Atmospheric
Chemistry
and Physics

Open Access

Radiative forcing in the ACCMIP historical and future climate simulations

D. T. Shindell[1], **J.-F. Lamarque**[2], **M. Schulz**[3], **M. Flanner**[4], **C. Jiao**[4], **M. Chin**[5], **P. J. Young**[6,*], **Y. H. Lee**[1], **L. Rotstayn**[7], **N. Mahowald**[8], **G. Milly**[1], **G. Faluvegi**[1], **Y. Balkanski**[9], **W. J. Collins**[10,**], **A. J. Conley**[2], **S. Dalsoren**[11], **R. Easter**[12], **S. Ghan**[12], **L. Horowitz**[13], **X. Liu**[12], **G. Myhre**[11], **T. Nagashima**[14], **V. Naik**[15], **S. T. Rumbold**[10], **R. Skeie**[11], **K. Sudo**[16], **S. Szopa**[9], **T. Takemura**[17], **A. Voulgarakis**[1,18], **J.-H. Yoon**[12], and **F. Lo**[8]

[1]NASA Goddard Institute for Space Studies and Columbia Earth Institute, New York, NY, USA
[2]National Center for Atmospheric Research (NCAR), Boulder, CO, USA
[3]Meteorologisk Institutt, Oslo, Norway
[4]Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI, USA
[5]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[6]Cooperative Institute for Research in Environmental Sciences, University of Colorado and NOAA Earth System Research Laboratory, Boulder, Colorado, USA
[7]Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, Aspendale, Vic, Australia
[8]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA
[9]Laboratoire des Sciences du Climat et de l'Environnement LSCE-IPSL, Gif-sur-Yvette, France
[10]Met Office, Hadley Centre, Exeter, UK
[11]Center for International Climate and Environmental Research Oslo (CICERO), Oslo, Norway
[12]Pacific Northwest National Laboratory, Richland, WA, USA
[13]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
[14]National Institute for Environmental Studies, Tsukuba-shi, Ibaraki, Japan
[15]UCAR/NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
[16]Department of Earth and Environmental Science, Graduate School of Environmental Studies, Nagoya University, Nagoya, Aichi, Japan
[17]Research Institute for Applied Mechanics, Kyushu University, Fukuoka, Japan
[18]Department of Physics, Imperial College London, London, UK
[*]now at: Lancaster Environment Centre, Lancaster University, Lancaster, UK
[**]now at: Department of Meteorology, University of Reading, Reading, UK

Correspondence to: D. T. Shindell (drew.t.shindell@nasa.gov)

Abstract. The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) examined the short-lived drivers of climate change in current climate models. Here we evaluate the 10 ACCMIP models that included aerosols, 8 of which also participated in the Coupled Model Intercomparison Project phase 5 (CMIP5).

The models reproduce present-day total aerosol optical depth (AOD) relatively well, though many are biased low. Contributions from individual aerosol components are quite different, however, and most models underestimate east Asian AOD. The models capture most 1980–2000 AOD trends well, but underpredict increases over the Yellow/Eastern Sea. They strongly underestimate absorbing AOD in many regions.

We examine both the direct radiative forcing (RF) and the forcing including rapid adjustments (effective radiative forcing; ERF, including direct and indirect effects). The models' all-sky 1850 to 2000 global mean annual average

total aerosol RF is (mean; range) $-0.26 \, \mathrm{W \, m^{-2}}$; -0.06 to $-0.49 \, \mathrm{W \, m^{-2}}$. Screening based on model skill in capturing observed AOD yields a best estimate of $-0.42 \, \mathrm{W \, m^{-2}}$; -0.33 to $-0.50 \, \mathrm{W \, m^{-2}}$, including adjustment for missing aerosol components in some models. Many ACCMIP and CMIP5 models appear to produce substantially smaller aerosol RF than this best estimate. Climate feedbacks contribute substantially (35 to $-58\,\%$) to modeled historical aerosol RF. The 1850 to 2000 aerosol ERF is $-1.17 \, \mathrm{W \, m^{-2}}$; -0.71 to $-1.44 \, \mathrm{W \, m^{-2}}$. Thus adjustments, including clouds, typically cause greater forcing than direct RF. Despite this, the multi-model spread relative to the mean is typically the same for ERF as it is for RF, or even smaller, over areas with substantial forcing. The largest 1850 to 2000 negative aerosol RF and ERF values are over and near Europe, south and east Asia and North America. ERF, however, is positive over the Sahara, the Karakoram, high Southern latitudes and especially the Arctic.

Global aerosol RF peaks in most models around 1980, declining thereafter with only weak sensitivity to the Representative Concentration Pathway (RCP). One model, however, projects approximately stable RF levels, while two show increasingly negative RF due to nitrate (not included in most models). Aerosol ERF, in contrast, becomes more negative during 1980 to 2000. During this period, increased Asian emissions appear to have a larger impact on aerosol ERF than European and North American decreases due to their being upwind of the large, relatively pristine Pacific Ocean. There is no clear relationship between historical aerosol ERF and climate sensitivity in the CMIP5 subset of ACCMIP models. In the ACCMIP/CMIP5 models, historical aerosol ERF of about -0.8 to $-1.5 \, \mathrm{W \, m^{-2}}$ is most consistent with observed historical warming. Aerosol ERF masks a large portion of greenhouse forcing during the late 20th and early 21st century at the global scale. Regionally, aerosol ERF is so large that net forcing is negative over most industrialized and biomass burning regions through 1980, but remains strongly negative only over east and southeast Asia by 2000. Net forcing is strongly positive by 1980 over most deserts, the Arctic, Australia, and most tropical oceans. Both the magnitude of and area covered by positive forcing expand steadily thereafter.

1 Introduction

While well-mixed greenhouse gases (WMGHGs) are the largest single driver of climate change since the preindustrial, aerosols and ozone are also important contributors. Best estimates in the IPCC AR4 of 1750 to 2006 radiative forcing were $2.63 \, \mathrm{W \, m^{-2}}$ from changes in WMGHGs, $0.35 \, \mathrm{W \, m^{-2}}$ for tropospheric ozone, $-0.42 \, \mathrm{W \, m^{-2}}$ for aerosol direct effects and $-0.70 \, \mathrm{W \, m^{-2}}$ for aerosol-cloud albedo effects (Forster et al., 2007). Aerosols and ozone are also distributed

unevenly, and their distinct radiative forcing patterns contribute to the regional pattern of climate change.

Despite their importance, climate model intercomparisons have traditionally neglected to document these agents. For example, simulations performed for the Climate Model Intercomparison Project (CMIP) phase 3 in support of the IPCC AR4 provided valuable insight into climate sensitivity, historical climate and climate projections, but did not examine aerosol or ozone forcing. This is important since the radiative forcings imposed in the simulations differed from model to model due to varying assumptions about emissions, differences in the behavior of physical processes affecting short-lived species, and differences in which processes and constituents were included at all (e.g. only 8 of 23 CMIP3 models included black carbon (BC) while less than half included future tropospheric ozone changes). Hence it is not straightforward to understand the relative importance of variations in climate sensitivity versus differences in the forcings.

The CMIP5 project (Taylor et al., 2012) compares historical and future climate simulations from coupled ocean-atmosphere models, but similarly provides little information on aerosol or ozone forcing. Hence there is a need for characterization of the forcings imposed in the CMIP5 historical and future simulations, for diagnostics to allow us to understand the causes of the differences in forcings from model to model, and for evaluation of the underlying simulated aerosols and ozone. ACCMIP attempts to meet these needs through a set of coordinated simulations primarily using the same composition models used by the CMIP5 groups (Lamarque et al., 2013). Here we describe the ACCMIP models' aerosol simulations, evaluate the simulated optical properties against observations, and present the resulting forcing. We examine both the conventional direct radiative forcing at the tropopause (RF) and the forcing including rapid adjustments (effective radiative forcing; ERF, including direct and indirect effects). We then combine the aerosol forcings with ACCMIP analyses of ozone and WMGHG forcing to estimate the total anthropogenic forcing through time.

2 ACCMIP model descriptions and experimental design

While intended primarily to examine the anthropogenic drivers of climate change in CMIP5, ACCMIP was open to the wider modeling community, and several groups participated that were not in CMIP5. We include both types of models, providing analyses of the CMIP5 subset of AC-CMIP models when appropriate. As this study focuses on aerosols, we do not include models that provided only gas-phase diagnostics. Key information about the aerosols included in the models and the experiments performed is presented in Table 1. We note that the GISS-E2-R-TOMAS and NCAR-CAM5.1 models include representations of aerosol

Table 1. Aerosol components included and simulations performed by the ACCMIP models.

| Model | Aerosol components included (d=dust, n=nitrate, su=sulfate, ss=sea-salt; underlined indicates included in AOD but not in forcing) | Aerosol-cloud interactions (L=liquid phase, I=ice or mixed phase; indicated aerosols as CCN) | Used in O₃ RF | Analogous CMIP5 runs | Historical | | | | RCP2.6 | | RCP4.5/6.0 | RCP8.5 | | Primary references |
					1850	1930	1980	2000	2030	2100	2010	2030	2100	
CICERO-OsloCTM2	BC, OA, d, n, su, soa, ss	None		None	♦	♦	♦	♦	♦	♦	♦	♦	♦	Skeie et al. (2011)
CSIRO-Mk3.6	BC, OA, d, su, soa, ss	L (all)	♦	CSIRO-Mk3.6	♦	♦	♦	♦					♦	Rotstayn et al. (2012)
GFDL-AM3	BC, OA, d, su, ss	L (su, ss, OA)	♦	GFDL-CM3	♦		♦	♦	♦	♦	♦	♦	♦	Donner et al. (2011);
GISS-E2-R	BC, OA, d, n, su, soa, ss	L (all)	♦	GISS-E2-R	♦	♦	♦	♦	♦	♦	♦	♦	♦	Shindell et al. (2013);
GISS-E2-R-TOMAS	BC, OA, d, su, ss	L (all)		None	♦	♦	♦	♦						Lee and Adams (2010)
HadGEM2	BC, OA, d, n[1], su, soa, ss	L (OA, n, su, soa, ss)	♦	HadGEM2-ES	♦		♦	♦		♦			♦	Collins et al. (2011)
LMDzORINCA	BC, OA, d, su, ss	L (all)	♦	IPSL-CM5A-LR	♦	♦	♦	♦	♦	♦	♦	♦	♦	Szopa et al. (2012)
MIROC-CHEM	BC, OA, d, n[2], su, soa[2], ss	L (all); I (BC, d)	♦	MIROC-ESM-CHEM	♦	♦	♦	♦	♦	♦	♦	♦	♦	Watanabe et al. (2011)
NCAR-CAM3.5	BC, OA, d, su, ss	None	♦	NCAR-CCSM4	♦	♦	♦	♦	♦	♦		♦	♦	Lamarque et al. (2012)
NCAR-CAM5.1	BC, OA, d, su, soa, ss	L (all); I (su, d)	♦	CESM1-CAM5.1-FV2	♦	♦	♦	♦						Liu et al. (2012); Ghan et al. (2013)

[1] HadGEM2 simulations for ACCMIP (and CMIP5) did not include dust or nitrate forcing, but nitrate was calculated in Bellouin et al. (2011), and those results have been included here when available (AOD and forcing).
[2] MIROC-CHEM nitrate and SOA were calculated, but not used for their CMIP5 simulations. AOD diagnostics for these two components were not available, but forcings were.

sizes, while all other models use a bulk approach in which size distributions are prescribed and only aerosol masses are computed. More detailed information on each model can be found in the ACCMIP overview paper of Lamarque et al. (2013).

