Evaluation of ACCMIP ozone simulations using a multi-constituent chemical reanalysis

Kazuyuki Miyazaki[1,2] and Kevin Bowman[2]

[1]Japan Agency for Marine-Earth Science and Technology, Yokohama 236-0001, Japan
[2]Jet Propulsion Laboratory-California Institute of Technology, Pasadena, CA, USA

Correspondence to: K. Miyazaki (kmiyazaki@jamstec.go.jp)

Abstract.

The Atmospheric Chemistry Climate Model Intercomparison Project (ACCMIP) ensemble ozone simulations for the present-day are evaluated by a state-of-the-art multi-constituent atmospheric chemical reanalysis that ingests multiple satellite data including Tropospheric Emission Spectrometer (TES), Microwave Limb Sounder (MLS), Ozone Mapping Instrument (OMI), and the Measurements of Pollution in the Troposphere (MOPITT). Validation of the chemical reanalysis against global ozonesondes shows good agreement throughout the free troposphere and lower stratosphere for both seasonal and year-to-year variations, with an annual mean bias of less than 0.9 ppb in the middle and upper troposphere at the tropics and mid-latitudes. The model evaluation using the reanalysis reveals that the ensemble mean overestimates ozone in the northern extratropics by 6–11 ppb while underestimating by up to 18 ppb in the southern tropics over the Atlantic in the lower troposphere. Most models underestimate the spatial variability of the annual mean concentration in the extratropics of both hemispheres in the lower troposphere. The ensemble mean also underestimates the seasonal amplitude by 25–70 % in the northern extratropics and overestimates the inter-hemispheric gradient by about 30 % in the lower and middle troposphere. These differences are less evident with the current sonde network, which is shown to provide biased regional and monthly ozone statistics, especially in the tropics. These systematic biases have implications for ozone radiative forcing and the response of chemistry to climate that can be further quantified as the satellite observational record extends to multiple decades.

1 Introduction

Tropospheric ozone is one of the most important air pollutants and the third most important greenhouse gases in the atmosphere (Forster et al., 2007; HTAP, 2010; Myhre et al., 2013; Stevenson et al., 2013) while also playing a crucial role in the tropospheric oxidative capacity through production of hydroxyl radicals (OH) by photolysis in the presence of water vapor (Logan et al., 1981; Thompson, 1992). Global tropospheric ozone is formed from secondary photochemical production of ozone precursors including hydrocarbons or carbon monoxide (CO) in the presence of nitrogen oxides (NO_x) modulated by additional processes including in-situ chemical loss, deposition to the ground surface, and inflow from the stratosphere. These ozone precursors are largely controlled by anthropogenic and natural emissions, e.g., mobile, industry, lightning, biomass burning sources. Representation of tropospheric ozone in chemical transport models (CTMs) and chemistry climate models (CCMs) is important

in estimating its impact on the atmospheric radiative budget (Bowman et al., 2013). A number of CTMs and CCMs have been developed and used to study variations in atmospheric environment and its impacts on climate (e.g. Shindell et al, 2006, 2013; Stevenson et al., 2006, 2013; Wild, 2007, Kawase et al., 2011; Young et al., 2013). However, current tropospheric ozone simulations still have large uncertainties because of the incomplete representation of model processes, as well as the large

5   uncertainty in precursor emissions. These in turn increase uncertainty in CCM projections.

Climate model evaluation has primarily been achieved by comparisons with observed concentrations or related variables, which requires a precise description of their geographical, vertical, and temporal variations. Various measurements, e.g., ground-based, ozonesonde, and satellite-retrieved measurements, have been employed for evaluating simulated fields. However, information obtained from individual measurements is limited, and evaluation of global ozone fields with a suite of satel-

10   lite measurements and in situ measurements is challenging because of limited vertical sensitivity profiles that differ among measurements, different overpass times, and mismatches in spatial and temporal coverage between the instruments.

First, surface measurements have a spatial representativeness that is much smaller than that of global models over polluted areas. Ozone climatology data sets have been established based on ozonesonde measurements for use in model evaluation (Logan et al., 1999; Considine et al., 2008). Tilmes et al. (2012) generated an ozone climatology using ozonesonde measurements

15   obtained between 1995 and 2011, which mostly consists of the same station data described by Logan (1999) and Thompson et al (2003), but covering a longer time period. Using the compiled data of Tilmes et al. (2012), Young et al. (2013) conducted an intensive validation of tropospheric ozone from multiple model simulations in the Atmospheric Chemistry and Climate Model Inter-comparison Project (ACCMIP). However, we consider that the spatial and temporal coverage of the ozonesonde network is insufficient to capture the temporally and spatially representative model bias. Third, satellite-retrieved measurements such

20   as those from the Tropospheric Emission Spectrometer (TES) (Herman and Kulawik, 2013) and the Infrared Atmospheric Sounding Interferometer (IASI) (Clerbaux et al., 2009) have great potential for evaluating global ozone fields (e.g., Aghedo et al., 2011). However, information obtained from currently available satellite measurements are still limited. Their vertical sensitivity is not enough to resolve detailed vertical structures in the troposphere as appeared in current global models, and they measure at only a particular overpass time, thus the diurnal variation information is missing. Meanwhile, the characteristics

25   of each measurement, such as observational error, vary with observational condition, but their influence is rarely taken into consideration in model evaluations.

Data assimilation is a technique for combining different observational data sets with a model, with consideration of the characteristics of individual measurements (e.g., Kalnay, 2003; Lahoz and Schneider, 2014). Advanced data assimilation allows the propagation of observational information in time and space and from a limited number of observed species to a wide

30   range of chemical components, and provides global fields that are physically and chemically consistent and in agreement with individual observations. Various studies have demonstrated the capability of data assimilation techniques in the analysis of chemical species in the troposphere and stratosphere (e.g. Stajner and Wargan, 2004; Jackson, 2007; Parrington et al., 2009; Kiesewetter et al., 2010; Flemming et al., 2011; Coman et al., 2012; Inness et al., 2013; Emili et al., 2014; Miyazaki et al., 2012a, 2012b, 2013, 2014, 2015, 2016; van der A et al., 2015; Gaubert et al., 2016).

Reanalysis is a systematic approach to creating a long-term data assimilation product. Meteorological reanalyses have been established at operational centers for many years and are widely used in climate and meteorological research (e.g., Hartmann et al., 2013). Tropospheric chemical reanalysis, however, is relatively new. Inness et al. (2013) performed an eight-year reanalysis of tropospheric chemistry for 2003–2010 using the integrated forecasting system with modules for atmospheric composition (C-IFS) with observations sensitive primarily to the upper troposphere, and highlighted the importance of estimating surface emissions. This chemical reanalysis is recently updated by Flemming et al. (2016). Miyazaki et al. (2015) simultaneously estimated concentrations and emissions for an eight-year tropospheric chemistry reanalysis for 2005–2012 obtained from an assimilation of multi-constituent satellite measurements, which had greater lower tropospheric sensitivity, using an ensemble Kalman filter (EnKF). Chemical reanalysis using the EnKF has been used to provide comprehensive information on atmospheric composition variability and elucidate variations in precursor emissions and to evaluate bottom-up emission inventories (Miyazaki et al., 2014, 2015, 2016).

In this study, we explore the new potential of chemical reanalysis for evaluation of tropospheric ozone profiles in multi-model chemistry climate simulations from ACCMIP (Lamarque et al., 2013). ACCMIP models have been used to calculate historic and future radiative and chemically important species and their coupling with the broader climate system (Bowman et al., 2013; Lee et al., 2013; Naik et al. ,2013; Stevenson et al., 2013; Shindell et al., 2013; Voulgarakis et al., 2013; Young et al., 2013). We characterize ACCMIP models in simulating global distributions and the seasonal variation of ozone from the lower troposphere to the lower stratosphere. We further discuss the limitation of the current ozonesonde network for evaluating temporally and spatially representative model errors. To the best of our knowledge, this is the first study to apply chemical reanalysis to the evaluation of global chemistry-climate models and consequently offers a similar potential as meteorological reanalysis for evaluation of climate models (Ana4MIPS, https://esgf.nccs.nasa.gov/projects/ana4mips/Background).

2 Methodology

2.1 Chemical data assimilation system

The data assimilation system is constructed based on a global CTM MIROC-Chem (Watanabe et al. 2011) and an EnKF described in Miyazaki et al. (2016), which can be consulted for more detailed information. We use the two-hourly global chemical reanalysis data for the period 2005–2009 when tropospheric ozone fields are strongly constrained by TES tropospheric ozone measurements. The availability of TES measurements is strongly reduced after 2010, which led to a degradation of the reanalysis performance, as demonstrated by Miyazaki et al. (2015).

2.1.1 Forecast model

The forecast model, MIROC-Chem (Watanabe et al., 2011), considers detailed photochemistry in the troposphere and stratosphere by simulating tracer transport, wet and dry deposition, and emissions, and calculates the concentrations of 92 chemical species and 262 chemical reactions (58 photolytic, 183 kinetic, and 21 heterogeneous reactions). Its tropospheric chemistry

considers the fundamental chemical cycle of O_x-NOx-HO_x-CH_4-CO along with oxidation of non-methane volatile organic compounds (NMVOCs) to properly represent ozone chemistry in the troposphere. Its stratospheric chemistry simulates chlorine and bromine containing compounds, CFCs, HFCs, OCS, N_2O, and the formation of polar stratospheric clouds (PSCs) and associated heterogeneous reactions on their surfaces. MIROC-Chem has a T42 horizontal resolution (2.8°) with 32 vertical levels from the surface to 4.4 hPa. It is coupled to the atmospheric general circulation model MIROC-AGCM version 4 (Watanabe et al., 2011). The simulated meteorological fields were nudged toward the six-hourly ERA-Interim (Dee et al., 2011) to reproduce past meteorological fields.

The a priori values for surface emissions of NO_x and CO were obtained from bottom-up emission inventories. Anthropogenic NO_x and CO emissions were obtained from the Emission Database for Global Atmospheric Research (EDGAR) version 4.2 (EC-JRC, 2011). Emissions from biomass burning were based on the monthly Global Fire Emissions Database (GFED) version 3.1 (van der Werf et al., 2010). Emissions from soils were based on monthly mean Global Emissions Inventory Activity (GEIA) (Graedel et al., 1993). Lightning NO_x (LNO_x) sources in MIROC-Chem were calculated based on the relationship between lightning activity and cloud top height (Price and Rind, 1992) and using the convection scheme of MIROC-AGCM. For black carbon (BC) and organic carbon (OC) and other precursor gases, surface and aircraft emissions are specified from the emission scenarios for Greenhouse Gas and Air Pollution Interactions and Synergies (GAINS) model developed by International Institute for Applied System Analysis (IIASA) (Klimont et al., 2009; Akimoto et al., 2015).

2.1.2 Data assimilation method

Data assimilation used here is based upon on an EnKF approach (Hunt et al., 2007). The EnKF uses an ensemble forecast to estimate the background error covariance matrix and generates an analysis ensemble mean and covariance that satisfy the Kalman filter equations for linear models. In the forecast step, a background ensemble, $\boldsymbol{x}_i^b(i=1,...,k)$, is obtained from the evolution of an ensemble model forecast, where $\boldsymbol{x}$ represents the model variable, b is the background state, and k is the ensemble size (i.e., 32 in this study). The background ensemble is then converted into the observation space, $\boldsymbol{y}_i^b = H(\boldsymbol{x}_i^b)$, using the observation operator H which is composed of a spatial interpolation operator and a satellite retrieval operator, which can be derived from an a priori profile and an averaging kernel of individual measurements (e.g., Eskes and Boersam, 2003; Jones et al, 2003). Using the covariance matrices of observation and background error as estimated from ensemble model forecasts, the data assimilation determines the relative weights given to the observation and the background, and then transforms a background ensemble into an analysis ensemble, $\boldsymbol{x}_i^a(i=1,...,k)$. The new background error covariance is obtained from an ensemble forecast with the updated analysis ensemble.

In the data assimilation analysis, a covariance localization is applied to neglect the covariance among unrelated or weakly related variables, which has the effect of removing the influence of spurious correlations resulting from the limited ensemble size. The localization is also applied to avoid the influence of remote observations that may cause sampling errors. The state vector includes several emission sources (surface emissions of NO_x and CO, and LNO_x sources) as well as the concentrations of 35 chemical species. The emission estimation is based on a state augmentation technique, in which the background error correlations determines the relationship between the concentrations and emissions of related species for each grid point.

Because of the simultaneous assimilation of multiple-species data and because of the simultaneous optimization of the concentrations and emission fields, the global distribution of various species, including OH, is modified considerably in our system. This propagates the observational information between various species and modulates the chemical lifetimes of many species (Miyazaki et al., 2012b; 2015; 2016).

2.1.3 Assimilated measurements

Assimilated observations were obtained from multiple satellite measurements. Tropospheric NO_2 column retrievals used are the version-2 DOMINO data product for Ozone Monitoring Instrument (OMI) (Boersma et al., 2011) and version 2.3 TM4NO2A data products for Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) and Global Ozone Monitoring Experiment-2 (GOME-2) (Boersma et al., 2004) obtained through the TEMIS website (www.temis.nl). The TES ozone data and observation operators used are version 5 level 2 nadir data obtained from the global survey mode (Bowman et al, 2006; Herman and Kulawik, 2013). This data set consists of 16 daily orbits with a spatial resolution of 5–8 km along the orbit track. The Microwave Limb Sounder (MLS) data used are the version 4.2 ozone and HNO_3 level 2 products (Livesey et al., 2011). We used data for pressures of less than 215 hPa for ozone and 150 hPa for HNO_3. The Measurement of Pollution in the Troposphere (MOPITT) CO data used are version 6 level 2 TIR products (Deeter et al., 2013).

2.2 ACCMIP models

The Atmospheric Chemistry Climate Model Intercomparison Project (ACCMIP) focuses on chemistry-climate interactions needed to compute the proper climate forcing for Climate Model Intercomparison Project (CMIP5) climate simulations (Taylor et al., 2012) as well as the impact of climate change on chemical species. The ACCMIP consists of a series of time slice experiments for the long-term changes in atmospheric composition between 1850 and 2100, as described by Lamarque et al. (2013). The experimental design was based on decadal time-slice experiments driven by decadal mean sea surface temperatures (SST). This study uses the 2000 decade simulation results from 15 models (1. CESM-CAM, 2. CICERO-OsloCTM2, 3. CMAM, 4. EMAC, 5. GEOSCCM, 6. GFDL-AM3, 7. GISS-E2-R, 8. GISS-E2-TOMAS, 9. HadGEM2, 10. LMDzOR-INCA, 11. MIROC-CHEM, 12. MOCAGE, 13. NCAR-CAM3.5, 14. STOC-HadAM3, 15. UM-CAM).

Different models vary greatly in complexity. The calculated chemical species vary from 16 to 120 species. Photolysis rates are computed with offline or online methods, depending on the model. Many models include a full representation of stratospheric ozone chemistry and the heterogeneous chemistry of polar stratospheric clouds, but several models specify stratospheric ozone. Methane concentration is prescribed for the surface or over the whole atmosphere in many models. Ozone precursor emissions from anthropogenic and biomass burning sources were taken from those compiled by Lamarque et al. (2010), and the same emissions were used in all the models. Natural emission sources such as isoprene emissions, and lightning and soil NO_x sources were not specified and were accounted for differently between models. There is a large range in soil NO_x emissions from 2.7 to 9.3 $TgNyr^{-1}$ and in LNO_x sources from 1.2 to 9.7 $TgNyr^{-1}$ for the 2000 conditions. The range of natural emissions is a significant source of model-to-model ozone differences (Young et al. 2013). A complete description of the models along with the experiment design can be found in Lamarque et al. (2013).

