in the Southern Annular Mode) is expected (Perlwitz et al., 2008; Son et al., 2008, 2010; SPARC-CCMVal, 2010). In addition, the projected strengthening of the Brewer–Dobson circulation could result into a decrease in tropical ozone and an increase in extratropical ozone in the lower stratosphere, with impacts on RF (Butchart et al., 2006, 2010; Eyring et al., 2007; Shepherd, 2008; SPARC-CCMVal, 2010). Chemistry-Climate Models (CCMs) or more generally Earth System Models (ESMs) with interactive chemistry simulate tropospheric and stratospheric ozone as well as the underlying key processes.

Here, we evaluate simulations performed with the ECHAM/MESSy Atmospheric Chemistry (EMAC) model, which is a numerical chemistry and climate simulation system that includes submodels describing tropospheric and middle atmosphere processes and their interaction with oceans, land and human influences (Jöckel et al., 2006). The focus of this study is to assess strengths and weaknesses in the representation of ozone in different setups of the EMAC model, to answer the question whether shorter nudged simulations can be used to evaluate the free-running version of the model, and to detect general biases in EMAC. We compare the conclusions from a previous evaluation of a model simulation in nudged mode that uses a Newtonian relaxation technique in the troposphere (Jöckel et al., 2006; Pozzer et al., 2007) to new simulations with different model setups and emissions datasets in free-running timeslice and nudged Quasi Chemistry-Transport Model (QCTM; Deckert et al., 2011) mode. The model is driven by prescribed input parameters such as SSTs and SICs, concentrations of long-lived greenhouse gases, and emissions from an-

thropogenic sources, biomass burning and natural processes (e.g., volcanic eruptions and lightning). The evaluation of tropospheric ozone is focused on ozone itself and its precursors (NO_x, CO and NMHCs). Additionally, an evaluation of basic climate parameters (temperature, ~~wind~~winds, geopotential height, specific humidity, and radiation) is performed to assess the different setups of EMAC simulations against each other.

This paper is organized as follows: The model and model simulations are described in Sects. 2 and 3, respectively. An overview of the evaluation diagnostics and performance metrics is given in Sect. 4, together with a short description of the ESMValTool. The observational data used for the model evaluation are ~~given~~ described in Sect. 5. The results of the evaluation are presented and discussed in Sect. 6. ~~Section~~Sect. 7 closes with a summary.

2 ECHAM/MESSy Atmospheric Chemistry (EMAC) model description

EMAC uses the Modular Earth Submodel System (MESSy; Jöckel et al., 2006) to link multi-institutional computer codes. The core atmospheric model is the 5th generation European Centre Hamburg general circulation model (ECHAM5; Roeckner et al., 2006). ~~For the present study, two~~ In the present paper, we evaluate four different EMAC simulations, which were conducted in the framework of different studies and are based on two different versions of EMAC~~(~~: ECHAM 5.3.02/MESSy 2.41 and ECHAM 5.3.01/MESSy 1.10~~) are applied~~ . All simulations are performed in the T42L90MA (Middle Atmosphere) resolution, i.e. with a spherical truncation of T42 (corresponding to a quadratic Gaussian grid of approximately $2.8° \times 2.8°$ in latitude and longitude), with 90 vertical hybrid pressure levels up to 0.01 hPa.

The ~~version~~ versions of MESSy used in this study ~~includes~~ include more than 30 submodels, with different functions and purposes. The submodels that are used in the simulations evaluated in this work are summarized in Table 1 and are mostly common to the two versions of MESSy considered in this work (2.41 and 1.10). Additional and more detailed information can be found in Jöckel et al. (2006, 2010) ~~,~~ and on the MESSy project web-page (www.messy-interface.org).

Gas-phase chemistry is calculated with the submodel MECCA (Sander et al., 2005), which deals with both tropospheric and stratospheric chemistry. The chemical mechanism is integrated in the entire model domain, i.e., consistently from the surface to the stratosphere. It is important to highlight that no arbitrary or artificial intermediate boundary conditions (for instance at the tropopause or between layers) are prescribed. ~~This means that in particular the stratosphere-to-troposphere transport of ozone is simulated self-consistently, i.e., with a single ozone tracer.~~ Chemical species are advected according to the algorithm of Lin and Rood (1996), which is part of ECHAM5. The chemical mechanism in the model setup used here consists of gas phase reactions (including ozone tropospheric chemistry, non-methane hydrocarbons up to isoprene and stratospheric chemistry for bromine and chlorine), photolysis reactions and heterogeneous reactions, involving more than 100 species overall. Additional heterogeneous, acid-base and aqueous-phase reactions are included in the submodel SCAV (Tost et al., 2006a). Interactive aerosols are not included in the current setup and are prescribed according to a climatology by Tanre et al. (1994) . The convection processes are simulated following the ~~Tiedtke scheme (Tiedtke, 1989)~~ Tiedtke (1989) s with the Nordeng (1994) closure, as in ECHAM5 (Roeckner et al., 2006). The radiation calculations take into account prognostic cloud cover, cloud water, cloud ice (from the CLOUD submodel) and prognostic specific humidity. Forcings from radiatively active gases (CO_2, CH_4, O_3, N_2O, $CFCl_3$ and CF_2Cl_2) are computed from the corresponding prognostic tracers within the RAD4ALL submodel (RAD in MESSy2). Therefore these constituents are consistently used for the coupling between chemistry and dynamics in both directions via radiative forcing and tracer transport. ~~Interactive aerosols are not included in the current setup and are prescribed according to a climatology by Tanre et al. (1994) .~~

3 Model simulations setups

The four EMAC simulations discussed in this study have the same resolution but differ from each other in their setup. Two nudged, transient simulations (EVAL2 and QCTM) driven by the same meteorology (including SSTs) and emission inventories are compared to two free-

running timeslice simulations (ACCMIP and TS2000). As a reference, we use the nudged experiment described in Jöckel et al. (2010), which is an update using version 2.41 of MESSy of the S2 experiment discussed by Jöckel et al. (2006) and Pozzer et al. (2007). The setup of this simulation (hereafter referred to as EVAL2) is described in Sect. 3.1. The other three simulations are performed using MESSy version 1.10. A second nudged experiment (hereafter called QCTM) is run using the so-called QCTM mode, developed by Deckert et al. (2011), and is described in Sect. 3.2. Two additional simulations in timeslice mode under 2000 conditions are carried out: TS2000, using observed climatological SSTs and SICs, and ACCMIP, using simulated climatological SSTs and SICs (Sects. 3.3 and 3.4, respectively). The basic features of these four simulations are summarized in Table 2.

In the following, the specific features that characterize each EMAC simulation are briefly summarized (see also Table S1). A more detailed description of the general model setup which applies to all the experiments is provided in the Supplement (Sect. S1). The four simulations were conducted as part of various projects. The specific requirements of each project (e.g., ACCMIP) motivated the different configurations that were applied.

3.1 Simulation in nudged mode: EVAL2

This simulation has been previously evaluated by Jöckel et al. (2010). It covers 12 years (1998–2009), with the first year used for spin-up and not considered in the model analysis. Boundary conditions are, as much as possible, taken from observations. It is performed in nudged mode ~~, applying weak nudging~~ towards observed meteorology, ~~in particular to~~ namely to the operational analysis data from the European Centre for Medium-Range Weather Forecast (ECMWF), through the Newtonian relaxation of four prognostic model variables: temperature, divergence, vorticity and the logarithm of surface pressure (van Aalst et al. 2004). SSTs are prescribed from ECMWF operational analysis data as well. The nudging is applied in the spectral representation, well adapted to atmospheric wave phenomena and the spherical geometry. It is important to note that we do not nudge the wave zero (i.e., the global mean) but only wave patterns. With the exception of the logarithm of the surface pressure, the nudging in this method is applied only in the free troposphere, so that stratospheric

dynamics is calculated freely and inconsistencies between the boundary layer representation of ECMWF and ECHAM5 models are avoided. The nudging (relaxation e-folding time in parentheses) of temperature (12 h), surface pressure (12 h), divergence (48 h), and vorticity (6 h) ~~in spectral representation~~ is applied between model levels 63 ($\sim$ 97 hPa) and 84 ($\sim$ 706 hPa), with reduced values between level 63 and 71 ($\sim$ 204 hPa), as for the S2 model simulation in Jöckel et al. (2006). Transition zones (intermediate stepwise reduced nudging coefficients) are applied between levels 58 ($\sim$ 62 hPa) and 62 ($\sim$ 89 hPa), between 65 ($\sim$ 116 hPa) and 70 ($\sim$ 185 hPa), and between 85 ($\sim$ 775 hPa) and 87 ($\sim$ 909 hPa). The nudging is not applied throughout the whole model domain, since previous EMAC studies (Jöckel et al., 2006; Lelieveld et al., 2007) showed that a better representation of the stratosphere can be achieved if the nudging is applied up to 100 hPa and not further above. Moreover, we forced the quasi-biennal oscillation (QBO) externally by relaxation (nudging) of the stratospheric equatorial eastward wind to observed equatorial eastward wind profiles (Giorgetta and Bengtsson, 1999).

As this experiment is designed to (approximately) reproduce the meteorology and the atmospheric composition of the individual years, transient (i.e. varying year by year) emission data are used where available. For anthropogenic non-traffic emissions, we use the CMIP5 emission inventory of Lamarque et al. (2010) for the year 2000, which provides fluxes on a $0.5^\circ \times 0.5^\circ$ grid. We used this source also for shipping emissions and rescaled the emissions using the scaling factors from Eyring et al. (2010) in order to get a transient set. For the road traffic sector we use the QUANTIFY dataset for the year 2000 (Hoor et al., 2009), which has a spatial resolution of ~~1~~ $1^\circ \times 1^\circ$. The aviation emissions come from Schmitt and Brunner (1997) and are available for the period 1960–2009, distributed on a $3.7^\circ \times 3.7^\circ$ grid. Biomass burning emissions are taken from the GFED ~~3.1~~ v3.1 inventory (van der Werf et al., 2010). These emissions are gridded with a resolution of $0.5^\circ \times 0.5^\circ$, for the period 1997–2009. For NH_3 we also ~~consider~~ use the EDGAR3.2FT database (van Aardenne et al., 2005). Emission totals for all species in each sector are summarized in Table S2 and compared to the other setups.

3.2 Simulation in nudged QCTM mode

The QCTM simulation covers a period of 10 years (1998–2007, 1 year spin-up) and is based on a setup for EMAC (Deckert et al., 2011) in which chemical effects are decoupled from the dynamics (i.e., any feedback from chemistry on dynamics is realized via climatologies of the relevant trace gases instead of on-line coupling). This configuration is particularly useful when analyzing the effect of small chemical perturbations (like the addition of a specific emission source, e.g., shipping) on the climate system. Investigating such effects is usually hampered by the internal variability of the model, which induces very low signal-to-noise ratios and makes extracting a significant signal extremely hard. In the QCTM mode, the meteorological differences between different experiments are eliminated and the signal-to-noise ratio can be significantly increased, thus enabling the study of small perturbations even with a limited number of simulated years.

The QCTM mode is realized by driving the radiation with external climatological fields for the radiatively active gases (CO_2, CH_4, O_3, N_2O and chlorofluorocarbons). Furthermore, chemical water vapour tendencies are only affected by offline methane oxidation and offline mixing ratios of nitric acid are used to calculate the repartitioning and sedimentation in polar stratospheric clouds.

Like EVAL2, this simulation was carried out to approximate meteorology and atmospheric composition for individual years, therefore it is performed in nudged mode and using transient emissions. We use the same nudging coefficients as for EVAL2. The emission setup is also identical to EVAL2, with the exception of aviation emissions which were taken from QUANTIFY (Hoor et al., 2009), resulting however in a similar globally-integrated amount of emitted NO_x. In addition, while using the same lightning NO_x parametrization, the resulting total emission was tuned to a lower value in this experiment (see Table S2).

3.3 Simulation in free-running mode: TS2000

In contrast to the nudged simulations (EVAL2 and QCTM), the TS2000 simulation is a time-slice experiment, performed in free-running mode over a period of 10 years under 2000 con-

ditions. The boundary conditions are similar except that emissions and SSTs are climatological mean datasets representing 2000 conditions, instead of transient datasets. The distributions of SSTs and SICs are prescribed using the HadISST1 dataset from the Met Office Hadley Centre (Rayner et al., 2003), containing monthly global fields on a $1^\circ \times 1^\circ$ grid and regridded to the model T42 resolution. Here we use a 10 year climatology from 1995 to 2004. The emission setup is ~~the same as in the the~~ similar to the QCTM experiment, but it considers only the year ~~2000.~~ 2000 and uses the CMIP5 dataset instead of GFED and QUANTIFY for the biomass burning and the land transport sector, respectively, and instead of EDGAR for the NH_3 emissions.

3.4 Simulation in free-running mode: ACCMIP

This timeslice simulation was performed in support of the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP; Lamarque et al., 2013). The simulation is identical to the TS2000 setup, except that slightly different emission inventories were used (see Table S1), in order to conform to the project requirements. This timeslice simulation is only one out of the ACCMIP series of experiments, covering the period 1850 to 2100. The corresponding EMAC simulations are evaluated and analyzed in a variety of ACCMIP papers (Fiore et al., 2012; Naik et al., 2013; Silva et al., 2013; Stevenson et al., 2013; Voulgarakis et al., 2013; Young et al., 2013). To allow a consistent use of SSTs/SICs that cover the full period without discontinuities, simulated SSTs/SICs from a long-term climate model simulation were prescribed instead of using observations as in TS2000. Monthly mean SSTs and SICs are prescribed as a 10 year climatological mean around the base year 2000 using the historical CMIP5 experiment carried out with the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) climate model, which is based on ECHAM5 like EMAC. A comparison of the CMCC SSTs to the climatology from the HAdIST data for the same period shows significant differences (up to $\sim 2\,K$) over large areas of the ocean (Fig. S1). Note that because of the too short period, this is not an evaluation of the CMCC SSTs/SICs but rather just documents the differences between the two datasets that are prescribed in the TS2000 and ACCMIP simulations.

4 Diagnostics, performance metrics and evaluation tool

In order to quantitatively assess and compare the ability of the different EMAC simulations in representing key features of observed climate and chemical composition, basic statistical measures are calculated in addition to the diagnostic plots that provide more detailed insights. For each diagnostic, the root mean square difference (RMSD), the overall mean bias, and the Taylor diagram are presented. The RMSD and bias metrics are calculated considering the space–time field (latitude, longitude plus annual cycle) where available, or only the annual cycle otherwise.

Following Gleckler et al. (2008), the RMSD and overall mean bias in the annual cycle of different mean climate parameters at a particular pressure level ~~is~~ are calculated within four different domains (global, tropics, NH extratropics and SH extratropics). The results of such quantitative evaluation are presented as portrait diagrams, where the RMSD gives positive values only (due to squaring), whereas the overall mean bias is ~~also~~ sensitive to the sign of the deviation, being positive (negative) when the model overestimates (underestimates) the observations. To compare the relative performance of the simulations, the RMSD and bias are normalized by dividing through their multi-model average (see Appendices A1 and A2 for details).

All diagnostics and performance metrics shown in this paper have been implemented into the Earth System Model Validation Tool (ESMValTool)~~, which is an extension of the previously developed CCMVal Diagnostic Tool for chemistry-climate models (Gettelman et al., 201 Earth System Models (ESMs)~~. This ensures that the analysis presented in this paper can be applied to other EMAC simulations and other ESMs in a routine manner. The ESMValTool was originally based on the previously-developed CCMVal Diagnostic Tool for chemistry-climate models (Gettelman et al., 2012) , but has significantly changed since then, both with respect to its structure and scientific focus.

The ESMValTool is designed to work on model output formatted according to the Climate Model Output Rewriter (CMOR) ~~standard~~ tables metadata (see http://www2-pcmdi.llnl.gov/cmor) ~~The CMOR standard~~ This metadata defines, for example, standard names for variables,

units, coordinates names and values, etc. A reformatting routine is implemented in the ESMValTool that converts the original EMAC model output to the format required for the tool. Applying this reformatting routine to new EMAC simulations is straightforward, so that new simulations can be easily compared to the results shown here. The reformatting routine can also serve as an example for converting the output of other ESMs.

The ESMValTool is developed as an international community tool by multiple institutions with the goal to enhance routine benchmarking and evaluation of ESMs. The priority of the effort so far has been to target specific scientific themes focusing on selected Essential Climate Variables (ECVs), tropical variability (e.g., Monsoon), Southern ocean, continental dry bias and soil hydrology-climate interactions, carbon dioxide (CO_2), aerosols and ozone, but the package is being developed in such a way that additional analysis can be easily added. In this way the standard for model evaluation can be built up over time. ~~For further information and updates, see the ESMValTool website at .~~

5 Observational data for model evaluation

A variety of different observations are used for the model evaluation. For most variables, we choose a reference and an alternative dataset in order to estimate differences and uncertainties in observations.

A summary of the main diagnostics applied in this study is given in Table 3, along with the variables, observations, the short names and period/domain for the performance metrics and corresponding references.

5.1 Temperature, winds, geopotential height and specific humidity

For global temperature, winds, geopotential height and specific humidity, meteorological reanalyses are the best available reference data. Reanalysis projects provide spatially complete and coherent records of atmospheric variables. Given the improvement of models, input data and assimilation methods, reanalyses have significantly improved in reliability,

cover longer time-periods and have increased in spatial and temporal resolution (Dee et al., 2011).

We use two different reanalysis datasets (ERA-Interim and NCEP/NCAR, see below) for the comparisons to simulated temperature, ~~wind~~winds, geopotential height and specific humidity. The differences between the climatologies derived from these fields are an indicator of the uncertainties in the meteorological analyses. ERA-Interim reanalysis is produced by the ECMWF and covers the period from 1979 to present (Dee et al., 2011). All observations used in the reanalysis undergo quality control, selection steps (e.g., to sort out duplicate reports or data that is known to have large errors) and bias corrections (Dee et al., 2011). We therefore consider ERA-Interim as the main reference dataset for meteorological fields in this work and analyse the period ~~1995–2005~~1996–2005.

In addition, the NCEP/NCAR reanalysis is applied, which covers the period from 1948 to present (Kalnay et al., 1996). Over the reanalysis period, developments in the observation system took place, particularly when satellite observations became available in the 1970s. Consistently with ERA-Interim, we analyse the period ~~1995–2005~~1996–2005.

For specific humidity, we follow Gleckler et al. (2008) and use observations from the Atmospheric Infrared Sounder (AIRS) experiment (Aumann et al., 2003) as our reference data set and ERA-Interim as alternative. AIRS data are available from the middle of 2002 to the middle of 2011. The data used in this work cover the years 2003 to 2010.

Vertical and meridional profiles of climatological zonal mean water vapour volume mixing ratios are compared to measurements taken by the HALogen Occultation Experiment (HALOE) on board of the Upper Atmosphere Research Satellite (UARS), launched in 1991 (Russell et al., 1993). Model climatologies are formed for the period 1991–2002 (Grooß and Russe 2005). HALOE data for H_2O ranges from about 11 to 65 km altitude and cover 80° S to 80° N in latitude within one year. For all measured species the accuracy of the HALOE retrievals decreases near the tropopause (Brühl et al., 1996; Harries et al., 1996; Park et al., 1996; Russell et al., 1996) and sparse coverage of the polar regions increases the uncertainty in the HALOE climatologies there.

5.2 Radiation

For evaluating radiation fluxes, our primary dataset is taken from the Surface Radiation Budget Project (SRB; GEWEX-news, 2011) and the alternative dataset is taken from the Clouds and the Earth's Radiant Energy System (CERES; Wielicki et al., 1996) experiment. The SRB dataset in its current version (3.0) covers the period from July 1983 to December 2007. Here we consider the time range 1995–2005. The dataset provides surface and top of the atmosphere (ToA) longwave and shortwave fluxes derived from a variety of satellite-observed parameters, like cloud parameters, ozone fields and reanalysis meteorology (GEWEX-news, 2011). The CERES experiment products include information about solar and longwave radiation for the surface and ToA between 2001 and ~~2005.~~ 2012.

5.3 Total column ozone

For the evaluation of total column ozone, we use the NIWA combined total column ozone dataset over the period 1998–2010 as reference dataset (Bodeker et al., 2005) and a combined dataset from the Global Ozone Monitoring Experiment (GOME) and GOME-2 for the same period as alternative (Loyola and Coldewey-Egbers, 2012; Loyola et al., 2009). The NIWA dataset is an assimilated database that combines TOMS (Total Ozone Mapping Spectrometer), GOME and SBUV (Solar Backscatter Ultra-Violet radiometer) data. In order to obtain a global homogeneous dataset, ground-based data from the Dobson spectrophotometer network are used, removing differences between the individual input data or filling existing gaps.

5.4 Tropospheric ozone

For the evaluation of tropospheric column ozone we use a global climatology based on the Aura Ozone Monitoring Instrument (OMI) and Microwave Limb Sounder (MLS) ozone measurements for the period 2005–2012 (Ziemke et al., 2006, 2011). The MLS/OMI gridded ozone climatology data are made available to the scientific community via the NASA Goddard Space Flight Center ozone and air quality ~~website~~ web-page (http://ozoneaq.gsfc.nasa.gov/).