All models used time-varying anthropogenic and biomass burning emissions of aerosol and tropospheric ozone precursors from Lamarque et al. (2010) for the historical period. For 2005 to 2100, anthropogenic and biomass burning emissions were created by four separate integrated assessment modeling (IAM) groups. The scenarios are called "representative concentration pathways" (RCPs) and are named by their nominal 2100 forcing relative to 1750: RCP2.6, 4.5, 6.0 and 8.5 (van Vuuren et al., 2011). In all the RCPs, aerosol-related anthropogenic and biomass burning emissions generally fall markedly in the distant future, with global mean 2100 emissions roughly 80 % lower than present-day for SO_2, 50 % for BC, and from 10–40 % for organic aerosol (OA). Ammonia emissions, however, increase by 10–80 % at 2100. Emissions were modified in two models, with GISS-E2-R scaling the biomass burning emissions of BC and OA by 1.4, and CSIRO-Mk3.6 scaling all BC by 1.25 and all OA by 1.5. Natural emissions varied across the models, and in many models also varied as climate changed. Further information on emissions in each model can be found in Lamarque et al. (2013). Concentrations of well-mixed greenhouse gases (CO_2, CH_4, N_2O and halocarbons) were prescribed according to the RCP projections that are from the reduced-complexity coupled carbon cycle climate model MAGICC6.3 which estimated mixing ratios based on the IAM RCP emissions (Meinshausen et al., 2011).

ACCMIP simulations were typically performed as timeslices for several historical and future times (Table 1). These used emissions from the given year, prescribed "climate" (sea surface temperatures (SSTs), sea-ice and WMGHGs) for that same time from companion CMIP5 simulations, and free-running atmospheric models. For the GISS-E2-R model, however, the CMIP5 runs included interactive chemistry and aerosols and hence the ACCMIP diagnostics were saved directly from the CMIP5 transient simulations. Similarly, the LMDzORINCA model archived many diagnostics from transient simulations (Szopa et al., 2012). Values are averaged over the available years of the timeslice (generally 5–10 yr) or the nearest 11 yr for the transients, and are area-weighted for global or regional means.

All models included changes in the ocean, and hence meteorology/climate, except NCAR-CAM5.1, CICERO-OsloCTM2 and CSIRO-Mk3.6. These simulations include not only the impacts of climate change on aerosols via processes such as altered wet or dry removal or oxidation, but also changes in emissions of dust and sea-salt aerosols in many models (Table 1). To separate these, additional simulations with fixed climate but altered emissions, or vice-versa, were performed. Climate was maintained at 1850 while later anthropogenic and biomass burning emissions were used, for example. We include climate change-induced aerosol forcing in our estimates when they were included in the groups' CMIP5 simulations.

3 Evaluation of present-day aerosols

Recent data is the most comprehensive and highest quality, so we first evaluate the present-day aerosol climatology using the ACCMIP 2000 timeslice. RF is the end result of a path from emissions to concentrations to aerosol optical properties to forcing. As RF is not directly observed, we examine the earlier stages, primarily focusing on aerosol optical depth (AOD) as the nearest observed quantity to RF. Analysis of ACCMIP BC surface concentrations is presented elsewhere (Lee et al., 2013).

Table 2. Monthly mean AOD (550 nm) compared with observations.

Model	R vs. MODIS	R vs. MISR	R vs. SurfObs	Bias vs. MODIS	Bias vs. MISR	Bias vs. SurfObs
CICERO-OsloCTM2	0.62	0.67	0.69	15	10	25
CSIRO-Mk3.6	0.64	0.66	N/A	13	7	N/A
CSIRO-Mk3.6-cs	0.62	0.64	N/A	21	16	N/A
GFDL-AM3	0.50	0.55	0.51	3	−3	15
GISS-E2-R	0.55	0.59	0.56	54	48	77
GISS-E2-R-cs	0.58	0.64	0.61	−11	−14	−8
GISS-E2-R-TOMAS*	0.55	0.62	0.56	−14	−16	19
HadGEM2	0.55	0.59	0.65	0	−4	−2
LMDzORINCA	0.54	0.56	0.54	−17	−20	−5
MIROC-CHEM	0.51	0.56	0.44	−28	−30	−22
NCAR-CAM3.5	0.60	0.64	0.58	−26	−29	−13
NCAR-CAM5.1	0.46	0.50	0.56	−16	−18	−22
10 Model mean	0.55	0.60	0.57	−7	−11	−1
Absolute biases				15	16	14

The suffix cs indicates clear-sky. Means use only the clear-sky versions of model diagnostics when available (and exclude all-sky from those models). SurfObs are the AeroNet surface observations. *GISS-E2-R-TOMAS results were virtually identical for all-sky and clear-sky. Biases are the normalized mean bias (in percent), and absolute biases are the average absolute value of the area-weighted biases.

3.1 Satellite AOD

We first compare the spatial distribution of annual mean 550 nm AOD in the models with observations from the MODIS (Remer et al., 2008) and MISR (Kahn et al., 2009) satellite instruments averaged over 2004–2006. We use the standard MODIS AOD product over most of the globe, but obtain greater coverage by infilling missing data with values produced by the Deep Blue algorithm that allows retrievals over bright surfaces such as deserts (Hsu et al., 2004). Differences between 2000 and 2004–2006 are expected to be generally small. Though not requested, separate all-sky and clear-sky AOD diagnostics were available from several models, and hence both were analyzed. We concentrate on clear-sky output when both are available since that is more comparable to observations.

The 10-model mean clearly captures many features seen in the satellite measurements (Fig. 1). High aerosol loadings over desert regions associated with mineral dust stand out in both models and observations, as does the band of locally enhanced AOD over the Southern Ocean associated with sea-salt. Areas with large anthropogenic aerosol emissions also have relatively high AOD, especially East Asia and the Indo-Gangetic plain. The multi-model mean underestimates both the magnitude and extent of AOD in these areas, however. Similarly, AOD in tropical South America, Africa and Indonesia, where biomass burning emissions are large, is underpredicted.

The multi-model mean captures the broadly lower AOD levels in Europe and North America relative to developing Asia. At finer scales, however, local features are sometimes poorly represented, for example over the Po Valley or the Mojave Desert. Over Australia, the multi-model mean is sim-

ilar to MISR, and does not show the extreme lows or highs seen by MODIS. The models underpredict AOD over Arctic land areas and the Southern Ocean.

We next quantify correlations and biases. All calculations use regridded $1° \times 1°$ monthly mean model fields sampled only when and where the satellite instruments report observations. Correlations between the models and the two satellite datasets are generally similar, but slightly higher for MISR (Table 2). Differences between all-sky and clear-sky are very small for the GISS-E2-R-TOMAS model, modest for CSIRO-Mk3.6, and quite large for GISS-E2-R (see Appendix A). We therefore use clear-sky data for GISS-E2-R and CSIRO-Mk3.6 hereafter. The models show a substantial diversity in their ability to capture the observed distribution and magnitude of AOD. Biases range between −30 % to +20 %, with most models being too low. Evaluating the mean AOD of ACCMIP models shows the same biases as the average of the individual models, but substantially higher correlations (0.69 versus MODIS, 0.75 versus MISR) than either the average or any individual model. An enhanced performance of the multi-model mean relative to any single model has also been seen in analyses of climate parameters (e.g. Reichler and Kim, 2008).

A portion of the negative bias in many models is due to missing nitrate and SOA. Adding the mean AODs from the other models to adjust for these missing components improves model agreement with observed global mean total AOD in most cases, though not always (Fig. 2).

Though global mean total AOD biases tend to be ~ 0.03 or less (within 20 % of observations), models find a realistic total AOD with a very diverse partitioning among components (Fig. 3). Both the absolute AOD and the fraction of the total contributed by primarily natural dust and sea-salt vary
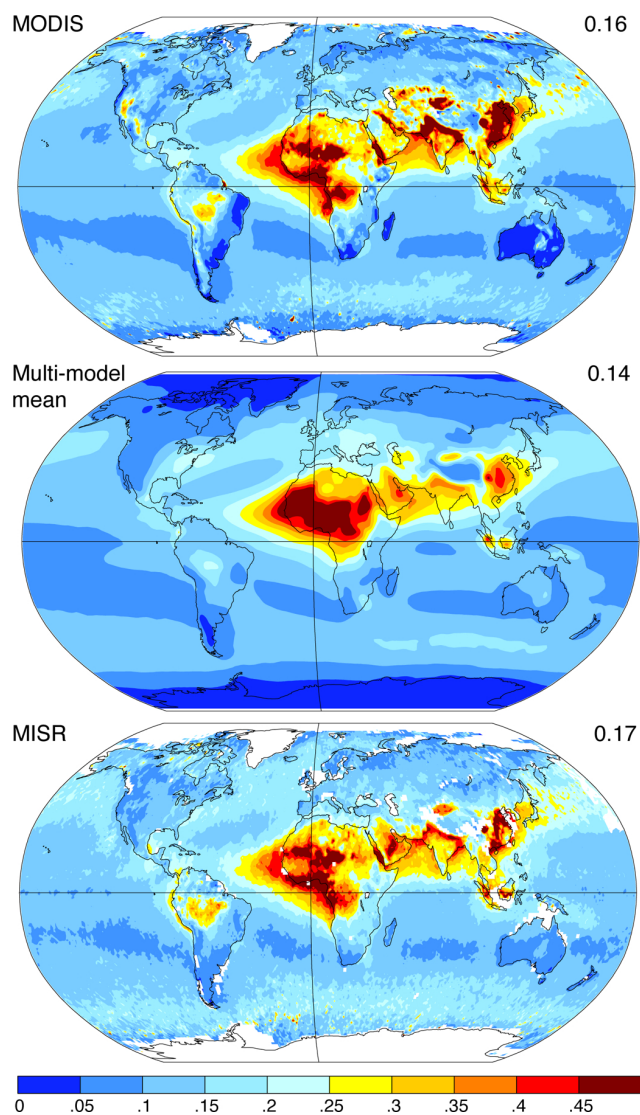
Fig. 1. Annual average aerosol optical depth (550 nm) seen by the MODIS (top) and MISR (bottom) instruments and the ACCMIP multi-model mean (10 models) for the simulations of conditions representative of ∼ 2000 (center). Global mean values are given in the upper right corner of each map. White indicates no data in the satellite measurements. Clear-sky model values are used when available (see Table 2).