2.3 Ozonesonde data

Ozonesonde observations were taken from the World Ozone and Ultraviolet Radiation Data Center (WOUDC) database (available at http://www.woudc.org). All available data from the WOUDC database are used for the validation. The accuracy of the ozonesonde measurement is about ±5 % in the troposphere (Smit and Kley, 1998).

5   To compare ozonesonde measurements with the data assimilation and ACCMIP models, all ozonesonde profiles have been interpolated to a common vertical pressure grid, with a bin of 25 hPa. The reanalysis and model fields were linearly interpolated to the time and location of each measurement using the two-hourly output data, with a bin of 25 hPa, and then compared with the measurements. The averaged profile is computed globally and for five latitudinal bands, SH high latitudes (55°S–90°S), SH mid-latitudes (15°S–55°S), tropics (15°S–15°N), NH mid-latitudes (15°N–55°N), and NH high latitudes (55°N–90°N).

10   3   Consistency between chemical reanalysis and ozonesonde observations

Miyazaki et al. (2015) validated an older version of the reanalysis (http://www.jamstec.go.jp/res/ress/kmiyazaki/reanalysis/) and showed good agreement with independent observations such as ozonesonde and aircraft measurements on regional and global scales and for both seasonal and year-to-year variations from the lower troposphere to the lower stratosphere for the 2005-2012 period. The mean bias against the ozonesonde measurements in the older dataset is -3.9 ppb at the NH high-latitudes,
15   -0.9 ppb at the NH mid-latitudes, 2.8 ppb in the tropics, -1.0 ppb at the SH mid-latitudes, -1.7 ppb at the SH high-latitudes between 850 and 500 hPa (Miyazaki et al., 2015). A major update from the system used in Miyazaki et al. (2015) to the system used in this study is the replacement of forecast model from CHASER (Sudo et al., 2002) to MIROC-Chem (Watanabe et al., 2011), which caused substantial changes in the a priori field and thus the data assimilation results of various species. In addition, we attempt to optimize the surface NO_x emission diurnal variability using data assimilation of OMI, SCIAMACHY,
20   and GOME-2 retrievals in the updated system (Miyazaki et al., 2016). Since the updated reanalysis ozone fields used in this study have not yet been validated in any publication, we first present the evaluation results of the chemical reanalysis using global ozonesonde observations for 2005–2009.

Figs 1 and 2 compare the reanalysis and the global ozonesonde observations, and the comparison result is summarized in Table 1. In order to confirm improvements in the reanalysis, results from a model simulation without any chemical data
25   assimilation (i.e., a control run) is also shown. The control run shows systematic biases, such as positive biases in the upper troposphere and lower stratosphere (UTLS) throughout the globe and negative biases in the lower and middle troposphere in the extratropics of both hemispheres. The positive bias in the UTLS is larger in the Southern Hemisphere (SH) than in the Northern Hemisphere (NH). The a priori systematic bias in this study is larger than that in our previous study (Miyazaki et al., 2015) in the UTLS, because of different model setting. However, the reanalysis fields were less sensitive to the a priori profiles
30   in the UTLS than in the lower and middle troposphere because of strong constraints by MLS measurements and long chemical lifetime of ozone in the UTLS.

The reanalysis shows improved agreements with the ozonesonde observations over the globe for the entire troposphere. The data assimilation removed most of the positive bias in the UTLS throughout the year and reduced the negative bias in the

lower and middle troposphere in the extratropics. At NH mid and high latitudes in the lower and middle troposphere, the data assimilation reduced the annual mean negative bias of the model by 45–90 %, which is attributed to the reduced bias in boreal spring–summer. The mean bias in the new dataset is smaller than that in the older dataset for most cases (e.g., from -3.9 to -2.9 ppb at the NH high-latitudes, -0.9 to -0.1 ppb at the NH mid-latitudes, -1.0 to -0.1 ppb at the SH mid-latitudes between 850 and 500 hPa). The simultaneous optimization of concentrations and emissions played important roles in improving the lower tropospheric ozone analysis, associated with the pronounced ozone production caused by NO_x increases, as demonstrated by Miyazaki et al. (2015). This advantage increases the ability of the chemical reanalysis to evaluate the simulated tropospheric ozone profiles, including the lower tropospheric ozone concentrations. Root-Mean-Square-Errors (RMSEs) are also reduced above the middle troposphere. The tropospheric concentrations show distinct seasonal and year-to-year variations, for which the temporal correlation is increased by the data assimilation globally, except at high latitudes in the lower troposphere (Table 1).

4   Evaluation of ACCMIP models

4.1   Global distribution

We use the global chemical reanalysis to evaluate the global ozone profiles in ACCMIP simulations. Fig. 3 compares the global distribution of the annual mean ozone concentration between the reanalysis and the ensemble mean of the ACCMIP models. The average over the multiple models can be expected to improve the robustness of the model simulation results, because some parts of the model errors may cancel each other out. As summarized in Table 2, the global spatial distributions are similar between the reanalysis and the ensemble mean, with a spatial correlation (r) greater than 0.94 from the lower troposphere to the lower stratosphere, except for the NH extratropical middle troposphere (r=0.57). The reanalysis and multi-model mean commonly reveal distinct inter-hemispheric differences, associated with a stronger downwelling across the tropopause and stronger emission sources of ozone precursors in the NH. The wave-1 pattern in the zonal ozone distribution in the tropics, with a minimum over the Pacific Ocean and maximum over the Atlantic (Thompson et al., 2003; Bowman et al, 2009; Ziemke et al., 2011), can also be commonly found in the reanalysis and the multi-model mean.

Large errors between the reanalysis and the multi-model mean in the troposphere are found in the NH extratropics and SH tropics (right panel in Fig. 3). The multi-model mean overestimates the zonal and annual mean concentrations by 6–11 ppb at 800 hPa and by 2–9 ppb at 500 hPa in the NH extratropics. The overestimation is larger over the oceans than over land at the NH mid-latitudes at 800 hPa. Both the mean RMSE and bias are larger at 800 hPa than at 500 hPa in the NH extratropics, whereas they are larger at 500 hPa in the NH tropics (Table 2). In the SH tropics, the multi-model mean underestimates the concentration over the eastern Pacific by up to 9 ppb, over the Atlantic by up to 18 ppb, and over the Indian Ocean by up to 8 ppb at 500 hPa. These negative biases are larger in the middle troposphere than in the lower troposphere for most places and also for the zonal means in the SH tropics (-15 % in the middle troposphere and -10 % in the lower troposphere) (Table 2). The positive bias in the NH and negative bias in the SH were common reported using OMI/MLS tropospheric ozone column

measurements (Young et al., 2013). At 200 hPa, the multi-model mean underestimates the zonal mean concentration by 20–30 ppb at high latitudes in both hemispheres, with a larger error in the SH than in the NH (Table 2).

Fig. 4 shows the Taylor diagram of the ACCMIP models against the reanalysis for three latitudinal bands for three levels. The relevant statistics at 500 hPa are summarized in Table 3, for which the tropics is separated into two hemispheres. In the

5 NH extratropics at 800 hPa, most models reproduced the spatial distribution (r = 0.8–0.95), while underestimating the spatial standard deviation (SD) by up to 50 %. Three exceptional models (1, 7, 8) show relatively poor agreements (r = 0.45–0.6 and SD underestimations by 50–60 %). At 500 hPa, there is a large diversity in the agreement. Only a few models (2, 4, 9, 11) show close agreement with the reanalysis (r > 0.8, SD error < 20 %). Notably, two models (12, 15) reveal too large spatial variabilities (SD error > 80 %), and five models (1, 6, 7, 8, 12) reveal small spatial correlation (r < 0.15). The regional mean bias is largely

10 positive (> 10 ppb) in several models (7, 8, 12) (Table 3). In this region, ozone distributions are modified by various processes, including vertical transport by convection and along conveyor belts, inflow from the stratosphere, long-range transports, and photochemical production (e.g, Lelieveld and Dentine, 2000; Oltmans et al., 2006; Sudo and Akimoto, 2007; Jonson et al., 2010). The evaluation results indicate that these processes occur differently among models. At 200 hPa, all the models simulate well the spatial distribution (r > 0.95), whereas the spatial variability differs between the models (SD error ranges from -50 %

15 to +30 %). There is relatively large variation in the stratospheric concentration, which results in the diversity in the UTLS.

In the tropics, the spatial correlation is greater than 0.8 at all levels for most models (except for 12, 15), as they capture the wave-1 structure. When dividing the tropics into two hemispheres (Table 3), only a few models (4, 12) reveal low spatial correlation (r < 0.8) for the SH tropics (30°S–EQ) at 500 hPa. The spatial correlation in the tropics is lower at 500 hPa than at 800 hPa for most models. The SD error is less than 40 % for all the models at 800 and 500 hPa, while mostly overestimating

20 the spatial variability at 800 hPa by up to 30 %. The mean bias is negative for most models at 500 hPa in the tropics in both hemispheres, with larger negative biases in the SH tropics (Table 3). Young et al (2013) noted that correlations between the biases for the NH and SH tropical tropospheric columns are strong. Similarly, our analysis using the reanalysis reveal a high correlation (0.91) between the NH and SH tropical biases at 500 hPa, suggests that similar processes are producing the model biases in the tropical middle troposphere between the hemispheres. At 200 hPa in the tropics, the SD error differs among

25 models, which could primarily be associated with the different representations of convective transports and ozone production by LNO_x sources (e.g., Lelieveld and Crutzen, 2007; Wu et al., 2007).

In the SH extratropics at 800 hPa, most models reproduce the spatial distribution (r > 0.9), while underestimating the SD by 15–70 %, except for model 15. The model performance is similar between 800 hPa and 500 hPa, with a smaller SD error at 500 hPa for most models. These high spatial correlations may be related to a lack of local precursor emissions in the SH.

30 At 500 hPa, a majority of the models underestimate the mean concentration (Table 3), with large negative biases (< -8 ppb) in several models (1, 2, 12, 14). At 200 hPa, the SD error varies from -80 % to + 65 %. The large diversity at 200 hPa may be related to the different representation of the tropopause and stratosphere–troposphere exchange (STE) among models. The poor agreement in model 8 is attributed to too-high concentrations at SH mid-latitudes and too-low concentrations at SH high latitudes in austral spring.

4.2 Seasonal variation

Fig. 5 compares the seasonal variation of zonal mean ozone concentration between the ACCMIP models, the reanalysis, and ozonesonde observations. The comparison between the reanalysis concentrations from the ozonesonde sampling (black dashed line) and the ozonesonde observations (blue solid line) shows that the reanalysis is in close agreement with the ozonesonde

5 observations over the globe, as described in Sec. 3. However, in the NH extratropics at 800 hPa, the reanalysis concentration is too low from boreal spring to summer by up to 4 ppb, which leads to an underestimation of the seasonal amplitude. In the NH tropics at 500 hPa, the reanalysis overestimates the concentration except in April. In the SH tropics at 500 and 800 hPa, the reanalysis slightly overestimates the concentrations throughout the year by up to 5 ppb. In the SH extratropics at 800 hPa, the reanalysis concentration is too low by up to 5 ppb from austral autumn to winter. The reanalysis concentration and seasonal

10 variation differs largely between the complete sampling (black bold line, where the concentrations were averaged over all grid points) and the ozonesonde sampling (black dashed line) for the globe. The impact of using the reanalysis instead of the ozonesonde network in characterizing the ozone seasonal variation is discussed in Section 5.

The global ozone concentrations averaged over all grid points are compared between the ACCMIP models and the reanalysis (black solid line vs. red solid line for the multi-model mean and thin colored lines for individual models). There is considerable

15 interannual variability in both the reanalysis and the ACCMIP models. We confirmed that the ACCMIP ensemble mean is mostly within the standard deviation (i.e., year-to-year variation) of the reanalysis (not shown). In the NH extratropics, the multi-model mean overestimates the monthly mean concentrations by 6–9 ppb at 800 hPa and by 3–6.5 ppb at 500 hPa. The multi-model mean reproduces the seasonal variation, whereas there is large diversity among the models. The increase from winter to spring differs among models at 500 hPa, which is probably associated with different representations of downwelling

20 from the stratosphere. Fig. 6 compares the seasonal amplitude. Most models overestimate the seasonal amplitude in the NH lower and middle troposphere, with a mean overestimation of 50–70 % at 800 hPa and 25–40 % at 500 hPa at NH high latitudes. At 200 hPa, the multi-model annual mean concentration is in good agreement with that of the reanalysis, whereas the seasonal amplitude is underestimated by most models at NH high latitudes, with a mean underestimation of 15–25 %.

In the NH tropics at 500 hPa, the multi-model mean underestimates the concentration by 1–4 ppb throughout the year, which

25 can be attributed to the anomalously low concentrations in several models. There is a large diversity among the models in this region. In the SH subtropics, the multi-model mean is lower by up to 5 ppb at 800 hPa and by up to 11 ppb at 500 hPa, with the largest errors occurring in austral spring. A majority of models overestimate the seasonal amplitude in the NH subtropics at 800 hPa (by about 10–40 %), whereas they mostly underestimate the amplitude in the SH tropics at 800 and 500 hPa. In the tropical upper troposphere in both hemispheres, a few models reveal anomalously high or low concentrations. Both the ozonesondes

30 and reanalysis reveal a sharp increase in ozone between March and April in the NH subtropics, which is not captured in the multi-model mean.

In the SH extratropics, the multi-model mean and the reanalysis are in good agreement at 800 hPa, whereas it largely underestimates the peak concentration in austral winter–spring at 500 hPa (by up to 7 ppb) and 200 hPa (by up to 35 ppb). The large diversity among the models and the large underestimation in the multi-model mean at 500 hPa in spring could be

attributed to the differing influence of stratospheric air. The seasonal amplitude is overestimated at 800 and 200 hPa by most models at SH high-latitudes.

4.3 Inter-hemispheric gradient

Fig. 7 compares the inter-hemispheric gradient (NH/SH ratio) of the annual mean ozone concentration. We calculated the
5 gradient across the equator; however, recognize a more careful definition of the boundary between two hemispheres would be required to isolate air masses originated from each hemisphere (e.g., Hamilton et al., 2008). The gradient is similar between the ozonesonde observations (blue solid line) and the reanalysis concentration from the ozonesonde sampling (black dashed line) throughout the troposphere. In these estimates, the NH mean concentration is higher than the SH mean by 60–70 % in the lower troposphere, by 30–40 % in the middle troposphere, and by 55–60 % around 200 hPa. Near the surface, the reanalysis
10 slightly overestimates the NH/SH ratio, mainly because of overestimated concentrations at the NH mid-latitudes.

By taking a complete sampling in the reanalysis (i.e., averaging over all model grid points for each hemisphere) (black solid line), the NH/SH ratio becomes smaller by about 25–30 %, 7–10 %, and 15–25 % in the lower troposphere, the middle troposphere, and around 200 hPa, respectively, compared to the average at the ozonesonde sampling sites (black dashed line). The difference is a consequence of ozonesonde stations located near large cities at NH mid-latitudes, and therefore tend to
15 observe higher ozone concentration than the hemispheric average. Around 200 hPa, the difference could also be attributed to the presence of atmospheric stationary waves and Asian monsoon circulation in the NH, which result in substantial spatial ozone variations in the UTLS (e.g., Wirth, 1993; Park et al., 2008) (c.f., Fig. 3). The annual mean NH/SH ratio based on the global reanalysis field estimated at the surface, 800 hPa, 500 hPa, and 200 hPa are 1.36, 1.42, 1.30, and 1.35, respectively.