For the comparison of ozone vertical profiles in the troposphere, we use a recently updated global climatology by Tilmes et al. (2012), based on ozone soundings over the last 15 years and focusing on the troposphere and the lower stratosphere. This is an important extension to the Logan (1999) climatology, since it covers the more recent years included in the simulated period of the experiments evaluated here. Vertical ozone profiles for 41 stations around the globe have been compiled and averaged for the years 1980–2009. The climatology provides information about the median and the width of the ozone probability distribution function, as well as the interannual variability of ozone between 1995 and 2009, in pressure- and tropopause-referenced altitudes. In addition to single stations, regional aggregates are included, combining stations with similar ozone characteristics. We use these regional aggregates for model evaluation and focus on the 1995–2009 time period, corresponding to the simulated period of our experiments.

In addition, we use ozone data from a collection of aircraft campaigns (Emmons et al., 2000). These data are particularly valuable because they include additional species, measured at the same location and time of ozone, allowing a more detailed analysis on ozone precursor species. These data are provided as global distribution and vertical profiles and were validated against ozonesondes and measurements onboard commercial aircraft. The ozone data cover only selected regions of the Earth and time periods vary for each region. The use of aircraft data for model evaluation might have some limitations, due to the fact that model and observations are not always temporally co-located. This could imply, for example, that observations taken in the vicinity of strong emission sources (as biomass burning) could be affected by large temporal variability and indicate large biases when compared to model simulations.

5.5 Ozone precursors

For the evaluation of ozone precursors, we use ~~again~~ the Emmons et al. (2000) dataset, which provides information about a variety of species, including CH_4, CO, NO_x and NMHCs.

For the evaluation of CO, we additionally use the observational data from the NOAA GLOBALVIEW dataset (4th annual update, GLOBALVIEW-CO2, 2010), over the ~~1998–2008~~

1999–2008 period. This dataset is provided by the Cooperative Atmospheric Data Integration Project for Carbon Monoxide which is coordinated by NOAA (National Oceanic and Atmospheric Administration), ESRL (Earth System Research Laboratory) and GMD (Global Monitoring Division). The goal of the GLOBALVIEW initiative was to get data products with a large spatial and temporal resolution to support carbon cycle modelling studies based on measurements from land-surface, ship, aircraft, and tower observations. The processing includes smoothing, interpolation and extrapolation following Masarie and Tans (1995), resulting in an extended record.

6 Results and discussion of model evaluation

6.1 Basic climate parameters

In the following subsections, we first evaluate how well the mean climate state in selected basic climate variables such as temperature, eastward and northward wind, geopotential height, specific humidity and radiation is represented in the four simulations. In the choice of the tropospheric diagnostics and performance metrics we closely follow those that were applied by Gleckler et al. (2008), with periods changed to represent 2000 conditions. Since the EVAL2 and the QCTM simulations are both nudged by meteorological reanalysis, a generally better agreement with meteorological reanalyses compared to the free-running timeslice experiments (TS2000 and ACCMIP) can be expected. However, ~~since the nudging is relatively weak,~~ differences could still occur, in particular in regions where the nudging parameters are small, i.e. outside the main nudging interval, which is between $\sim 97\,\mathrm{hPa}$ and $\sim 706\,\mathrm{hPa}$ (see Sect. 3.1).

For the calculation of the eastward and northward wind components, a 10 % correction to the original EMAC output has been applied here, to account for a recently reported error in the output of the horizontal wind components. This error affects only the way the output is written and not the actual model performance and internal consistency (see Appendix B for more details).

6.1.1 Temperature

Temperature (ta) is evaluated by investigating the climatological mean annual cycle at the four selected pressure levels 850, 200, ~~300~~ 30 and 5 hPa (Fig. 1) and the annual mean zonally averaged temperature differences between each EMAC simulations and the reference data set (ERA-Interim, Fig. 2) and the alternative dataset (NCEP).

The annual cycle is in general well reproduced by all simulations at all levels and in all regions, with the exception of the 200 hPa level in the tropics. At 850 hPa, all EMAC simulations are in good agreement with ERA-Interim and NCEP/NCAR and lie generally within the interannual variability of the meteorological reanalyses, with the exception of ACCMIP which shows a positive bias ($\sim$ 1 K) in the tropical NH summer months (JJA). Such overestimation can be explained by the positive bias of the tropical SSTs in the prescribed dataset when compared to HadISST1 (see Fig. S1)~~, which for~~ . For the ACCMIP simulation the prescribed SST dataset is taken from a historical simulation with the CMCC climate model (see Sect. 3.4). Both the meteorological reanalyses and the model simulations are characterized by a very small interannual variability at this level (Fig. 1).

At 200 hPa, all EMAC simulations have a cold bias of around 5 K in all regions compared to the meteorological reanalyses and are well outside the interannual variability. This bias is particularly pronounced in the tropics in the two nudged simulations, whereas in the extratropics of both hemispheres the nudged simulations are in slightly better agreement with ERA-Interim than the free-running timeslice simulations. Note that such bias can not be due to differences between ERA-Interim and the ECMWF data used to nudge the EVAL2 and QCTM experiments. As shown by Dee et al. (2011) , the difference between the RMS forecast error produced by ERA-Interim and the ECMWF forecasting system that was operational in 1989 is only about 0.2 K at 200 hPa. It is also important to recall that we did not nudge the global mean temperature but only patterns (see Sect. 3.1).

Stratospheric temperatures at 30 and 5 hPa (Fig. 1, lower rows) are within one standard deviation of ERA-Interim in the extratropics in all simulations, with the exception of the summer months in the NH. In the tropics, a cold bias of around 2 K is simulated. At 5 hPa in the

tropics, ACCMIP and TS2000 show a better agreement with the observations than the other experiments. In general, temperature is much better simulated in the lower troposphere, where the simulated deviations from ERA-Interim are of similar magnitude than the differences between the two reanalysis datasets, which are anyway small and suggest therefore low uncertainties in the reference and alternative dataset. It is also interesting to note that the QCTM simulation has a global average temperature at 30 hPa that is quite different from the other simulations. Since the QCTM experiment uses prescribed ozone and water vapour for the model radiation, this might be a sign of the impacts of the interactions between chemistry and radiation.

The above mentioned biases are also visible in the zonally-averaged temperature profiles in Fig. 2. EMAC simulates the common features of the temperature distribution, characterized by high temperatures at ground levels in the tropics, by a decrease of temperature with altitude and towards the poles and by a further increase with altitude towards the tropopause, reasonably well (within $\sim$ 1–2 K in most parts of the simulated domain).

A warm bias can be identified in the polar SH stratosphere (50–100 hPa) in the free-running experiments and is particularly strong in TS2000. This is related to a too weak representation of the polar vortex and an underestimation of the ozone hole, which are both particularly prominent in the TS2000 simulation (see further discussion in Sect. 6.2.1). In addition to the annual mean, the seasonal mean temperatures for this simulation are shown in Fig. S2, confirming that this warm bias is mainly present in the JJA and SON seasons~~and thus related to the representation of the~~, coinciding with the polar vortex and the ozone hole.

All experiments are characterized by a cold bias in the extratropical lower stratosphere. This feature is common to many of the CMIP3 and CCMVal models (IPCC, 2007; SPARC-CCMVal, 2010) and is related to the wet bias (an overestimation of the water vapour concentrations) that affects all four EMAC simulations. This wet bias is shown in Fig. 3, which displays the annual cycle of water vapour in the EMAC simulations compared to HALOE data at 200 hPa in the SH extratropics. Water vapour is a greenhouse gas and therefore absorbs and emits infrared radiation. In the stratosphere, the emission of infrared radiation into space is larger

than the absorption of upwelling infrared radiation from the troposphere. This causes a net cooling effect. Overall, too high concentrations of water vapour in the extratropical lower stratosphere lead to too high infrared radiative cooling, which results into too low temperatures. This relation between the cold bias and the wet bias in the extratropical lower stratosphere has been shown in previous studies, for example in Stenke et al. (2008) for the ECHAM4.L39(DLR) E39 model. We shall note, however, that HALOE is believed to be biased low in these regions (see, e.g., Hegglin et al., 2013) . Temperature biases are also evident above the tropopause in the tropics. This bias was already examined by Jöckel et al. (2006), who related it to a slightly too strong ~~Brewer–Dobson Circulation~~ Brewer–Dobson circulation in the EMAC model, indicating deficiencies related to the wave forcing and adiabatic cooling/warming rates.

The temperature of the tropical tropopause layer is an important aspect of model representation since it has strong implications for the water vapour distribution in the stratosphere. The lower-stratospheric water vapour mixing ratios are generally a function of the model temperature near the tropical tropopause at 100 hPa (Gettelman et al., 2009), because low temperatures at the tropical tropopause cause condensation, dry the air and therefore less water vapour enters the stratosphere. This is similar to the behaviour in the four EMAC simulations, where smaller biases in temperatures compared to ERA-Interim at 100 hPa (EVAL2 and QCTM) relate to smaller biases in water vapour at this level compared to HALOE observations (Fig. 4). ~~Not surprisingly, the~~ The nudged simulations EVAL2 and QCTM represent the simulated annual cycle and absolute temperature values at 100 hPa (which is the upper limit at which nudging is applied) remarkably well compared to ERA-Interim, while TS2000 and ACCMIP show about 1 to 4 K lower values and a reasonable annual cycle. Correspondingly, the water vapour at 100 hPa is close to HALOE in the nudged simulations (within the 1σ interannual variability, except in September and October) and is lower than HALOE in the two free-running simulations throughout most of the year. The phase of the annual water vapour cycle in the tropics at 100 hPa is well captured by all model simulations, but as for temperature, ~~it is slightly weaker~~ its amplitude is slightly lower than ERA-Interim for the free-running simulations TS2000 and ACCMIP (Fig. 4).

The relative performance of the four simulations in reproducing temperature at the four pressure levels (850, 200, 30 and 5 hPa) and in the four domains (global, tropics, NH and SH extratropics) is summarized by the portrait diagrams in Fig. 5 (root mean square difference) and Fig. 6 (overall mean bias). In general, nudged simulations (EVAL2 and QCTM) perform slightly better than the free-running ones (TS2000 and ACCMIP) in the lower levels, where the nudging is indeed stronger. The performance of the four experiments is nevertheless quite similar. The model performance with respect to the two meteorological reanalyses considered for the temperature (lower and upper triangles in the portrait diagrams) is comparable, although there are some noticeable differences ~~,~~ (especially near the tropical tropopause, see Fig. 2), revealing that uncertainties exist in the reanalyses as well. The results of the Taylor diagram (Fig. 7, first row) show a good representation of the temperature by all model experiments. Most points lie above a correlation $R = 0.9$, indicating that the temperature pattern is very well captured, and deviations from the observational reference point (marked with REF on the x-axis) are mostly small. Most of points lie also very close the dashed arc corresponding to a normalized standard deviation equal to 1, which indicates a good match of the pattern variations between models and reanalysis data. A slightly worse performance is attained by the two free-running experiments at the 200 hPa level, with correlation values around 0.7–0.8, larger deviations from the reference point and discrepancies in the normalized standard deviation values. In general, the global domain and the extra-tropical regions are better reproduced than the tropics. The points corresponding to NCEP agree well with ERA-Interim in terms of correlation and pattern variations, but show some slight deviations from the REF point. This again suggests the existence of uncertainties in the meteorological reanalyses, which, analogously to EMAC, are largest in the tropics.

6.1.2 Eastward wind

The eastward wind (ua) as simulated by EMAC is in good agreement with both reanalysis datasets at 850 hPa in the tropics and extratropics, where all simulations reproduce the

annual cycle ~~very~~ pretty well (Fig. ~~S3~~8). As expected, the nudged simulations (EVAL2 and QCTM) perform better at this level.

The agreement is still good at 200 hPa, with the nudged simulations performing better than the free-running in the tropics. TS2000 and ACCMIP, on the other hand, slightly overestimate the eastward wind by about ~~3~~3–~~4~~ $m\,s^{-1}$ in this region. All simulations reproduce the annual cycle quite precisely at this level.

In the stratosphere, where the nudging is much weaker, all the simulations show a similar behaviour, and no significant improvement is obtained from the nudged simulations with respect to the ~~free~~ free-running ones. On the contrary, the QCTM simulation has some problems in reproducing the annual cycle in the tropics in particular at the 5 hPa level. The other simulations reproduce the annual cycle quite well and are within the interannual variability of the observations. In the extratropics, a small negative bias is found in winter for all simulations, in particular at 30 hPa.

Figures 9 and ~~S4~~ S3 show the difference plots of the seasonal mean of the eastward wind in DJF and JJA, respectively. A generally good agreement between the EMAC simulations and ERA-Interim is simulated, and especially the summertime stratospheric easterlies are well represented in all simulations. Some weaknesses are found however in the simulations of westerlies. In DJF (Fig. 9), the subtropical jet is underestimated at about 60° S in the free-running simulations (TS2000 and ACCMIP), while the nudged simulations capture the jet. On the other hand, the nudged simulations underestimate the polar night jet in the northern polar regions. Such underestimation might be related to a weak representation of the polar vortex in the NH. The temperature profiles for DJF (not shown) for the nudged simulations show indeed a warm bias in this specific region, which might be an indication for a too weak polar vortex. In JJA (Fig. ~~S4~~S3), the westwind jet at 60° S is ~~severely~~ underestimated by the free-running simulations throughout the entire atmosphere, while the nudged simulations underestimate westerlies in the stratosphere. The underestimation of the west wind jets in the free-running simulations ~~reveal a too weak~~ is an indication of an underestimation of the polar vortex. ~~As shown in Sect. 6.1.1,~~ This is also supported by the warm bias in the seasonal mean of the temperature ~~(Fig. S2) showed a warm bias~~ in this region discussed

in Sect. 6.1.1 and shown in Fig. S2. ~~Both, the weaker winds and the higher temperatures are an indication of an underestimation of the polar vortex.~~

The better performance of the nudged simulations with respect to the free-running simulations in the lower troposphere (850 hPa) is revealed by the portrait diagrams (Figs. 5 and 6). The eastward wind is generally underestimated in the extratropics and in the global domain (with the notable exception of ACCMIP at 200 hPa), whereas it is overestimated in the tropics, especially in the stratosphere. As found for the temperature, there are differences in the model performance with respect to the two meteorological renalyses considered for the evaluation, which reveals potential uncertainties in the observational datasets, particularly in the tropics. These considerations are further supported by the Taylor diagram (Fig. 7), which shows an excellent representation of the eastward wind globally and in the extratropics by all model simulations. In the tropical domain, on the other hand, variations in the phase and amplitude are significantly larger.

6.1.3 Northward wind, geopotential height and specific humidity

Northward wind, geopotential height and specific humidity are evaluated mainly to assess whether there are some serious limitations in the ~~simulation~~ representation of the mean climate by the model and only discussed briefly.

The northward wind (va) at the four selected levels (850, 200, 30, and 5 hPa) mostly lies within the interannual variability of the ERA-Interim reanalysis, with differences between ERA-Interim and NCEP being in the same order or larger than differences to the model simulations (Fig. ~~S5~~S4). The annual mean zonally averaged plot (Fig. ~~S6~~S5) show that, in general, the major features are well reproduced by all model setups. The portrait diagrams (Figs. 5 and 6) further confirm the expected generally better performance of the nudged simulations compared to the free running ones. In the overall mean bias diagram, northward winds are found to be either overestimated or underestimated depending on the considered observational datasets.

The comparison of simulated geopotential height (zg) with observations shows a generally good agreement (see Figs. ~~S7 and S8~~S6 and S7), with relative differences of the

order of a few percent. The annual cycle is mostly captured. Differences of the same order, however, can also be found when comparing ERA-Interim with NCEP data, revealing some uncertainties in the meteorological reanalyses as well.

The annual cycle of the specific humidity (hus) is mostly captured by the EMAC simulations (Fig. ~~S9~~S8), with the exception of the tropical domain, in particular at the 30 hPa level. Following Gleckler et al. (2008), instead of the 200 hPa level we consider 400 hPa, since this is more significant for the evaluation of specific humidity in the troposphere. In the ~~extratropics~~extratropical troposphere, the annual cycle shows a clear maximum in the summer months, following the change in incoming solar radiation during the year~~which affects water evaporation~~, which affects temperature (see Fig. 1) and consequently the amount of water vapour that the air can hold. In the tropics, on the other hand, the annual cycle shows a much smaller variation with time, since in this region the change in incoming radiation during the year is much less pronounced. The nudged simulations, which are driven by ECMWF operational analysis data, are generally closer to ERA-Interim than to AIRS data, while the free-running simulations simulate monthly mean values closer to the AIRS data in the lower troposphere. The general pattern of the specific humidity profile climatology (Fig. ~~S10~~S9) is characterized by a maximum over the equator at the surface, decreasing with latitude and altitude, and is well reproduced by all simulations.

6.1.4 Radiation

Climatological mean maps of outgoing longwave clear-sky radiation at the ToA (rlutcs) are shown in Fig. ~~S11~~S10, compared with SRB and CERES. The observational data (Fig. ~~S11~~S10, upper row, left) displays its highest values in the tropics (about 300 $W\,m^{-2}$) and two clear minima over the poles (around 150 $W\,m^{-2}$ at the South and 200 $W\,m^{-2}$ at the North). The EMAC simulations capture these features as can be seen in the differences plots (Fig. ~~S11~~S10). Compared to SRB, variations smaller than 20 $W\,m^{-2}$ are found everywhere on the globe, with a clear overestimation over the South polar regions (about 10 $W\,m^{-2}$, 5–10 %), which is stronger in the free-running simulations. The other parts of the globe show a general underestimation (maximum biases of about 30 $W\,m^{-2}$, 10–20 %) which is

stronger in the ~~nudged simulations~~ACCMIP simulation. A similar difference pattern results from the comparison between EMAC and CERES (not shown).

The outgoing longwave all-sky radiation at the ToA (rlut) is compared again to SRB and CERES (Fig. ~~S12~~S11). The observations show a maximum value over the tropics (250–300 $W\,m^{-2}$) and two extended minima over the polar regions (about 150 $W\,m^{-2}$ for the South and 200 $W\,m^{-2}$ for the North). In general, the radiation values are lower than for clear-sky conditions (Fig. ~~S11~~S10), as expected due to the presence of clouds. All EMAC simulations show a similar pattern of deviations compared to SRB, with the free-running experiments characterized by the largest differences (about 20–30 $W\,m^{-2}$). Biases of about 10–20 $W\,m^{-2}$ in the tropics were also found for the CMIP3 models when compared to ERBE data (IPCC, 2007), although some had very large deviations (up to about 50 $W\,m^{-2}$). The larger bias in the free running simulations could be due to the fact that cloud and convective parameters have been optimised for the free running mode (see, e.g., Mauritsen et al., 2012) and applied also for the EVAL2 and QCTM. If nudging systematically alters the cloud properties, the radiative balance will be altered as well. The two free-running experiments are indeed characterized by a similar globally-averaged cloud cover (64 %) which is higher than in EVAL2 (57 %) and QCTM (60 %).

Another important quantity for the evaluation of the radiation budget is the reflected shortwave all-sky radiation (rsut, Fig. ~~S13~~S12). The net shortwave radiation is primarily determined by solar incoming radiation and by the presence of clouds. The general pattern is therefore a combination of the variation of incoming solar radiation with latitude/season and of cloud cover. The EMAC simulations reproduce this pattern well~~(Fig. S13).~~ . The observations show their highest values (around 120–150 $W\,m^{-2}$) over regions of high surface albedo or significant cloud cover (deserts, snow covered areas, Himalaya and Sahara), while the strongly absorbing ocean surface is characterized by lower values (60–80 $W\,m^{-2}$). The comparison of EMAC simulations with SRB and CERES data shows a positive bias at mid-latitudes and in polar regions, with the highest deviations (30–40 $W\,m^{-2}$, 10–20 %) in northern higher latitudes (Alaska, North-East Russia), which are particularly present in the EVAL2 simulation. Negative biases are found in the tropics and subtropics, up to about 20–

$30\,W\,m^{-2}$ (20–30 %) in the ITCZ. This pattern is consistent with the general tendency of EMAC to underestimate low cloud fraction in the tropics and to overestimate it in the extratropics in comparison with ISCCP satellite data (Räisänen and Järvinen, 2010). These results are summarized in the performance metrics plots (Figs. 5 and 6).

6.2 Ozone and ozone precursors

In this paper we focus on tropospheric ozone, and consider the stratosphere only in the context of total column ozone. Biases in tropospheric ozone found in all four EMAC simulations led to two additional simulations (ACCMIP-S1 and ACCMIP-S2) to explore related model uncertainties. These two simulations are included in the figures, but discussed separately in Sect. 6.2.5.