Fig. 2. Global mean area-weighted annual average AOD (550 nm) seen by the MODIS and MISR instruments and in the indicated models sampling when and where MODIS and MISR provide observations. Solid bars show model total AOD values while light shaded areas show the multi-model mean contribution from nitrate or nitrate and SOA in those models that do not include those species.
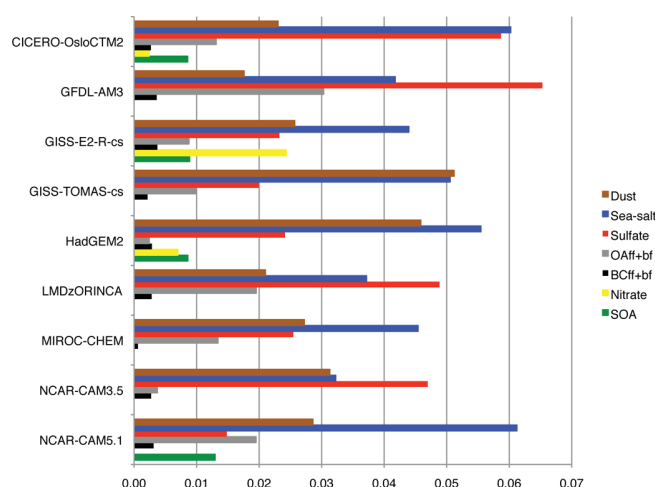

Fig. 3. Global mean annual average AOD (550 nm) by component. Additional AOD of 0.06 is classified as aerosol water in CAM5.1. Clear-sky AOD is used for the GISS models.

by more than a factor of 2 across models. Primarily anthropogenic aerosols show similarly large variations, with individual component AODs varying by as much as a factor of 4 across models despite nearly identical sources.

Satellite observations cannot readily distinguish between different aerosol types. Instead, we compare monthly MODIS fine-mode fraction data (Remer et al., 2005) with the modeled sum of all aerosols except sea-salt and dust, as these are predominantly large particles. Though some sea-salt and dust particles are small and are included in the MODIS fine mode product, there is considerable uncertainty

associated with separating out such small particles (Bellouin et al., 2008; Yu et al., 2009). Models should therefore be biased low. Over the oceans, this is the case for all but three models, though the magnitude varies greatly (Table 3). Spatial and temporal correlations are fairly low, however. For the entire globe, half the models show positive biases, and correlations are again relatively poor. However, we place greater value on the comparison over ocean regions, where MODIS data are more reliable.

To further evaluate individual aerosol components, we compare observed and modeled annual AOD in locations within the top decile (10 %) of component mass density in each model (Lee and Adams, 2010) (Table 4). Note that locations vary from model to model. Where sulfate mass density maximizes, AOD is positively biased in CICERO-OsloCTM2 and GFDL-AM3, and negatively biased in MIROC-CHEM. Some models have comparable magnitude

Table 3. Comparison of models with MODIS fine mode AOD over all areas and over ocean areas.

Model	R	Bias	R ocean	Bias ocean
CICERO-OsloCTM2	0.45	31	0.56	2
GFDL-AM3	0.42	57	0.35	23
GISS-E2-R-cs	0.47	3	0.42	−15
GISS-E2-R-TOMAS	0.47	−51	0.45	−64
HadGEM2	0.38	−41	0.44	−55
LMDzORINCA	0.36	7	0.50	−17
MIROC-CHEM	0.29	−60	0.30	−65
NCAR-CAM3.5	0.47	−19	0.48	−31
NCAR-CAM5.1	0.35	62	0.22	53

biases against one satellite dataset, but little bias against the other (GISS-E2-R-TOMAS, NCAR-CAM5.1). Examining global mean sulfate AOD, indeed CICERO-OsloCTM2 and GFDL-AM3 have values substantially higher than other models, but the other models' totals do not correspond closely to their biases in sulfate-rich areas. In particular, the GISS-E2-R-TOMAS model is among the lowest for global mean sulfate AOD despite its positive biases in sulfate-rich areas versus MISR, while MIROC-CHEM is mid-range in global mean despite it's negative bias.

In OA-rich locations, several models appear to have too little AOD while GFDL-AM3 appears to have too much. These biases have no clear relationship to global mean OA AOD values other than GFDL-AM3 having the largest OA AOD. Dust and sea-salt are primarily naturally occurring aerosols, and thus their absolute amount tells us little about their contribution to forcing. In areas with largest dust loading, the GISS-E2-R-TOMAS model overestimates AOD (and has the highest global mean dust AOD), while MIROC-CHEM underestimates in comparison with MISR. The latter may be part of the reason that model has the lowest global mean AOD of any of the models in ACCMIP (though it is only marginally lower than some others). Note that differences between the satellite datasets are particularly large for dusty areas. The AOD in sea-salt rich regions tends to be fairly well simulated, though there are moderate biases in some models. Sea-salt biases do correspond fairly well with global mean sea-salt AOD values, with models that have a global mean sea-salt AOD in the 0.043 to 0.056 range agreeing best with measurements.

3.2 AeroNet AOD

The most quantitatively reliable large-scale network of AOD observations are AeroNet sun photometers. We compare the 2000 simulations with an AeroNet climatology spanning 2000–2009 based on measurements from 388 stations located below 1000 m in altitude. Coverage is largely limited to continental areas (other than a few island stations) and is quite sparse in some regions. Comparisons are again made

only in months with data available (roughly 10 months per year on average).

Correlations reflect the models' ability to capture both the spatial pattern of AOD and its seasonal cycle. Values range from 0.44 to 0.69 for the ACCMIP models (Table 2). For comparison, correlations in the AeroCom phase I models ranged from 0.29 to 0.77, with 6 of the 7 model having values between 0.52 and 0.77 (Textor et al., 2006). The AeroCom phase II models report correlations from 0.26 to 0.78, with 8 of the 10 models between 0.60 and 0.78 (M. Schulz, personal communication, 2012). In the ACCMIP set, 7 of the 9 models have correlations between 0.54 and 0.69. The slightly lower correlations in the ACCMIP models could reflect the use of meteorological reanalyses in AeroCom versus free-running models in ACCMIP, the use of less sophisticated aerosol physics in some ACCMIP/CMIP5 models, the use of daily data in AeroCom as opposed to monthly mean in ACCMIP, or a combination of these factors. Biases in the models tend to be smaller with respect to AeroNet than in comparison with satellite datasets, perhaps indicating that primarily anthropogenic aerosols over land are better simulated than aerosols over remote oceans. Overall, evaluation against ground-based and satellite datasets is fairly consistent, however.

Separating the analysis into temporal and spatial components, we find that the models all capture the monthly variations of AOD in every continental-sized region fairly well, with nearly all correlations between 0.6 and 0.8. There is no relationship between the quality of the seasonality in the models and the quality of the spatial structure. Hence we focus the remainder of the analysis on the spatial pattern, which differs more strongly across models, comparing the annual mean values in the model against observations.

Looking regionally, we find that the spatial correlations in the models are highest for North Africa (0.46–0.80), with North America second (0.48–0.67 in all but one model). Spatial correlations vary widely across models for East Asia (0.09–0.57), but are consistently low over Europe (0.31–0.45 in all but one model). Biases in the models tend to be comparatively small over North Africa, Europe and North America (Fig. 4). Nearly all models show large negative biases over East Asia, however. The two models that do not show a large negative bias over East Asia show the largest positive biases over both Europe and North America, indicating they are systematically higher than the other models rather than matching East Asia observations better due to regional differences. Root-mean-square (RMS) differences relative to AeroNet are also typically fairly small for Europe and North America, and substantially larger for East Asia. This suggests that the models have the greatest difficulty in capturing AOD over East Asia (of the regions analyzed), though the spatial pattern over Europe is also problematic (although aerosol loading is lower, so biases and RMS errors remain relatively small). For areas with more limited AeroNet coverage, biases over remote island stations in the central Pacific and

Table 4. Evaluation of model AOD in comparison with satellite observations sorted by locations dominated by sulfate, OA, dust and sea-salt.

Top decile of sulfate mass density							
	Versus MISR			Versus MODIS			Global mean AOD
Model	bias	lmnb	lmne	bias	lmnb	lmne	
CICERO-OsloCTM2	0.09	0.14	0.15	0.07	0.12	0.16	0.058
GFDL-AM3	0.05	0.10	0.15	0.06	0.11	0.17	0.065
GISS-E2-R-cs	−0.01	−0.02	0.11	−0.04	−0.05	0.14	0.023
GISS-E2-R-TOMAS-cs	0.05	−0.02	0.18	0.01	−0.08	0.20	0.020
HadGEM2	0.00	−0.01	0.11	−0.01	−0.01	0.12	0.024
MIROC-CHEM	−0.06	−0.12	0.18	−0.07	−0.14	0.20	0.026
NCAR-CAM3.5	−0.02	−0.03	0.11	−0.04	−0.04	0.13	0.047
NCAR-CAM5.1	0.00	−0.04	0.16	−0.06	−0.12	0.20	0.014

Top decile of OA mass density							
Model	bias	lmnb	lmne	bias	lmnb	lmne	AOD
CICERO-OsloCTM2	0.03	0.04	0.11	0.03	0.06	0.13	0.013
GFDL-AM3	0.05	0.07	0.12	0.05	0.09	0.14	0.031
GISS-E2-R-cs	−0.05	−0.11	0.16	−0.05	−0.09	0.18	0.009
GISS-E2-R-TOMAS-cs	0.00	−0.17	0.29	0.00	−0.15	0.28	0.010
HadGEM2	0.02	0.00	0.12	0.01	−0.01	0.14	0.002
MIROC-CHEM	−0.12	−0.28	0.29	−0.12	−0.26	0.28	0.014
NCAR-CAM3.5	−0.10	−0.23	0.26	−0.10	−0.23	0.25	0.004
NCAR-CAM5.1	−0.07	−0.17	0.22	−0.08	−0.16	0.24	0.019