Most models overestimate the NH/SH ratio compared with the reanalysis, with a mean overestimation (black solid line vs.
20 red solid line) of 34 % at the surface and 22-30 % in the free troposphere, attributing to both too-high concentrations in the NH extratropics and too-low concentrations in the SH subtropics in most models (c.f., Figs. 3 and 5). The multi-model mean reveals annual mean NH/SH ratios of 1.71, 1.73, 1.54, and 1.49 at the surface, 800 hPa, 500 hPa, and 200 hPa, respectively. The large systematic error in the NH/SH ratio suggests that, for instance, the radiative heating distribution in chemistry–climate simulations are largely uncertain in most models, and such comprehensive information for different altitudes in the troposphere
25 cannot be obtained using any individual measurements, as is further discussed in Section 6.3.

5 Impact of sampling on model evaluation

As presented in the previous section, the chemical reanalysis provides comprehensive information on global ozone distributions for the entire troposphere which is useful for validating global model performance. It was also demonstrated that the inter-hemispheric gradient of ozone measured with the ozonesonde and complete sampling method produced different results, and
30 the model-reanalysis difference strongly depended on the choice of the sampling method. As these networks have been the primary basis for CCM evaluation (e.g., Stevenson et al., 2006; Huijnen et al., 2010; Young et al., 2013), the implications of this sampling bias need to be quantified. The current ozonesonde network does not cover the entire globe and is not homogeneously

distributed between the hemispheres, ocean and land, and urban and rural areas. Also, the sampling interval of ozonesonde observations is typically a week or longer, which does not reflect the influence of diurnal and day-to-day variations. Model errors are also expected to vary greatly in time and space at various scales. Therefore, the implications of model differences at ozonesonde locations to regional and seasonal processes is uncertain. This section evaluates how changes in evaluated model

5   performance could be obtained by using the complete sampling chemical reanalysis fields instead of the existing ozonesonde network on simulated regional ozone fields.

    Sampling bias is an error in a computed quantity that arises due to unrepresentative (i.e., insufficient or inhomogeneous) sampling, which induces spurious features in the average estimates (e.g., Aghedo et al., 2011; Foelsche et al 2011; Toohey et al., 2013; Sofieva et al., 2014). Sampling bias may occur when the atmospheric state within the time-space domain over which

10   the average is calculated is not uniformly sampled. In regions where variability is dominated by short-term variations, limited sampling may lead to a random sampling error. The primary technique for sampling bias estimation is to subsample model or reanalysis fields based on the sampling patterns of the measurements and then to quantify differences between the mean fields based on the measurement sampling and those derived from the complete fields. Sampling bias cannot be negligible, even for satellite measurements (Aghedo et al. 2011; Toohey et al., 2013; Sofieva et al., 2014).

15   To estimate sampling biases of the ozonesonde network in the ACCMIP model evaluation, two evaluation results of mean model bias are compared using the chemical reanalysis. The first evaluation was conducted based on the complete sampling; the second evaluation used the ozonesonde sampling (in both space and time) that is based on the completion by Tilmes et al. (2012). By using the two-hourly reanalysis fields, we can address possible biases due to the limited model sampling (i.e., monthly model outputs were used). Note that the relatively coarse resolution of the reanalysis may lead to an underestima-

20   tion of the sampling bias in the model evaluation, because the variability of a sampled field depends on the resolution of the measurement. Tilmes et al. (2012) stated that regional aggregates of individual ozonesonde measurements with similar characteristics are more representative for larger regions; however, this may not mean that evaluation results using the compiled data generate model errors that are representative of actual monthly mean for a surrounding area. The model evaluation results are shown for the 11 regions illustrated in Fig. 8 and summarized in Table 3. Japan was excluded from the evaluation because

25   data from only one station was available for the reanalysis period. The 11 areas surrounding the ozonesonde stations were considered for complete atmospheric sampling (rectangles in Fig. 8), for which small margins were considered around the stations to prevent overestimation of the ozonesonde network limitation. It was confirmed that the discrepancy between the two evaluations generally increases with the size of the area. In contrast, for the SH mid- and high latitudes, the defined areas cover the entire range of longitudes, because of generally less variabilities in the SH than in the SH.

30   The reality of the reanalysis fields is important for reasonable estimates of the true sampling bias of the real atmosphere. As discussed in Section 3, there is good agreement in the evaluated model performance using the reanalysis and the ozonesonde measurements at the ozonesonde sampling, except for the lower troposphere. This result supports the use of the reanalysis data at the ozonesonde locations. The performance of the ACCMIP model as compared with the ozonesonde measurements is mostly consistent with that shown by Young et al. (2013), although the ozonesonde data periods differ – 1997-2011 was used

35   by Young et al (2013) and 2005-2009 was used in this study.

5.1 Mean error and its distribution

The model evaluation results for the two cases differ greatly for many regions, as shown by Fig. 9 and summarized in Table 5. For the NH Polar Regions, Tilmes et al. (2012) stated that separating the regions into eastern and western sectors reduces the variability in ozone within each region because long-range transports of pollution from low and mid-latitudes into high

5 latitudes shows longitudinal variations in the NH (e.g., Stohl, 2006). Comparisons further suggest that, except for the UTLS in winter (December–February (DJF)), the evaluated model performance using the ozonesonde measurements are representative of the surrounding regional and seasonal mean model performance. For the two NH polar regions at 200 hPa in DJF, the validation based on the ozonesonde sampling reveals a large negative sampling bias in the model bias as compared with regional and monthly means. Large negative model biases against the ozonesonde observations have been reported by Young

10 et al. (2013) for 250 hPa, whereas results from this study suggest that these errors are larger than those from regional and seasonally representative model bias. At 500 hPa, the ozonesonde network reveals a negative sampling bias for the NH polar east in DJF. Thus, the positive bias reported in Young et al. (2013) for the NH polar east at 500 hPa may be lower than regional and seasonally representative model biases. The large discrepancy between the two estimates in the UTLS model performance can be attributed to the large variability of ozone distribution and associated model errors on a regional and seasonal scale.

15 For Canada, large differences (>30 %) exist in the two evaluations in the lower troposphere and for the UTLS in DJF and for the middle troposphere in MAM. The ozonesonde measurements reveal a large negative sampling bias in the model evaluation in DJF at 200 hPa (-4 % in the complete sampling and -25 % in the ozonesonde sampling), while they reveal a negative sampling bias (by about 50 %) at 500 hPa in MAM. Similar differences between the two evaluations are found for Western Europe at 500 hPa and at 200 hPa in DJF. These results suggest that, for instance, the positive bias for Western Europe

20 estimated by Young et al (2013) may be lower than regional and seasonally representative model bias, even for such a small area. The smaller discrepancy between the two estimates for Western Europe as compared for Canada for most cases could be associated with the better coverage of the ozonesonde measurements for Western Europe. Even for the small area of the eastern United States, the two validations differ largely in the UTLS (e.g., -9 % in the ozonesonde sampling and +6 % in the complete sampling at 200 hPa in MAM) and at 500 hPa in MAM, June–August (JJA), and September–November (SON). In

25 the NH subtropics, the two evaluations disagree largely in the middle and upper troposphere in JJA and SON.

The tropical stations were separated into the three sub-regions: Western Pacific and East Indian Ocean, equatorial America, and the Atlantic Ocean and Africa. These regions reflect the different dominant tropical processes including biomass burning and lightning over the Atlantic and Africa. The large variability of tropical ozone and its associated model error, together with the sparse ozonesonde network in these regions, results in large discrepancies between the two evaluations in the tropical

30 regions. At 500 hPa, the ozonesonde measurements reveal a large (by 40–50 %) negative sampling bias in March–May (MAM) and a positive sampling bias in DJF over the Western Pacific and East India, whereas it shows a large negative sampling bias (by 110 %) in MAM over the equatorial Americas. The probability distribution function (PDF) estimated using monthly mean reanalysis and model fields also differs largely between the two samplings (Fig. 10). Over the Western Pacific and East Indian Ocean in SON at 500 hPa, the multi-model mean shows a sharp peak around 54–58 ppb, in contrast to the broad

distribution seen in the reanalysis with two peaks around 65 ppb and 35–45 ppb for the complete sampling (left bottom panel in Fig. 10). This information is useful to characterize model errors and for process-oriented model validation. On the other hand, the validation based on the ozonesonde sampling (left top panel) does not show any clear pattern and does not support model evaluation. Note that the influence of inter-annual variability was not considered in the analysis because the monthly

5   climatological data were used by averaging over ten years for the models and five years for the reanalysis.

Although the variability of ozone is generally smaller in the SH than in the NH because of smaller local precursor emissions, large sampling biases exist even at SH mid- and high-latitudes due to the sparse ozonesonde network. In the SH mid latitudes, for example, the sign of the evaluated bias is opposite between the two cases at 200 hPa in DJF (-2.8 ppb in the complete sampling and +25.1 ppb in the ozonesonde sampling). In the SH high latitudes, evaluation results differ largely throughout the

10 year in the middle troposphere. Based on the complete sampling, the ozone PDF is broadly distributed with a peak around 38 ppb at 500 hPa in SON at the SH high latitudes (right bottom panel in Fig. 10), while the multi-model mean underestimates high concentrations (>47 ppb) and shows a sharp peak of about 35 ppb. The PDF generated by the ozonesonde sampling does not provide a strong information on the distribution of the ozone (right top panel). These results highlight the advantage of using the reanalysis data for evaluating regional and seasonally representative model performance, and for characterizing these

15 distributions.

5.2   Seasonal variation

The seasonal cycle of tropospheric ozone is determined by various factors such as local photochemical production and atmospheric transport (e.g., Monks, 2000). Carslaw (2005), Bloomer et al. (2010), and Parrish et al. (2013) found multi-decadal changes in the amplitude and phase of the seasonal cycle at NH mid-latitudes. It was suggested that these changes can be

20 attributed to changes in atmospheric transport patterns combined with spatial and temporal changes in emissions. CTMs have been used to explore the causal mechanisms; however, they failed to simulate several important features of the observed seasonal cycles (e.g., Ziemke et al., 2006; Stevenson et al., 2006; Parrish et al. 2014; Young et al., 2013). Accurate validation of the seasonal cycle is thus important for evaluating general model performance.

Table 6 compares the relative error in the seasonal amplitude obtained from the multi-mean model with that of the reanalysis

25 for the complete and ozonesonde samplings. The evaluation based on the ozonesonde sampling results in a larger overestimation of the seasonal amplitude in the NH lower troposphere for most cases (+13.4–+63.4 % in the sonde sampling and -19.0–+40.2 % in the complete sampling). The large discrepancies can be attributed to large spatial variability in the seasonal variations of ozone and its model errors within each defined region and also the existence of short-term variability that is not completely captured by the ozonesonde sampling. For the Eastern US and Western Europe at 800 hPa, the sign of the bias

30 is opposite between the two estimates. In contrast, at 200 hPa in the NH, results between the two evaluations are similar, suggesting spatial homogeneity in the seasonal cycle and its model errors within each region in the NH.

In the tropics, the estimated errors of the seasonal amplitude largely differ between the two samplings throughout the troposphere, suggesting that information obtained from the sparse ozonesonde network cannot be applied to characterize regional model errors in the seasonal cycle, even within the small defined area. Because of the large spatial variability, detailed

validations using the chemical reanalysis (e.g., for each grid point) would be helpful. Also, in the SH high latitudes, large disagreements in the seasonal amplitude exist at 800 and 200 hPa.

6 Discussions

6.1 Reanalysis uncertainty

5 Although the reanalysis dataset provides comprehensive information for global model evaluations, its performance still needs to be improved, especially for the lower troposphere, as also discussed by Miyazaki et al. (2015). Performance can be improved by ingesting more datasets including meteorological sounders such as IASI (Clerbaux et al., 2009), AIRS (Chahine et al., 2006), and CrIS (Glumb et al., 2002). Application of a bias correction procedure for multiple measurements, which is common in numerical weather prediction (e.g., Dee, 2005), is needed to improve reanalysis accuracy. Recently developed retrievals with

10 high sensitivity to the lower troposphere (e.g. Deeter et al., 2013; Fu et al., 2016) and the optimization of additional precursor emissions would be helpful to improve analysis of the lower troposphere. The relatively coarse resolution of the model could cause large differences between the simulated and observed concentrations at urban sites and may degrade the reanalysis.

The statistical information obtained from the reanalysis and the multi-model simulations can be used to suggest further developments for the models and observations. The analysis ensemble spread from EnKF can be regarded as uncertainty

15 information about the analysis mean fields, indicating requirements for additional observational constraints. As shown in Fig. 11 (left panels), the relative reanalysis uncertainty is large over the tropical areas of the oceans at 800 hPa (>20 %), over the Southern Ocean at 500 hPa (10–20 %), and over the tropics of the Pacific Ocean and the Antarctic at 200 hPa (>16 %). Conversely, the reanalysis uncertainty is small from the tropics to mid-latitudes in both hemispheres at 500 hPa (<11 %). These variations may be related to changes in observation errors, the number of assimilated measurements, as to model errors.

20 ### 6.2 Model uncertainty

The variability across the ensemble models (i.e., ensemble spread) identifies where the models are most consistent or uncertain (center panels in Fig. 11). As discussed by Young et al. (2013), the relative spread among the ACCMIP models is large over the tropical areas of the oceans in the lower and middle troposphere, a reflection of the important differences among the models in various processes such as convective processes, lightning sources, biogenetic emission sources with related chemistry. The large

25 relative spread (>20%) at the NH mid-latitudes and in the SH at 200 hPa may be associated with the different representations of the tropopause and STE among models. In contrast, the relative spread is small around 20–40°N at 500 hPa (< 10 %). The simultaneous enhancement of the analysis uncertainty (c.f., Section 6.1), together with the model spread, indicates low robustness of the validation results for some tropical regions over the oceans in the lower troposphere, and over the tropics in the Pacific Ocean as well as the Antarctic at 200 hPa. On the other hand, the ACCMIP model standard deviation with respect

30 to the reanalysis could be used to identify the averaged uncertainty of ACCMIP models (right panels in Fig. 11). The standard

deviation is large at NH high latitudes and over the tropical ocean areas at 800 hPa, over the SH tropics at 500 hPa, and in the SH extratropics at 200 hPa (> 25 %).

6.3 Implications into model improvements and climate studies

Numerous studies have identified decadal-scale changes in global tropospheric ozone using observations, such as the shift in the seasonal cycle at NH mid-latitudes and trends observed over many regions (e.g., Parrish et al., 2014; Cooper et al., 2014). A long-record of the reanalysis allows detailed structures in simulated inter-annual and long-term variations to be evaluated in association with changes in human activities and natural processes. It is noted that the influence of ENSO was not well simulated in ACCMIP due to a decadal-averaged SST boundary condition, which limits the evaluation of inter-annual variations and could lead to bias in the ACCMIP models and reanalysis comparisons.