6.2.1 Total column ozone

Zonal mean total column ozone (toz) climatologies from the different EMAC simulations are compared to the NIWA assimilated data and observations from the GOME instrument in Figs. 10 and ~~??~~S13. The well-known features of highest column ozone values in NH spring, low ozone values in the tropics, with a small seasonal cycle, a column-ozone maximum in the mid-latitudes of the SH in late winter/early spring and the ozone hole above the Antarctic are well represented the EMAC simulations, but significant quantitative differences compared to observations do exist. The ozone hole is underestimated in all EMAC simulations, in particular in TS2000, where the ozone hole is only marginally present and underestimated by around 75–100 DU. In NH winter, EMAC simulations overestimate column ozone in the high latitudes by ~~around 100~~about 50–100 DU compared to NIWA observations, but differences of about ~~50~~30–40 DU also exist between the two observational datasets, with GOME showing higher values in this region. At about 50–60° S, the mid-latitude maximum in total column ozone in autumn is produced by all EMAC simulations, but is more pronounced than in the NIWA and GOME observations: This positive bias ranges between 47 (EVAL2) and 59 DU (TS2000) compared to NIWA, and between 49 and 61 DU compared

to EOC-GOME. In the tropics the EMAC simulations show good agreement with NIWA and GOME observations. The above features are also reflected in the zonal mean total ozone values for the different seasons and the annual mean (Fig. **??**).

~~The main difference between the two free-running simulations, TS2000 and ACCMIP, is in the prescribed SSTs. A key impact of SSTs on stratospheric ozone is attributed to S13). The~~ differences in the ~~meridional circulation and the underlying wave forcing. The SST signal propagates to the stratosphere by modulation of the planetary wave activity and thus the Brewer–Dobson Circulation (Garny et al., 2009) . Compared to ACCMIP, the south polar ozone hole is similarly well represented in the two nudged simulations (in particular in EVAL2) which use ECMWF data for the SSTs and do not apply nudging in the stratosphere. Both the ECMWF and ACCMIP SSTs seem to enforce dynamical conditions that allow more realistic ozone destruction in the south polar vortex during southern spring.~~ representation of the ozone hole among the four simulations are not statistically significant (to a 95 % confidence level, not shown).

Stratospheric ozone is mainly affected by emissions from long-lived species (CO_2, CH_4, N_2O, chlorofluorocarbons, hydrochlorofluorocarbons, halons, and H_2) which are prescribed from the Advanced Global Atmospheric Gases Experiment (AGAGE; Prinn et al., 2000) observations as lower boundary conditions in all four simulations. Differences in emissions affecting tropospheric ozone contribute to the differences in total column ozone between ACCMIP vs. the other three simulations. Despite different emissions and different dynamics, total column ozone is generally biased high in all four EMAC simulations. This is evident also in Fig. 11 (lower left panel). The reason for this bias will be investigated in follow-up studies, since this paper focuses on tropospheric ozone (see discussion in the following sections). The correlation on the other hand is above $R = 0.8$–0.9 except in the SH polar region (see Fig. 12, upper left panel), indicating that the pattern is very well captured, and deviations from the observational reference point (marked with REF on the ~~r~~x axis) are mostly small.

6.2.2 Tropospheric column ozone

The geographical pattern and annual cycle of tropospheric column ozone (toztrop) from the EMAC simulations is compared to MLS/OMI measurements on board the Aura satellite in Figs. 13 and 14, respectively. All EMAC simulations tend to overestimate tropospheric column ozone, in particular in the NH mid-latitudes, with deviations around 10–20 DU. This is evident also in the near-global mean values given at the top right of each panel in Fig. 13 and in the overall mean bias metric (Fig. 11, lower left panel). It should be noted that Ziemke et al. (2011) reported root-mean-square uncertainties of about 5 DU in local measurements of total column ozone from OMI/MLS using ozonesondes as reference. They interpreted differences of 10 DU and higher as significant, while smaller values were essentially considered at noise level. It should also be noted that the calculation of tropospheric column ozone is sensitive to the tropopause height in the observations and in the model. In the MLS/OMI dataset, the vertically integrated MLS ozone profiles are subtracted from OMI total column ozone to derive the tropospheric column (Ziemke et al., 2011). The tropopause pressure separates tropospheric from stratospheric column ozone and is taken from NCEP using the WMO tropopause definition as in the EMAC simulations. Different temperatures in the EMAC simulations will shift the tropopause with respect to NCEP. If the tropopause is shifted towards too high (low) altitudes, this results in an overestimation (underestimation) of tropospheric column ozone. The tropospheric ozone column in EMAC is particularly sensitive to the tropopause definition~~(see Table 3 in Stevenson et al., 2013)~~, which could explain some of the differences between the observations and the EMAC simulations (see, e.g., Table 3 in Stevenson et al., 2013, although this refers to the changes in troposp However, the high bias of tropospheric ozone column in EMAC-ACCMIP is also confirmed by a comparison to other ACCMIP models, using a different tropopause definition (see Table 3 in Young et al., 2013), with a slight low bias of the tropospheric ozone column for the SH extratropics, but pronounced high biases in the tropics and NH extratropics (see Table 4 in Young et al., 2013). The EMAC ACCMIP simulation has one of the highest tropospheric

ozone burdens of all models in the ACCMIP-Hist2000 simulations (see Table 1 and Fig. 2a in Young et al., 2013).

The near-global mean in EMAC EVAL2 (36.8 DU) is equally high than the one in ACCMIP (36.1 DU), and tropospheric column ozone is still too high in TS2000 (~~33.6~~33.7 DU) compared to the MLS-OMI dataset (28.5 DU). However, the bias significantly reduces in the QCTM simulation (29.6 DU). QCTM and EVAL2 are both nudged towards the same dynamics and do not differ significantly in their chemistry schemes. As noted in Sect. 3.2, the emissions setup in QCTM is identical to EVAL2 except for the aviation inventories, which however have only slight differences in the total emissions (see Table S2). The differences in tropospheric ozone therefore most likely stem from lightning emissions. While both simulations use the Price and Rind (1994) parametrization, they use different scaling factors aiming at at total value of 11.0 $\mathrm{Tg\,NO\,yr^{-1}}$ for EVAL2 and 3.8 $\mathrm{Tg\,NO\,yr^{-1}}$ for QCTM. The latter value is close to the lower limit of the estimated range from observations (4.3–17.1 $\mathrm{Tg\,NO\,yr^{-1}}$; Schumann and Huntrieser, 2007). This explains the differences in NO_x between the two simulations (see also Sect. 5.2.4) and corresponding differences in tropospheric ozone. For the configurations compared here, a lower NO_x emission from lightning results in a better representation of tropospheric column ozone. TS2000 and ACCMIP use a different lightning parameterization (Grewe et al., 2001) , resulting in about 10.7 and 12.4 $\mathrm{Tg\,NO\,yr^{-1}}$ lightning emissions, respectively. Aviation emissions of NO_x, on the other hand, are quite similar among the four simulations, in the range 1.4 to 2.0 $\mathrm{Tg\,NO\,yr^{-1}}$ for EVAL2 (transient) and 1.8 $\mathrm{Tg\,NO\,yr^{-1}}$ for the others, therefore they cannot explain the differences in tropospheric column ozone.

In agreement with observations, lower values are simulated in the tropics and in the SH compared to NH mid-latitudes. However, significant differences in the pattern are simulated, with correlation values around $R = 0.85$ or lower (Fig. 12, bottom-left panel). The local maximum between Africa and South America, a region affected by biomass burning emission, is reproduced in all simulations although it is slightly underestimated by the QCTM simulation and overestimated by all other.

The annual cycle is overall well reproduced by the EMAC simulations, showing two distinct maxima during spring in the SH and during spring/summer in the NH. This seasonal increase in tropospheric column ozone is due to an increase of photo-chemical production and stratosphere–troposphere exchange (de Laat et al., 2005; Ziemke et al., 2006). It varies among the EMAC simulations also because of the difference in emissions. Furthermore, single year emissions in the time slice model simulations (TS2000 and ACCMIP) compared to transient emissions in the nudged EVAL2 and QCTM simulations lead to some differences in emission totals of ozone precursors (see Table S2), with subsequent impacts on tropospheric ozone formation.

6.2.3 Vertical ozone profiles

Similar to Fig. 6 in Young et al. (2013), Fig. 15 compares EMAC to ozonesonde data from Tilmes et al. (2012) in three regions (tropics, NH and SH extratropics) and at three altitude levels (250, 500, and 700 hPa). The ~~Tilmes et al. (2012) regional ozonesonde data contain~~ tropical region is represented by 9 ~~regions: 6 are located in the NH extratropics, 1 in the SH tropics and 2 in the SH extratropics.The tropical region contains mainly data from the SH tropics and only one station, close to the equator in the NH tropics (about~~ stations, NH and SH extratropics by 22 and 5 ~~N), therefore this region is considered to be the SH tropics.~~ stations, respectively (the geographical location of the stations is depicted in Fig. S14). The annual cycle in ozone is pretty well reproduced by the model for all the regions and levels, except for the tropics at 500 and 250 hPa. The comparison shows that the high bias in tropospheric column ozone in the ACCMIP simulation that was identified in the previous section stems mainly from the 250 and 500 hPa level in the tropics, whereas at 700 hPa and in the NH and SH extratropics the agreement with the ozonesonde data is good. This is similar for the EVAL2 and TS2000 simulation, but the QCTM simulation actually shows a small but negative bias in the tropics at the two levels. As discussed in Sect. 6.2.2, the difference in the simulation can likely be attributed to the difference in lightning NO_x emissions.

Simulated vertical profiles of ozone are also compared to in-situ measurements from aircraft campaigns, which have been mapped onto a $5° \times 5°$ grid by Emmons et al. (2000),

with additional data from more recent campaigns (http://gctm.acd.ucar.edu/data). For the present analysis, a subset of campaigns as selected by Pozzer et al. (2007) was chosen. The same time of the year and the same regions as in the campaigns were sampled in the simulations. However, the actual flight tracks and measurement time of the day were not considered in sampling the simulation output. Furthermore, simulations and measurements may be from different years. Even though the sampling methodology of simulation and in-situ data already implies some averaging, we do not expect exact matches between individual trace gas profiles. Nevertheless, there is a very good overall agreement, with the model results mostly within 90 % interval of the observational data (Fig. 16). All four EMAC simulations yield similar ozone profiles over the different locations, with EVAL2 generally producing the highest ozone mixing ratios and QCTM the lowest, as in the above comparison. Ozone precursor emissions vary substantially from year to year and the time periods between the EMAC simulations and the observations are not always the same. This could explain some of the disagreement between model and observations and indeed the three campaigns where the model performance appears to be not very good (Fiji, S-Atlantic and ~~Brazil-Coast~~E-Brazil-Coast) have been conducted from 1992 to 1996, about a decade before the period simulated in the EMAC simulations considered here. Furthermore, these regions are quite sensitive to biomass burning emissions, which can vary quite strongly (van der Werf et al., 2008).

6.2.4 Ozone precursors

Similar to ozone, simulated vertical profiles of ozone precursors are compared to in-situ measurements for aircraft campaigns by Emmons et al. (2000). Campaigns closest to those used in the ozone evaluation are shown, if ozone precursor data are not available for a certain campaign.

Nitrogen oxides serve as catalyst in the photochemical cycles relevant for the production and destruction of tropospheric ozone. Ozone production depends non-linearly on NO_x concentrations, but higher NO_x concentrations mostly result in higher ozone mixing ratios in the troposphere. At very high NO_x concentrations, ozone production becomes less efficient,

because it is then limited by the abundance of NMHCs (Fowler et al., 2008) . The vertical profiles simulated by EMAC show a similar shape as the observational data, and lie within the 90 % observational interval in most cases (Fig. 17). The spread among the mixing ratios simulated by the four EMAC simulations is very small in the lower and middle troposphere (up to about 7 km), whereas there are larger differences (of the order of 100 pmol mol^{-1}) among the simulations in the upper troposphere which could be related to lightning emissions, as discussed before. TS2000 and ACCMIP usually simulate the highest mixing ratios. The higher NO_x emissions of EVAL2 in comparison to QCTM are consistent with the results for ozone mixing ratios, given that NO_x is one of the main substances increasing ozone production via photochemical reactions in the troposphere. ~~Between the two free-running simulations, ACCMIP shows slightly higher concentrations, which was also noted for ozone.~~

The hydroxyl radical (OH) is another important species in the photochemistry of ozone, as the HO_x catalytic cycle is coupled to the NO_x cycle. However, the hydroxyl radical is a very short-lived species and direct observations of OH are very sensitive to local small-scale conditions, limiting the informative value of comparisons with coarse resolution simulation data. Furthermore, estimates of global mean OH concentration are not very well constrained ~~(Gottschaldt et al., 2013; Naik et al., 2013) . Alternatively, methane () and~~ (Lawrence et al., 2001; Gottschaldt et al., 2013; Naik et al., 2013) . As an indicator of tropospheric oxidation capacity, we analyse tropospheric lifetime of methane and methylchloroform (MCF) following the method by Lawrence et al. (2001) . Methane and MCF lifetimes are calculated with respect to the reactions

$$CH_4 + OH \rightarrow CH_3 + H_2O, \quad \text{(R1)}$$

$$CH_3CCl_3 + OH \rightarrow H_2O + 3Cl, \quad \text{(R2)}$$

and using the reaction rate coefficients from Atkinson (2003) and Sander et al. (2003) , respectively. The results are summarized in Table 4: The four simulations show very similar

values for the methane lifetime, in the range 7.9–9.1 yr, and for the MCF lifetime, 4.8–5.5 yr.

Another indirect indicator for tropospheric oxydation capacity is carbon monoxide (CO)~~may serve as indirect indicators for tropospheric oxidation capacity here, as they are important sinks for~~ , which is tightly coupled to OH~~in the troposphere.~~ : The reaction between CO and OH ($CO + OH \rightarrow CO_2 + H$) in the troposphere constitutes a sink of 90–95 % for CO and of about 41 % for OH (von Kuhlmann et al., 2003) with more CO generally leading to smaller OH mixing ratios in the troposphere.

~~For the ozone precursor~~ The annual cycle of CO ~~,~~ mixing ratio is compared to the NOAA GLOBALVIEW observations in various locations (Fig. 18). The ACCMIP simulation shows always a higher CO mixing ratio with respect to the other simulations, because of the large contribution from biomass burning and traffic sources (Table S2). The annual cycle is reasonably well reproduced by all model simulations at all 9 locations considered here, although a general underestimation of CO mixing ratios by the model is clearly visible. As reported by several other studies (Shindell et al., 2006, 2008; Monks et al., 2014) , models tend to underestimate tropospheric CO, especially in the NH winter and spring, although the reason for this bias is not fully understood. We further evaluate CO by comparison with the vertical profiles from Emmons et al. (2000) ~~are compared to EMAC simulations~~ in Fig. 19. Again, we find CO mixing ratios ~~are~~ to be too low in all simulations, often outside the uncertainty ranges, with deviations of about of 50–100 $\mathrm{nmol\,mol^{-1}}$ in the lower troposphere. Also in this case, ACCMIP has higher total CO emissions ~~, because of the large contribution from biomass burning and traffic sources (Table S2)~~than the other simulations. Remarkably, the differences between the simulations are negligible compared to the standard deviation from the observations. In more polluted regions like China, the model simulations deviate more from the observations, especially in the lower and middle troposphere, while they improve in the upper levels, where ~~again~~ the effect of emissions is much smaller. A similar problem was pointed out also by Pozzer et al. (2007), who concluded that this is probably due to underestimated fossil-fuel emissions in this region. The simulations evaluated here still underestimate CO mixing ratios, although they have higher emissions from

anthropogenic sources than in Pozzer et al. (2007). Also the ACCMIP simulation, which has twice as high CO emissions from traffic sources, does not perform better than the other simulations in this region. The representation of CO in South ~~West~~ Eest Asia is a long-standing problem in many model simulations and will require more extensive analysis in the future. ~~These considerations are further supported by the analysis of the annual cycle in various locations as observed by NOAA GLOBALVIEW (Fig. 18). The ACCMIP simulation again shows always a higher mixing ratio with respect to the other simulations. The annual cycle is reasonably well reproduced by all model simulations in all the nine locations considered here, although a general underestimation of mixing ratios~~ However, as mentioned in Sec. 5.4, we stress again that the use of aircraft data to evaluate model simulation for specific years might be affected by limitations due to the fact that aircraft data climatologies often covers time periods which do not correspond to those simulated by the model~~is clearly visible. Also in this case, in remote locations, Christmas Island, Easter Island, Mariana Island and Hawaii, the agreement between model and observations is better than in more populated regions (Canada, Germany and United States). This was also noted by Pozzer et al. (2007) .~~.

Vertical profiles of CH_4 mixing ratios in six selected regions (not shown) hardly reveal any disagreement among the EMAC simulations. This is not surprising, since the CH_4 lower boundary conditions are prescribed from the same observed data (AGAGE) in all the EMAC simulations.

Non-methane hydrocarbons (NMHCs) also affect ozone chemistry through a large number of complex reactions. Several species of this family (ethylene (C_2H_4), ethane (C_2H_6), propene (C_3H_6), propane (C_3H_8) and acetone (CH_3COCH_3)) are compared to the observational data of Emmons et al. (2000) in the Supplement (Figs. ~~S14–S18~~S15–S19). A reasonable agreement is found only for some NMHCs in a few locations. In general, all model simulations have problems in reproducing the NMHCs. Discrepancies between model and observations cannot always be attributed to emissions, as indicated by vertical profiles in remote regions or by model simulations with similar emission totals that lead to different results. The geographical distribution of the emissions might influence the representation of

these species in the model simulations. Another issue is the speciation fraction adopted for the different NMHCs compounds. Emission datasets usually provide total NMHC emissions, which have then to be speciated into individual components, consistently with the chemical mechanism of the model. Here we adopt the speciation fractions by von Kuhlmann et al. (2003). An underestimate of ethane (Fig. S16) gainst surface data have been also shown by Emmons et al. (2014) in the NH for several models.

6.2.5 Sensitivity simulations

The high bias in tropospheric column ozone identified in particular in the ACCMIP simulation motivated two additional sensitivity simulations to explore related model uncertainties. Both are identical to the ACCMIP simulation and cover the same time period (10 years under 2000 conditions), except for a code modification in the EMAC scavenging submodel SCAV (ACCMIP-S1), and an additional modification in the chemical mechanism (ACCMIP-S2).

The SCAV modification avoids the use of unrealistically high convective liquid and ice water contents for scavenging, which is expected to result in reduced uptake and less subsequent removal of nitric acid, particularly in the tropical upper troposphere/lower stratosphere (UTLS). The ACCMIP-S1 simulation serves two purposes: (1) comparing to the otherwise identical ACCMIP simulation, in order to estimate the uncertainty imposed by the reduced uptake on the results in all other simulations~~.~~; (2) ~~As~~ as a reference for the sensitivity simulation ACCMIP-S2, which is also performed with the updated scavenging code. The code modification for ACCMIP-S1 results in less and more realistic convective cloud water and cloud ice concentrations, and consequently less scavenging of HNO_3 and other species. Less scavenging of HNO_3 by cloud particles means that more HNO_3 is available for gas phase reactions. This essentially increases the abundance of NO_x, which in most parts of the free troposphere would lead to higher ozone mixing ratios. However, less scavenging also means less redistribution of reactive nitrogen in the atmosphere. Convection is strongest in the tropics and thus the differences between ACCMIP and ACCMIP-S1 are most pronounced there (Fig. 15, top row). For the considered altitudes (250, 500 and 700 hPa), ACCMIP-S1 produces ~~higher~~ however similar ozone mixing ratios than AC-

CMIP for most months. The global mean tropospheric ozone column slightly decreases in ACCMIP-S1 compared to ACCMIP (from 36.1 to 35.3 DU). ACCMIP-S1 performs slightly better than ACCMIP for ~~most~~ some comparisons to observations (Figs. 11 and 12). This is also reflected by a slightly better representation of most ozone precursors, but overall the differences between ACCMIP and ACCMIP-S1 are small. Thus the effects of this update should not strongly affect the conclusions drawn from EVAL2, QCTM, TS2000 and ACCMIP.

ACCMIP-S2 is a sensitivity simulation to quantify the uncertainty imposed by a possible HNO_3-forming channel of the $HO_2 + NO$ reaction (Butkovskaya et al., 2007) on the results of the other simulations in the present study. None of the other simulations included this reaction channel, which is not implemented into the recent JPL catalogue (Sander et al., 2011). The reaction rate coefficient of the channel is uncertain. It may depend on pressure and temperature only (Butkovskaya et al., 2007), or additionally on water vapour concentration (Butkovskaya et al., 2009). ACCMIP-S2 includes the additional dependence on water vapour concentration as described by Gottschaldt et al. (2013). This provides an upper estimate for the effects of the reaction, because water vapour enhances the HNO_3-forming channel. Apart from the above modification to the chemical mechanism, ACCMIP-S2 is identical to ACCMIP-S1. The additional reaction has the biggest absolute impact on ozone in the altitude range of about 10 hPa, where atmospheric ozone mixing ratios have a maximum. Compared to a simulation without the reaction, ozone increases at around 10 hPa, and decreases throughout the troposphere (see also Gottschaldt et al., 2013). The relative impact of the reaction is largest in the lower parts of the troposphere. Both, total and tropospheric column ozone decrease when the additional HNO_3-forming channel is included (compare ACCMIP-S1 and ACCMIP-S2 in Figs. 10, 13 and 14). Due to the dependency of the reaction rate coefficient on temperature and water vapour concentration, effects on tropospheric column ozone are largest in the tropics. The pronounced high ozone bias of the other simulations in this region is significantly reduced as a result. RMSD and overall mean bias decrease in ACCMIP-S2 compared to ACCMIP and ACCMIP-S1 (Fig. 11, left panels). In particular, the positive bias in total and tropospheric column ozone in ACCMIP reduces in all regions, and gets negative in the SH extratropics for tropospheric column

ozone. The better performance of ACCMIP-S2 (and QCTM) in tropospheric column ozone is also visible in the Taylor diagram (lower left panel in Fig. 12).