Top decile of dust mass density							
Model	bias	lmnb	lmne	bias	lmnb	lmne	AOD
GFDL-AM3	−0.02	−0.03	0.10	0.13	0.29	0.31	0.014
GISS-E2-R-cs	−0.01	−0.02	0.12	0.15	0.29	0.31	0.018
GISS-E2-R-TOMAS-cs	0.13	0.10	0.16	0.28	0.42	0.44	0.039
MIROC-CHEM	−0.05	−0.10	0.13	0.08	0.19	0.24	0.021
NCAR-CAM3.5	−0.03	−0.05	0.11	0.12	0.27	0.29	0.026
NCAR-CAM5.1	0.03	0.03	0.15	0.18	0.33	0.35	0.018

Top decile of sea-salt mass density							
Model	bias	lmnb	lmne	bias	lmnb	lmne	AOD
CICERO-OsloCTM2	0.04	0.09	0.10	0.05	0.13	0.13	0.061
GFDL-AM3	−0.02	−0.05	0.06	0.00	0.00	0.04	0.043
GISS-E2-R-cs	−0.02	−0.06	0.07	0.00	−0.01	0.05	0.045
GISS-E2-R-TOMAS-cs	−0.01	−0.04	0.06	0.00	0.00	0.04	0.051
HadGEM2	−0.03	−0.08	0.10	−0.01	−0.03	0.06	0.056
MIROC-CHEM	0.01	0.03	0.06	0.03	0.07	0.08	0.046
NCAR-CAM3.5	−0.04	−0.14	0.14	−0.03	−0.12	0.12	0.033
NCAR-CAM5.1	0.05	0.13	0.18	0.05	0.14	0.18	0.062

lmnb is log-mean normalized bias (the sum of the log (base 10) of the modeled to observed AOD ratio at each point divided by the number of points), lmne is log-mean normalized error (same except except using the absolute value of the ratio). Global mean values for the given component are for AOD where MODIS data is available (values where MISR is available are within 0.001 for sulfate, OA and sea-salt, but can differ substantially for dust).

over Southern Hemisphere mid-latitude stations are nearly always positive, while biases over the Russian and especially the Western Hemisphere Arctic are typically negative (see Appendix B).

As with the screening versus satellite AOD based on the dominant mass component, we can use the spatial compari-son against AeroNet to constrain the dominant aerosol com-ponent in some locations. Examining the zonal mean AOD by component (Fig. 5), we see that the largest contributor to AOD in much of the NH mid-latitudes is sulfate in all models except NCAR-CAM5.1. Several of the models over-predict the total AOD from about 40–60° N, with a strong
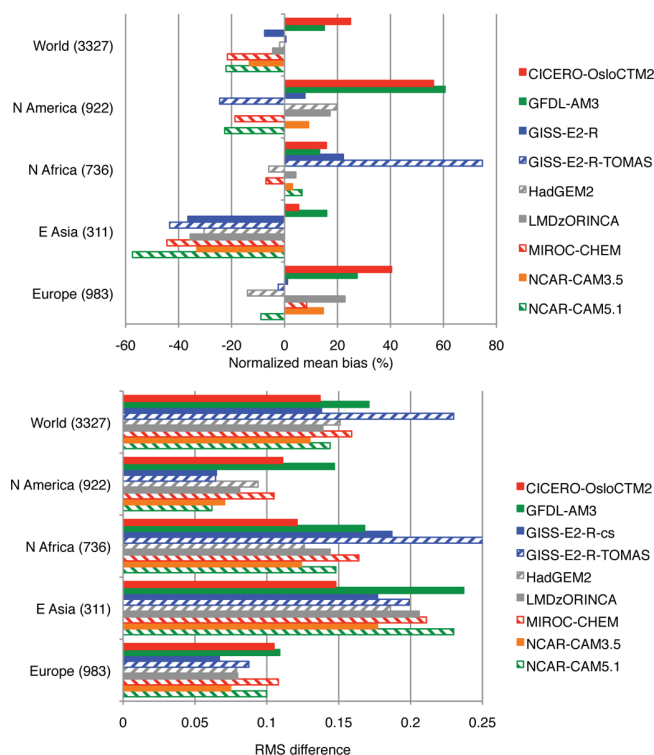
Fig. 4. Regional normalized mean biases (%; top) and root-mean-square differences (bottom) in AOD compared with AeroNet. The value in parentheses after each region's name is the number of valid monthly observations within the region. The GISS-E2-R-TOMAS RMS value for North Africa goes off the scale to 0.41.

suggestion that this is due to too much sulfate AOD. This is especially the case for the CICERO-OsloCTM2 and GFDL-AM3 models, consistent with the analysis against satellite AOD in sulfate-rich areas. Further towards the equator, dust becomes a large contributor to AOD with a peak around 20° N in most models. Dust is by far the largest component of AOD in several models, but not in others, and there is no clear relationship between the models' skill in capturing observed total AOD and the magnitude of the dust AOD. The GISS-E2-R-TOMAS model's dust AOD is larger than the total observed AOD around 15° N, indicating that there is too much dust in that model and accounting for the positive bias in North African total AOD (Fig. 4; and consistent with the screening by dust mass against satellite data in Table 4, though that model actually has the highest spatial correlation with AeroNet in that region), but in other cases the zonal mean AOD does not constrain the dust loading.

At low Southern latitudes there is a peak in AOD from OA in the two GISS models, LMDzORINCA, and CICERO-OsloCTM2, which show low OA AOD elsewhere, while the values are very, very low at all latitudes in HadGEM2, NCAR-CAM3.5 and MIROC-CHEM. In contrast, organic AOD is high in NCAR-CAM5.1 across a broad area of the tropics and even into the extratropics in both hemispheres,

and is also broadly distributed in GFDL-AM3 though with a smaller magnitude than in NCAR-CAM5.1. Hence as in the case of dust (and partly due to the overlap in dust and OA at low latitudes), the total AOD does not provide a tight constraint on the contribution from OA. We note, however, that OA is the only component that typically exhibits a local peak AOD at 20° S, where the observations also show a local maximum, suggesting that the strong influence of OA on AOD at this latitude in some models is likely realistic (sulfate and dust, in contrast, tend to have a local minimum at this latitude). The sea-salt AOD peak seen in most models from 50–60° S appears to be typically too large, so that nearly all models overpredicts total AOD at the edge of the AeroNet record. Data availability becomes very limited at these latitudes, however, so this result should be treated with caution.

3.3 Satellite and AeroNet AAOD

Though both contribute to AOD, scattering aerosols exert a negative RF, while absorbing aerosols can cause positive RF. We therefore also analyzed the absorbing aerosol optical depth (AAOD). Note that there appeared to be a problem with the HadGEM2 AAOD diagnostic, so it was excluded from this analysis.

We first compared the models' 2000 timeslice with satellite observations from the Ozone Monitoring Instrument (OMI) (Torres et al., 2007) averaged over 2005–2007 (Fig. 6). The multi-model mean shows maximum values over Africa, Arabia, South and East Asia, and tropical South America. OMI measurements show a similar pattern, but values are generally much larger than in the models in areas with substantial AAOD. Model values are especially low over tropical South America, the Persian Gulf, and much of South and Southeast Asia. The observations also suggest substantial transport over the ocean from Southern Hemisphere continents, which is not captured in the models. This might indicate that absorbing aerosol lifetimes are too short in the models, though distributions seem reasonable in the Northern Hemisphere. Comparison with remote in situ measurements indicates that most models overestimate BC (Schwarz et al., 2010), suggesting that the oceanic AAOD underestimates may be due to an underestimate of the Southern Hemisphere sources or lifetime of dust. Quantitatively, the models have fairly poor spatial correlations with OMI, and underestimate AAOD by roughly a factor of two (Table 5). The multi-model mean shows negative biases in AAOD in every region in comparison with OMI. Underestimates are particularly large in South America, South and Southeast Asia, East Asia and Southern Hemisphere Africa, where every model shows markedly less AAOD than the satellite measurements. OMI data indicates large Arctic AAOD values, but satellite AAOD retrievals are especially challenging over bright surfaces, so these observations may not be reliable. Correlations with OMI are substantially better excluding the Arctic (Table 5).
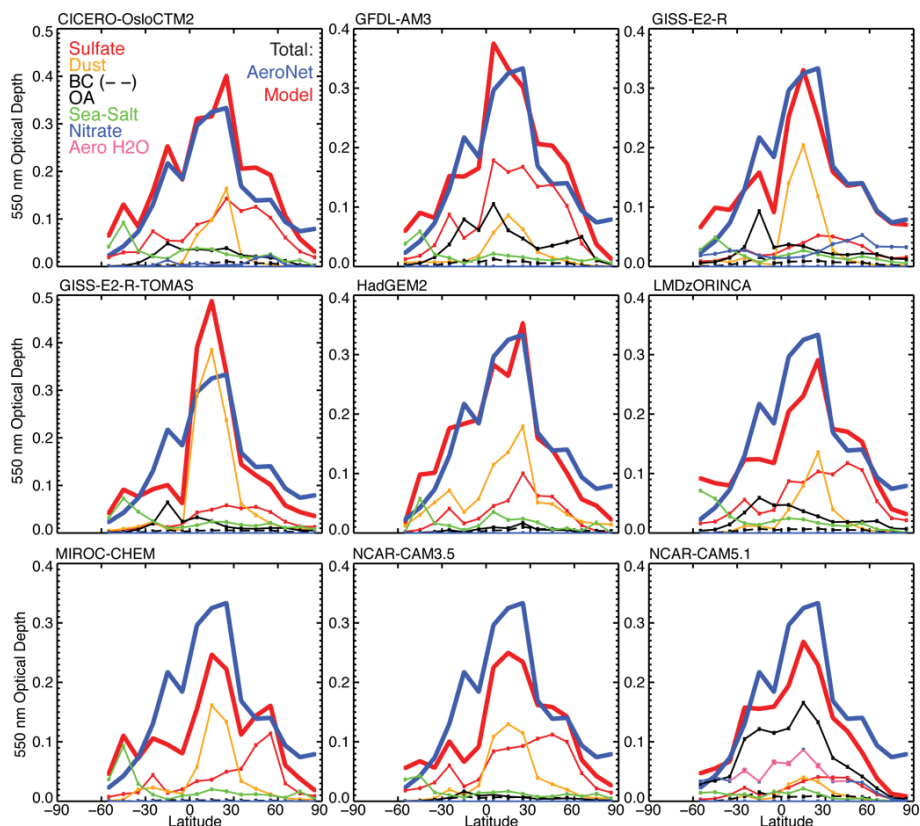
Fig. 5. Zonal mean annual average total AOD in AeroNet observations and models (thick lines) along with AOD by aerosol component (thin lines). Aero H_2O is aerosol water in NCAR-CAM5.1. Clear-sky AOD is used for the GISS-E2-R model.