Process-oriented validations using the reanalysis would be useful for understanding the uncertainty in simulated ozone fields and associated mechanisms. The ACCMIP models reveal large variations in short-lived species such as OH and ozone precursors (Naik et al., 2013; Voulgarakis et al., 2013), whereas information obtained from direct in-situ measurements cannot be applied for investigating global distributions because of the limited coverage of the measurements and the large spatial variability of concentrations. Validation of various species using the chemical reanalysis product can be used to identify potential sources of error in the simulated ozone fields. Meanwhile, the global monthly products of precursor emissions from the chemical reanalysis calculations (Miyazaki et al., 2012a, 2014, 2016) can be used to validate emission inventories and LNO_x source parameterizations used in model simulations. As changes in tropospheric ozone burden associated with different future scenarios show a broadly linear relation to changes in NO_x emissions (Stevenson et al., 2006), evaluations using up-to-date estimated emissions (Miyazaki et al., 2016) may prove useful to partly validate emissions for each scenario.

The performance of the simulated radiative forcing is largely influenced by representation of ozone in model simulations (Bowman et al., 2013; Shindell et al., 2013; Stevenson et al., 2013). Bowman et al (2013) suggested that overestimation of the OLR in the tropical seas of the east Atlantic Ocean and over Southern Africa is associated with model ozone errors, a persistent feature in all ACCMIP models, which was also found in this study using the reanalysis. Validation of short-lived species is also important for evaluating the radiative forcing because simulated OH fields influence simulated climates through for instance their influences on methane (Voulgarakis et al., 2013). Thus, detailed information on model errors in ozone and other short-lived species could be used to improve estimates of radiative forcing in climate studies. Meanwhile, model biases for present-day ozone may be correlated with biases in other time periods. Young et al. (2013) showed that ACCMIP models with high, present day ozone burdens also had high burdens for the other periods of time, including the preindustrial period. Thus, the validation of present-day ozone fields using the reanalysis have the potential to evaluate preindustrial to present day ozone radiative forcing.

7   Conclusions

We conducted a ten-year tropospheric chemistry reanalysis by assimilating multiple chemical species from the OMI, MLS, TES, MOPITT, SCIAMACHY, and GOME-2 to provide a gridded, chemically consistent estimate of concentrations and precursor emissions. This study explores the potential of atmospheric chemical reanalysis to evaluate global tropospheric ozone of multi-model chemistry-climate model simulations.

The reanalysis product provides comprehensive and unique information on the weakness of the individual models and multi-model mean. We found that the ACCMIP multi-model mean overestimates ozone concentration in the NH extratropics throughout the troposphere (by 6–11 ppb and 800 hPa and by 2–9 ppb at 500 hPa for the zonal and annual mean concentration), and underestimates it in the SH tropics in the lower and middle troposphere by about 9 ppb over the eastern Pacific, by up to 18 ppb over the Atlantic, and by up to 8 ppb over the Indian Ocean. Most models underestimate the spatial variability of the annual mean concentration in the NH extratropics at 800 hPa (by up to 50 %) and in the SH extratropics at 800 and 500 hPa (by up to 70 %). The multi-model mean overestimates the seasonal amplitude in the NH by 50–70 % in the lower troposphere and by 25–40 % in the middle troposphere, whereas the seasonal amplitude is underestimated by 15–25 % at 200 hPa in the NH extratropics. The seasonal amplitude in the NH extratropics shows great diversity among models. The NH/SH ratio is overestimated by 22–30 % in the free troposphere in the multi-model mean; this can be attributed to both a concentration high bias in the NH and a concentration low-bias in the SH in most models.

We quantified the ozonesonde network sampling bias and how reanalysis can help extend the range of that network as a kind of "transfer standard". For instance, the ozonesonde sampling bias in the evaluated model bias is largely negative (positive) in MAM (in DJF) by 40–50 % over the Western Pacific and East India and largely negative by 110 % in MAM over the equatorial Americas at 500 hPa. Although the spatial and temporal variability is generally smaller in the SH than in the NH, the ozonesonde sampling bias cannot be negligible for capturing the regionally and monthly representative model errors even in the SH. The evaluation of the seasonal cycle of tropospheric ozone is also largely limited by the ozonesonde sampling bias. The evaluation based on the ozonesonde sampling introduces a larger overestimation of the seasonal amplitude than that based on the complete sampling for most of the surrounding areas in the NH lower troposphere, whereas the two estimates are largely different for the entire tropical regions. Therefore, there is an advantage of the reanalysis data for evaluating actual regionally and seasonally representative model performance required for model improvements. However, the network provides critical independent validation of the reanalysis, which can provide a much broader spatial constraint on chemistry-climate model performance.

The proposed model validation approach provides regionally and temporally representative model performance; this could ensure more accurate predictions for the chemistry–climate system. In future studies, validation of multiple species concentrations and precursor emissions from reanalysis would be useful in identifying error sources in model simulations. In particular, the response of tropospheric composition to changing emissions over decadal time scales is still not captured in CCMs relative to a few remote sites (Parrish et al, 2014). Recent increases in emissions from China have been linked to changes in tropospheric ozone concentrations (Verstraeten et al, 2015). Over the next decade, a new constellation of of low Earth Orbiting

sounders, e.g., IASI, AIRS, CrIS, Sentinel-5p (TROPOMI), Sentinel-5 and geostationary satellites (Sentinel-4, GEMS, and TEMPO) will provide even more detailed knowledge of ozone and its precursors (Bowman, 2013). Assimilating these datasets into a decadal chemical reanalysis will be a more direct means of quantifying the response of atmospheric composition to emissions at climate relevant time scales, which should be a more direct test on chemistry-climate change scenarios. We also

5 plan to apply the proposed evaluation approach to a more recent model inter-comparison project, the Chemistry-Climate Model Initiative (CCMI).

Acknowledgements. We acknowledge the use of data products from the NASA AURA and EOS Terra satellite missions. We also acknowledge the free use of tropospheric NO_2 column data from the SCIAMACHY, GOME-2, and OMI sensors from www.temis.nl. This work was supported through JSPS KAKENHI grant numbers 15K05296, 26220101, and 26287117, and Coordination Funds for Promoting AeroSpace

10 Utilization by MEXT, JAPAN.

References

Aghedo, A. M., Bowman, K. W., Shindell, D. T., and Faluvegi, G.: The impact of orbital sampling, monthly averaging and vertical resolution on climate chemistry model evaluation with satellite observations, Atmos. Chem. Phys., 11, 6493-6514, doi:10.5194/acp-11-6493-2011, 2011.

5 Akimoto, H., J. Kurokawa, K. Sudo, T. Nagashima, T. Takemura, Z. Klimont, M. Amann, K. Suzuki: SLCP co-control approach in East Asia: Tropospheric ozone reduction strategy by simultaneous reduction of NOx/NMVOC and methane, Atmos. Environ., 122, 588-595, doi:10.1016/j.atmosenv.2015.10.003, 2015.

Bloomer, B. J., Vinnikov, K. Y., and Dickerson, R. R.: Changes in seasonal and diurnal cycles of ozone and temperature in the eastern US, Atmos. Environ., 44, 2543–2551, doi:10.1016/J.Atmosenv.2010.04.031, 2010.

10 Boersma, K.F., H.J. Eskes and E.J. Brinksma, Error Analysis for Tropospheric NO2 Retrieval from Space, J. Geophys. Res. 109 D04311, doi:10.1029/2003JD003962, 2004.

Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmos. Meas. Tech., 4, 1905–1928, doi:10.5194/amt-4-1905-2011, 2011.

15 Bowman, K. W., Rodgers, C. D., Sund-Kulawik, S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Luo, M., Eldering, A., Shephard, M. W., Worden, H., Clough, S. A., Brown, P. D., Rinsland, C. P., Lampel, M., Gunson, M., and Beer, R., Tropospheric emission spectrometer: Retrieval method and error analysis, IEEE Geosci. Remote Sens., 44, 1297–1307, doi:10.1109/TGRS.2006.871234, 2006.

Bowman, K. W., Jones, D. B. A., Logan, J. A., Worden, H., Boersma, F., Chang, R., Kulawik, S., Osterman, G., Hamer, P., and Worden, J.: The zonal structure of tropical O_3 and CO as observed by the Tropospheric Emission Spectrometer in November 2004 - Part 2: Impact of

20 surface emissions on O3 and its precursors, Atmos. Chem. Phys., 9, 3563-3582, doi:10.5194/acp-9-3563-2009, 2009.

Bowman, K. W., Shindell, D. T., Worden, H. M., Lamarque, J.F., Young, P. J., Stevenson, D. S., Qu, Z., de la Torre, M., Bergmann, D., Cameron-Smith, P. J., Collins, W. J., Doherty, R., Dalsøren, S. B., Faluvegi, G., Folberth, G., Horowitz, L. W., Josse, B. M., Lee, Y. H., MacKenzie, I. A., Myhre, G., Nagashima, T., Naik, V., Plummer, D. A., Rumbold, S. T., Skeie, R. B., Strode, S. A., Sudo, K., Szopa, S., Voulgarakis, A., Zeng, G., Kulawik, S. S., Aghedo, A. M., and Worden, J. R.: Evaluation of ACCMIP outgoing longwave radiation from

25 tropospheric ozone using TES satellite observations, Atmos. Chem. Phys., 13, 4057-4072, doi:10.5194/acp-13-4057-2013, 2013.

Carslaw, D. C.: On the changing seasonal cycles and trends of ozone at Mace Head, Ireland, Atmos. Chem. Phys., 5, 3441-3450, doi:10.5194/acp-5-3441-2005, 2005.

Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnet, C., Blaisdell, J., Chen, L., Divakarla, M., Fetzer, E. J., Goldberg, M., Gautier, C., Granger, S., Hannon, S., Irion, F. W., Kakar, R., Kalnay, E., Lambrigtsen, B. H., Lee, S.-Y., Le Marshall, J., McMillan, W. W.,

30 McMillin, L., Olsen, E. T., Revercomb, H., Rosenkranz, P., Smith, W. L., Staelin, D., Strow, L. L., Susskind, J., Tobin, D., Wolf, W., and Zhou, L.: AIRS: Improving Weather Forecasting and Providing New Data on Greenhouse Gases, Bull. Am. Meteorol. Soc., 87, 911–926, 2006.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F.: Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, Atmos. Chem.

35 Phys., 9, 6041–6054, doi:10.5194/acp-9-6041-2009, 2009.

Coman, A., Foret, G., Beekmann, M., Eremenko, M., Dufour, G., Gaubert, B., Ung, A., Schmechtig, C., Flaud, J.-M., and Bergametti, G.: Assimilation of IASI partial tropospheric columns with an Ensemble Kalman Filter over Europe, Atmos. Chem. Phys., 12, 2513–2532, doi:10.5194/acp-12-2513-2012, 2012.

Considine, D. B., Logan, J. A., and Olsen, M. A.: Evaluation of near-tropopause ozone distributions in the Global Modeling Initiative
5   combined stratosphere/troposphere model with ozonesonde data, Atmos. Chem. Phys., 8, 2365-2385, doi:10.5194/acp-8-2365-2008, 2008.

Cooper, O., D. D. Parrish, J. R. Ziemke, N. V. Balashov, M. Cupeiro, I. Galbally, S. Gilge, Larry W. Horowitz, N. R. Jensen, J. F. Lamarque, and V. Naik, et al., July 2014: Global distribution and trends of tropospheric ozone: An observation-based review. Elementa: Science of the Anthropocene, 2, 000029, DOI:10.12952/journal.elementa.000029.

Dee, D. P.: Bias and data assimilation, Q. J. R. Meteorol. Soc., 131, 3323–3343, doi:10.1256/qj.05.137, 2005.

10   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., Bid- lot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, 2011.

15   Deeter, M. N., Martinez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Pittman, J. V., Daube, B. C., and Wofsy, S. C.: Validation of MOPITT Version 5 thermal-infrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000–2011, J. Geophys. Res.-Atmos., 118, 6710–6725, doi:10.1002/jgrd.50272, 2013.

EC-JRC/PBL (European Commission, Joint Research Center/Netherlands Environmental Assessment Agency), Emission Database for Global Atmospheric Research version 4.2, avail- able at: http://edgar.jrc.ec.europa.eu (last access: June 2016), 2011.

20   Emili, E., Barret, B., Massart, S., Le Flochmoen, E., Piacentini, A., El Amraoui, L., Pannekoucke, O., and Cariolle, D.: Combined assimilation of IASI and MLS observations to constrain tropospheric and stratospheric ozone in a global chemical transport model, Atmos. Chem. Phys., 14, 177–198, doi:10.5194/acp-14-177-2014, 2014.

Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, Atmos. Chem. Phys., 3, 1285–1291, doi:10.5194/acp-3-1285-2003, 2003.

25   Flemming, J., Inness, A., Jones, L., Eskes, H. J., Huijnen, V., Schultz, M. G., Stein, O., Cariolle, D., Kinnison, D., and Brasseur, G.: Forecasts and assimilation experiments of the Antarctic ozone hole 2008, Atmos. Chem. Phys., 11, 1961–1977, doi:10.5194/acp-11-1961-2011, 2011.

Flemming, J., A. Benedetti, A. Inness, R. Engelen, L. Jones, V. Huijnen, S. Remy, M. Parrington, M. Suttie, A. Bozzo, V.-H. Peuch, D. Akritidis, and E. Katragkou, The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003–2015, Atmos. Chem.
30   Phys. Discuss., doi:10.5194/acp-2016-666, 2016, in review.

Foelsche, U., Scherllin-Pirscher, B., Ladstädter, F., Steiner, A. K., and Kirchengast, G.: Refractivity and temperature climate records from multiple radio occultation satellites consistent within 0.05%, Atmos. Meas. Tech., 4, 2007-2018, doi:10.5194/amt-4-2007-2011, 2011.

Fu, D., Bowman, K. W., Worden, H. M., Natraj, V., Worden, J. R., Yu, S., Veefkind, P., Aben, I., Landgraf, J., Strow, L., and Han, Y.: High-resolution tropospheric carbon monoxide profiles retrieved from CrIS and TROPOMI, Atmos. Meas. Tech., 9, 2567-2579,
35   doi:10.5194/amt-9-2567-2016, 2016.

Gaubert, B., A. F. Arellano Jr., J. Barré, H. M. Worden, L. K. Emmons, S. Tilmes, R. R. Buchholz, F. Vitt, K. Raeder, N. Collins, J. L. Anderson, C. Wiedinmyer, S. Martinez Alonso, D. P. Edwards, M. O. Andreae, J. W. Hannigan, C. Petri, K. Strong, N. Jones: Toward

a chemical reanalysis in a coupled chemistry-climate model: An evaluation of MOPITT CO assimilation and its impact on tropospheric composition, J. Geophys. Res. Atmos., 121, 7310–7343, doi:10.1002/2016JD024863, 2016.

Glumb, R. J., D. C. Jordan, and P. Mantica: Development of the Crosstrack Infrared Sounder (CrIS) sensor design, Proc. SPIE, 4486,411–424, doi:10.1117/12.455124, 2002.

5 Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D. W., Haywood, J. L., J., Lowe, D. C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R.: Changes in atmospheric constituents and in radiative forcing, in: Climate change 2007: The physical science basis, edited by: Solomon, S., Cambridge University Press, New York, 2007.

Graedel, T. E., Bates, T. S., Bouwman, A. F., Cunnold, D., Dignon, J., Fung, I., Jacob, D. J., Lamb, B. K., Logan, J. A., Marland, G., Middleton, P., Pacyna, J. M., Placet, M., and Veldt, C.: A compilation of inventories of emissions to the atmosphere, Global Biogeochem.
10 Cy., 7, 1–26, 1993.