The effects on ozone precursors are mainly determined by a decreased oxidizing capacity in an atmosphere with the additional HNO_3-forming channel (Gottschaldt et al., 2013). Most notably this is reflected in the annual cycle of CO (Fig. 18), where all other simulations are biased low, but ACCMIP-S2 is mostly biased high. ~~This could also have an impact on~~ The reaction with OH ~~, as increased~~ is a major sink of CO ~~generally leads to~~ in the troposphere, which leads to higher CO mixing ratios in the less oxidizing atmosphere of ACCMIP-S2. There is also a secondary effect from reduced OH ~~concentrations and to an increase in methanelifetime (Naik et al., 2013)~~ on CO, as mixing ratios of CO precursors depend on the oxidizing capacity too. One of these precursors is methane, which has a ~ 50 % longer lifetime in ACCMIP-S2 than in ACCMIP-S1 (Table 4). Compared to ACCMIP-S1, ACCMIP-S2 agrees better ~~with the observations for Hohenpeissenberg~~ at the Alert and Terceira Island stations, but worse ~~for Niwot Ridge, Easter Island and Christmas Island.~~ at the others (Fig. 18). The value of this inconclusive result is further limited by the fact that discrepancies between observations and simulations also reflect uncertainties in the CO emission inventories. However, we note that the effect of $HO_2 + NO \rightarrow HNO_3$ on CO is bigger than the effects of different CO emissions (Table S2). Comparing to Emmons et al. (2000), ACCMIP-S2 generally performs better than ACCMIP-S1, except for NO_x and CH_4. Note that CH_4, CO and NO_x mixing ratios strongly depend on the emissions and thus also reflect uncertainties in the inventories used. Furthermore, there are other uncertainties of reaction kinetics in atmospheric ozone chemistry (Taraborrelli et al., 2012), which need to be explored in subsequent studies.

Overall, introducing the HNO_3-forming channel of the HO_2 + NO reaction has a stronger influence on ozone-related performance metrics than most of other differences between the six simulations and significantly reduces the high bias in tropospheric column ozone. This is an indication that including this reaction channel is important for a realistic simulation of ozone, but further experimental evidence is required. In some altitudes and regions, in particular at lower levels in the tropics, the performance worsens, pointing to a possible

required modification in the reaction rate dependence on water vapour as included in the simulation here.

7 Conclusions

Four present-day simulations with different setups of the ECHAM/MESSy Atmospheric Chemistry (EMAC) model have been evaluated in this study through a comprehensive comparison to observations. In particular, results from a previous EMAC evaluation of a model simulation with ~~a weak~~ nudging towards realistic meteorology in the troposphere by Jöckel et al. (2010) have been compared to new simulations with different model setups and emissions datasets in free-running timeslice and nudged Quasi Chemistry-Transport Model (QCTM) mode (Deckert et al., 2011). The latter two configurations are important for chemistry-climate projections and the quantification of individual sources (e.g., transport sector) that lead to small chemical perturbations of the climate system, respectively. The goal of this work was to compare the EMAC simulations to each other with a focus on how well ozone and selected climate parameters are represented in the different setups (nudged vs. free-running) and simulations with different boundary conditions (emissions, sea surface temperatures and sea ice concentrations).

The two nudged simulations (EVAL2 and QCTM) are transient and driven by the same SSTs and (transient where available) emission inventories (with the exception of aviation). The previously evaluated EVAL2 simulation that covers the time period 1998–2009 (Jöckel et al., 2010) serves as the reference simulation. In the QCTM simulation (QCTM, 1998–2007) the chemistry is decoupled from radiation and dynamics, thus omitting feedback mechanisms between these fundamental aspects of a chemistry-climate model. The setups of the free-running timeslice simulations (TS2000 and ACCMIP) differ from each other in the emission inventories and the SSTs. To follow the specification of the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), emissions from Lamarque et al. (2010) and simulated SSTs and SICs from the CMCC climate model are used as input parameters in the ACCMIP simulation. The boundary conditions in the TS2000 simulation are

more similar to the nudged simulations, except that emissions and SSTs are climatological means instead of transient datasets. All four EMAC simulations are carried out using the same resolution (T42L90MA).

In addition to a qualitative evaluation showing figures for a variety of different selected diagnostics, a quantitative evaluation has been performed to summarize the results. In particular, the normalized root-mean square difference (RMSD) between model simulation and observations as well as the overall mean bias have been calculated consistently for climate parameters and ozone for certain domains and height-levels. Where possible, an alternative observational dataset was used in addition to the reference dataset to consider observational uncertainty that is introduced by differences between different instruments or meteorological reanalyses. In addition, Taylor diagrams which are a common graphical summary to evaluate climate models have been shown. These diagrams display the normalized standard deviation, the centred RMSD and the pattern correlation between the model simulations and the observations.

The main differences due to the setup of the simulations (free-running vs. nudged) are introduced through differences in the meteorology. The evaluation of the mean state of basic climate parameters is therefore important in addition to the evaluation of ozone. This study show that the mean state of temperature, eastward wind, northward wind, geopotential height, specific humidity, and radiation is in general well represented by the four EMAC simulations. Some differences exist in specific regions and altitudes which are related to the different setups. In particular we find a cold bias ($\sim$ 3–7 K) in the extratropical lowermost stratosphere in the free-running simulations (TS2000 and ACCMIP). This feature is common to many of the CMIP3 and CCMVal models (IPCC, 2007; SPARC-CCMVal, 2010). This cold bias is related to the wet bias, an overestimation of water vapour in this region by around a factor of 2–10, depending on the season, leading to a too high infrared radiative cooling. The nudged simulations show the same wet bias but due to the relaxation of the temperature towards realistic meteorology, a reduced cold bias in the extratropical lowermost stratosphere. In addition, the subtropical jet ($\sim$ 10–15 m s^{-1}) at 60° S in DJF from the ground up to around 50 hPa is underestimated in the free-running simulations.

The evaluation of tropospheric ozone and ozone precursors (NO_x, CO and NMHCs) showed that the differences among the four model simulations which are related to the model setup are generally small. A common bias is the underestimation of the ozone hole. More significant differences exist due to the use of different boundary conditions (~~emissions and SSTs/SICs). These lead to differences between the ACCMIP and the other three simulations. In particular, the role of the prescribed SSTs on the ability of the model to realistically simulate the ozone hole was identified through a comparison of the two free-running simulations, with one (TS2000) prescribing observed and the other one (ACCMIP) modeled SSTs. The TS2000 simulation has major problems representing the ozone hole with observed SSTs, whereas the ACCMIP simulation that used modeled SSTs showed a much better agreement with the observations, similar to the representation of the ozone hole in the nudged simulations.~~ in particular emissions). For the temperature at 850 hPa in the tropics, the ACCMIP simulation shows a warm bias compared to ERA-Interim and NCEP due to the bias in the tropics in the prescribed modeled SSTs.

Tropospheric column ozone is generally overestimated compared to satellite observations, but the annual cycle of total column ozone is well represented. The high bias in tropospheric column ozone motivated two additional simulations that are identical to the ACCMIP simulation except for a code modification to avoid unrealistically high convective cloud water and ice contents for scavenging (ACCMIP-S1), and an additional modification in the chemical mechanism (ACCMIP-S2). ACCMIP-S2 includes a possible HNO_3-forming channel of the HO_2 + NO reaction (Butkovskaya et al., 2007) which has a stronger influence on ozone-related performance than most other differences among the six simulations and significantly reduces the high bias in tropospheric column ozone. While there is missing experimental confirmation for this additional reaction channel, our model study suggests that including it could be important for a realistic simulation of ozone particularly in the UTLS in the tropics. In some altitudes and regions, in particular at lower levels in the tropics, the performance worsens, pointing to a possible required modification in the reaction rate dependence on water vapour as included in the simulation here. A similar improvement in

tropospheric column ozone is also achieved by the nudged QCTM simulation, which uses a lower lightning NO_x emission compared to other experiments.

Biases in ozone precursors exist but are strongly dependent on the inventory used. For example, the evaluation of CO showed an underestimation compared to observations in all EMAC simulations, particularly in regions with anthropogenic influence. ~~Better results were achieved in remote locations.~~ The ACCMIP simulation with its different emission inventory from Lamarque et al. (2010) that includes a factor of 2 higher CO emissions than the inventory used in the other three simulations is generally in better agreement with the observations for CO. This stresses again the importance of accurate emission inventories for chemistry-climate modelling.

Evaluating ozone and ozone precursors with aircraft data has been proven as important in this and many previous studies. It would be important to update existing climatologies like the one by Tilmes et al. (2012) and Emmons et al. (2000) on a regular basis with newer campaigns. In addition to comparing to climatologies of aircraft data, a more direct comparison to particular campaigns should be envisaged. However, more local measurements exhibit the problem of a mismatch of spatial and temporal scales between observations and models. Sampling the model output along the flight path during the model simulation (see for example the S4D routine in Jöckel et al., 2010) and extracting the corresponding data, as planned as part of the Chemistry-Climate Model Initiative (CCMI, Eyring et al., 2013b), would facilitate and improve this comparison.

In addition, with growing complexity of chemistry-climate and Earth system models, we advocate routine evaluation of models to be facilitated by common software tools that are made available to the community. All diagnostics and performance metrics shown in this paper are now implemented in the Earth System Model Validation Tool (ESMValTool). They can be routinely reproduced and applied to new EMAC simulations or other ESMs such as those participating in CCMI (Eyring et al., 2013b) or the Coupled Model Intercomparison Project (Meehl et al., 2014).

Appendix A: Statistical measures for quantitative model evaluation

A1 Root-mean-square difference and overall mean bias

The root-mean-square difference (RMSD), which is commonly used to quantify performance of climate and numerical weather forecast models, is defined as follows:

$$\mathrm{RMSD}_m = \sqrt{\frac{1}{W}\sum_{ijk} w_{ijk}(M_{ijk} - O_{ijk})^2} \tag{A1}$$

where the index m indicates the model simulation. The fields M_{ijk} and O_{ijk} are the model and observation fields, respectively. They can be a function of latitude (i), longitude (j) and time (k), depending on the diagnostic considered. The weights w_{ijk} are proportional to the ~~grid area when the considered diagnostic is a latitude-longitude map, or~~ area of the (i,j) gridbox and to the length of each month ~~when the diagnostic is a monthly time series (or monthly climatology)~~k. The weight are normalized by their sum W.

This metric has been considered (among others) by Taylor (2001), Jöckel et al. (2006), Gleckler et al. (2008), Reichler and Kim (2008), Karpechko et al. (2010) and Yokoi et al. (2011). While Taylor (2001) and Yokoi et al. (2011) did not consider any weighting, Gleckler et al. (2008), Reichler and Kim (2008) and Karpechko et al. (2010) use the weighting described above. Additionally, Reichler and Kim (2008) weighted the sum also by a factor indirectly proportional to the variance from the observation (thus stressing the variables with lower variance), and Karpechko et al. (2010) by a factor indirectly proportional to the uncertainty in the observed variable (thus laying stress on more accurate observations). Jöckel et al. (2006) apply a weighting depending on the model error (standard deviation from the averaged value) and the measurement error (combination of instrumental error and variance), thereby giving more importance to values with a smaller total error. Although the latter weightings are reasonable, they are only applicable consistently if all the errors and uncertainties are known.

Additionally the overall mean bias is calculated according to:

$$b_m = \langle M_{ijk} \rangle - \langle O_{ijk} \rangle = \frac{1}{W} \sum_{ijk} w_{ijk} (M_{ijk} - O_{ijk}) \tag{A2}$$

Hence, the overall mean bias b_m is the difference of the weighted means of the model and of observational fields.

Finally, in order to be able to focus on relative performance among the different EMAC simulations, we normalize the RMSD and the overall mean bias by dividing through the average across the m individual model simulations (i.e., the model-mean-metrics $\overline{\text{RMSD}}$ and $\overline{b}$ defined below) similar to Reichler and Kim (2008) and Karpechko et al. (2010):

$$\overline{\text{RMSD}} = \frac{1}{M} \sum_{m=1}^{M} \text{RMSD}_m \tag{A3}$$

$$\overline{b} = \frac{1}{M} \sum_{m=1}^{M} |b_m| \tag{A4}$$

For the interpretation of the portrait diagrams, it should be noted that the RMSD gives positive values only (due to squaring), whereas the overall mean bias is also sensitive to the sign of the deviation, being positive (negative) when the model overestimates (underestimates) the observations. The better the model performance and thus agreement with observations are, the smaller the absolute statistical values. This is different to Gleckler et al. (2008), who normalized their metrics by subtracting the model-mean-metric before dividing by the latter ($(b_m - \overline{b})/\overline{b} = b_m/\overline{b} - 1$) in which case a value of 0 means that the model is equal to the multi-model-mean metric. It also differs from Yokoi et al. (2011), who normalized the metrics by subtracting the model-mean and then dividing by the inter-model standard deviation.

A2 Taylor diagram

A further possibility to graphically summarize how closely a set of modelled patterns matches observations is provided by the so called Taylor diagram, which was originally proposed by

Taylor (2001) and used in many studies (IPCC, 2001; SPARC-CCMVal, 2010) to judge the relative skills of many different models. Since then, it is routinely employed in comparison investigations (e.g., Jöckel et al., 2006; Gleckler et al., 2008). The Taylor diagram is a polar grid plot where the radial coordinate refers to the standard deviation of the model (test) field:

$$\sigma_m = \sqrt{\frac{1}{W}\sum_{ijk} w_{ijk}(M_{ijk} - \langle M_{ijk}\rangle)^2}, \tag{A5}$$

normalized to the standard deviation of the observed (reference) field:

$$\sigma_o = \sqrt{\frac{1}{W}\sum_{ijk} w_{ijk}(O_{ijk} - \langle O_{ijk}\rangle)^2} \tag{A6}$$

The angular coordinate is defined by the inverse cosine of the correlation R_m between model and observation which is defined as follows:

$$R_m = \frac{\frac{1}{W}\sum_{ijk} w_{ijk}\tilde{M}_{ijk}\tilde{O}_{ijk}}{\sigma_m \sigma_o} \tag{A7}$$

where $\tilde{M}_{ijk} = M_{ijk} - \langle M_{ijk}\rangle$ and $\tilde{O}_{ijk} = O_{ijk} - \langle O_{ijk}\rangle$.

Thus, each model and each diagnostic will provide a distinct point on the diagram. The closer the position of this point to the reference position of the observation ($\sigma_m/\sigma_o = 1$, $R = 1$ and thus $\arccos(R) = 0$), the better the agreement to the observation. The distance between this point and the reference position is thereby the centred root mean square difference cRMSD_m:

$$\mathrm{cRMSD}_m = \sqrt{\frac{1}{W}\sum_{ijk} w_{ijk}(\tilde{M}_{ijk} - \tilde{O}_{ijk})^2}, \tag{A8}$$

Note that the statistics given above are not independent, particularly, adding the centred RSMD and the overall mean bias b quadratically gives the mean square difference:

$$\mathrm{RMSD}_m^2 = b_m^2 + \mathrm{cRMSD}_m^2 \tag{A9}$$

The correlation R_m reaches its maximum value of 1 if the two fields have the same pattern of variation ($\tilde{M}_{ijk} \propto \tilde{O}_{ijk}$ for all i, j and k), however, it does not reveal whether the two fields have the same amplitude of variation (i.e. the proportionality constant between the variation patterns). This amplitude of variation is, however, determined by the standard deviations σ_m and σ_o. Thus, the Taylor diagram contains all the information needed to compare the pattern and amplitude of variation of two fields, and helps indicate which of them (the pattern or the amplitude) has a bigger share in the centred root mean square difference cRMSD.

A3 Welch's t test

Additionally to the already mentioned statistics, the Welch's t test (an extension of Student's t test for samples with unequal variances) is applied to the diagnostics where differences between a model and a reference is calculated. This test shows whether two uni-variate variables have an equal mean via a null hypothesis (von Storch and Zwiers, 1984).

The difference of the mean between two variables X_1 and X_2 is expressed in dimensionless units as:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{A10}$$

where σ indicates the standard deviation with respect to interannual variability and n the number of years in each variable.

Appendix B: Correction of the EMAC model output for the eastward and northward wind variables

An error affecting the EMAC model output of the horizontal wind components has been recently reported (M. Kunze, personal communication, 2014). For the corresponding quantities, the intermediate state within the leapfrog time filter was output instead of the finalised value. According to the applied leapfrog time filter, this introduces an error of about 10 %,

independent of model level and region. This has been verified in a dedicated test simulation. For the analysis in this paper, we therefore apply a +10 % correction to both, ua and va.

It is important to stress that this error does not affect the internal consistency of the model dynamics in any way, but concerns only the way the output is written. This error will be corrected in the upcoming release of EMAC.

Code availability

The Modular Earth Submodel System (MESSy) is continuously further developed and applied by a consortium of institutions. The usage of MESSy and access to the source code is licenced to all affiliates of institutions which are members of the MESSy Consortium. Institutions can be a member of the MESSy Consortium by signing the MESSy Memorandum of Understanding. More information can be found on the MESSy Consortium web-page (http://www.messy-interface.org).

The ESMValTool is currently under development and will be publicly released only at a later stage. A stable version of the tool can be made available upon request for development purposes. Interested users and developers are welcome to contact the lead author. For further information and updates, see the ESMValTool web-page at http://www.pa.op.dlr.de/ESMVal

The Supplement related to this article is available online at doi:10.5194/gmdd-0-1-2015-supplement.

Acknowledgements. This work was funded by the German Aerospace Center (DLR) Earth System Model Validation (ESMVal) project. The implementation of the performance metrics and diagnostics into the Earth System Model Validation Tool (ESMValTool) was also supported by the European Commission's 7th Framework Programme, under Grant Agreement number 282672, Earth system Model Bias Reduction and assessing Abrupt Climate change (EMBRACE) project. We thank Diego Loyola (DLR, Germany) for providing the GOME data and Chiara Cagnazzo and the entire CMCC climate model group from the Centro Euro-Mediterraneo sui Cambiamenti Climatici (Bologna, Italy), for providing us with SSTs and SICs from their CMIP5 model simulations. Thanks also go to Martin Evaldsson (SMHI, Sweden) for his valuable technical support on the ESMValTool ~~and~~, to Hella Garny (DLR, Germany) for her helpful suggestions on the manuscript draft, and to Markus Kunze (FUB, Germany) for reporting the EMAC bug in the output of the wind components.

The service charges for this open access publication

have been covered by a Research Centre of the Helmholtz Association.

References

Atkinson, R.: Kinetics of the gas-phase reactions of OH radicals with alkanes and cycloalkanes,Atmos. Chem. Phys., 3, 2233–2307, doi:10.5194/acp-3-2233-2003, 2003.

Aumann, H., Chahine, M., Gautier, C., Goldberg, M., Kalnay, E., McMillin, L., Revercomb, H., Rosenkranz, P., Smith, W., Staelin, D., Strow, L., and Susskind, J.: AIRS/AMSU/HSB on the Aqua mission: design, science objectives, data products, and processing systems, IEEE T. Geosci. Remote, 41, 253–264, doi:10.1109/TGRS.2002.808356, 2003.

Bodeker, G. E., Shiona, H., and Eskes, H.: Indicators of Antarctic ozone depletion, Atmos. Chem. Phys., 5, 2603–2615, doi:10.5194/acp-5-2603-2005, 2005.

Brühl, C., Drayson, S. R., Russell, J. M., Crutzen, P. J., McInerney, J. M., Purcell, P. N., Claude, H., Gernandt, H., McGee, T. J., McDermid, I. S., and Gunson, M. R.: Halogen occultation experiment ozone channel validation, J. Geophys. Res., 101, 10217–10240, doi:10.1029/95JD02031, 1996.

Butchart, N., Scaife, A., Bourqui, M., Grandpré, J., Hare, S., 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., Cionni, I., Eyring, V., Shepherd, T. G., Waugh, D. W., Akiyoshi, H., Austin, J., Brühl, C., Chipperfield, M. P., Cordero, E., Dameris, M., Deckert, R., Dhomse, S., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Li, F., Mancini, E., McLandress, C., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Sassi, F., Scinocca, J. F., Shibata, K., Steil, B., and Tian, W.: Chemistry-climate model simulations of twenty-first century stratospheric climate and circulation changes, J. Climate, 23, 5349–5374, doi:10.1175/2010JCLI3404.1, 2010.