Table 5. Correlation and bias of modeled AAOD with respect to OMI observations.

Model	R vs. OMI global	NMB vs. OMI global	R vs. OMI $90°$ S to $60°$ N	NMB vs. OMI $90°$ S to $60°$ N
CICERO-OsloCTM2	0.35	−55	0.60	−49
CSIRO-Mk3.6	0.35	−52	0.58	−47
GFDL-AM3	0.31	−29	0.57	−19
GISS-E2-R	0.39	−45	0.63	−41
GISS-E2-R-TOMAS	0.41	−72	0.64	−69
LMDzORINCA	0.31	−33	0.55	−25
MIROC-CHEM	0.31	−85	0.54	−83
NCAR-CAM3.5	0.29	−37	0.50	−28
NCAR-CAM5.1	0.30	−37	0.51	−29
Average 9 models	0.34	−50	0.57	−43

NMB is normalized mean bias (in percent).

It is difficult to measure AAOD accurately from space, and the OMI measurements may have substantial biases. In particular, the reported AAOD exceeds the total AOD in many locations. We therefore also compare with AeroNet observations (see Appendix C), which likewise shows that the models severely underestimate AAOD over South and Southeast Asia, South America, Southern Hemisphere Africa and East Asia. Since dust dominates the total BC+dust AOD over

South Asia and parts of East Asia, part of the low bias there could by due to underestimates of dust loading as well as BC. The low biases in modeled AAOD over South America and Southeast Asia occur in areas where BC's contribution to total AAOD is quite large, however. Though AAOD retrievals from ground-based lidar may themselves have biases (e.g. due to required removal of NO_2 absorption), it seems likely that the at least a portion of the modeled AAOD bias

Table 6. Comparison of regional trends in AOD (550 nm) from ∼ 1980 to ∼ 2000 with AVHRR observations (630 nm) over water.

Model	Europe	Eastern North America	Great Lakes	South and East Asia	Yellow/Eastern Sea
CICERO-OsloCTM2	−0.059	−0.031	−0.041	0.073	0.101
GFDL-AM3	−0.073	−0.029	−0.055	0.097	0.099
GISS-E2-R-cs	−0.015	−0.004	−0.005	0.035	0.053
HadGEM2	−0.048	−0.013	−0.020	0.037	0.030
LMDzORINCA	−0.040	−0.011	−0.031	0.048	0.065
NCAR-CAM3.5	−0.050	−0.027	−0.037	0.044	0.036
NCAR-CAM5.1	−0.027	−0.013	−0.018	0.012	0.005
7-model average	−0.045	−0.018	−0.030	0.049	0.056
Observations					
NOAA AVHRR	−0.042	−0.023	−0.047	0.039	0.104
GACP AVHRR	−0.045	−0.046	−0.053	−0.010	0.009
NCDC AVHRR-CDR	−0.057	−0.039	no data	0.007	0.011

Regions are defined as Europe (35–70° N; 20–40° E); eastern North America (25–50° N; 100–60° W); the Great Lakes (40–52° N; 76–92° W); South and East Asia (10–30° N; 70–110° E plus 20–40° N; 110–130° E); and the Yellow/Eastern Sea (30–42° N; 118–124° E). Values are shown for sampling models according to where the NOAA AVHRR product has observations. Sampling based on GACP locations yields values that are within 0.006 in all cases, and within 0.002 for eastern North America, South and East Asia and the Great Lakes.

versus the two datasets analyzed is real. The model biases could stem from underestimates in the emissions inventory, the absorption efficiency of BC, or both factors.

4 AOD trends

Long-term data on aerosol AOD is fairly limited. The most complete long-term satellite record is from the Advanced Very High Resolution Radiometer (AVHRR) and extends back to 1981. We compare the difference between the AC-CMIP 1980 and 2000 timeslices with differences between AVHRR over 1981–1987 (excluding March 1982–December 1984 to avoid the effects of the El Chichon volcanic eruption) and 1997–2003. We primarily analyze average monthly mean AVHRR data produced by the Global Aerosol Climatology Project (GACP) (Geogdzhayev et al., 2005) and by NOAA (Zhao et al., 2008) over oceans and lakes, where the retrievals are most reliable for trend analysis. Both AVHRR datasets show substantial decreases over most of the globe (Fig. 7). Decreasing trends in AOD are most pronounced around Europe, especially in the NOAA product, and off of eastern North America and over the Great Lakes (though GACP shows large decreases over most of the Northern Hemisphere). AOD also decreases off West Africa, though as interannual variability in dust emissions is large, our relatively short averaging periods may not be representative for long-term trends there.

The models show decreases in AOD over continental Europe and eastern North America that extend out over the nearby oceans. Quantitative comparisons were made by sampling the models where satellite data were reported for at least eight months during at least two years of both the 1981–1987 and 1997–2003 periods. Tests show that values in most

locations are extremely similar using thresholds of 9 or 10 months. The multi-model mean decreases near Europe and eastern North America are in good agreement with NOAA AVHRR, though North American trends are too small in comparison with GACP AVHRR (Table 6). A very recent reprocessing of the AVHRR dataset by the same NOAA group has produced the NCDC AVHRR-CDR dataset (Chan et al., 2013). This dataset has less coverage (i.e. no Great Lakes data), but at least nominally improves over the prior processing. This version shows more negative trends everywhere relative to the earlier NOAA analysis, and the models underpredict AOD decreases near both Europe and North America with respect to this analysis (Table 6).

The models show large increases in AOD in east Asia, south Asia and Indonesia between 1980 and 2000 (Fig. 7). Observed trends in nearby oceans show strong increases in the NOAA product, but only weak increases in the AVHRR-CDR product and decreases in GACP. It seems counterintuitive that AOD would decrease near rapidly developing Asian countries during this period, and thus we favor the NOAA analyses. The multi-model mean increases are similar to the older NOAA values averaged over a broad area of near-Asian oceans, but are substantially smaller than those seen the Yellow/Eastern Sea region where observed trends maximize (Table 6). The modeled increases, however, extend over a broader area of the Indian and western Pacific Oceans. Modeled trends are much too large near Asia relative to the new AVHRR-CDR product. It is difficult to reconcile modeled increases being far too large with the consistent underprediction of 2000 total AOD in south and east Asia (Fig. 1). Hence the older NOAA AVHRR trends may be the most plausible in this region. There are clearly substantial limitations to AVHRR data, however, as highlighted by the large differences between the three datasets, with many
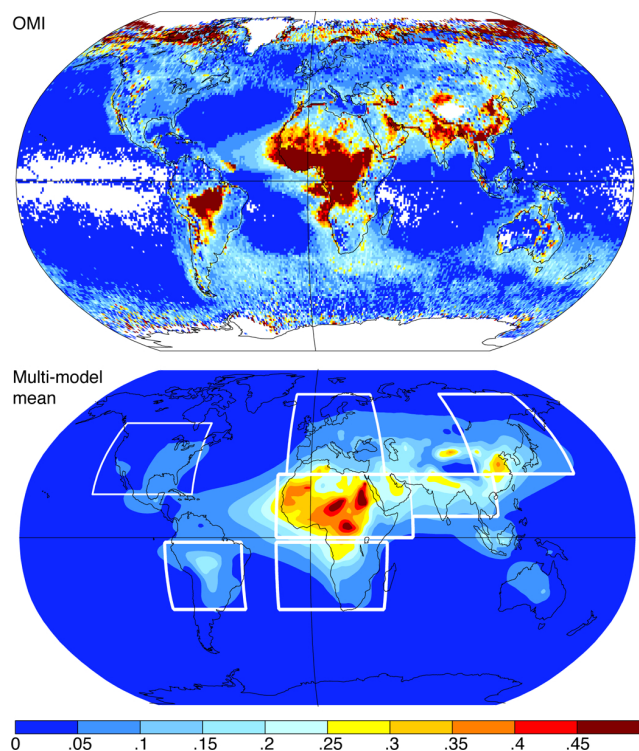
Fig. 6. Annual average AAOD in measurements from the OMI instrument (top) and in the ACCMIP multi-model mean (bottom). The multi-model mean is from 9 models, with HadGEM2 excluded. OMI retrieval is based on OMAERUVd.003 daily products from 2005–2007 that were obtained through and averaged using GIOVANNI (Acker and Leptoukh, 2007). White indicates no data in the satellite record. White borders in the lower panel show areas included in regional AAOD analyses.

assumptions about clouds and aerosols required to translate observed radiances into AOD (Li et al., 2009). The trend comparison suggests that in regions with relatively robust observed trends the modeled AOD trends are fairly reasonable, while near Asia trends are more difficult to constrain from observations but the model results seem plausible.

Near Europe and eastern North America, GISS-E2-R simulations show particularly small AOD trends. The difference relative to other models stems from large increases in nitrate aerosol AOD in that model in these areas. This suggests that nitrate increases in GISS-E2-R (and the resulting forcing) are likely too strong. Near Asia, the GFDL-AM3 and CICERO-OsloCTM2 models best match the AOD increases in the NOAA AVHRR analysis over the Yellow/Eastern Sea, but greatly overestimate trends over the broader South and East Asian coastal region where most other models perform well (Table 6), making it difficult to evaluate overall model skill.
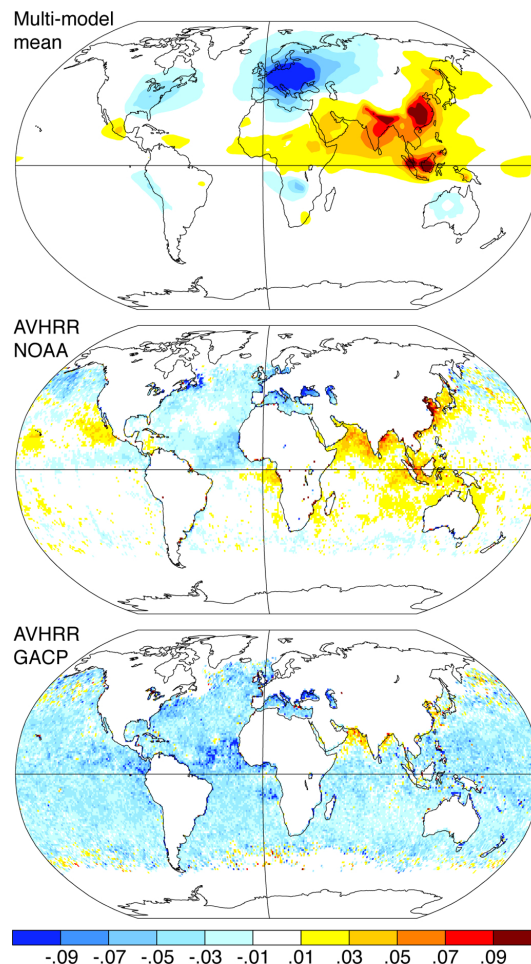
Fig. 7. Modeled and observed annual average AOD trends from $\sim$ 1980 to $\sim$ 2000. The multi-model mean (top) compares the 2000 and 1980 timeslice simulations. Observations are based on AVHRR data and show the difference between the 1997–2003 and 1981–1987 (excluding March 1982–December 1984) averages based on the NOAA product (center) and the GACP product (bottom).