Hamilton, J. F., Allen, G., Watson, N. M., Lee, J. D., Saxton, J. E., Lewis, A. C., Vaughan, G., Bower, K. N., Flynn, M. J., Crosier, J., Carver, G. D., Harris, N. R. P., Parker, R. J., Remedios, J. J., and Richards, N. A. D.: Observations of an atmospheric chemical equator and its implications for the tropical warm pool region, J. Geophys. Res., 113, D20313, doi:10.1029/2008JD009940, 2008.

Hartmann, D.L., A.M.G. Klein Tank, M. Rusticucci, L.V. Alexander, S. Brönnimann, Y. Charabi, F.J. Dentener, E.J. Dlugokencky, D.R.
15 Easterling, A. Kaplan, B.J. Soden, P.W. Thorne, M. Wild and P.M. Zhai: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Herman, R. L. and Kulawik, S. S. (Eds.): Tropospheric Emission Spectrometer TES Level 2 (L2) Data User's Guide, D-38042, version 6.0,
20 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, available at: http://tes.jpl.nasa.gov/ documents, 2013.

HTAP: Part A: Ozone and Particulate Matter, edited by: Den- tener, F., Keating, T., and Akimoto, H., prepared by the Task Force on Hemispheric Transport of Air Pollution acting within the framework of the Convention on Long-range Transboundary Air Pollution, United Nations, New York and Geneva, 2010.

Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F.,
25 Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédĺec, P., and Pĭz, H.-W.: The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0, Geosci. Model Dev., 3, 445-473, doi:10.5194/gmd-3-445-2010, 2010.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, Physica D, 230, 112–126, 2007.

30 Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073–4109, doi:10.5194/acp-13-4073-2013, 2013.

35 Jackson, D. R.: Assimilation of EOS MLS ozone observations in the Met Office data-assimilation system, Q. J. Roy. Meteorol. Soc., 133, 1771–1788, doi:10.1002/qj.140, 2007.

Jones, D. B. A., K. W. Bowman, P. I. Palmer, J. R. Worden, D. J. Jacob, R. N. Hoffman, I. Bey, and R. M. Yantosca: Potential of observations from the Tropospheric Emission Spectrometer to constrain continental sources of carbon monoxide, J. Geophys. Res.-Atmospheres, 108(D24), 4789, doi: 10.1029/2003JD003702, 2003.

Jonson, J. E., Stohl, A., Fiore, A. M., Hess, P., Szopa, S., Wild, O., Zeng, G., Dentener, F. J., Lupu, A., Schultz, M. G., Duncan, B. N., Sudo, K., Wind, P., Schulz, M., Marmer, E., Cuvelier, C., Keating, T., Zuber, A., Valdebenito, A., Dorokhov, V., De Backer, H., Davies, J., Chen, G. H., Johnson, B., Tarasick, D. W., Stübi, R., Newchurch, M. J., von der Gathen, P., Steinbrecht, W., and Claude, H.: A multi-model analysis of vertical ozone profiles, Atmos. Chem. Phys., 10, 5759-5783, doi:10.5194/acp-10-5759-2010, 2010.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 341 p., 2003.

Kawase, H., T. Nagashima, K. Sudo, and T. Nozawa, Future changes in tropospheric ozone under Representative Concentration Pathways (RCPs), Geophys. Res. Lett., 38, L05801, doi:10.1029/ 2010GL046402, 2011.

Klimont, Z., Cofala, J., Xing, J., Wei, W., Zhang, C., Wang, S., Kejun, J., Bhandari, P., Mathur, R., Purohit, P., Rafaj, P., Chambers, A., Amann, M., and Hao, J.: Projections of SO2, NOx and carbonaceous aerosols emissions in Asia, Tellus, 61B, 602–617, 2009.

Kiesewetter, G., Sinnhuber, B. M., Vountas, M., Weber, M., and Burrows, J. P.: A long-term stratospheric ozone data set from assimilation of satellite observations: high-latitude ozone anomalies, J. Geophys. Res., 115, D10307, doi:10.1029/2009JD013362, 2010.

Lahoz, W. A. and Schneider, P.: Data assimilation: making sense of Earth Observation, Front. Environ. Sci., 2, 16, doi:10.3389/fenvs.2014.00016, 2014.

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., Eyring, V., Young, P. J., Cionni, I., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y., McKenzie, I., Nagashima, T., Naik, V., Plummer, D., Rumbold, A., Skeie, R., 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.

Lee, Y. H., Lamarque, J.-F., Flanner, M. G., Jiao, C., Shindell, D. T., Berntsen, T., Bisiaux, M. M., Cao, J., Collins, W. J., Curran, M., Edwards, R., Faluvegi, G., Ghan, S., Horowitz, L. W., McConnell, J. R., Ming, J., Myhre, G., Nagashima, T., Naik, V., Rumbold, S. T., Skeie, R. B., Sudo, K., Takemura, T., Thevenon, F., Xu, B., and Yoon, J.-H.: Evaluation of preindustrial to present-day black carbon and its albedo forcing from Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 2607-2634, doi:10.5194/acp-13-2607-2013, 2013.

Lelieveld, J. and P. J. Crutzen, Role of deep cloud convection in the ozone budget of the troposphere, Science, 264, 1759-1761, 1994.

Lelieveld, J., and F. J. Dentener, What controls tropospheric ozone?, J. Geophys. Res., 105(D3), 3531–3551, doi:10.1029/1999JD901011, 2000.

Livesey, N. J., Read, W. G., Froidevaux, L., Lambert, A., Manney, G. L., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Cofield, R. E., Cuddy, D. T., Fuller, R. A., Jarnot, R. F., Jiang, J. H., Knosp, B. W., Stek, P. C., Wagner, P. A., and Wu, D. L.: Aura Microwave Limb Sounder (MLS), Version 3.3 Level 2 data quality and description document, Tech. Rep. JPL D-33509, Jet Propul. Lab., Pasadena, CA, 2011.

Logan, J. A., Prather, M. J., Wofsy, S. C., and McElroy, M. B.: Tropospheric chemistry: a global perspective, J. Geophys. Res., 86, 7210–7354, 1981.

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(D13), 16115–16149, doi:10.1029/1998JD100096, 1999.

5   Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO_2 columns, Atmos. Chem. Phys., 12, 2263–2288, doi:10.5194/acp-12-2263-2012, 2012a.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of tropospheric chemical composition and emissions, Atmos. Chem. Phys., 12, 9545–9579, doi:10.5194/acp-12-9545-2012, 2012b.

10   Miyazaki, K. and Eskes, H.: Constraints on surface NO_x emissions by assimilating satellite observations of multiple species, Geophys. Res. Lett., 40, 4745–4750, doi:10.1002/grl.50894, 2013.

Miyazaki, K., Eskes, H. J., Sudo, K., and Zhang, C.: Global lightning NO_x production estimated by an assimilation of multiple satellite data sets, Atmos. Chem. Phys., 14, 3277–3305, doi:10.5194/acp-14-3277-2014, 2014.

Miyazaki, K., Eskes, H. J., and Sudo, K.: A tropospheric chemistry reanalysis for the years 2005–2012 based on an assimilation of OMI,
15   MLS, TES, and MOPITT satellite data, Atmos. Chem. Phys., 15, 8315-8348, doi:10.5194/acp-15-8315-2015, 2015.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NO_x emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-529, in review, 2016.

Monks, P. S.: A review of the observations and origins of the spring ozone maximum, Atmos. Environ., 34, 3545–3561, 2000.

Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima,
20   A. Robock, G. Stephens, T. Takemura and H. Zhang: Anthropogenic and Natural Radiative Forc- ing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013

Naik, V., Voulgarakis, A., Fiore, A. M., Horowitz, L. W., Lamarque, J.-F., Lin, M., Prather, M. J., Young, P. J., Bergmann, D., Cameron-
25   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.

30   Oltmans, S., Lefohn, A. S., Harris, J. M., Galbally, I., Scheel, H. E., Bodeker, G., Brunke, E., Claude, H., Tarasick, D., Johnson, B. J., Simmonds, P., Shadwick, D., Anlauf, K., Hayden, K., Schmidlin, F., Fujimoto, T., Akagi, K., Meyer, C., Nichol, S., Davies, J., Re-dondas, A., and Cuevaso, E.: Long-term changes in tropospheric ozone, Atmos. Environ., 40, 3156–3173, 2006.

Park, M., Randel, W. J., Emmons, L. K., Bernath, P. F., Walker, K. A., and Boone, C. D.: Chemical isolation in the Asian monsoon anticyclone observed in Atmospheric Chemistry Experiment (ACE-FTS) data, Atmos. Chem. Phys., 8, 757–764, doi:10.5194/acp-8-757-2008, 2008.

35   Parrington, M., Jones, D. B. A., Bowman, K. W., Thompson, A. M., Tarasick, D. W., Merrill, J., Oltmans, S. J., Leblanc, T., Witte, J. C., and Millet, D. B.: Impact of the assimilation of ozone from the Tropospheric Emission Spectrometer on surface ozone across North America, Geophys. Res. Lett., 36, L04802, doi:10.1029/2008GL036935, 2009.

Parrish, D. D., K. S. Law, J. Staehelin, R. Derwent, O. R. Cooper, H. Tanimoto, A. Volz-Thomas, S. Gilge, H.-E. Scheel, M. Steinbacher, and E. Chan: Lower tropospheric ozone at northern midlatitudes: Changing seasonal cycle, Geophys. Res. Lett., 40, 1631–1636, doi: 10.1002/grl.50303, 2013.

Parrish, D.D., J.-F. Lamarque, V. Naik, L. Horowitz, D.T. Shindell, J. Staehelin, R. Derwent, O.R. Cooper, H. Tanimoto, A. Volz-Thomas, S. Gilge, H.-E. Scheel, M. Steinbacher, M. Frohlich: Long-term changes in lower tropospheric baseline ozone concentrations: Comparing chemistry-climate models and observations at northern mid-latitudes, J. Geophys. Res. Atmos., 119, doi:10.1002/2013JD021435, 2014.

Price, C. and Rind, D.: A simple lightning parameterization for calculating global lightning distributions, J. Geophys. Res., 97, 9919–9933, doi:10.1029/92JD00719, 1992.

Shindell, D., G. Faluvegi, A. Lacis, J. Hansen, R. Ruedy, and E. Aguilar, Role of tropospheric ozone increases in 20th-century climate change, J. Geophys. Res., 111, D08302, doi:10.1029/2005JD006348, 2006.

Shindell, D. T., Lamarque, J.-F., Schulz, M., Flanner, M., Jiao, C., Chin, M., Young, P. J., Lee, Y. H., Rotstayn, L., Mahowald, N., Milly, G., Faluvegi, G., Balkanski, Y., Collins, W. J., Conley, A. J., Dalsoren, S., Easter, R., Ghan, S., Horowitz, L., Liu, X., Myhre, G., Nagashima, T., Naik, V., Rumbold, S. T., Skeie, R., Sudo, K., Szopa, S., Takemura, T., Voulgarakis, A., Yoon, J.-H., and Lo, F.: Radiative forcing in the ACCMIP historical and future climate simulations, Atmos. Chem. Phys., 13, 2939-2974, doi:10.5194/acp-13-2939-2013, 2013.

Smit, H. G. J. and Kley, D.: Ozone Sonde Intercomparison Experiment (JOSIE), WMO Global Atmosphere Watch report series, No. 130 (Technical Document No. 926), World Meteorological Organization, Geneva, 1998.

Stajner, I. and Wargan, K.: Antarctic stratospheric ozone from the assimilation of occultation data, Geophys. Res. Lett., 31, L18108, doi:10.1029/2004GL020846, 2004.

Stevenson, D., Dentener, F. J., Schultz, M. G., Ellingsen, K., van Noije, T. P. C., Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglus- taine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J. F., Lawrence, M. G., Montanaro, V., Muller, J. F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo, K., and Szopa, S.: Multi-model ensemble simulations of present-day and nearfuture tropospheric ozone, J. Geophys. Res., 111, D08301, doi:10.1029/2005JD006338, 2006.

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.

Stohl, A., Characteristics of atmospheric transport into the Arctic troposphere, J. Geophys. Res., 111, D11306, doi:10.1029/2005JD006888, 2006.

Sudo, K., M. Takahashi, J. Kurokawa, and H. Akimoto, CHASER: A global chemical model of the troposphere 1. Model description, J. Geophys. Res., 107, 10.1029/2001JD001113, 2002.

Sudo, K., and H. Akimoto, Global source attribution of tropospheric ozone: Long-range transport from various source regions, J. Geophys. Res., 112, D12302, doi:10.1029/2006JD007992, 2007.

Sofieva, V. F., Kalakoski, N., Pivrinta, S.-M., Tamminen, J., Laine, M., and Froidevaux, L.: On sampling uncertainty of satellite ozone profile measurements, Atmos. Meas. Tech., 7, 1891–1900, doi:10.5194/amt-7-1891-2014, 2014.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, 2012.

Thompson, A. M.: The oxidizing capacity of the Earth's atmosphere: probable past and future changes, Science, 256, 1157–1165, 1992.

Thompson, A. M., Witte, J. C., McPeters, R. D., Oltmans, S. J., Schmidlin, F. J., Logan, J. A., Fujiwara, M., Kirch- hoff, V. W.J. H., Posny, F., Coetzee, G. J. R., Hoegger, B., Kawakami, S.,Ogawa, T., Johnson, J. B., Vomel, H., and Labow, G.: Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998–2000 tropical ozone climatology 1. Comparison with Total Ozone Mapping Spectrometer (TOMS) and ground-based measurements, J. Geophys. Res., 108, 8238, doi:10.1029/2001JD000967, 2003.

Thompson, A. M., Miller, S. K., Tilmes, S., Kollonige, D. W., Witte, J. C., Oltmans, S. J., Johnson, B. J., Fujiwara, M., Schmidlin, F. J., Coetzee, G. J. R., Komala, N., Maata, M., Mo- hamad, M. B., Nguyo, J., Mutai, C., Ogino, S.-Y., Da Silva, F. R., Paes Leme, N. M., Posny, F., Scheele, R., Selkirk, H. B., Sh- iotani, M., Stuebi, R., Levrat, G., Calpini, B., Thouret, V., Tsuruta, H., Valverde Canossa, J., Voemel, H., Yonemura, S., Andres Diaz, J., Thanh, N. T. T., and Ha, H. T. T.: Southern Hemisphere Additional Ozonesondes (SHADOZ) ozone climatology (2005–2009): tropospheric and Tropical Tropopause Layer (TTL) profiles with comparisons to OMI- based ozone products, J. Geophys. Res.-Atmos., 117, D23301, doi:10.1029/2011JD016911, 2012.

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.

Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Ael, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krger, K., Kyrl, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, J. Geophys. Res. Atmos., 118, 11847–11862, doi:10.1002/jgrd.50874, 2013.

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.

van der A, R. J., Allaart, M. A. F., and Eskes, H. J.: Extended and refined multi sensor reanalysis of total ozone for the period 1970–2012, Atmos. Meas. Tech., 8, 3021-3035, doi:10.5194/amt-8-3021-2015, 2015.

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

Verstraeten, W. W., Neu, J. L., Williams, J. E., Bowman, K. W., Worden, J. R., and Boersma, K. F.: Rapid increases in tropospheric ozone production and export from China, Nat. Geosci, 8, 690–695, doi:10.1038/ngeo2493, 2015.