Butkovskaya, N., Kukui, A., and Le Bras, G.: HNO_3 forming channel of the $HO_2 + NO$ reaction as a function of pressure and temperature in the ranges of 72–600 Torr and 223–323 K, J. Phys. Chem. A, 111, 9047–9053, doi:10.1021/jp074117m, 2007.

Butkovskaya, N., Rayez, M.-T., Rayez, J.-C., Kukui, A., and Le Bras, G.: Water vapor effect on the HNO_3 yield in the $HO_2 + NO$ reaction: experimental and theoretical evidence, J. Phys. Chem. A, 113, 11327–11342, doi:10.1021/jp811428p, pMID: 19780600, 2009.

Cionni, I., Eyring, V., Lamarque, J. F., Randel, W. J., Stevenson, D. S., Wu, F., Bodeker, G. E., Shepherd, T. G., Shindell, D. T., and Waugh, D. W.: Ozone database in support of CMIP5 simulations: results and corresponding radiative forcing, Atmos. Chem. Phys., 11, 11267–11292, doi:10.5194/acp-11-11267-2011, 2011.

de Laat, A. T. J., Aben, I., and Roelofs, G. J.: A model perspective on total tropospheric O_3 column variability and implications for satellite observations, J. Geophys. Res., 110, D13303, doi:10.1029/2004JD005264, 2005.

Deckert, R., Jöckel, P., Grewe, V., Gottschaldt, K.-D., and Hoor, P.: A quasi chemistry-transport model mode for EMAC, Geosci. Model Dev., 4, 195–206, doi:10.5194/gmd-4-195-2011, 2011.

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, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, 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., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Emmons, L. K., Hauglustaine, D. A., Müller, J.-F., Carroll, M. A., Brasseur, G. P., Brunner, D., Staehelin, J., Thouret, V., and Marenco, A.: Data composites of airborne observations of tropospheric ozone and its precursors, J. Geophys. Res., 105, 20497–20538, doi:10.1029/2000JD900232, 2000.

Emmons, L. K., Arnold, S. R., Monks, S. A., Huijnen, V., Tilmes, S., Law, K. S., Thomas, J. L., Raut, J.-C., Bouarar, I., Turquety, S., Long, Y., Duncan, B., Steenrod, S., Strode, S., Flemming, J., Mao, J., Langner, J., Thompson, A. M., Tarasick, D., Apel, E. C., Blake, D. R., Cohen, R. C., Dibb, J., Diskin, G. S., Fried, A., Hall, S. R., Huey, L. G., Weinheimer, A. J., Wisthaler, A., Mikoviny, T., Nowak, J., Peischl, J., Roberts, J. M., Ryerson, T., Warneke, C., and Helmig, D.: The POLARCAT Model Intercomparison Project (POLMIP): overview and evaluation with observations, Atmos. Chem. Phys. Discuss., 14, 29331–29393, doi:10.5194/acpd-14-29331-2014, 2014.

Eyring, V., Waugh, D. W., Bodeker, G. E., Cordero, E., Akiyoshi, H., Austin, J., Beagley, S. R., Boville, B. A., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M. P., Dameris, M., Deckert, R., Deushi, M., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Mancini, E., Manzini, E., Marsh, D. R., Matthes, S., Nagashima, T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Scinocca, J. F., Seme-

niuk, K., Shepherd, T. G., Shibata, K., Steil, B., Stolarski, R. S., Tian, W., and Yoshiki, M.: Multimodel projections of stratospheric ozone in the 21st century, J. Geophys. Res., 112, D16303, doi:10.1029/2006JD008332, 2007.

Eyring, V., Isaksen, I. S. A., Berntsen, T., Collins, W., Corbett, J. J., Endresen, O., Grainger, R. G., Moldanova, J., Schlager, H., and Stevenson, D. S.: Transport impacts on atmosphere and climate: shipping, Atmos. Environ., 44, 4735–4771, doi:10.1016/j.atmosenv.2009.04.059, 2010.

Eyring, V., Arblaster, J. M., Cionni, I., Sedlàček, J., Perlwitz, J., Young, P. J., Bekki, S., Bergmann, D., Cameron-Smith, P., Collins, W. J., Faluvegi, G., Gottschaldt, K.-D., Horowitz, L. W., Kinnison, D. E., Lamarque, J.-F., Marsh, D. R., Saint-Martin, D., Shindell, D. T., Sudo, K., Szopa, S., and Watanabe, S.: Long-term ozone changes and associated climate impacts in CMIP5 simulations, J. Geophys. Res., 118, 5029–5060, doi:10.1002/jgrd.50316, 2013a.

Eyring, V., Lamarque, J.-F., Hess, P., Arfeuille, F., Bowman, K., Chipperfield, M. P., Duncan, B., Fiore, A., Gettelman, A., Giorgetta, M. A., Granier, C., Hegglin, M., Kinnison, D., Kunze, M., Langematz, U., Luo, B., Martin, R., Matthes, K., Newman, P. A., Peter, T., Robock, A., Ryerson, T., Saiz-Lopez, A., Salawitch, R., Schultz, M., Shepherd, T. G., Shindell, D., Stähelin, J., Tegtmeier, S., Thomason, L., Tilmes, S., Vernier, J.-P., Waugh, D. W., and Young, P. J.: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) community simulations in support of upcoming ozone and climate assessments, SPARC Newsletter, 40, 48–66, 2013b.

Fiore, A. M., Naik, V., Spracklen, D. V., Steiner, A., Unger, N., Prather, M., Bergmann, D., Cameron-Smith, P. J., Cionni, I., Collins, W. J., Dalsøren, S., Eyring, V., Folberth, G. A., Ginoux, P., Horowitz, L. W., Josse, B., Lamarque, J.-F., MacKenzie, I. A., Nagashima, T., O'Connor, F. M., Righi, M., Rumbold, S. T., 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.

Fowler, D., Amann, M., Anderson, R., Ashmore, M., Cox, P., Depledge, M., Derwent, D., Grennfelt, P., Hewitt, N., Hov, O., Jenkin, M., Kelly, F., Liss, P., Pilling, M., Pyle, J., Slingo, J., Stevenson, D.: Ground-level ozone in the 21st century: future trends, impacts and policy implications, Royal Society Policy Document 15/08, The Royal Society, 2008.

Garny, H., Dameris, M., and Stenke, A.: Impact of prescribed SSTs on climatologies and long-term trends in CCM simulations, Atmos. Chem. Phys., 9, 6017–6031, doi:10.5194/acp-9-6017-2009, 2009.

Gettelman, A., Birner, T., Eyring, V., Akiyoshi, H., Bekki, S., Brühl, C., Dameris, M., Kinnison, D. E., Lefevre, F., Lott, F., Mancini, E., Pitari, G., Plummer, D. A., Rozanov, E., Shibata, K., Stenke, A.,

Struthers, H., and Tian, W.: The Tropical Tropopause Layer 1960–2100, Atmos. Chem. Phys., 9, 1621–1637, doi:10.5194/acp-9-1621-2009, 2009.

Gettelman, A., Eyring, V., Fischer, C., Shiona, H., Cionni, I., Neish, M., Morgenstern, O., Wood, S. W., and Li, Z.: A community diagnostic tool for chemistry climate model validation, Geosci. Model Dev., 5, 1061–1073, doi:10.5194/gmd-5-1061-2012, 2012.

GEWEX-news: February 2011, vol. 21, No. 1, available at: http://www.gewex.org/gewex_nwsltr.html (last access: 2 October 2014), 2011.

Giorgetta, M. A. and Bengtsson, L.: Potential role of the quasi-biennial oscillation in the stratosphere–troposphere exchange as found in water vapor in general circulation model experiments, J. Geophys. Res., 104, 6003–6019, doi:10.1029/1998JD200112, 1999.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, J. Geophys. Res., 113, D06104, doi:10.1029/2007JD008972, 2008.

GLOBALVIEW-CO2: Cooperative Atmospheric Data Integration Project – Carbon Dioxide, CD-ROM, NOAA ESRL, Boulder, Colorado, available at: ftp://ftp.cmdl.noaa.gov (last access: 2 October 2014), 2010.

Gottschaldt, K., Voigt, C., Jöckel, P., Righi, M., Deckert, R., and Dietmüller, S.: Global sensitivity of aviation NO_x effects to the HNO_3-forming channel of the $HO_2 + NO$ reaction, Atmos. Chem. Phys., 13, 3003–3025, doi:10.5194/acp-13-3003-2013, 2013.

Grewe, V., Brunner, D., Dameris, M., Grenfell, J., Hein, R., Shindell, D., and Staehelin, J.: Origin and variability of upper tropospheric nitrogen oxides and ozone at northern mid-latitudes, Atmos. Environ., 35, 3421–3433, doi:10.1016/S1352-2310(01)00134-0, 2001.

Grooß, J.-U. and Russell III, James M.: Technical note: A stratospheric climatology for O_3, H_2O, CH_4, NO_x, HCl and HF derived from HALOE measurements, Atmos. Chem. Phys., 5, 2797–2807, doi:10.5194/acp-5-2797-2005, 2005.

Harries, J. E., Russell, J. M., Tuck, A. F., Gordley, L. L., Purcell, P., Stone, K., Bevilacqua, R. M., Gunson, M., Nedoluha, G., and Traub, W. A.: Validation of measurements of water vapor from the Halogen Occultation Experiment (HALOE), J. Geophys. Res., 101, 10205–10216, doi:10.1029/95JD02933, 1996.

Hegglin, M. I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D., Remsberg, E, Rozanov, A.,Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, R., and Wigel,K.: SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, J. Geophys. Res., 118, 11824–11846, doi:10.1002/jgrd.50752, 2013, 2013.

Hoor, P., Borken-Kleefeld, J., Caro, D., Dessens, O., Endresen, O., Gauss, M., Grewe, V., Hauglustaine, D., Isaksen, I. S. A., Jöckel, P., Lelieveld, J., Myhre, G., Meijer, E., Olivie, D., Prather, M., Schnadt Poberaj, C., Shine, K. P., Staehelin, J., Tang, Q., van Aardenne, J., van Velthoven, P., and Sausen, R.: The impact of traffic emissions on atmospheric ozone and OH: results from QUANTIFY, Atmos. Chem. Phys., 9, 3113–3136, doi:10.5194/acp-9-3113-2009, 2009.

IPCC: Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Houghton, J. T., Ding, Y., Griggs. D. J., Noguer, M., van der Linden. P. J., Dai, X., Maskell, K., and Johnson, C. A., Cambridge University Press, Cambridge, UK, 2001.

IPCC: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK and New York, USA, 2007.

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., Kerkweg, A., Buchholz-Dietsch, J., Tost, H., Sander, R., and Pozzer, A.: Technical Note: Coupling of chemical processes with the Modular Earth Submodel System (MESSy) submodel TRACER, Atmos. Chem. Phys., 8, 1677–1687, doi:10.5194/acp-8-1677-2008, 2008.

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

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Karpechko, A. Yu., Gillett, N. P., Hassler, B., Rosenlof, K. H., and Rozanov, E.: Quantitative assessment of Southern Hemisphere ozone in chemistry-climate model simulations, Atmos. Chem. Phys., 10, 1385–1400, doi:10.5194/acp-10-1385-2010, 2010.

Kerkweg, A., Buchholz, J., Ganzeveld, L., Pozzer, A., Tost, H., and Jöckel, P.: Technical Note: An implementation of the dry removal processes DRY DEPosition and SEDImentation in the Modular Earth Submodel System (MESSy), Atmos. Chem. Phys., 6, 4617–4632, doi:10.5194/acp-6-4617-2006, 2006a.

Kerkweg, A., Sander, R., Tost, H., and Jöckel, P.: Technical note: Implementation of prescribed (OFFLEM), calculated (ONLEM), and pseudo-emissions (TNUDGE) of chemical species in the Modular Earth Submodel System (MESSy), Atmos. Chem. Phys., 6, 3603–3609, doi:10.5194/acp-6-3603-2006, 2006b.

Kirner, O., Ruhnke, R., Buchholz-Dietsch, J., Jöckel, P., Brühl, C., and Steil, B.: Simulation of polar stratospheric clouds in the chemistry-climate-model EMAC via the submodel PSC, Geosci. Model Dev., 4, 169–182, doi:10.5194/gmd-4-169-2011, 2011.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmos. Chem. Phys., 10, 7017–7039, doi:10.5194/acp-10-7017-2010, 2010.

Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P. J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S. T., Schulz, M., Skeie, R. B., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., and Zeng, G.: The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations and climate diagnostics, Geosci. Model Dev., 6, 179–206, doi:10.5194/gmd-6-179-2013, 2013.

Landgraf, J. and Crutzen, P. J.: An efficient method for online calculations of photolysis and heating rates, J. Atmos. Sci., 55, 863–878, doi:10.1175/1520-0469(1998)055<0863:AEMFOC>2.0.CO;2, 1998.

Lawrence, M. G., Jöckel, P. and von Kuhlmann, R.: What does the global mean OH concentration tell us?, Atmos. Chem. Phys., 1, 37–49, doi:10.5194/acp-1-37-2001, 2001.

Lelieveld, J., Brühl, C., Jöckel, P., Steil, B., Crutzen, P. J., Fischer, H., Giorgetta, M. A., Hoor, P., Lawrence, M. G., Sausen, R., and Tost, H.: Stratospheric dryness: model simulations and satellite observations, Atmos. Chem. Phys., 7, 1313–1332, doi:10.5194/acp-7-1313-2007, 2007.

Lin, S.-J., and Rood, R. B.: Multidimensional flux-form semi-Lagrangian transport schemes, Mon. Weather Rev., 124, 2046–2070, doi:10.1175/1520-0493(1996)124<2046:MFFSLT>2.0.CO;2, 1996.

Logan, J. A.: An analysis of ozonesonde data for the troposphere: recommendations for testing 3-D models and development of a gridded climatology for tropospheric ozone, J. Geophys. Res., 104, 16115–16149, doi:10.1029/1998JD100096, 1999.

Loyola, D. and Coldewey-Egbers, M.: Multi-sensor data merging with stacked neural networks for the creation of satellite long-term climate data records, EURASIP J. Adv. Sig. Pr., 2012, 1–10, doi:10.1186/1687-6180-2012-91, 2012.

Loyola, D. G., Coldewey-Egbers, R. M., Dameris, M., Garny, H., Stenke, A., Van Roozendael, M., Lerot, C., Balis, D., and Koukouli, M.: Global long-term monitoring of the ozone layer – a prerequisite for predictions, Int. J. Remote Sens., 30, 4295–4318, doi:10.1080/01431160902825016, 2009.

Masarie, K. A. and Tans, P. P.: Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record, J. Geophys. Res., 100, 11593–11610, doi:10.1029/95JD00859, 1995.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H. and Tomassini, L.: Tuning the climate of a global model, J. Adv. Model. Earth Syst., 4, M00A01, doi:10.1029/2012MS000154, 2012.

Meehl, G. A., Moss, R., Taylor, K. E., Eyring, V., Stouffer, R. J., Bony, S., and Stevens, B.: Climate model intercomparisons: preparing for the next phase, Trans. Am. Geophys. Union, 95, 77–78, doi:10.1002/2014EO090001, 2014.

Monks, S. A., Arnold, S. R., Emmons, L. K., Law, K. S., Turquety, S., Duncan, B. N., Flemming, J., Huijnen, V., Tilmes, S., Langner, J., Mao, J., Long, Y., Thomas, J. L., Steenrod, S. D., Raut, J. C., Wilson, C., Chipperfield, M. P., Schlager, H. and Ancellet, G.: Multi-model study of chemical and physical controls on transport of anthropogenic and biomass burning pollution to the Arctic, Atmos. Chem. Phys. Discuss., 14, 25281–25350, doi:10.5194/acpd-14-25281-2014, 2014.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment., Nature, 463, 747–756, doi:10.1038/nature08823, 2010.

Naik, V., Voulgarakis, A., Fiore, A. M., Horowitz, L. W., Lamarque, J.-F., Lin, M., Prather, M. J., Young, P. J., Bergmann, D., Cameron-Smith, P. J., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R., Eyring, V., Faluvegi, G., Folberth, G. A., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., van Noije, T. P. C., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R., Shindell, D. T., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., and Zeng, G.: Preindustrial to present-day changes in tropospheric hydroxyl radical and methane lifetime from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 5277–5298, doi:10.5194/acp-13-5277-2013, 2013.

Nordeng, T.-E.: Extended Versions of the Convective Parametrization Scheme at ECMWF and Their Impact on the Mean and Transient Activity of the Model in the Tropics, European Centre for Medium-Range Weather Forecasts, Reading, UK, 1994.

Park, J. H., Russell, J. M., Gordley, L. L., Drayson, S. R., Benner, D. C., McInerney, J. M., Gunson, M. R., Toon, G. C., Sen, B., Blavier, J.-F., Webster, C. R., Zipf, E. C., Erdman, P., Schmidt, U., and Schiller, C.: Validation of halogen occultation experiment CH_4 measurements from the UARS, J. Geophys. Res., 101, 10183–10203, doi:10.1029/95JD02736, 1996.

Perlwitz, J., Pawson, S., Fogt, R. L., Nielsen, J. E., and Neff, W. D.: Impact of stratospheric ozone hole recovery on Antarctic climate, Geophys. Res. Lett., 35, L08714, doi:10.1029/2008GL033317, 2008.

Pozzer, A., Jöckel, P., Sander, R., Williams, J., Ganzeveld, L., and Lelieveld, J.: Technical Note: The MESSy-submodel AIRSEA calculating the air-sea exchange of chemical species, Atmos. Chem. Phys., 6, 5435–5444, doi:10.5194/acp-6-5435-2006, 2006.

Pozzer, A., Jöckel, P., Tost, H., Sander, R., Ganzeveld, L., Kerkweg, A., and Lelieveld, J.: Simulating organic species with the global atmospheric chemistry general circulation model ECHAM5/MESSy1: a comparison of model results with observations, Atmos. Chem. Phys., 7, 2527–2550, doi:10.5194/acp-7-2527-2007, 2007.

Price, C. and Rind, D.: Modeling global lightning distributions in a general circulation model, Mon. Weather Rev., 122, 1930–1939, doi:10.1175/1520-0493(1994)122<1930:MGLDIA>2.0.CO;2, 1994.

Prinn, R. G., Weiss, R. F., Fraser, P. J., Simmonds, P. G., Cunnold, D. M., Alyea, F. N., Doherty, S. O., Salameh, P., Miller, B. R., Huang, J., Wang, R. H. J., Hartley, D. E., Harth, C., Steele, L. P., Sturrock, G., Midgley, P. M., and McCulloch, A.: A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE, J. Geophys. Res., 105, 17751–17792, doi:10.1029/2000JD900141, 2000.

Räisänen, P., and Järvinen, H.: Impact of cloud and radiation scheme modifications on climate simulated by the ECHAM5 atmospheric GCM, Q. J. Roy. Meteor. Soc., 136, 1733–1752, doi:10.1002/qj.674, 2010.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108, 4407, doi:10.1029/2002JD002670, 2003.

Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, B. Am. Meteorol. Soc., 89, 303–311, doi:10.1175/BAMS-89-3-303, 2008.

Roeckner, E., Brokopf, E., Esch, M., Giorgetta, M. A., Hagema, S., and Kornblueh, L.: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model, J. Climate, 19, 3771–3791, doi:10.1175/JCLI3824.1, 2006.

Russell, J. M., Gordley, L. L., Park, J. H., Drayson, S. R., Hesketh, W. D., Cicerone, R. J., Tuck, A. F., Frederick, J. E., Harries, J. E., and Crutzen, P. J.: The halogen occultation experiment, J. Geophys. Res., 98, 10777–10797, doi:10.1029/93JD00799, 1993.

Russell, J. M., Deaver, L. E., Luo, M., Park, J. H., Gordley, L. L., Tuck, A. F., Toon, G. C., Gunson, M. R., Traub, W. A., Johnson, D. G., Jucks, K. W., Murcray, D. G., Zander, R., Nolt, I. G., and Webster, C. R.: Validation of hydrogen chloride measurements made by the halogen occultation experiment from the UARS platform, J. Geophys. Res., 101, 10151–10162, doi:10.1029/95JD01696, 1996.

Sander, S. P., Finlayson-Pitts, B. J., Friedl, R. R., Golden, D. M., Huie, R. E., Kolb, C. E., Kurylo, M. J., Molina, M. J., Moortgat, G. K., Orkin, V. L., and Ravishankara, A. R.: Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation Number 14, JPL Publication 02-25, Jet Propulsion Laboratory, Pasadena, USA, 2003.

Sander, R., Kerkweg, A., Jöckel, P., and Lelieveld, J.: Technical note: The new comprehensive atmospheric chemistry module MECCA, Atmos. Chem. Phys., 5, 445–450, doi:10.5194/acp-5-445-2005, 2005.

Sander, R., Baumgaertner, A., Gromov, S., Harder, H., Jöckel, P., Kerkweg, A., Kubistin, D., Regelin, E., Riede, H., Sandu, A., Taraborrelli, D., Tost, H., and Xie, Z.-Q.: The atmospheric chemistry box model CAABA/MECCA-3.0, Geosci. Model Dev., 4, 373–380, doi:10.5194/gmd-4-373-2011, 2011.