5 Historical and future aerosol forcing

Radiative forcing is a useful metric for evaluating the contribution of a given factor to climate change over a particular time and for comparing the influence of multiple factors. RF is not a perfect indicator of the eventual global mean temperature response to a sustained forcing, but it is generally reasonably close. Notable exceptions occur with aerosols, however, especially for aerosol-cloud interactions and BC albedo forcing (Forster et al., 2007; Koch et al., 2011; Flanner et al., 2007).

RF is calculated from the difference in flux at the tropopause between a pair of radiative transfer calculations with reference (usually zero) aerosols in the first and actual aerosols in the second. Changes in this flux difference over time are the RF. This diagnoses the so-called "direct" aerosol

RF, but does not capture the various effects of aerosols on clouds. To diagnose those, we use additional simulations that isolated flux changes due to all aerosol effects (see Sect. 5.2). BC surface albedo forcings were calculated using a combination of the NCAR Community Land Model 4 (Lawrence et al., 2011) and models of snow and sea-ice interactions with aerosols (Flanner et al., 2007; Holland et al., 2012) that determined the forcing due to black carbon and dust deposition as reported by each ACCMIP model (see Appendix D).

5.1   Aerosol radiative forcing

5.1.1   Global mean preindustrial to present-day RF

We first examine total and component global mean annual average aerosol RF between 1850 and 2000 (Fig. 8; Table 7). As the sample size is sometimes small, we show individual model results rather than a "best estimate" and range. The mean and standard deviation of sulfate RF is $-0.40 \pm 0.13 \, \mathrm{W \, m^{-2}}$, and this range encompasses 6 of the 9 models. The only models exhibiting sulfate RF larger than this range are CICERO-OsloCTM2 and CSIRO-Mk3.6. The CICERO model showed the largest positive AOD biases in sulfate-rich regions in comparison with observations (Table 4; CSIRO did not provide speciated AOD data). The only model with sulfate RF smaller than this range is NCAR-CAM5.1. That model's AOD was biased low in sulfate regions, but did not stand out from the other models. We thus estimate a most probable range for sulfate RF of -0.18 to $-0.44 \, \mathrm{W \, m^{-2}}$ (i.e. the range of all models except CICERO and CSIRO). As biases in present-day AOD do not necessarily correlate with biases in forcing, we cannot rule out larger negative forcings, however.

Carbonaceous aerosol forcing was diagnosed according to emission sources: BC from fossil and biofuel (ff+bf) sources, OA ff+bf, and biomass burning (BB) total carbonaceous forcing. This source apportionment was used since OA is always co-emitted along with BC emissions, but the BC/OA ratio is typically much lower for biomass burning. Evaluation by emission source is consistent with AeroCom (Schulz et al., 2006) and AR4 (Forster et al., 2007), although it leaves us with a mixture of pollutant-based and sector-based analyses.

For BCff+bf, mean and standard deviation of RF is $0.24 \pm 0.09 \, \mathrm{W \, m^{-2}}$, with a full range from 0.14 to $0.38 \, \mathrm{W \, m^{-2}}$. As shown in Sect. 3, the models underestimate AAOD in comparison with OMI by 52 %, while biases with respect to AeroNet are even larger in most parts of the world, although we acknowledge that there are substantial uncertainties regarding the AAOD measurements themselves. This suggests that the BCff+bf RF could similarly be greatly underestimated, though the magnitude of the RF bias is unclear as some of the AAOD bias may come from biomass burning BC or dust, and biases in BC's present-day climatology do not translate directly into biases in time-dependent forcing.
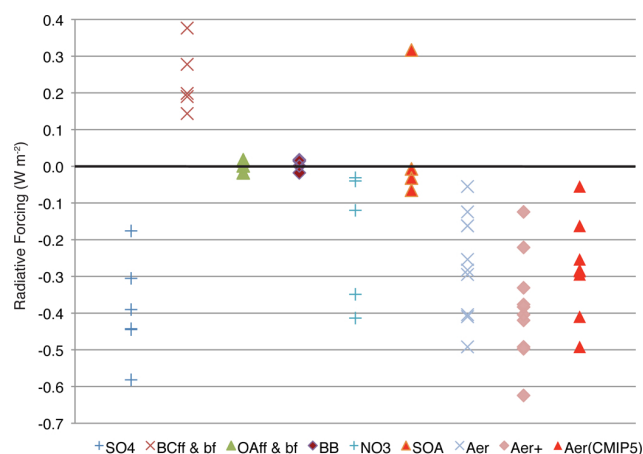

Fig. 8. Global mean aerosol RF (direct) by component and total. Individual points are from different models. Aer are totals. Aer+ are totals including adjustment by adding in forcing due to missing nitrate and SOA (see text). Aer(CMIP5) is the subset of ACCMIP models that also participated in CMIP5. The number of models for each component is: sulfate 9, BCff+bf 5, OAff+bf 4, BB 4, nitrate 5, SOA 4, Aer 10, Aer+10, and Aer(CMIP5) 8. Note that HadGEM2 nitrate (Bellouin et al., 2011) and NCAR-CAM3.5 nitrate are included in those models' Aer+ but not the Aer or Aer(CMIP) values. MIROC-CHEM nitrate and SOA are included in their Aer and Aer+ values, but not in their Aer(CMIP) values.

Evaluation against BC deposition recorded in ice cores also reveals substantial biases in these models (Lee et al., 2013), though again these cannot be clearly related to RF.

Few model results are available for the remaining components. The mean RF from ff+bf OA is $-0.04 \, \mathrm{W \, m^{-2}}$, with a range from -0.01 to $-0.08 \, \mathrm{W \, m^{-2}}$. The models with the weakest OA forcing substantially underestimate AOD in OA-rich regions (Tables 4 and 7). This suggests that the OA RFs from -0.04 to $-0.08 \, \mathrm{W \, m^{-2}}$ reported by the models with smaller OA-rich region biases might be more realistic (though there are uncertainties going from AOD to RF due to incomplete knowledge of OA optical properties and temporal evolution).

The mean RF for biomass burning BC+OA is $0.00 \, \mathrm{W \, m^{-2}}$, with a range from -0.02 to $0.02 \, \mathrm{W \, m^{-2}}$. This RF is thus quite small and the limited number of models show fairly similar results. Three additional models reported total carbonaceous aerosol RF (Table 7). The MIROC-CHEM model has the greatest bias with respect to OMI AAOD (Table 5), so is likely substantially low. The other models show similar totals to the sector-specific sums. Thus we are fairly confident that the reported ranges for the carbonaceous aerosols forcings are robust across these models.

Results for the remaining components, SOA and nitrate, show substantial spread. For SOA, the substantial positive forcing in MIROC-CHEM is quite different from the other models, which show small negative forcings. Emissions of
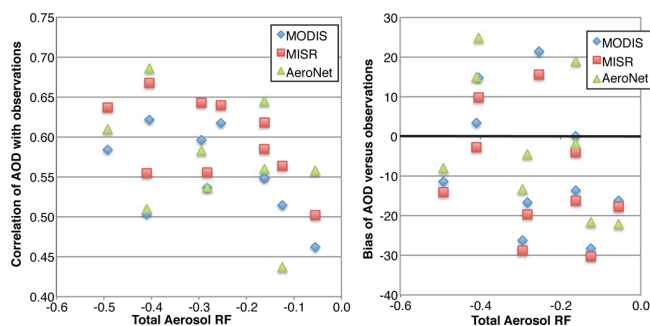
Fig. 9. Comparison of model skill in capturing the spatial pattern and seasonal cycle (left) or magnitude of annual average (right, percent) AOD observations with modeled RF due to all aerosols.

biogenic SOA precursors are coupled to land-use changes in MIROC-CHEM, while other models use fixed present-day vegetation distributions. Thus only MIROC-ESM-CHEM incorporates decreases in forest area leading to reduced emissions of SOA precursors, and hence a positive SOA RF. Thus the "outlier" may in fact be the most realistic model. This highlights the role of "structural" uncertainties concerning which physical processes are represented in models in addition to "scientific uncertainties" represented by the range of results across models incorporating similar processes.

Total aerosol RF was reported in 10 models, and the mean, standard deviation, and range are -0.26, 0.14 and -0.06 to $-0.49\,\mathrm{W\,m^{-2}}$, respectively. We also examine the total aerosol RF accounting for missing components (nitrate and SOA) in some models, which we call Aer+. Mean values for missing components are taken from the other ACCMIP models. We exclude the MIROC-CHEM SOA, however, as this includes land-use changes that have thus far been assessed in only a single model and so we do not know how representative that result is. We also weight GISS-E2-R nitrate by 0.5 to account for its biases against AVHRR trends in nitrate-rich areas (Sect. 4). Multi-model means are $-0.05\,\mathrm{W\,m^{-2}}$ for SOA and $-0.16\,\mathrm{W\,m^{-2}}$ for nitrate. The mean, standard deviation, and range for Aer+ are -0.39, 0.14 and -0.12 to $-0.62\,\mathrm{W\,m^{-2}}$, respectively. Since accounting for missing components improves the agreement between models and satellite AOD in nearly all cases (Sect. 3), we consider the missing component-adjusted values to be more realistic.