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.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, Geosci. Model Dev., 4, 845–872, 2011.

Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, Atmos. Chem. Phys., 7, 2643-2660, doi:10.5194/acp-7-2643-2007, 2007.

Wirth, V., Quasi-stationary planetary waves in total ozone and their correlation with lower stratospheric temperature, J. Geophys. Res., 98(D5), 8873–8882, doi:10.1029/92JD02820, 1993.

WMO (World Meteorological Organization): Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project – Report No. 55, 416 pp., Geneva, Switzerland, 2014.

Wu, S., L. J. Mickley, D. J. Jacob, J. A. Logan, R. M. Yantosca, and D. Rind: Why are there large differences between models in global budgets of tropospheric ozone? J. Geophys. Res., 112, D05302, doi:10.1029/2006JD007801, 2007.

Ziemke, J. R., S. Chandra, B. N. Duncan, L. Froidevaux, P. K. Bhartia, P. F. Levelt, and J. W. Waters: 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. Chemical reanalysis (or control run in brackets) minus ozonesonde comparisons of mean ozone concentrations in 2005–2009. RMSE is the root-mean-square error. Units of bias and RMSE are ppb. T-Corr is the temporal correlation.

	90–55° S			55–15° S			15S–15° N			15–55° N			55–90° N		
	Bias	RMSE	T-Corr	Bias	RMSE	T-Corr	Bias	RMSE	T-Corr	Bias	RMSE	T-Corr	Bias	RMSE	T-Corr
850–500	-1.9	4.3	0.84	-0.1	5.3	0.92	3.2	7.1	0.90	-0.1	6.8	0.87	-2.9	5.9	0.82
hPa	(-4.4)	(4.9)	(0.94)	(-0.9)	(6.0)	(0.85)	(2.3)	(7.4)	(0.87)	(-1.0)	(7.4)	(0.84)	(-5.3)	(6.5)	(0.87)
500–200	-1.0	18.3	0.67	-0.7	13.6	0.96	0.9	8.4	0.90	0.2	17.6	0.96	-9.9	29.3	0.97
hPa	(29.0)	(33.9)	(0.65)	(24.6)	(26.8)	(0.90)	(0.8)	(9.7)	(0.74)	(17.8)	(24.6)	(0.89)	(17.2)	(37.4)	(0.94)
200–90	37.3	88.4	0.95	15.0	53.1	0.99	4.8	20.3	0.93	8.6	62.6	0.99	13.1	101.1	0.99
hPa	(410.5)	(312.7)	(0.41)	(251.6)	(194.5)	(0.97)	(50.0)	(51.8)	(0.85)	(193.0)	(158.0)	(0.98)	(304.4)	(247.3)	(0.94)

Table 2. ACCMIP model mean minus reanalysis comparisons of the mean ozone concentrations. Units of bias and RMSE are ppb. S-Corr is the spatial correlation coefficient.

	90–30° S			30° S–Eq			Eq–30° N			30–90° N		
	Bias	RMSE	S-Corr	Bias	RMSE	S-Corr	Bias	RMSE	S-Corr	Bias	RMSE	S-Corr
800 hPa	0.0	2.0	0.99	-3.2	4.5	0.94	2.4	3.5	0.97	7.6	7.9	0.97
500 hPa	-3.5	4.0	0.99	-7.0	7.7	0.95	-3.1	4.1	0.96	4.7	5.4	0.57
200 hPa	-20.7	23.9	0.99	-2.0	4.2	0.99	-0.5	2.9	0.99	-15.7	20.1	1.00

Table 3. ACCMIP models minus reanalysis comparisons of the mean ozone concentrations at 500 hPa. Units of bias are ppb.

	90–30° S		30° S–Eq		Eq–30° N		30–90° N	
	Bias	S-Corr	Bias	S-Corr	Bias	S-Corr	Bias	S-Corr
1. CESM-CAM	-8.5	0.99	-14.6	0.90	-9.5	0.91	4.8	0.12
2. CICERO-OsloCTM2	-11.4	0.99	-8.1	0.89	-6.7	0.94	-5.9	0.82
3. CMAM	-0.8	0.99	-7.7	0.85	-6.6	0.91	-0.5	0.73
4. EMAC	0.5	0.96	2.0	0.77	3.5	0.90	1.6	0.87
5. GEOSCCM	-2.1	0.98	-7.4	0.90	-4.7	0.91	5.2	0.59
6. GFDL-AM3	4.4	0.99	-2.4	0.95	0.9	0.95	8.0	0.03
7. GISS-E2-R	-0.2	0.98	-3.6	0.87	1.0	0.91	14.5	0.04
8. GISS-E2-TOMAS	5.9	0.96	1.3	0.83	4.6	0.90	17.2	-0.07
9. HadGEM2	-6.6	0.98	-12.8	0.91	-7.8	0.90	-0.8	0.86
10. LMDzORINCA	-5.0	0.98	-6.5	0.94	-4.6	0.95	2.8	0.60
11. MIROC-CHEM	-4.2	0.98	-0.5	0.92	1.3	0.93	-2.0	0.87
12. MOCAGE	-9.6	0.93	-12.2	0.47	-4.0	0.82	11.8	-0.11
13. NCAR-CAM3.5	-4.0	0.99	-10.4	0.93	-4.3	0.93	4.3	0.45
14. STOC-HadAM3	-8.7	0.96	-8.9	0.86	-4.0	0.94	2.4	0.38
15. UM-CAM	-2.8	0.96	-13.4	0.85	-5.8	0.85	7.5	0.79

Table 4. Regions and observation sites used in model evaluation in Section 5. The 11 regions are defined following Tilmes et al. (2012). See also Fig. 8.

Region	Station (Lat/Lon)
NH polar West (60° N–90° N, 120° W–40° W)	Alert (83/-62), Eureka (80/-86), Resolute (74/-95)
NH polar East (60° N–90° N, 40° W–30° E)	NyAlesund (79/12), Scoresbysund (71/-22), Lerwick (60/-1)
Canada (53° N–60° N, 120° W–50° W)	Churchill (59/-94), Edmonton (53/-114), Goosebay (53/-60)
Western Europe (45° N–55° N, 0° E–24° E)	Legionowo (52/21), Lindenberg (52/14), Debilt (52/5),
	Uccle (51/4), Praha (50/15), Hohenpbg (48/11), Payerne (47/7)
Eastern US (32° N–38° N, 129° E–142° W)	Wallops Island (40/-76), Huntsville (35/87)
NH subtropics (15° N–29° N, 110° E–150° W)	Hilo (19/-155), Hongkong (22/114), Naha (26/128)
W. Pacific/E. Indian (20° S–6° S, 110° E–160° W)	Fiji (-18/178), Watukosek (-8/113), Samoa (-14/-171)
Equatorial Americas (4° S–9° N, 100° W–45° W)	Paramaribo (6/-55), Sancristobal (-1,-90)
Atlantic/Africa (11° S–2° N, 40° W–40° E)	Nairobi (-1/37), Natal (-5/-35), Ascension (-8/-14)
SH mid latitudes (60° S–40° S, all longitudes)	Lauder (-45/170), Macquarie (-55/159)
SH high latitudes (60° N–80° N, all longitudes)	Marambio (-64/57), Syowa (-69/40), Neumayer (-71/-8)

Table 5. Median of the ACCMIP models minus reanalysis at 500 hPa (in % relative to the reanalysis concentrations). Results are shown for the regional average (Regional) and at the ozonesonde sampling (Sonde). Relative differences between the two estimates larger than 30 % are shown in bold.

	DJF		MAM		JJA		SON	
	Regional	Sonde	Regional	Sonde	Regional	Sonde	Regional	Sonde
NH polar West	12.6	13.4	13.4	14.3	15.5	16.2	17.5	15.4
NH polar East	**5.9**	**1.9**	12.3	10.0	16.5	13.8	17.7	14.6
Canada	6.7	5.0	**7.2**	**3.9**	**7.9**	**5.6**	12.9	12.7
Western Europe	**3.1**	**1.3**	**6.1**	**4.4**	3.0	2.4	8.6	7.0
Eastern US	3.9	4.6	**1.6**	**3.0**	**1.4**	**-0.3**	**3.8**	**2.9**
NH subtropics	-9.5	-10.3	-9.0	-10.3	**-5.5**	**-0.3**	**-3.5**	**-8.6**
W. Pacific/E. Indian	**-27.3**	**-16.3**	**-16.1**	**-23.4**	-12.1	-12.6	**-16.3**	**-27.4**
Equatrial Americas	-10.9	-9.9	**-4.5**	**-9.6**	-15.6	-19.6	-21.9	-24.4
Atlantic/Africa	-26.6	-23.8	-17.8	-17.5	-15.7	-18.2	-23.6	-25.7
SH mid latitudes	**-12.1**	**-3.6**	**-6.2**	**-10.7**	-12.7	-13.3	-16.8	-17.6
SH high latitudes	**-10.3**	**-4.3**	**2.2**	**-3.4**	**-3.6**	**-9.2**	**-11.7**	**-4.7**

Table 6. ACCMIP multi-model mean minus reanalysis comparisons of the seasonal amplitude of regional mean ozone concentration (in %) for the regional average (Regional) and at the ozonesonde sampling (Sonde). The seasonal amplitude is estimated as a difference between maximum and minimum monthly mean concentrations.