Schmitt, A. and Brunner, B.: Emissions from aviation and their development over time, in: Pollutants from Air Traffic – Results of Atmospheric Research 1992–1997, edited by: Schumann, U.,

Chlond, A., Ebel, A., Kärcher, B., Pak, H., Schlager, H., Schmitt, A., and Wendling, P., DLR-Mitteilung 97-04, Deutsches Zentrum für Luft- und Raumfahrt, Cologne, Germany, 37–52, 1997.

Schumann, U. and Huntrieser, H.: The global lightning-induced nitrogen oxides source, Atmos. Chem. Phys., 7, 3823–3907, doi:10.5194/acp-7-3823-2007, 2007.

Shepherd, T. G.: Dynamics, stratospheric ozone, and climate change, Atmos. Ocean, 46, 117–138, doi:10.3137/ao.460106, 2008.

Shindell, D. T., Faluvegi, G., Stevenson, D. S., Krol, M. C., Emmons, L. K., Lamarque, J.-F., Pétron, G., Dentener, F. J., Ellingsen, K., Schultz, M. G., Wild, O., Amann, M., Atherton, C. S., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Lawrence, M. G., Montanaro, V., Müller, J.-F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Strahan, S. E., Sudo, K., Szopa, S., Unger, N., van Noije, T. P. C. and Zeng, G.: Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, J. Geophys. Res., 111, D19, doi:10.1029/2006JD007100, 2006.

Shindell, D. T., Chin, M., Dentener, F., Doherty, R. M., Faluvegi, G., Fiore, A. M., Hess, P., Koch, D. M., MacKenzie, I. A., Sanderson, M. G., Schultz, M. G., Schulz, M., Stevenson, D. S., Teich, H., Textor, C., Wild, O., Bergmann, D. J., Bey, I., Bian, H., Cuvelier, C., Duncan, B. N., Folberth, G., Horowitz, L. W., Jonson, J., Kaminski, J. W., Marmer, E., Park, R., Pringle, K. J., Schroeder, S., Szopa, S., Takemura, T., Zeng, G., Keating, T. J., and Zuber, A.: A multi-model assessment of pollution transport to the Arctic, Atmos. Chem. Phys., 8, 5353–5372, doi:10.5194/acp-8-5353-2008, 2008.

Silva, R. A., West, J. J., Zhang, Y., Anenberg, S. C., Lamarque, J.-F., Shindell, D. T., Collins, W. J., Dalsøren, S., Faluvegi, G., Folberth, G., Horowitz, L. W., Nagashima, T., Naik, V., Rumbold, S., Skeie, R., Sudo, K., Takemura, T., Bergmann, D., Cameron-Smith, P., Cionni, I., Doherty, R. M., Eyring, V., Josse, B., MacKenzie, I. A., Plummer, D., Righi, M., Stevenson, D. S., Strode, S., Szopa, S., and Zeng, G.: Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change, Environ. Res. Lett., 8, 034005, doi:10.1088/1748-9326/8/3/034005, 2013.

Son, S.-W., Polvani, L. M., Waugh, D. W., Akiyoshi, H., Garcia, R., Kinnison, D., Pawson, S., Rozanov, E., Shepherd, T. G., and Shibata, K.: The impact of stratospheric ozone recovery on the Southern Hemisphere westerly jet, Science, 320, 1486–1489, doi:10.1126/science.1155939, 2008.

Son, S.-W., Gerber, E. P., Perlwitz, J., Polvani, L. M., Gillett, N. P., Seo, K.-H., Eyring, V., Shepherd, T. G., Waugh, D., Akiyoshi, H., Austin, J., Baumgaertner, A., Bekki, S., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M. P., Cugnet, D., Dameris, M., Dhomse, S., Frith, S., Garny, H., Garcia, R., Hardiman, S. C., Jöckel, P., Lamarque, J. F., Mancini, E., Marchand, M., Michou, M., Nakamura, T., Morgenstern, O., Pitari, G., Plummer, D. A., Pyle, J., Rozanov, E., Scinocca, J. F., Shibata, K., Smale, D., Teyssèdre, H., Tian, W., and Yamashita, Y.: Impact of stratospheric ozone on Southern Hemisphere circulation change: a multimodel assessment, J. Geophys. Res., 115, D00M07, doi:10.1029/2010JD014271, 2010.

SPARC-CCMVal: SPARC Report on the Evaluation of Chemistry-Climate Models, edited by: Eyring, V., Shepherd, T. G., and Waugh, D. W., SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, 2010.

Stenke, A., Grewe, V., and Ponater, M.: Lagrangian transport of water vapor and cloud water in the ECHAM4 GCM and its impact on the cold bias, Clim. Dynam., 31, 491–506, doi:10.1007/s00382-007-0347-5, 2008.

Stevenson, D. S., Young, P. J., Naik, V., Lamarque, J.-F., Shindell, D. T., Voulgarakis, A., Skeie, R. B., Dalsoren, S. B., Myhre, G., Berntsen, 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.

Tanre, D., Geleyn, J.-F., and Slingo, J.: First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model, in: Aerosols and Their Climatic Effects, edited by: Gerber, H. and Deepak, A., A. Deepak Publ., 133–177, 1994.

Taraborrelli, D., Lawrence, M. G., Crowley, J. N., Dillon, T. J., Gromov, S., Grosz, C. B. M., Vereecken, L., and Lelieveld, J.: Hydroxyl radical buffered by isoprene oxidation over tropical forests, Nat. Geosci., 5, 190–193, doi:10.1038/ngeo1405, 2012.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, doi:10.1029/2000JD900719, 2001.

Thompson, D. W. J. and Solomon, S.: Interpretation of recent Southern Hemisphere climate change, Science, 296, 895–899, doi:10.1126/science.1069270, 2002.

Thompson, D. W. J. and Solomon, S.: Recent stratospheric climate trends as evidenced in radiosonde data: global structure and tropospheric linkages, J. Climate, 18, 4785–4795, doi:10.1175/JCLI3585.1, 2005.

Thompson, D. W. J., Baldwin, M. P., and Solomon, S.: Stratosphere–troposphere coupling in the Southern Hemisphere, J. Climate, 62, 708–715, doi:10.1175/JAS-3321.1, 2005.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Weather Rev., 117, 1779–1800, doi:10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Conley, A., Schultz, M. G., Saunois, M., Thouret, V., Thompson, A. M., Oltmans, S. J., Johnson, B., and Tarasick, D.: Technical Note: Ozonesonde climatology between 1995 and 2011: description, evaluation and applications, Atmos. Chem. Phys., 12, 7475–7497, doi:10.5194/acp-12-7475-2012, 2012.

Tost, H.: Global modelling of cloud, convection and precipitation influences on trace gases and aerosols, Ph.D. thesis, University of Bonn, Germany, available at: http://hss.ulb.uni-bonn.de/2006/0731/0731.htm (last access: 2 October 2014), 2006.

Tost, H., Jöckel, P., Kerkweg, A., Sander, R., and Lelieveld, J.: Technical note: A new comprehensive SCAVenging submodel for global atmospheric chemistry modelling, Atmos. Chem. Phys., 6, 565–574, doi:10.5194/acp-6-565-2006, 2006a.

Tost, H., Jöckel, P., and Lelieveld, J.: Influence of different convection parameterisations in a GCM, Atmos. Chem. Phys., 6, 5475–5493, doi:10.5194/acp-6-5475-2006, 2006b.

van Aalst, M. K., van den Broek, M. M. P., Bregman, A., Brühl, C., Steil, B., Toon, G. C., Garcelon, S., Hansford, G. M., Jones, R. L., Gardiner, T. D., Roelofs, G. J., Lelieveld, J., and Crutzen, P.J.: Trace gas transport in the 1999/2000 Arctic winter: comparison of nudged GCM runs with observations, Atmos. Chem. Phys., 4, 81–93, doi:10.5194/acp-4-81-2004, 2004.

van Aardenne, J., Dentener, F., Olivier, J., Peters, J., and Ganzeveld, L.: The EDGAR3.2 Fast Track 2000 dataset (32FT2000), Tech. rep., Joint Research Centre, Institute for Environment and Sustainability (JRC-IES), Climate Change Unit, Ispra (VA), Italy, 2005.

van der Werf, G. R., Randerson, J. T., Giglio, L., Gobron, N., and Dolman, A. J.: Climate controls on the variability of fires in the tropics and subtropics, Global Biogeochem. Cy., 22, GB3028, doi:10.1029/2007GB003122, 2008.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contri-

bution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.

von Kuhlmann, R., Lawrence, M. G., Crutzen, P. J., and Rasch, P. J.: A model for studies of tropospheric ozone and nonmethane hydrocarbons: model description and ozone results, J. Geophys. Res., 108, 4294, doi:10.1029/2002JD002893, 2003.

von Storch, H. and Zwiers, F.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, UK, 1984.

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.

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Louis Smith, G., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): an earth observing system experiment, B. Am. Meteorol. Soc., 77, 853–868, doi:10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2, 1996.

WMO: Scientific Assessment of Ozone Depletion 2010, Global Ozone Research and Monitoring Project-Report No. 52, Geneva, Switzerland, 2011.

Yokoi, S., Takayabu, Y. N., Nishii, K., Nakamura, H., Endo, H., Ichikawa, H., Inoue, T., Kimoto, M., Kosaka, Y., Miyasaka, T., Oshima, K., Sato, N., Tsushima, Y., and Watanabe, M.: How well do coupled models simulate today's climate?, J. Appl. Meteorol. Clim., 50, 1666–1675, doi:10.1175/2011JAMC2643.1, 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., Dalsøren, 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.

Ziemke, J. R., Chandra, S., Duncan, B. N., Froidevaux, L., Bhartia, P. K., Levelt, P. F., and Waters, J. W.: Tropospheric ozone determined from Aura OMI and MLS: Evaluation of measurements and comparison with the Global Modeling Initiative's Chemical Transport Model, J. Geophys. Res., 111, D19303, doi:10.1029/2006JD007089, 2006.

Ziemke, J. R., Chandra, S., Labow, G. J., Bhartia, P. K., Froidevaux, L., and Witte, J. C.: A global climatology of tropospheric and stratospheric ozone derived from Aura OMI and MLS measurements, Atmos. Chem. Phys., 11, 9237–9251, doi:10.5194/acp-11-9237-2011, 2011.

Table 1. List of the MESSy submodels used in the simulations. See Jöckel et al. (2006, 2010) for more details.

Submodel	Description	Reference
AIRSEA	Atmosphere–ocean exchange of trace gases	Pozzer et al. (2006)
CLOUD	Cloud scheme from ECHAM5	Roeckner et al. (2006)
CONVECT	Convection	Tost et al. (2006b)
CVTRANS	Convective transport of tracers	Tost (2006)
DRYDEP (DDEP in MESSy2)	Dry deposition of trace gases and aerosol	Kerkweg et al. (2006a)
H2O	Stratospheric water vapour and feedback	Jöckel et al. (2006)
HETCHEM	Heterogeneous reaction rates	Jöckel et al. (2006)
JVAL	Photolysis rate coefficients	Landgraf and Crutzen (1998)
LNOX	Lightning NO_x emissions	Price and Rind (1994) Grewe et al. (2001)
MECCA	Tropospheric and stratospheric chemistry	Sander et al. (2005)
OFFLEM (OFFEMIS in MESSy2)	Offline (prescribed) emissions of trace gases and aerosol	Kerkweg et al. (2006b)
ONLEM (ONEMIS in MESSy2)	Online emissions of trace gases and aerosol	Kerkweg et al. (2006b)
PSC	Polar stratospheric clouds	Kirner et al. (2011)
PTRAC	User-defined prognostic tracers	Jöckel et al. (2008)
QBO	Newtonian relaxation of quasi-biennial oscillation	Giorgetta and Bengtsson (1999) Jöckel et al. (2006)
RAD4ALL (RAD in MESSy2)	Radiation scheme from ECHAM5	Roeckner et al. (2006)

Table 2. Overview of the four EMAC simulations evaluated in this study. All experiments have a spin-up year at the beginning of the simulated period which is not considered in the analysis.

Name	Resolution	Analysed time period	Running mode	SST/SIC
EVAL2	T42L90MA	1999–2009	Nudged, coupled	ECMWF
QCTM	T42L90MA	1999–2007	Nudged, QCTM	ECMWF
TS2000	T42L90MA	10 years under 2000 conditions	Free-running timeslice, coupled	HadISST1
ACCMIP	T42L90MA	10 years under 2000 conditions	Free-running timeslice, coupled	CMCC

Table 3. List of the diagnostics applied in this work and for which a quantitative evaluation based on performance metrics has been applied. The climatological mean field considers both the time (annual cycle) and the space (latitude-longitude) coordinate, or only time in some cases. Regions are defined as follows: Glob (90° N–90° S), Trop (20° N–20° S), NHext (20–90° N), SHext (20–90° S), NHmidlat (35–60° N), SHmidlat (35–60° S), NHpolar (60–90° N), SHpolar (60–90° S). The short name of the diagnostics follows the CMOR ~~standard~~metadata definition. The references in the last column are labeled as follows: Bo2005 (Bodeker et al., 2005), GC2010 (GLOBALVIEW-CO2, 2010), Gl2008 (Gleckler et al., 2008), Lo2009 (Loyola et al., 2009), Lo2012 (Loyola and Coldewey-Egbers, 2012), Ti2000 (Tilmes et al., 2012), Zi2006 (Ziemke et al., 2006), Zi2011 (Ziemke et al., 2011).

Name	Clim. mean field	Level	Region	Observations	Short name	Reference
Temperature	space–time	850 hPa 200 hPa 30 hPa 5 hPa	Glob Trop NHext SHext	Era-Interim (1996–2005) NCEP (1996–2005)	ta_<reg>-<lev>	Gl2008
Eastward wind	space–time	850 hPa 200 hPa 30 hPa 5 hPa	Glob Trop NHext SHext	Era-Interim (1996–2005) NCEP (1996–2005)	ua_<reg>-<lev>	Gl2008
Northward wind	space–time	850 hPa 200 hPa 30 hPa 5 hPa	Glob Trop NHext SHext	Era-Interim (1996–2005) NCEP (1996–2005)	va_<reg>-<lev>	Gl2008
Geopotential height	space–time	850 hPa 500 hPa 30 hPa 5 hPa	Glob Trop NHext SHext	Era-Interim (1996–2005) NCEP (1996–2005)	zg_<reg>-<lev>	Gl2008
Specific humidity	space–time	850 hPa 400 hPa 30 hPa 5 hPa	Glob Trop NHext SHext	AIRS (2003–2010) ERA-Interim (1996–2005)	hus_<reg>-<lev>	Gl2008

Table 3. Continued.

Name	Clim. mean field	Level	Region	Observations	Short name	Reference
Outgoing LW radiation (clear-sky)	space-time	–	Glob Trop NHExt SHExt	SRB (1995–2005) CERES (2001–2012)	rlutcs_<reg>	Gl2008
Outgoing LW radiation (all-sky)	space-time	–	Glob Trop NHExt SHExt	SRB (1995–2005) CERES (2001–2012)	rlut_<reg>	Gl2008
Reflected SW radiation (all-sky)	space-time	–	Glob Trop NHExt SHExt	SRB (1995–2005) CERES (2001–2012)	rsut_<reg>	Gl2008
Total column ozone	space–time	–	Glob Trop NHmidlat SHmidlat NHpolar SHpolar	NIWA (1998–2010) GOME (1998–2010)	toz_<reg>	Bo2005 Lo2009 Lo2012
Tropospheric column ozone	space–time	–	Glob Trop NHext SHext	MLS/OMI (2005–2012)	tropoz_<reg>	Zi2006 Zi2011
Tropospheric ozone	time	700 hPa 500 hPa 250 hPa	Trop NHext SHext	Ozonesondes (1995–2009)	vmro3_<reg>-<lev>	Ti2000
Surface CO	time	surface	various	GLOBALVIEW (1999–2008)	vmrco_<reg>	GC2010

Table 4. Estimated methane and MCF lifetimes for the EMAC simulations.

Simulation	CH_4 lifetime [yr]	MCF lifetime [yr]
EVAL2	8.44	5.06
QCTM	9.10	5.53
TS2000	7.85	4.76
ACCMIP	7.99	4.85
ACCMIP-S1	8.22	4.98
ACCMIP-S2	12.51	7.57

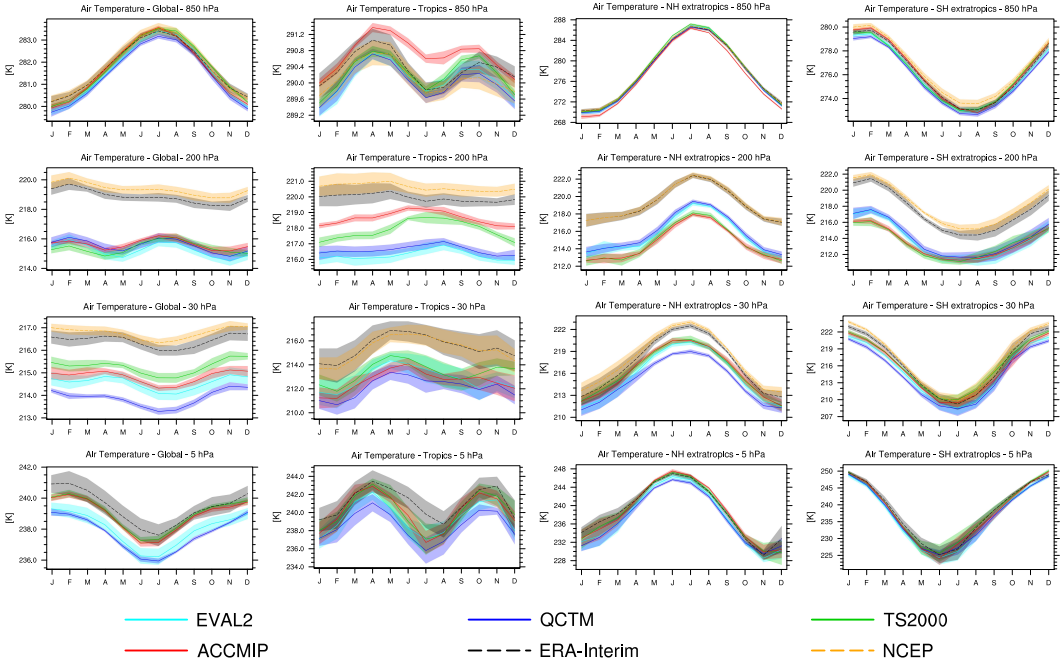

Figure 1. Annual cycle of temperature climatology at 850, 200, 30 and 5 hPa averaged globally, over the tropics (20° S–20° N), NH extratropics (20–90° N) and SH extratropics (20–90° S) for the EMAC simulations, in comparison to ERA-Interim and NCEP/NCAR reanalysis data. Shaded areas indicate the $\pm 1\sigma$ interannual variability.

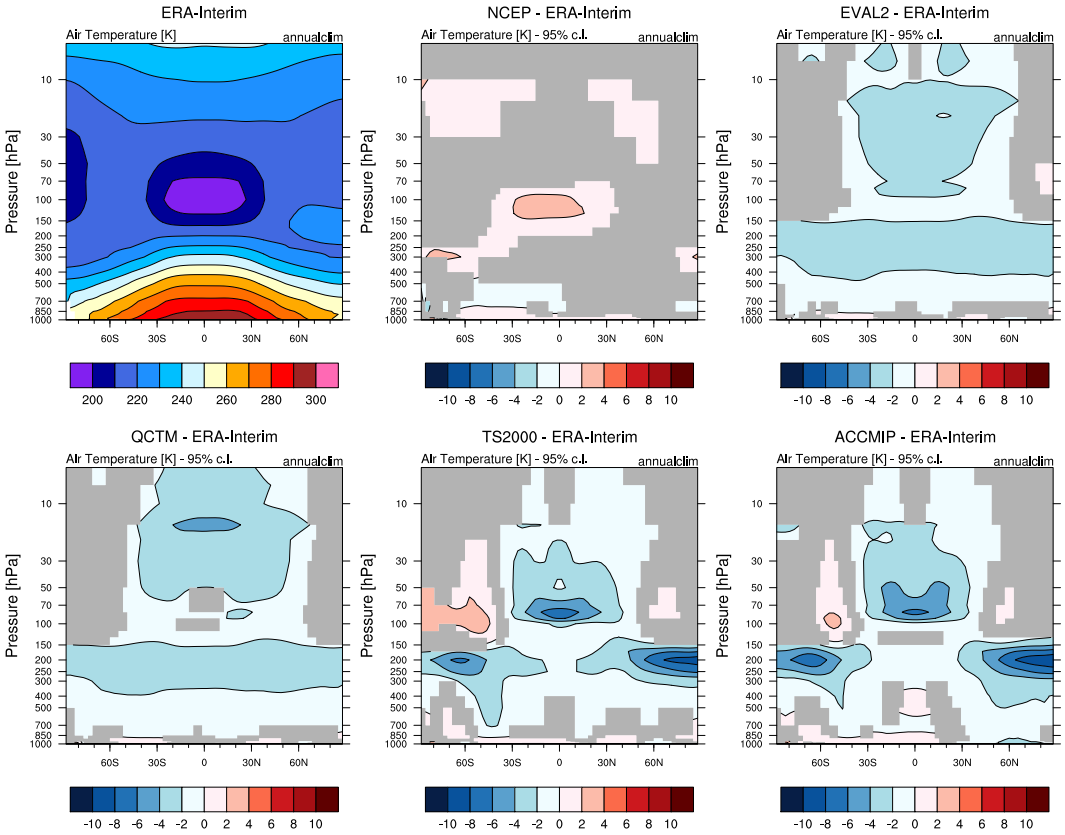

Figure 2. Annual mean of zonally averaged temperature profile for the EMAC simulations in comparison to ERA-Interim and NCEP/NCAR reanalysis data. The upper left plot shows ERA-Interim absolute values; all other plots show differences between the model simulations (or NCEP/NCAR) and ERA-Interim. Differences between the two fields which are not statistically significant according to the t test (95 % confidence level) are masked out in gray.