We also test if there is a relationship between model skill and aerosol RF. The models with the highest correlation (over 0.60) against the satellite datasets or AeroNet do have a narrower RF range, -0.16 to $-0.49\,\mathrm{W\,m^{-2}}$, than the full set of models (Fig. 9). The models with smallest biases ($< 15\,\%$) have the same forcing range. Screening the models by correlation or high bias with respect to MODIS fine-mode fraction over the oceans gives identical results. The range for this subset of quality-screened models is -0.33 to $-0.62\,\mathrm{W\,m^{-2}}$ accounting for missing components (Aer+). Screening based on NOAA AVHRR AOD trends (near-Europe and near-South

and East Asia where the multi-model mean agrees with observations), leaving out models more than one standard deviation from the multi-model mean, gives a range of -0.16 to $-0.40\,\mathrm{W\,m^{-2}}$, or -0.33 to $-0.50\,\mathrm{W\,m^{-2}}$ for Aer+. Note, however, that several models are absent from the AVHRR analysis, which may contribute to the reduced range. Hence the full range based on all screening is encompassed by $-0.42 \pm 0.09\,\mathrm{W\,m^{-2}}$ (Aer+). This RF is almost identical to the multi-model mean Aer+, but has less than half the range of the full set of models. Thus although there are large differences in RF per unit AOD in aerosol models, and hence screening by AOD would not obviously lead to a reduced RF range, this is the case in these models. Note that the screening does not take into account uncertainties in emissions, which could alter the relative agreement with observations of the various models.

The evaluation against observations thus indicates that $-0.42 \pm 0.09\,\mathrm{W\,m^{-2}}$ is the best estimate of the total 1850 to 2000 aerosol RF. This is similar to the IPCC AR4 estimate of $-0.50 \pm 0.40\,\mathrm{W\,m^{-2}}$, but has a much smaller uncertainty range. Our range does not account for uncertainty in the underlying emissions, however. Additionally, positive forcing from fossil+biofuel BC is likely underestimated (Sect. 3.3). We performed a similar analysis of fossil+biofuel BC forcing in comparison with model skill in representing AAOD. Unfortunately, the requisite data was available from only 5 models, and the analysis did not show robust relationships between skill and RF at either the global scale or for regions with greatest BC/dust AAOD ratios. Hence a model's ability to reproduce present-day climatological AAOD provides a poor test of its long-term BC RF. Without better understanding of the causes of the AAOD underestimate, it is not yet clear how best to adjust forcing from BC or co-emitted species to correct for model biases.

The CMIP5 subset of eight ACCMIP models has an aerosol RF mean and standard deviation of $-0.28 \pm 0.13\,\mathrm{W\,m^{-2}}$ (Fig. 8). Thus the aerosol RF actually driving those CMIP5 climate simulations tends to be underestimated in comparison with our best estimate (Aer+), primarily owing to the lack of nitrate and/or SOA in several models.

As mentioned previously, several models included changes in dust and sea-salt aerosols, and nearly all models included the effect of climate change on aerosols via the imposed SST and sea-ice trends (except NCAR-CAM5.1, CICERO-OsloCTM2 and CSIRO-Mk3.6). The impact of climate change can be isolated in four models by comparing the full 1850 to 2000 changes against the influence of emissions alone (the simulation with 1850 climate and 2000 emissions differenced with the simulation with all conditions at 1850). The effect on aerosol RF is $-0.02\,\mathrm{W\,m^{-2}}$ in HadGEM2, $-0.07\,\mathrm{W\,m^{-2}}$ in GFDL-AM3, $-0.17\,\mathrm{W\,m^{-2}}$ in GISS-E2-R and $0.07\,\mathrm{W\,m^{-2}}$ in MIROC-CHEM. The range is clearly quite large, at least in part because models include different processes influencing aerosols. In GISS-E2-R, for example, sulfate and nitrate aerosols can form coatings on

Table 7. Evaluation of 1850 to 2000 RF by component in each model.

Model	SO$_4$	BCff and bf	OAff and bf	BB	Carb. sum	NO$_3$	SOA	Aer direct	Aer direct plus missing	Aer ERF	O$_3$
CICERO-OsloCTM2	−0.58	0.38	−0.08	−0.02	0.27	−0.03	−0.07	−0.40	−0.40	–	–
CSIRO-Mk3.6	−0.58	–	–	–	0.32	n	n	−0.25	−0.42	−1.41	–
GFDL-AM3	–	–	–	–	–	n	n	−0.41	−0.62	−1.44	0.41
GISS-E2-R	−0.31	0.28	−0.04	0.00	0.24	−0.41	−0.03	−0.49	−0.49	−1.10	0.17
GISS-E2-R-TOMAS	−0.30	–	–	–	–	n	n	−0.16	−0.37	−0.76	–
HadGEM2	−0.35	0.19	–	–	–	−0.12	n	−0.16	−0.33	−1.22	0.23
LMDzORINCA	−0.44	–	–	–	0.16	n	n	−0.28	−0.50	−0.71	0.35
MIROC-CHEM	−0.39	–	–	–	0.10	−0.35	0.32	−0.12	−0.12	−1.24	0.39
NCAR-CAM3.5	−0.44	0.14	−0.01	0.02	0.15	−0.03	n	−0.29	−0.37	n	0.44
NCAR-CAM5.1	−0.18	0.20	−0.02	0.01	0.19	n	−0.01	−0.06	−0.22	−1.44	–
Mean	−0.40	0.24	−0.04	0.00	0.21	−0.19	0.05	−0.26	−0.39	−1.17	0.33
Std dev	0.13	0.09	0.03	0.02	0.08	0.18	0.18	0.14	0.14	0.29	0.11
Median	−0.39	0.20	−0.03	0.01	0.19	−0.12	−0.02	−0.27	−0.39	−1.23	0.37

Note that HadGEM2 nitrate results are from (Bellouin et al., 2011) and are included in the HadGEM2 Aer+missing but not the Aer values. MIROC-CHEM nitrate and SOA are included in their Aer values (for this model, a different setup was used for calculations of forcing by individual components, so the total cannot be compared with the sum of individual components). Carb stands for carbonaceous. A dash indicates value not available, n indicates component not included.

dust and sea-salt particles, changing their lifetimes. Similarly, HadGEM2 alone included sea-salt but not dust changes in their forcing calculation. From 35 to −58 % of the 1850 to 2000 total aerosol RF is attributable to the influence of climate feedbacks on aerosols rather than aerosol direct or precursor emissions.

5.1.2 Temporal evolution of global mean RF

In the ACCMIP timeslices, total aerosol RF becomes increasingly negative in all models from 1850 to 1930, and again from 1930 to 1980 (Fig. 10). From 1980 to 2000, however, the total aerosol negative RF becomes weaker in six of the nine models. This is due to pollution controls that limited emissions, especially of sulfur dioxide. Sulfate RF weakens or stays approximately constant from 1980 to 2000 in all models for which data is available. In contrast, fossil+biofuel BC RF grows more positive throughout the 20th century in all models, contributing to the weakening total aerosol negative RF between 1980 and 2000. Unlike fossil+biofuel BC RF, BC albedo forcing peaks in 1980. This is due to regional shifts in the location of BC emissions from higher latitudes, where they can more easily reach Arctic snow and ice covered areas, to lower latitude developing nations (further details are presented in Lee et al. (2013) and Appendix D).

Total aerosol RF from 1980 to 2000 becomes more negative in the GFDL-AM3 and GISS-E2-R models, by −0.05 and −0.10 W m^{-2}, respectively. GFDL-AM3 did not diagnose RF by aerosol component. In GISS-E2-R, sulfate contributes 0.01 W m^{-2}, carbonaceous aerosols 0.05 W m^{-2}, nitrate −0.13 W m^{-2} and SOA −0.01 W m^{-2}. Hence RF becoming more negative is primarily attributable to nitrate in

that model. Note that this cannot be the case for GFDL, which did not include nitrate.

In the RCP2.6 and RCP8.5 emission scenarios, which span the range of RCP projections in terms of total forcing/climate impacts, all aerosol RF declines greatly in most models. There is little difference between the scenarios in MIROC-CHEM. CICERO-OsloCTM2 and LMDzORINCA show substantially greater declines in negative aerosol RF at 2030 under RCP2.6, but the differences narrow at 2100. In contrast, differences between the scenarios increase throughout the 21st century in GFDL-AM3. In that model, total aerosol RF stays approximately constant under RCP8.5. Total aerosol RF becomes increasingly more negative in the future in GISS-E2-R. This results from increased nitrate negative RF and reduced BCff+bf positive RF, which together outweigh the reduced negative RF from other scattering aerosols. Most of the increase is due to nitrate, with 2100 versus 2000 RF of −0.36 W m^{-2} under RCP2.6 and −0.49 W m^{-2} under RCP8.5. Bellouin et al. (2011) report HadGEM2 2100 versus 2000 nitrate forcings of −0.4 W m^{-2} and −0.5 W m^{-2} for RCPs 2.6 and 8.5, respectively, including direct and cloud albedo effects. Their ratio of historical nitrate direct forcing to nitrate direct plus cloud albedo forcing is 0.71, suggesting that the direct forcing is ∼ 30 % lower than these RCP RF values (those estimates are shown in Fig. 10). Hence the HadGEM2 results seem fairly consistent with the GISS-E2-R nitrate projections, though slightly smaller. Future nitrate forcing was not available from other models. Nitrate aerosols become increasingly important because sulfur dioxide emissions are greatly reduced, and sulfate and nitrate precursors compete for a limited supply of ammonium, and because the RCPs assume that pollution controls are effective for industry, vehicles and power
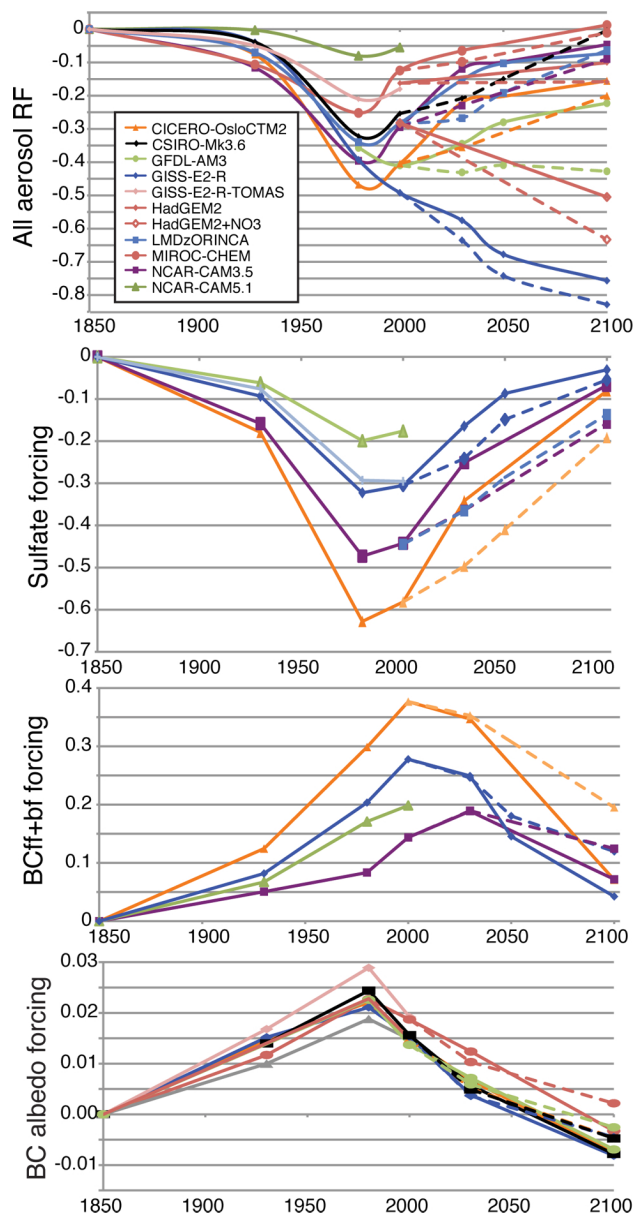
Fig. 10. Timeseries of global mean annual average RF from all aerosols (top), sulfate (2nd row), fossil+bio fuel BC (3rd row) and BC albedo (bottom) in the indicated models. RF was calculated at the years with symbols and is interpolated for other years. Open symbols for HadGEM2 show the results incorporating nitrate aerosols as described in Bellouin et al. (2011). RCP8.5 is shown as dashed lines, RCP2.6 as solid lines.