	800 hPa		500 hPa		200 hPa	
	Regional	Sonde	Regional	Sonde	Regional	Sonde
NH polar West	40.2	52.0	-7.0	4.5	-24.2	-20.9
NH polar East	14.2	13.4	10.5	9.4	-16.3	-24.2
Canada	1.0	27.1	-11.2	-22.9	-22.1	-14.6
Western Europe	-12.2	38.0	1.0	7.1	-18.5	-18.6
Eastern US	-19.0	71.3	-8.6	-13.3	-11.1	-15.7
NH subtropics	10.5	63.4	-27.0	16.7	-48.2	-46.0
W. Pacific/E. Indian	10.5	-42.9	-14.3	-24.2	-16.3	-33.6
Equatrial Americas	-27.3	-5.4	-64.1	-23.9	-37.2	-76.5
Atlantic/Africa	-14.7	-1.3	-13.1	3.3	-32.6	-15.2
SH mid latitudes	7.8	-1.5	-45.3	-47.5	-40.2	-32.6
SH high latitudes	40.0	4.6	-31.1	-36.3	83.6	-20.8
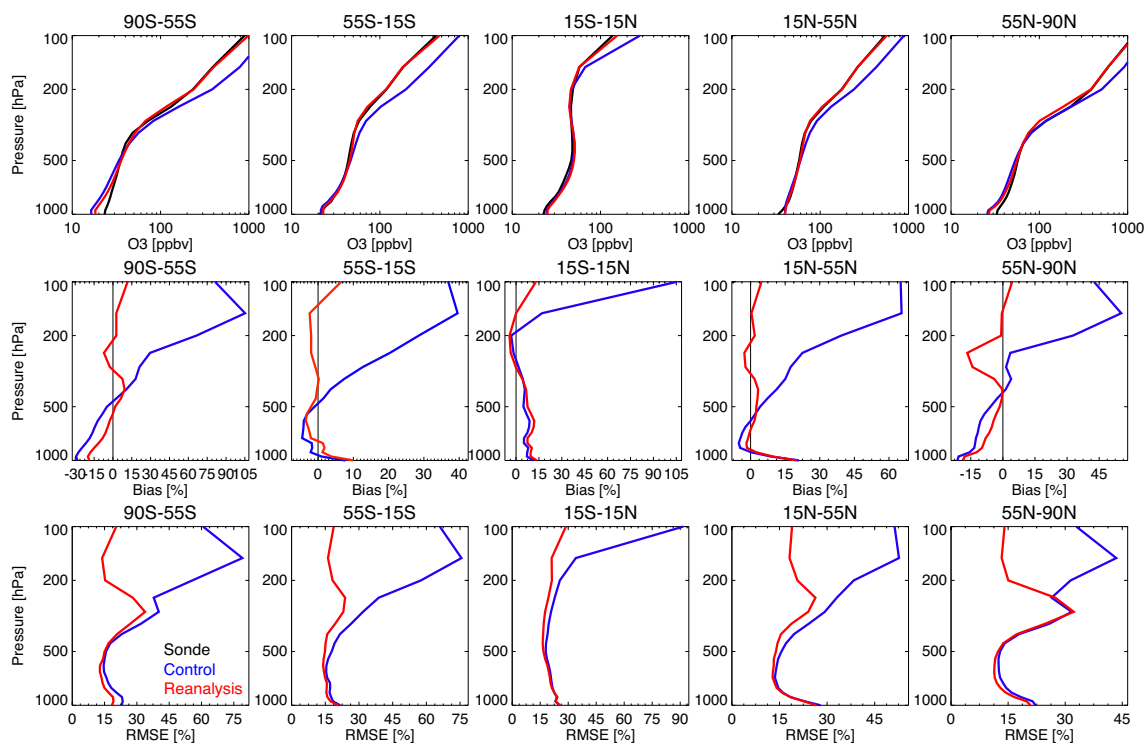
Figure 1. Comparison of vertical ozone profiles from ozonesondes (black), control run (blue), and reanalysis (red) averaged for the period 2005–2009. Top row shows mean profile; middle and bottom rows show mean difference and RMSE between control run and observations (blue) and between the reanalysis and the observations (red). From left to right, results are shown for SH high latitudes (55–90°S), SH mid-latitudes (15–55°S), tropics (15°S–15°N), NH mid-latitudes (15–55°N), and NH high latitudes (55–90°N). All ozonesonde observations taken from the WOUDC database were used in the comparison.
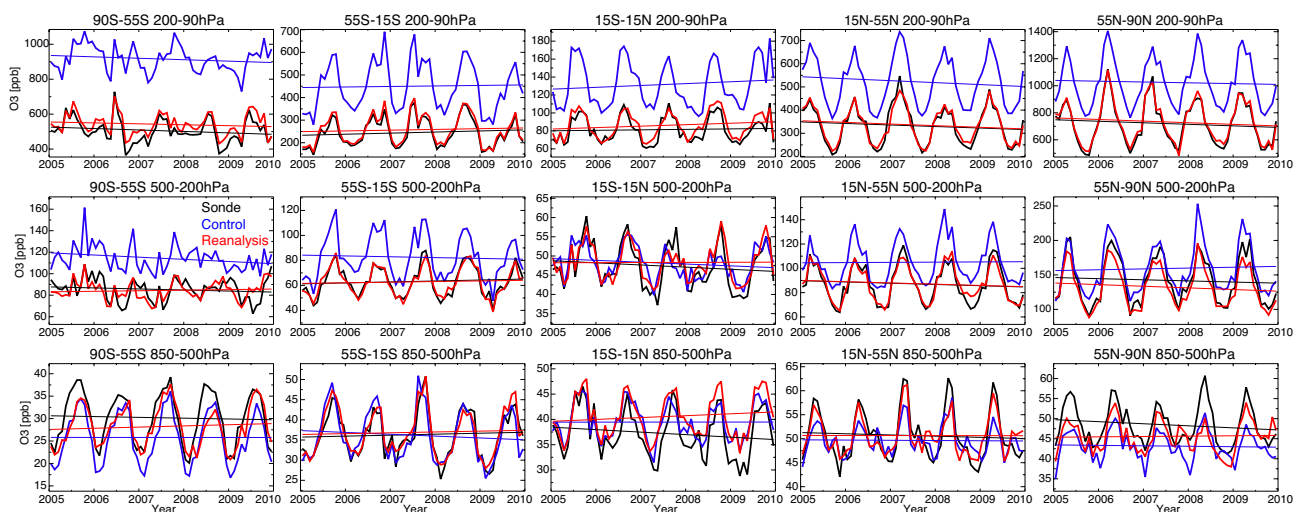
Figure 2. Time series of monthly mean ozone concentrations obtained from ozonesondes (black), control run (blue), and reanalysis (red) averaged between 850 and 500 hPa (top), 500 and 200 hPa (middle), and 200 and 90 hPa (bottom) for 2005–2009. From left to right the results are shown for SH high latitudes (55–90°S), SH mid-latitudes (15–55°S), tropics (15°S–15°N), NH mid-latitudes (15–55°N), and NH high latitudes (55–90°N).
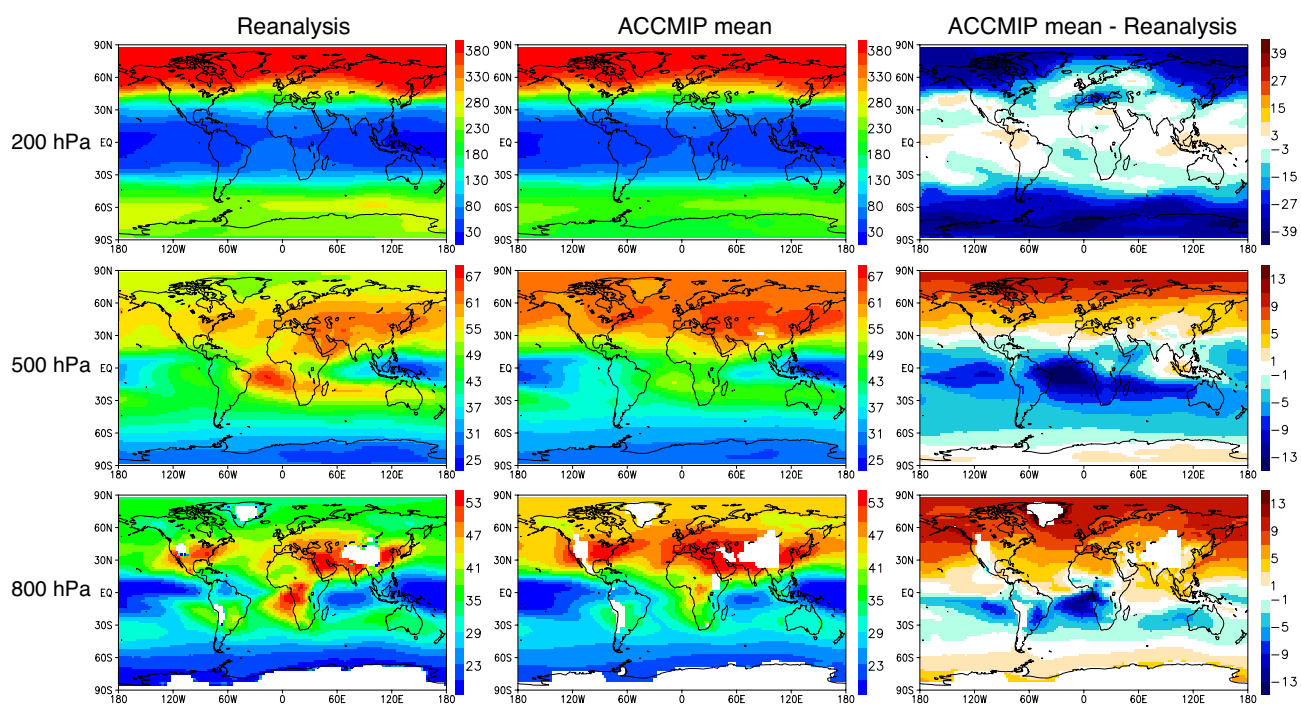
Figure 3. Global distributions of annual mean ozone concentrations obtained from reanalysis (left), ACCMIP model mean (center), and difference between ACCMIP model mean and reanalysis (right). From top to bottom, results are shown for global distributions at 200 hPa, 500 hPa, and 800 hPa. Units are ppb.
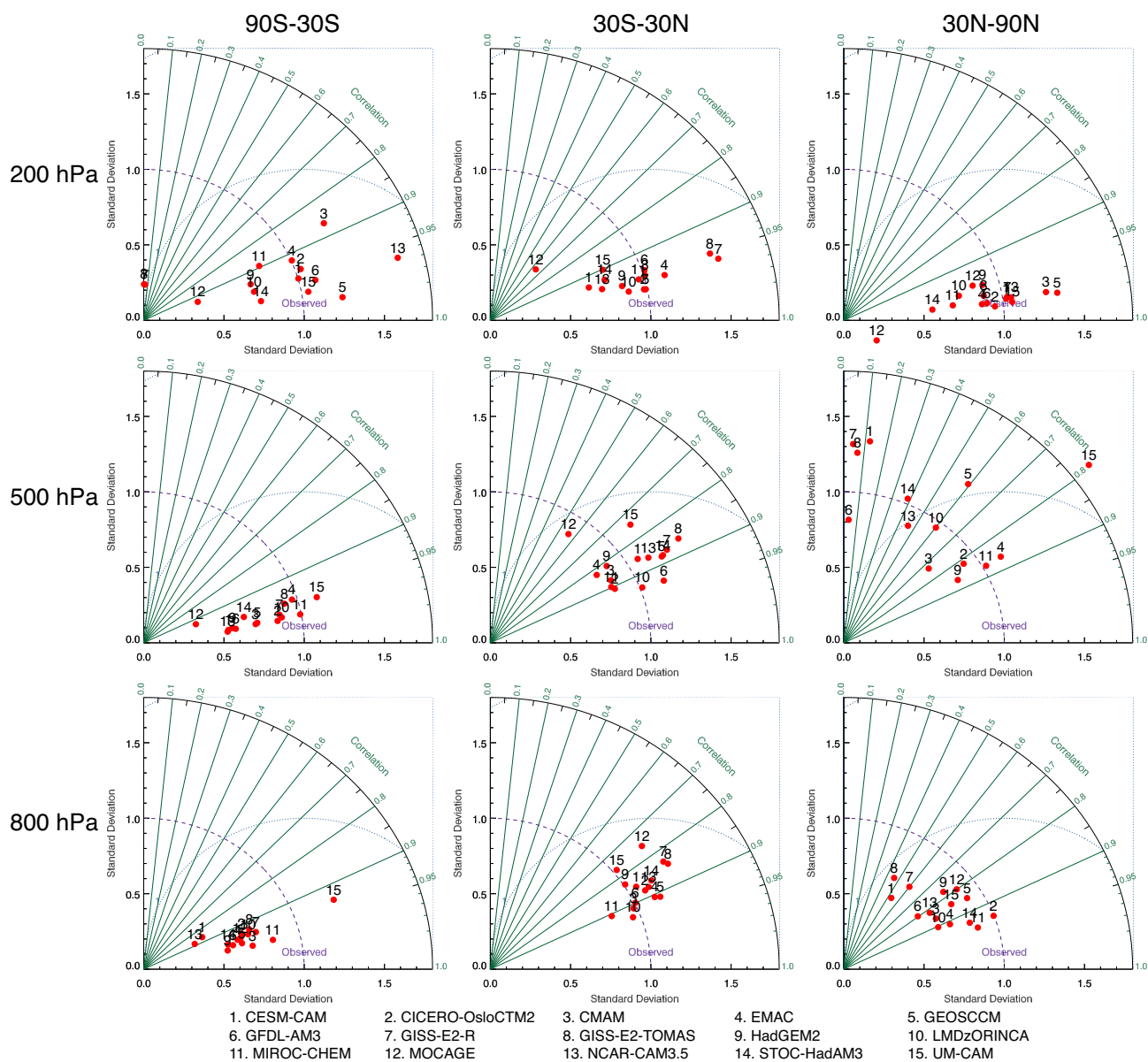
Figure 4. Taylor diagrams showing standard deviation normalized with respect to that of reanalysis (x-axis) and spatial correlation coefficient (y-axis) for the comparison of annual mean ozone concentrations between ACCMIP models and reanalysis for SH extratropics (90°S–30°S, left), tropics and subtropics (30°S–30°N, center), and NH extratropics (30°N–90°N, right) at 200 hPa (top), 500 hPa (middle), and 800 hPa (bottom).
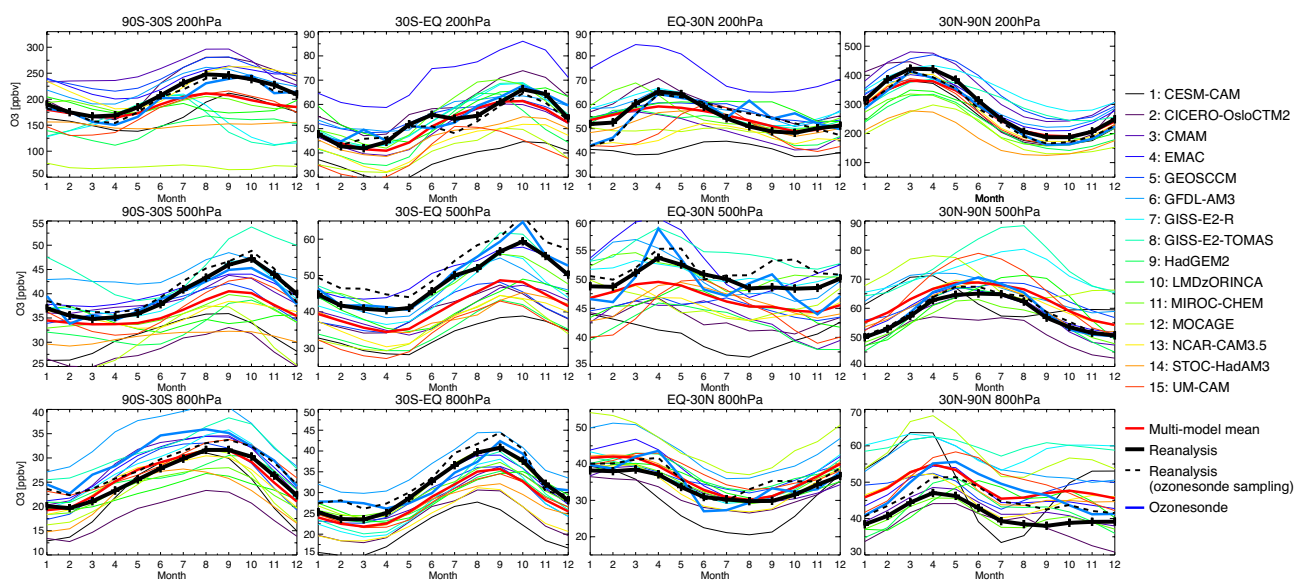