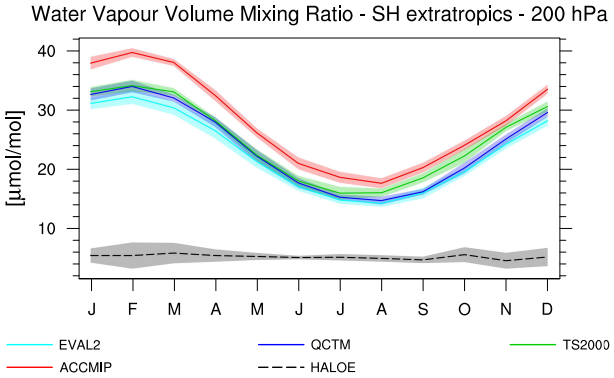

Figure 3. Annual cycle of water vapour climatology at 200 hPa averaged over the SH extratropics (20–90° S) for the EMAC simulations in comparison to HALOE data. Shaded area indicates the $\pm 1\sigma$ interannual variability.

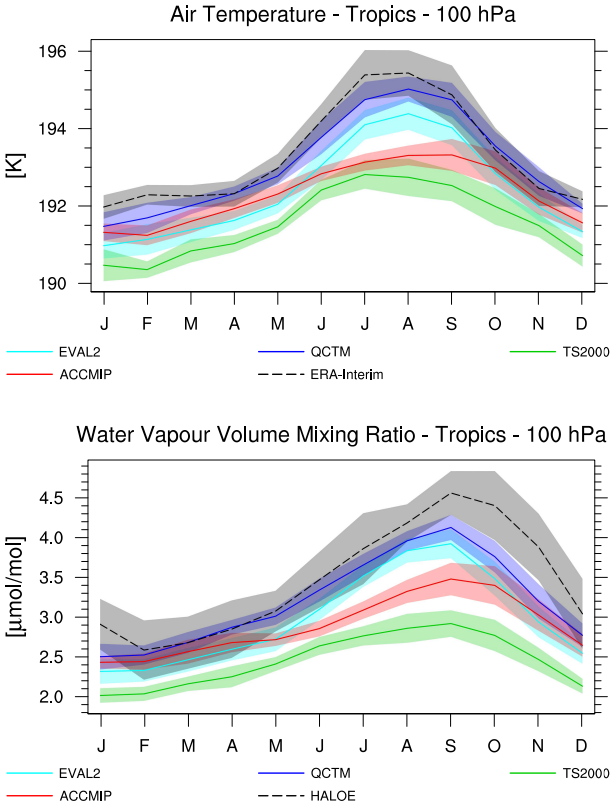

Figure 4. Annual cycle of temperature (top) and water vapour (bottom) climatology at 100 hPa averaged over the tropics (20° N–20° S) for the EMAC simulations, in comparison to ERA-Interim reanalysis and HALOE data, respectively. Shaded areas indicate the $\pm 1\sigma$ interannual variability.

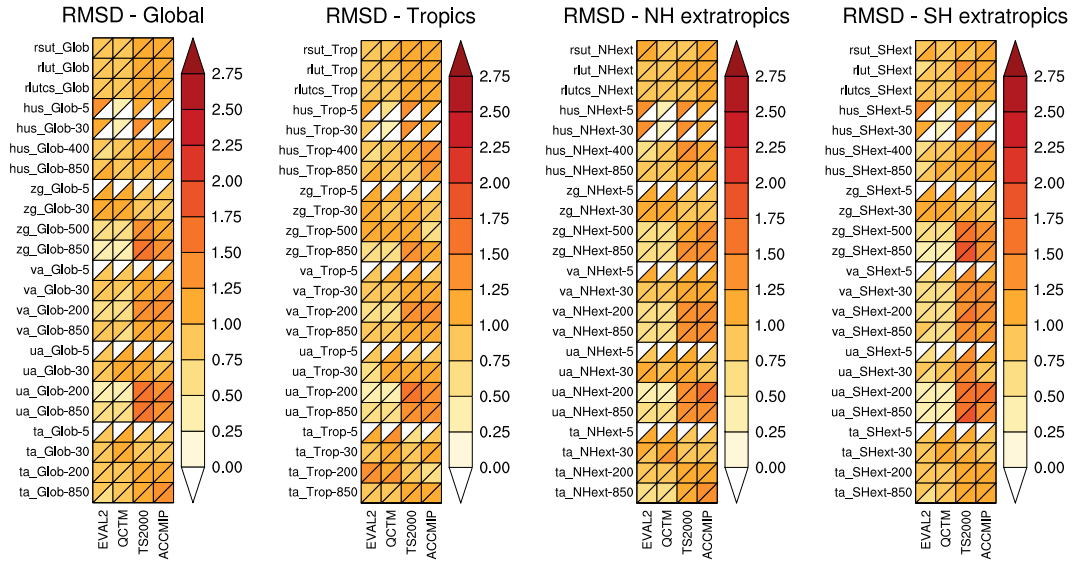

Figure 5. Root-mean square difference of the chosen basic climate parameters over the global domain, the tropics, and the NH and SH extratropics (from left to right, respectively). Columns and rows of each panel represent the EMAC simulations and the given diagnostics (see Table 3), respectively. Where an alternative dataset is available, the diagram boxes are split in two parts, showing the model performance compared to the primary (lower triangle) and alternative (upper triangle) dataset. Where no observations are available, the triangles are marked white.

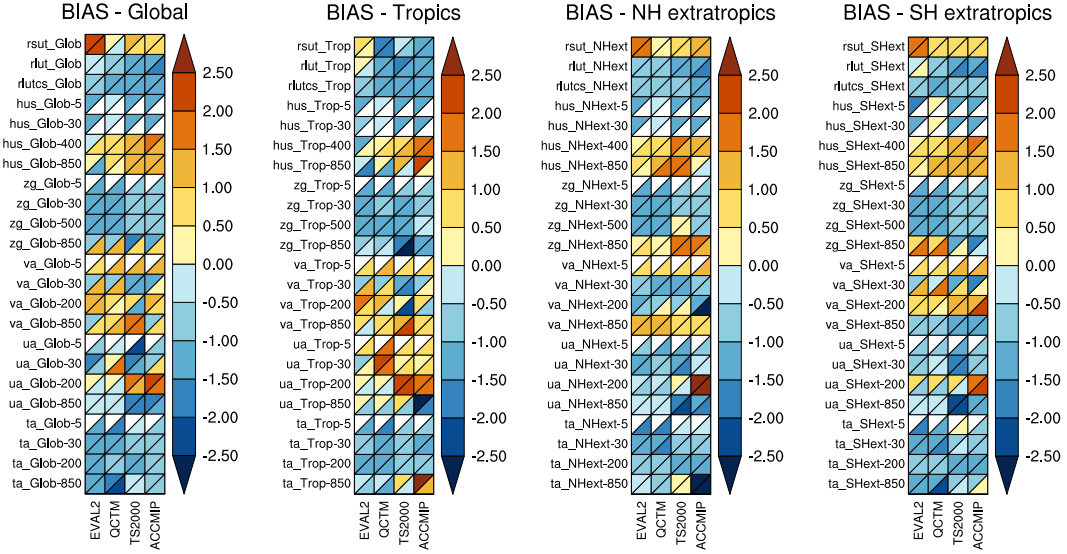

Figure 6. As in Fig. 5, for the overall mean bias.

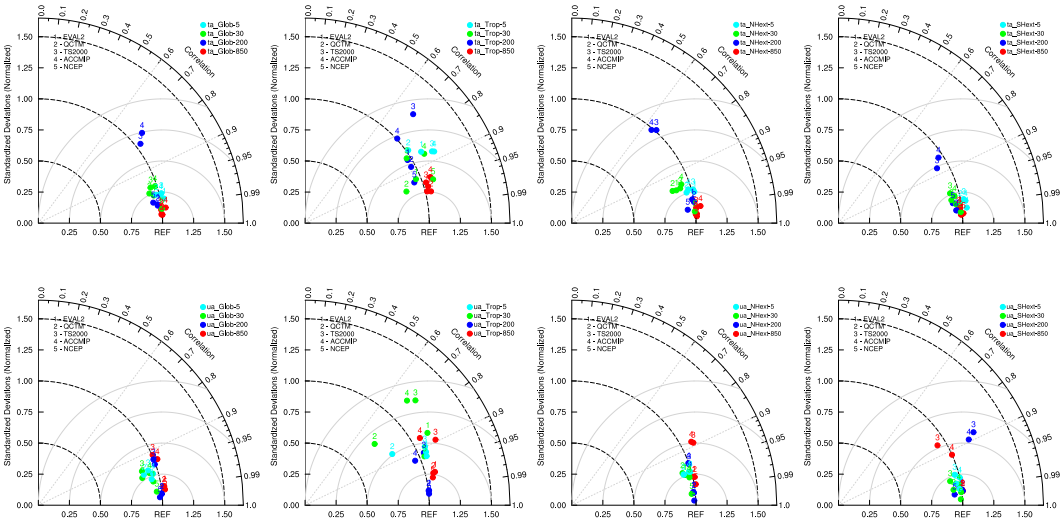

Figure 7. Taylor diagrams of temperature (top row) and eastward wind (bottom row) over the four chosen domains (global, tropics, NH and SH extratropics, from left to right, respectively) and height-levels (850, 200, 30, and 5 hPa).

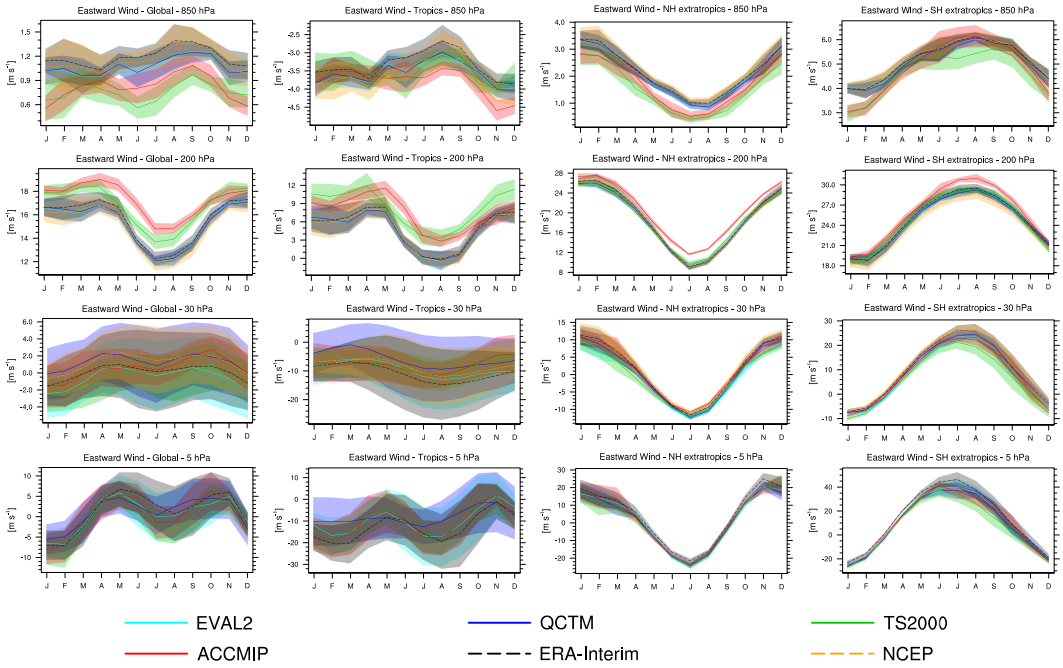

Figure 8. ~~Seasonal (DJF) mean of zonally averaged eastward wind profile for the EMAC simulations~~ As in ~~comparison to ERA-Interim and NCEP/NCAR reanalysis data~~Fig.~~The upper left plot shows ERA-Interim absolute values; all other plots show differences between the model simulations (or NCEP/NCAR) and ERA Interim. Differences between the two fields which are not statistically significant according to the t test (95confidence level) are masked out in gray~~1, for eastward wind.

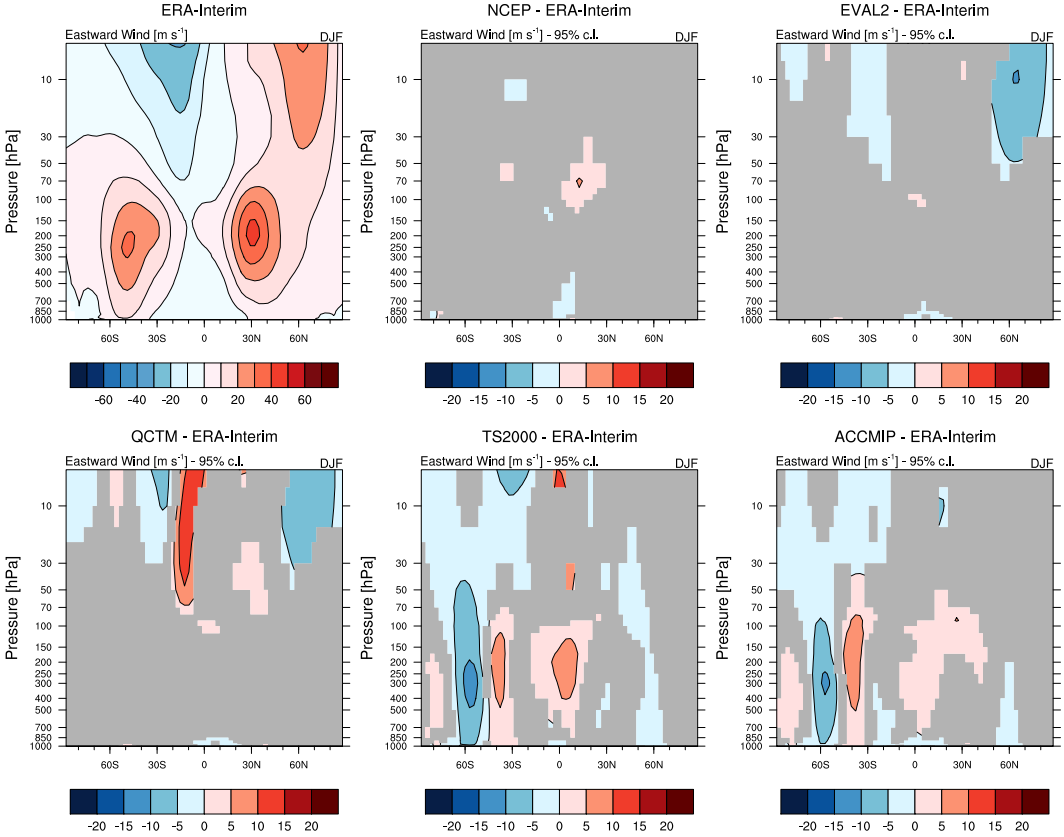

Figure 9. As in Fig. 2, for seasonal (DJF) mean of zonally averaged eastward wind profile.

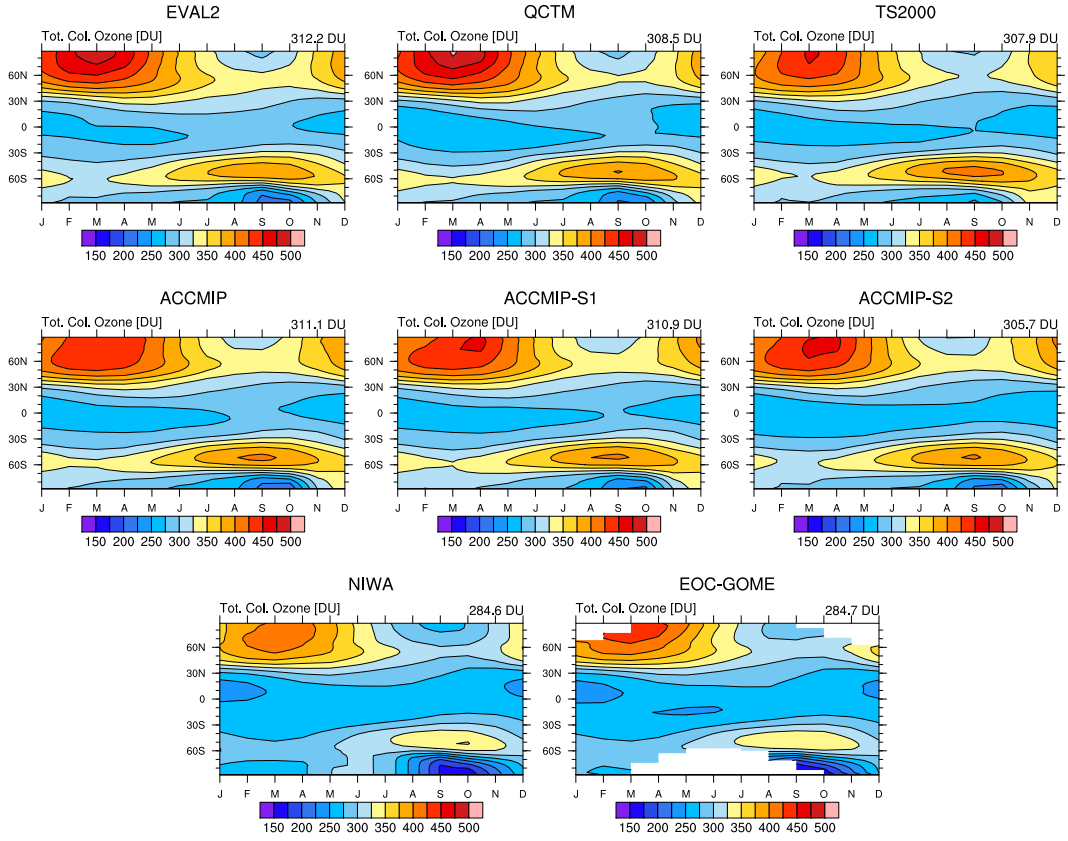

Figure 10. Total column ozone climatology for the EMAC simulations compared to the NIWA combined total column ozone database and GOME data. The values on top of each panel show the global (area-weighted) average, calculated after regridding the data to the horizontal grid of the model and ignoring the grid cells without available observational data in the GOME dataset.

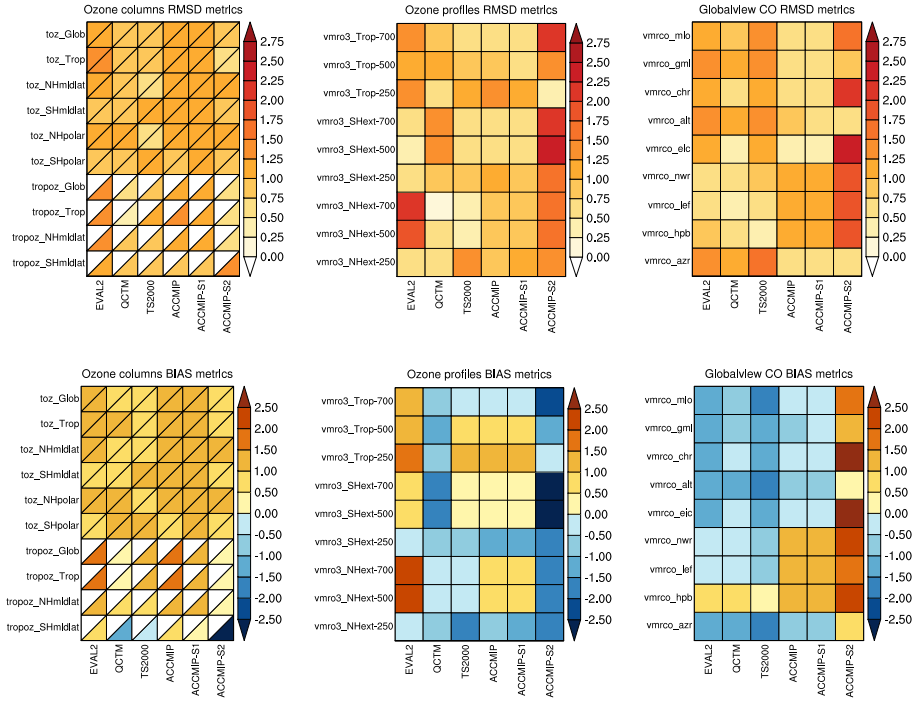
Ozone columns RMSD metrics
toz_Glob
toz_Trop
toz_NHmidlat
toz_SHmidlat
toz_NHpolar
toz_SHpolar
tropoz_Glob
tropoz_Trop
tropoz_NHmidlat
tropoz_SHmidlat
EVAL2
QCTM
TS2000
ACCMIP
ACCMIP-S1
ACCMIP-S2
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Ozone profiles RMSD metrics
vmro3_Trop-700
vmro3_Trop-500
vmro3_Trop-250
vmro3_SHext-700
vmro3_SHext-500
vmro3_SHext-250
vmro3_NHext-700
vmro3_NHext-500
vmro3_NHext-250
Globalview CO RMSD metrics
vmrco_mlo
vmrco_gml
vmrco_chr
vmrco_alt
vmrco_eic
vmrco_nwr
vmrco_lef
vmrco_hpb
vmrco_azr
Ozone columns BIAS metrics
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
Ozone profiles BIAS metrics
Globalview CO BIAS metrics

Figure 11. ~~Zonal~~ Root-mean square difference (top) and overall mean bias (bottom) for total and tropospheric column ozone ~~climatology for DJF~~(left), ~~MAM, JJA, SON~~ ozone profiles (middle) and surface CO diagnostics (right). Columns and rows of each panel represent the ~~annual mean for the~~ EMAC simulations (including the sensitivity experiments) and the given diagnostics (see Table 3), respectively. Where an alternative dataset is available, the diagram boxes are split in two parts, showing the model performance compared to ~~NIWA~~ the primary (lower triangle) and ~~GOME~~alternative (upper triangle) dataset. ~~Shaded areas indicate~~ Where no observations are available, the ~~$\pm 1\sigma$ interannual variability~~triangles are marked white.