Figure 5. Comparison of seasonal variation of ozone concentration between the reanalysis (black lines), individual ACCMIP models (thin colored lines), ACCMIP ensemble mean (red solid line), and ozonesonde observations (blue solid line) averaged between 90°S–30°S (1st column from left), 30°S–Equator (2nd column), Equator–30°N (3rd column), and 30°N–90°N (4th column). From top to bottom, results are shown for concentrations at 200 hPa, 500 hPa, and 800 hPa. Individual model results are shown by colored thin lines. The reanalysis result is shown for the average over all model grid points (black solid line) and over the ozonesonde sampling sites/time (black dashed line).
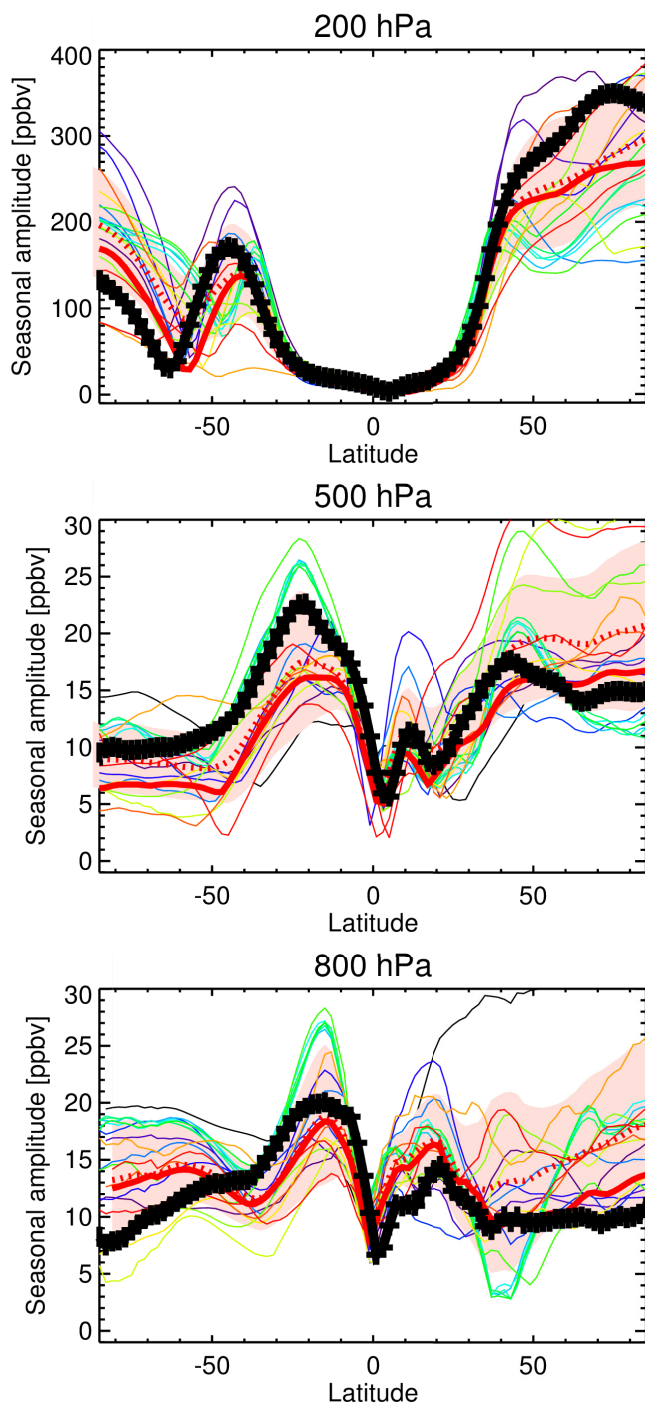
Figure 6. Seasonal amplitude (peak-to-peak difference) estimated from the reanalysis (black solid line) and ACCMIP models (thin colored lines). The $\pm 1\sigma$ deviation among ACCMIP models (i.e., model spread) is shown in pink. The seasonal amplitude derived from the multi-model mean fields (red solid line) and the multi-model mean of the seasonal amplitude from each model (red dashed line) are also shown. From top to bottom, results are shown for 200 hPa, 500 hPa, and 800 hPa.
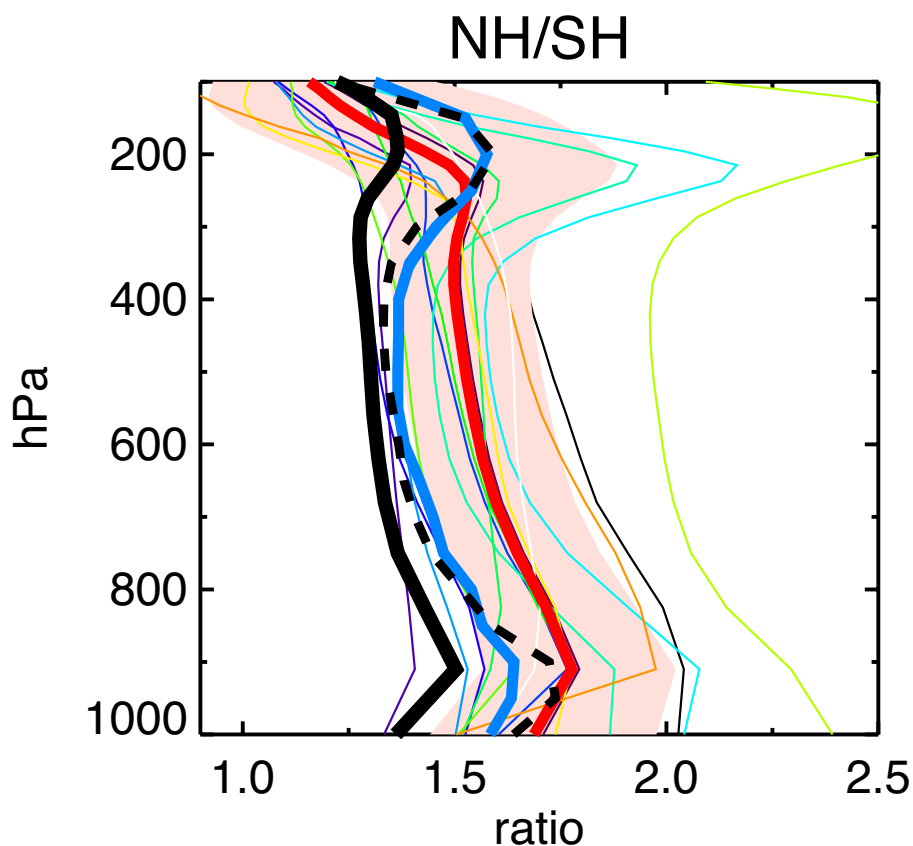
Figure 7. Vertical profile of inter-hemispheric gradient of annual mean ozone concentrations estimated from the reanalysis (black lines), ACCMIP ensemble mean (red solid line), ACCMIP models (thin colored lines), and ozonesonde observations (blue solid line). The reanalysis result is shown for the average over all model grid points (black solid line) and over the ozonesonde samplings (black dashed line). The $\pm 1\sigma$ deviation among the ACCMIP models is shown in pink.
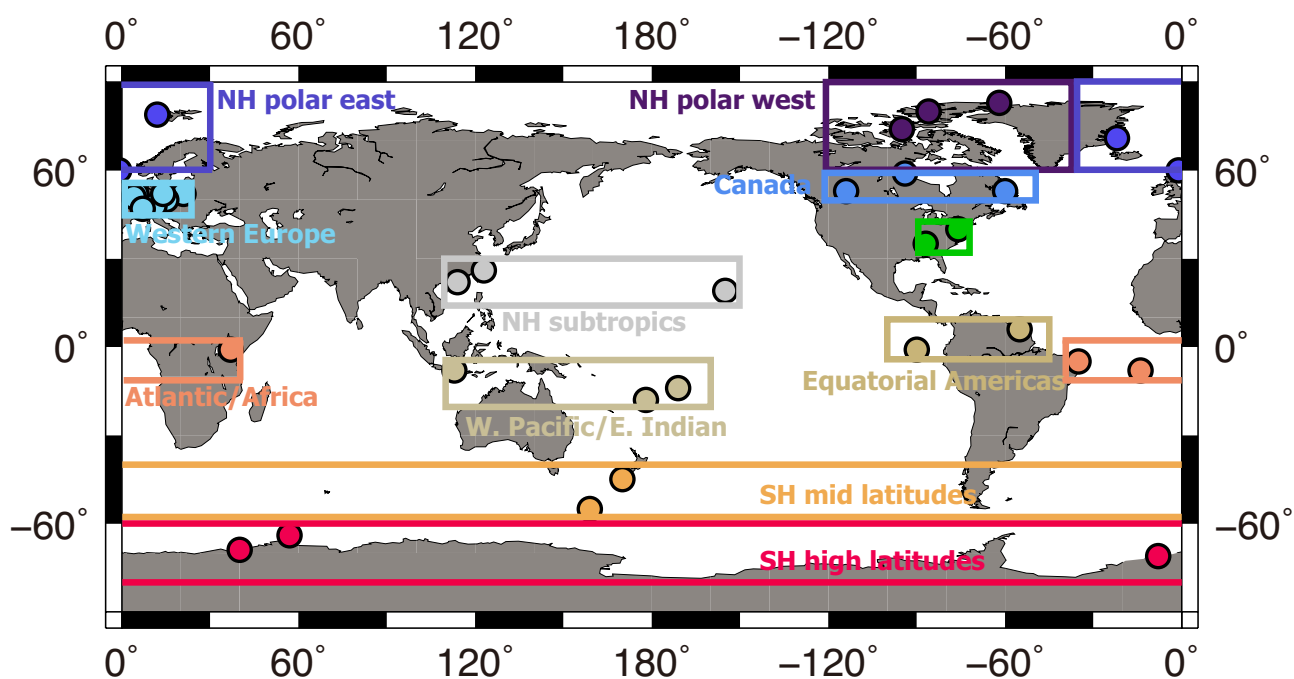
Figure 8. Regions and observation sites used in model evaluation. The 11 regions are defined following Tilmes et al. (2012). See also Table 3.
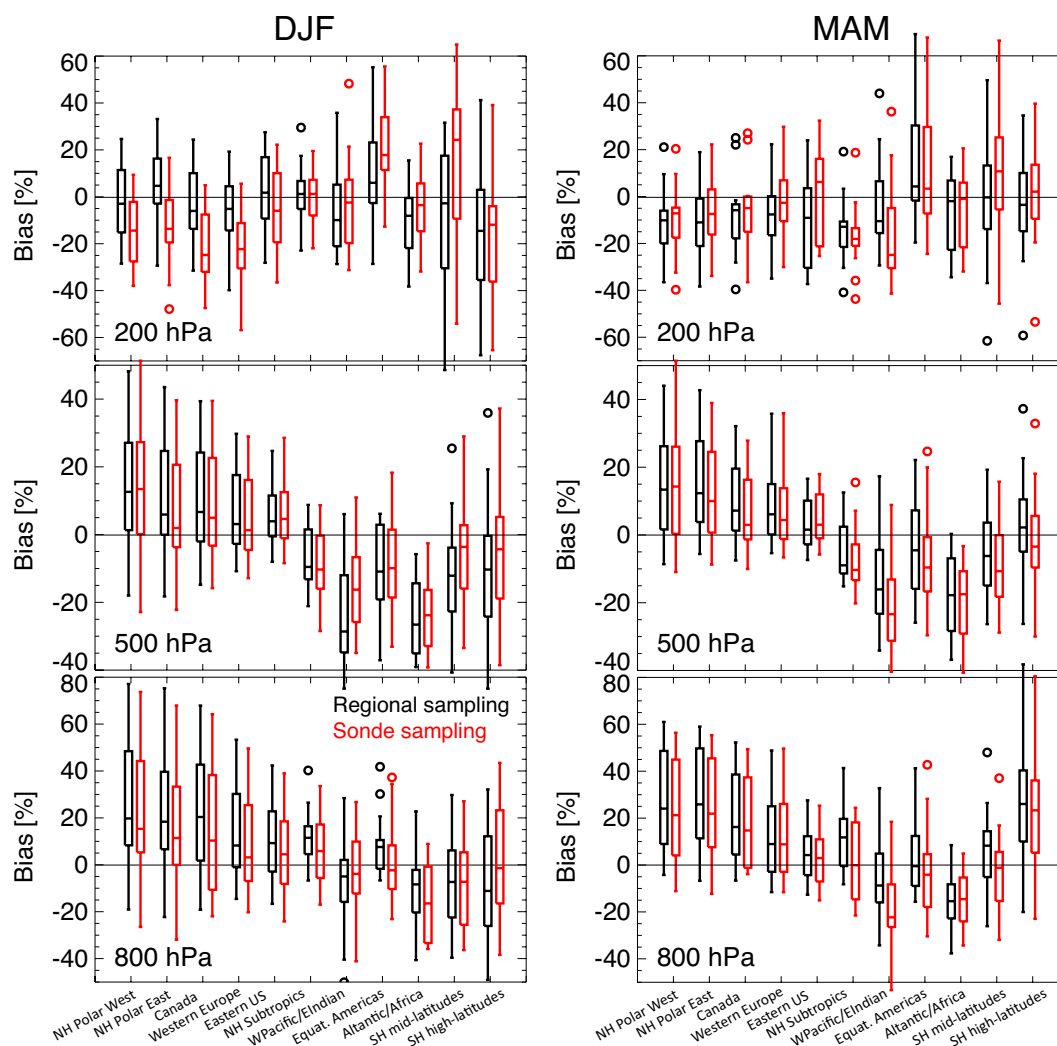
Figure 9. Box plots of relative model–reanalysis difference for seasonal mean concentration for DJF (left) and MAM (right) at 200 hPa (top), 500 hPa (middle), and 800 hPa (bottom). Results are shown for ACCMIP model simulations for 11 regions (c.f., Table 3 and Fig. 8). Black box shows model minus reanalysis difference for regional mean concentration (averaged over all model grid points); red box shows model minus reanalysis at the ozonesonde samplings.
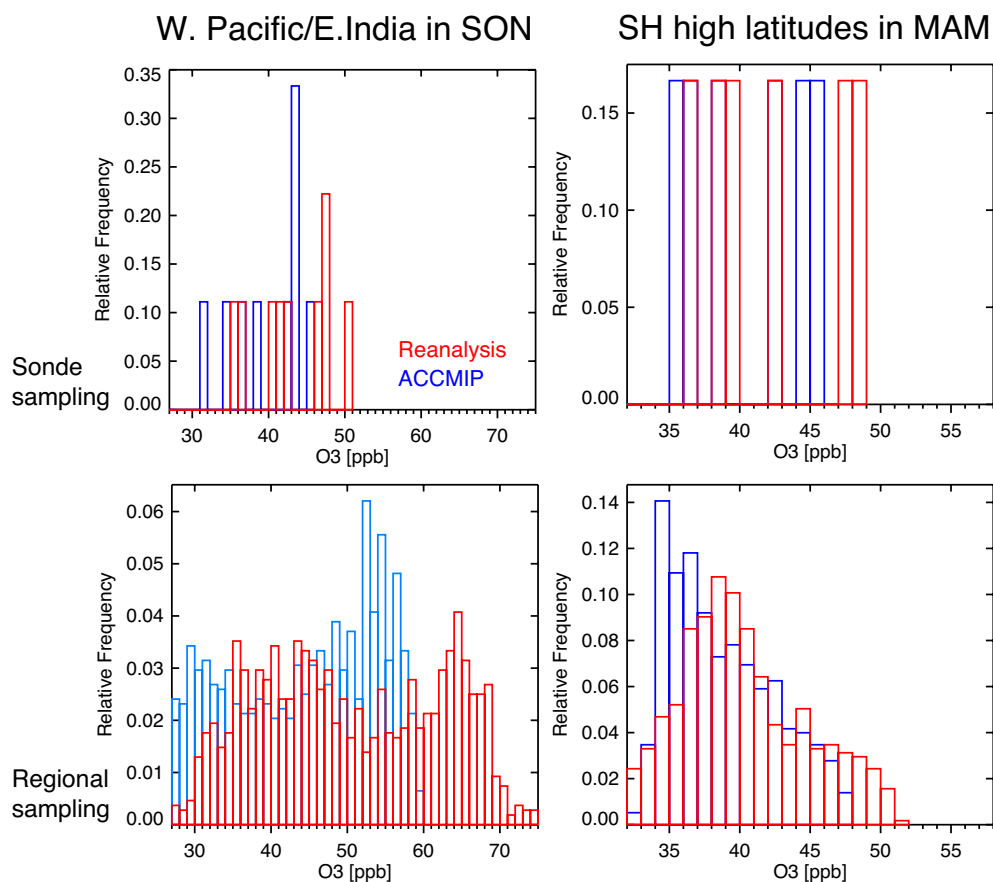
Figure 10. Probability distribution functions (PDFs) of ozone concentration obtained from the ACCMIP multi-model mean (blue) and the reanalysis (red) at 500 hPa for W. Pacific/E. India in SON (left) and for the SH high latitudes in MAM (right). The plots are shown for all model and reanalysis grid point (bottom) and for the ozonesonde sampling (top) within each defined region.
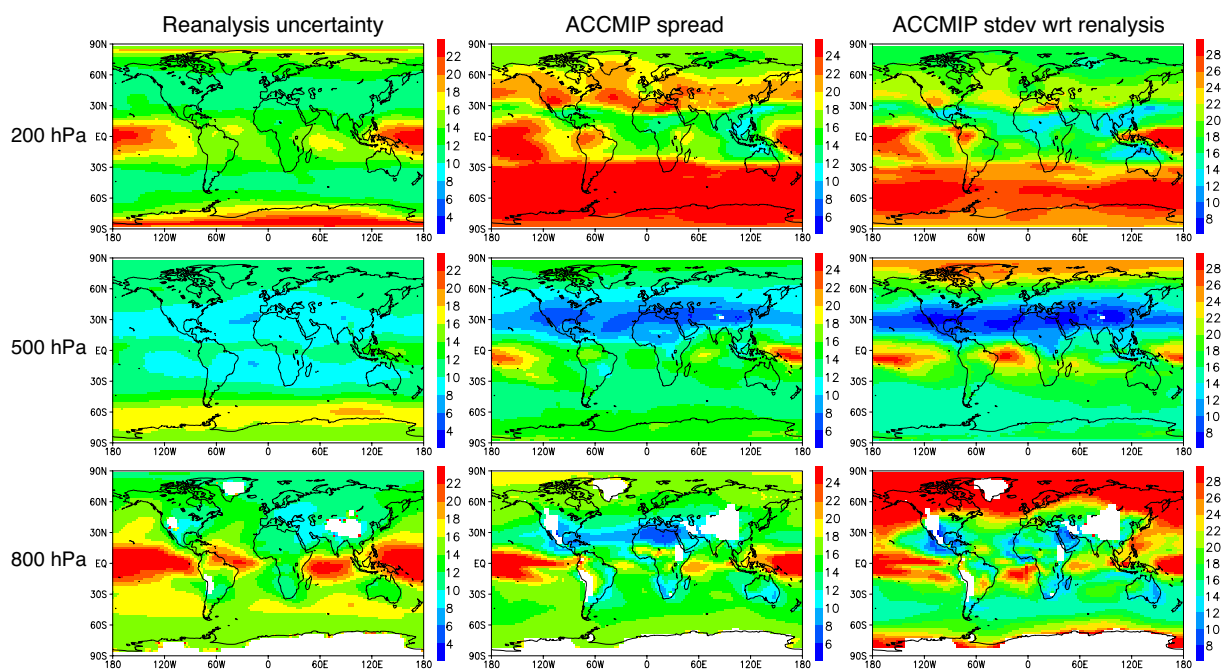
Figure 11. Global distributions of relative value (in %) of reanalysis uncertainty (left), standard deviation among the ACCMIP models (center), and ACCMIP model standard deviation with respect to the reanalysis for the annual mean concentration (right). From top to bottom, results are shown for global distributions at 200 hPa, 500 hPa, and 800 hPa.