~~Root mean square difference (top) and overall mean bias (bottom) for total and tropospheric column ozone (left), ozonesondes (middle) and surface CO diagnostics (right). Columns and rows of each panel represent the EMAC simulations (including the sensitivity experiments) and the given diagnostics (see Table 3), respectively. Where an alternative dataset is available, the diagram boxes are split in two parts, showing the model performance compared to the primary (lower triangle) and alternative (upper triangle) dataset. Where no observations are available, the triangles are marked white.~~

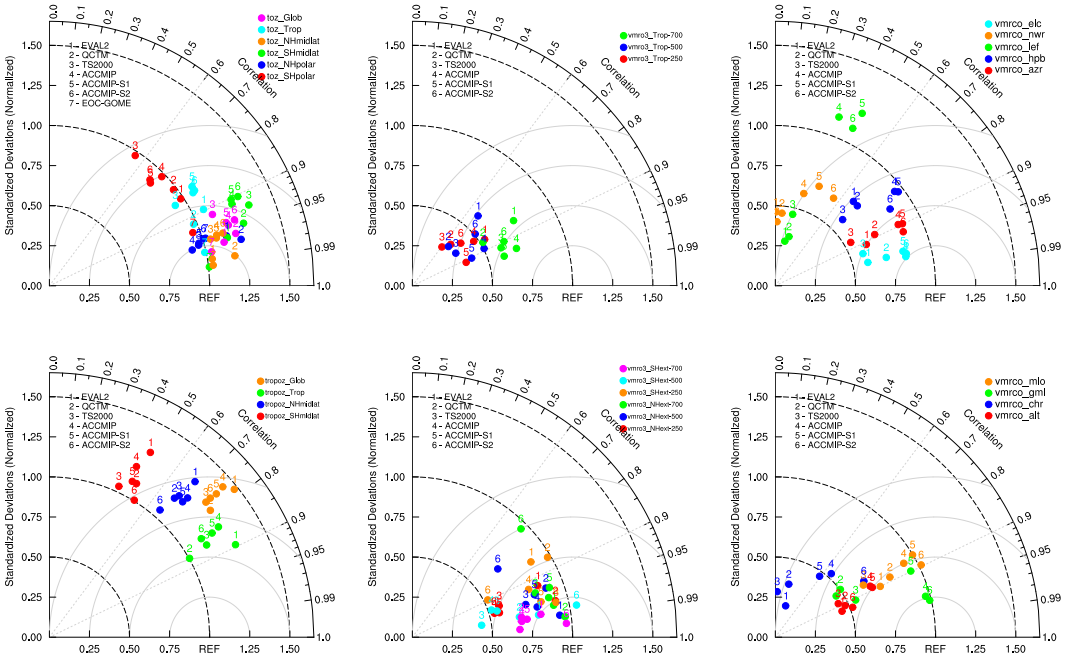

Figure 12. Taylor diagrams for total and tropospheric column ozone (left), ozone profiles (middle) and surface CO diagnostics (right).

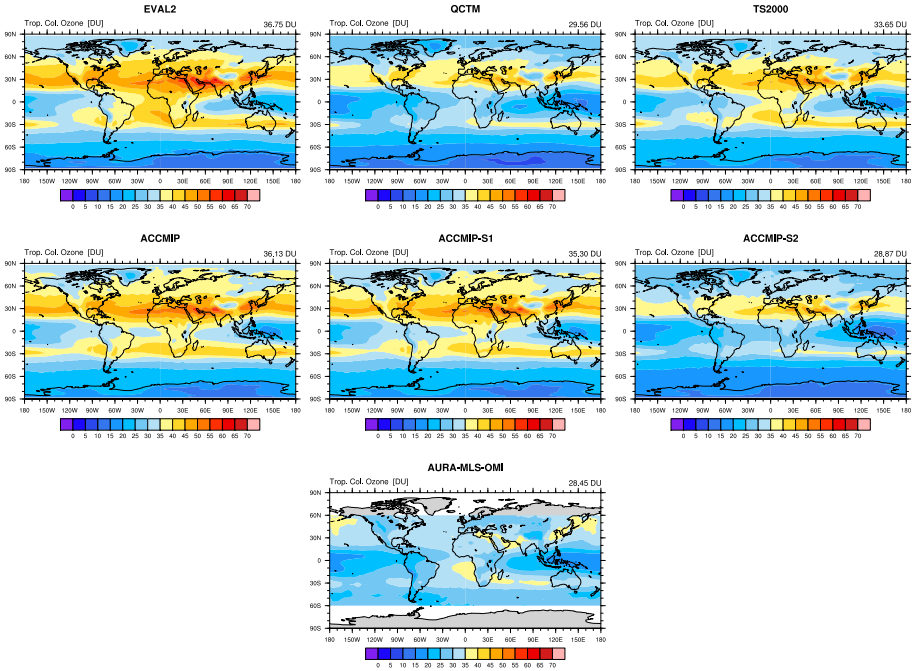

Figure 13. Tropospheric column ozone in the EMAC simulations compared to MLS/OMI observations. The values on top of each panel show the global (area-weighted) average, calculated after regridding the data to the horizontal grid of the model and ignoring the grid cells without available observational data.

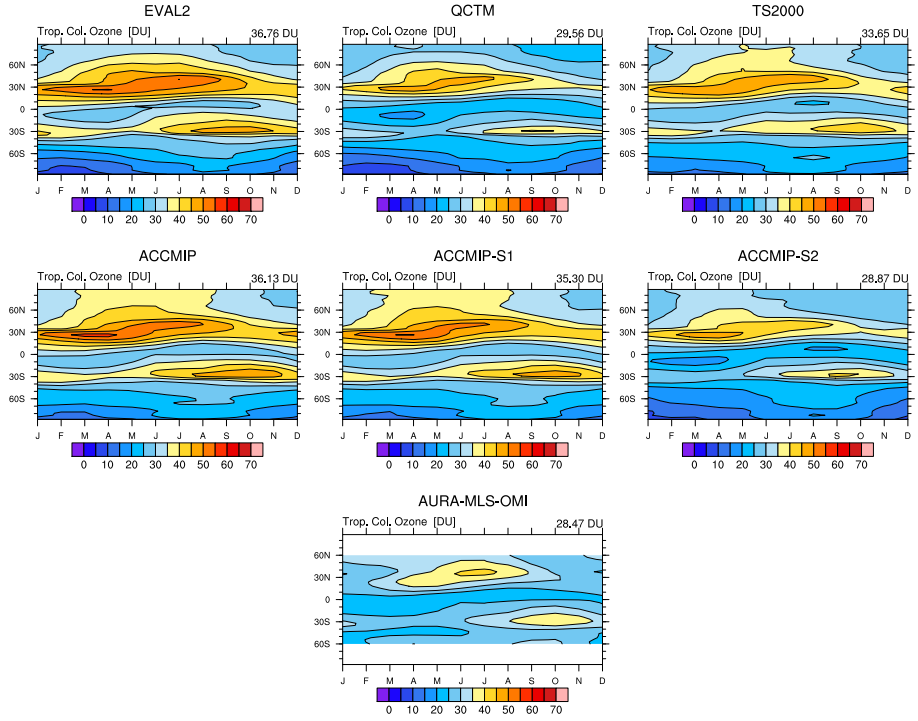

Figure 14. Annual cycle of the tropospheric column ozone climatology in the EMAC simulations compared to MLS/OMI observations. The values on top of each panel show the global (area-weighted) average, calculated after interpolating the observations on the model grid and ignoring the grid cells without available observational data.

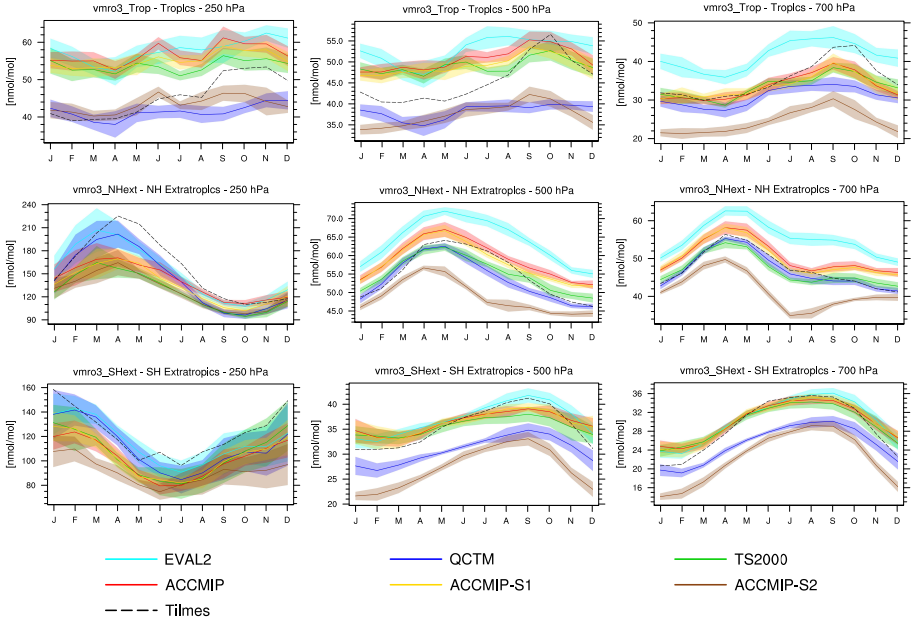

Figure 15. Annual cycle of ozone climatology in three regions (tropics, NH and SH extrastropics) at three pressure levels (250, 500 and 700 hPa) for the EMAC simulations compared with ozonesondes data by Tilmes et al. (2012). Model and observational data are grouped into four latitude bands and sampled at three, with the models sampled at the ozonesonde locations before averaging together. The tropical region is represented by 9 stations, NH and SH extratropics by 22 and 5 stations, respectively (see Fig. S14). The shaded areas indicate the $\pm 1\sigma$ interannual variability (for EMAC only).

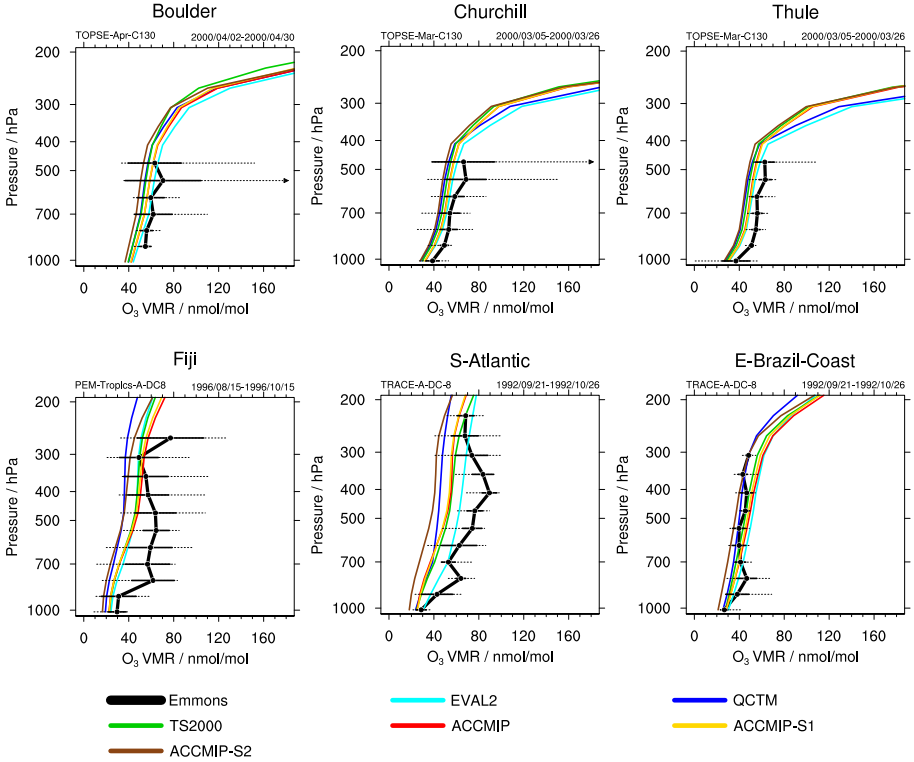

Figure 16. Ozone vertical profile climatology from selected aircraft campaign observations by Emmons et al. (2000) and corresponding simulated values by the EMAC simulations. Profiles represent mean values. The EMAC simulations are averaged over the same regions and time of year as the observations, but for different years. Solid whiskers indicate ± 1 standard deviation and dotted whiskers show minimum and maximum, both for the observational data.

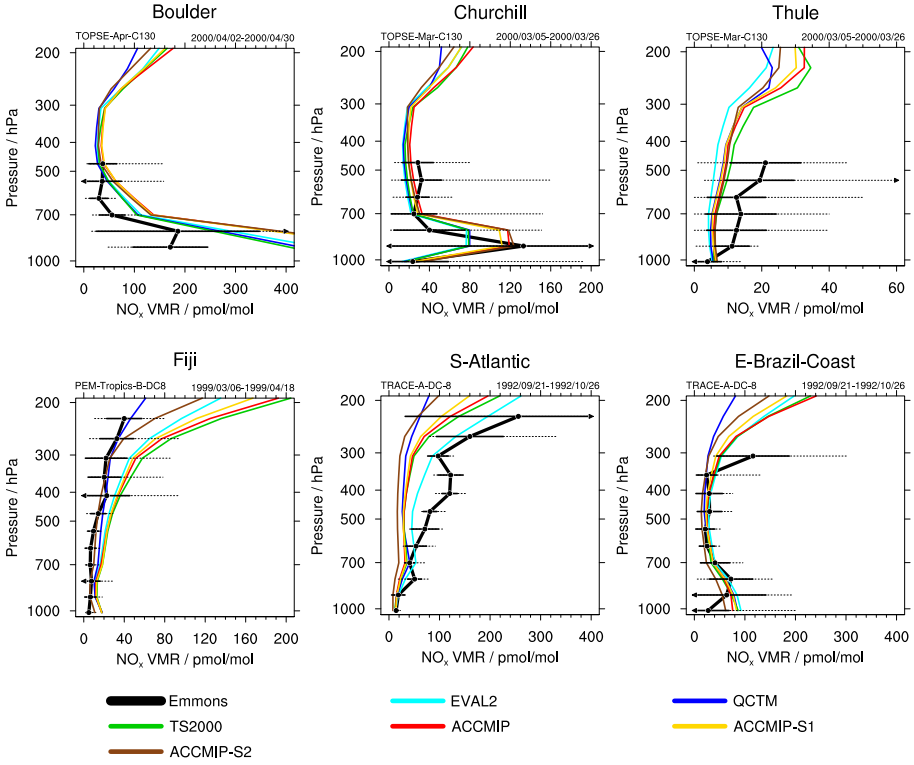

Figure 17. Similar to Fig. 16, for NO_x.

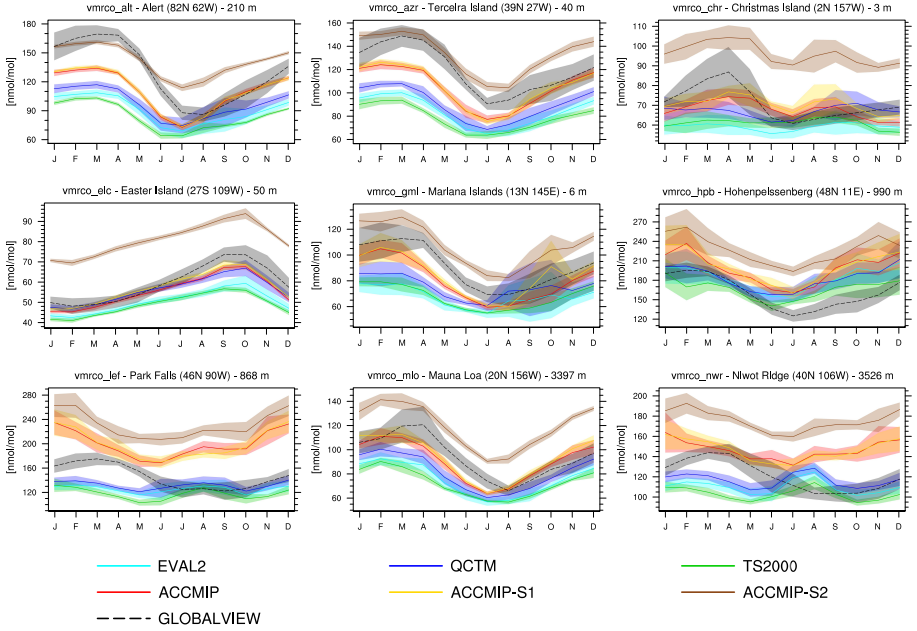

Figure 18. Annual cycle of CO surface concentration climatology for the EMAC simulations and NOAA GLOBALVIEW data, at nine different stations world wide. The shaded areas indicate the $\pm 1\sigma$ interannual variability.

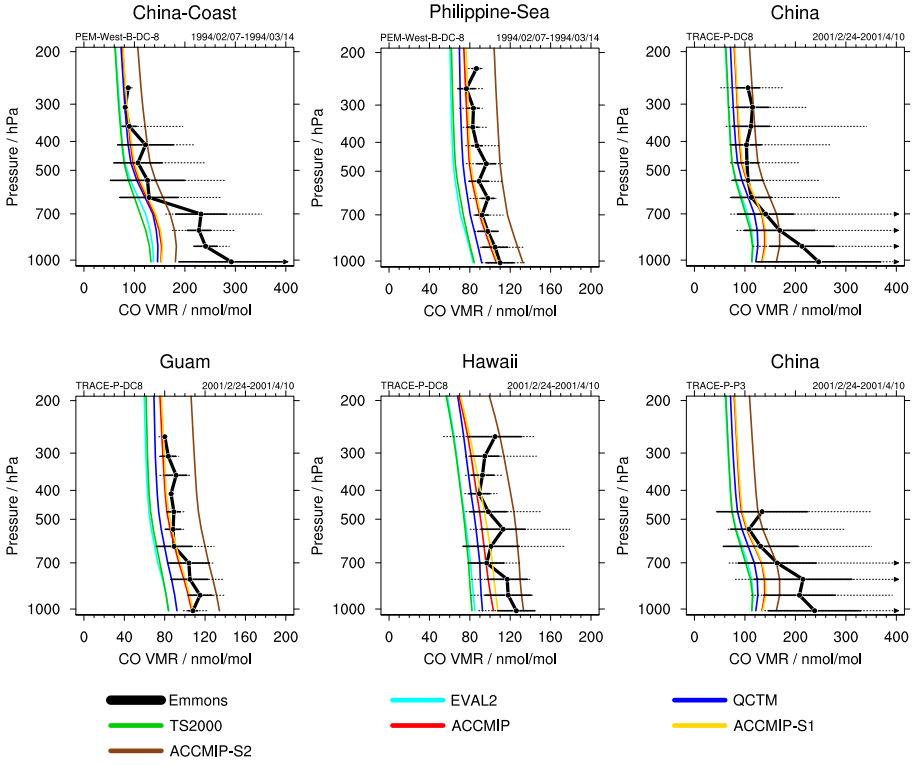

Figure 19. Similar to Fig. 16, for CO.

~~Annual cycle of surface concentration climatology for the EMAC simulations and NOAA GLOBALVIEW data, at nine different stations world wide. The shaded areas indicate the $\pm 1\sigma$ interannual variability.~~