generation, but ammonia emissions from agriculture increase during the 21st century (van Vuuren et al., 2011). Note that in some models, dust and sea-salt were included in future RF (Table 1). While diagnostics were not available for these natural aerosols in most models, in GISS-E2-R they were remarkably stable, contributing less than $0.03 \, \mathrm{W \, m^{-2}}$ to 2100 versus 2000 RF.

Analysis of 2000 to 2010 shows small aerosol RF. Across all four RCPs, results from five models (though not all models ran all scenarios) show aerosol RF from -0.013 to $0.033 \, \mathrm{W \, m^{-2}}$, with mean and median values of $0.007 \, \mathrm{W \, m^{-2}}$ and $-0.001 \, \mathrm{W \, m^{-2}}$, respectively. For RCPs 4.5 and 6.0, with at least three model results available (Table 1), values are not consistent in sign among the models.

5.1.3 Geographic pattern of RF

The distribution of aerosol RF is highly inhomogeneous. This has important consequences for both global and especially regional climate change, as feedbacks are non-uniform and aerosol impacts tend to maximize in areas with greatest forcing (e.g. Rotstayn and Lohmann, 2002; Ming and Ramaswamy, 2009; Boer and Yu, 2003; Shindell et al., 2010). We therefore next analyze the spatial pattern of aerosol RF.

The 1850 to 2000 sulfate forcing is greatest over the most industrialized and heavily populated areas, especially east and south Asia, Europe and eastern North America where negative forcing exceeds $-1 \, \mathrm{W \, m^{-2}}$ (Fig. 11). Forcing extends past these regions, where the emissions of sulfur dioxide are largest, out over the nearby oceans and over the Middle East, due to atmospheric transport. The variation across models is greatest in these same regions, and is fairly uniform across regions.

Fossil+biofuel BC RF is, like sulfate, very large over east and south Asia where it is more than $1 \, \mathrm{W \, m^{-2}}$, but is comparatively small over Europe and North America. The spatial distribution of fossil+biofuel OA RF is generally very similar to fossil+biofuel BC RF, unsurprisingly, but has the opposite sign (except over eastern North America) and a much smaller magnitude.

Unlike other components, biomass burning aerosol and SOA RF show regions of both substantial positive and negative forcing. For biomass burning, this stems from varying regional fire frequency trends, with decreases in the southeastern United States and increases in Indonesia. For SOA, the large positive forcings come from the MIROC-CHEM model's incorporation of changing land-use. As the other models do not include this factor, there is an extremely large standard deviation in this RF. While global means always mask regional patterns, the existence of both positive and negative forcing means that the global mean can be particularly misleading for these two components.

Nitrate aerosol RF shows local maxima over East Asia, Europe and eastern North America, and to a lesser extent over south Asia. This distribution is similiar to that of sulfate. There is a broader distribution of small Southern Hemisphere forcing values, however. This stems from both MIROC-CHEM and especially from GISS-E2-R, which efficiently lofts ammonia in convective plumes from tropical sources to the upper troposphere, where it then spreads to both hemispheres. This leads to unexpectedly large nitrate aerosol abundances in the upper troposphere over much of
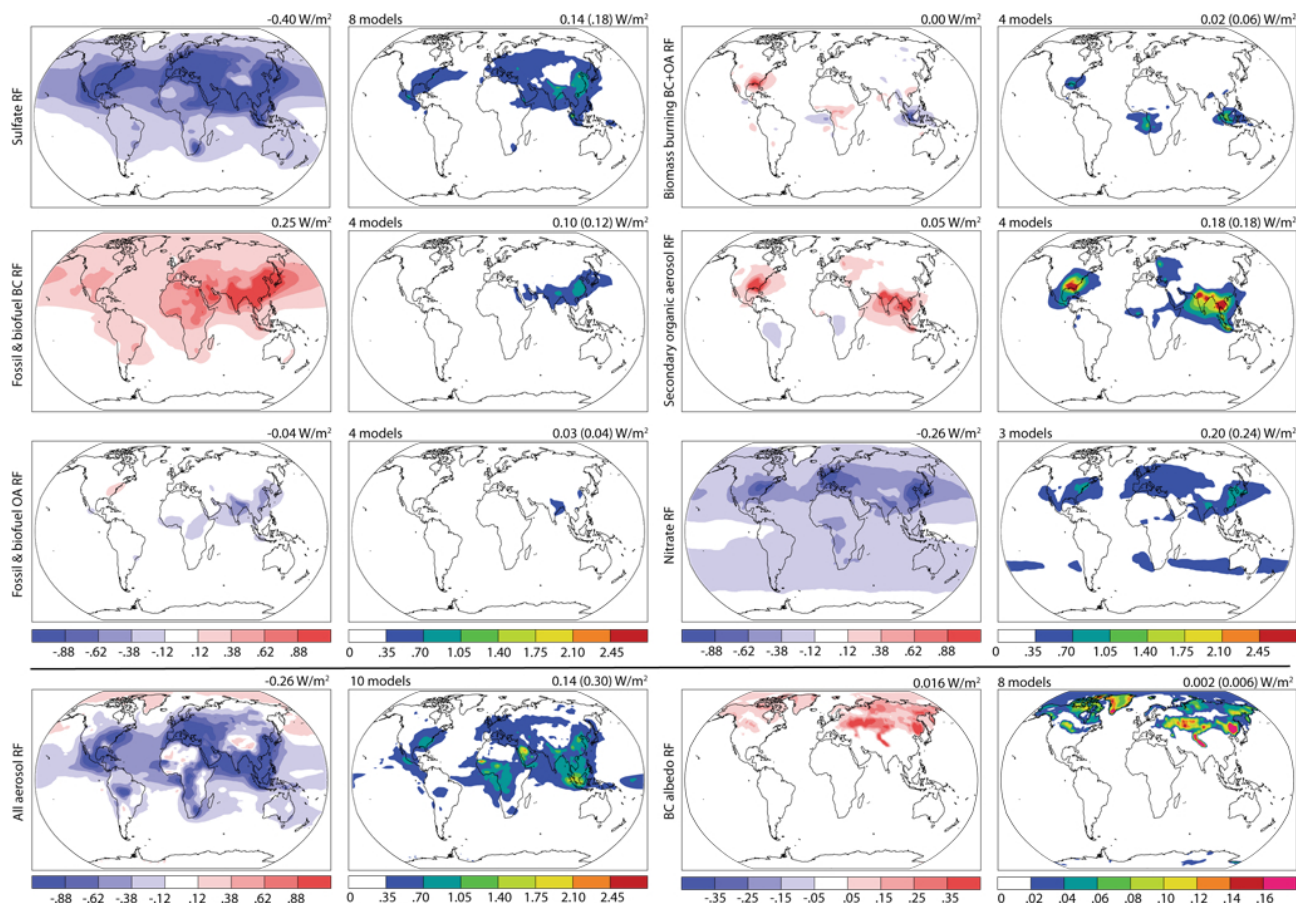
Fig. 11. Industrial era multi-model mean (left columns) and standard deviation (right columns) of RF for each aerosol component (top three rows), all aerosols (lower left) and BC albedo (lower right; note change in scale). Global means are in the upper right corner, with standard deviation of the global mean values from each model given first followed by the global mean of the standard deviation at each point.

the world, though given the paucity of measurements it's difficult to evaluate this forcing. As noted previously, GISS-E2-R appears to underestimate AOD trends near North America and Europe owing to overly large increases in nitrate. Hence the multi-model mean nitrate forcing over Southern Hemisphere mid-latitudes, coming in large part from that model, may also be too large. There is great divergence between models in this region. In comparison, standard deviations are of similar magnitude over industrialized areas, but forcing is much larger. Models differ in industrialized regions as well though, and it is difficult to determine which is more realistic overall. For example, over China, a model that captures observed nitrate well found annual mean nitrate RF of $-0.95\,\mathrm{W\,m^{-2}}$ (relative to zero nitrate) (Zhang et al., 2012). For ACCMIP models, GISS-E2-R finds a very similar value $(-1.08\,\mathrm{W\,m^{-2}})$, while CICERO-OsloCTM2 has about half as much year 2000 nitrate and MIROC-CHEM roughly double the GISS-E2-R amount. Given the small number of models having reported nitrate RF, it is clear that uncertainties are especially large for this aerosol component.

The total aerosol RF is strongly negative over most Northern Hemisphere land areas below $60°$ N. Over the Sahara, the Tibetan plateau, and the Arctic, however, very high surface albedo reduces the effect of scattering aerosols while increasing the effect of absorbing aerosols, leading to net positive forcing. There is a local maximum in negative forcing over the Amazon that is largely attributable to SOA, and large values over western Central Africa driven by SOA, biomass burning and nitrate. This suggests that forcing in these regions may be underestimated (i.e. not negative enough), as many models do not include all these aerosols. Large model-to-model variations in all aerosol RF over west Africa and Indonesia stem from SOA and biomass burning aerosols, while large diversity over Arabia is due to forcing there arising primarily from long-range transport which can vary substantially between models, with contributions from dust changes included in some models as well. Substantial forcing also extends over many oceanic regions.

BC albedo forcing, which is not included in the all aerosol RF, is largest in western Russia, the Karakoram and
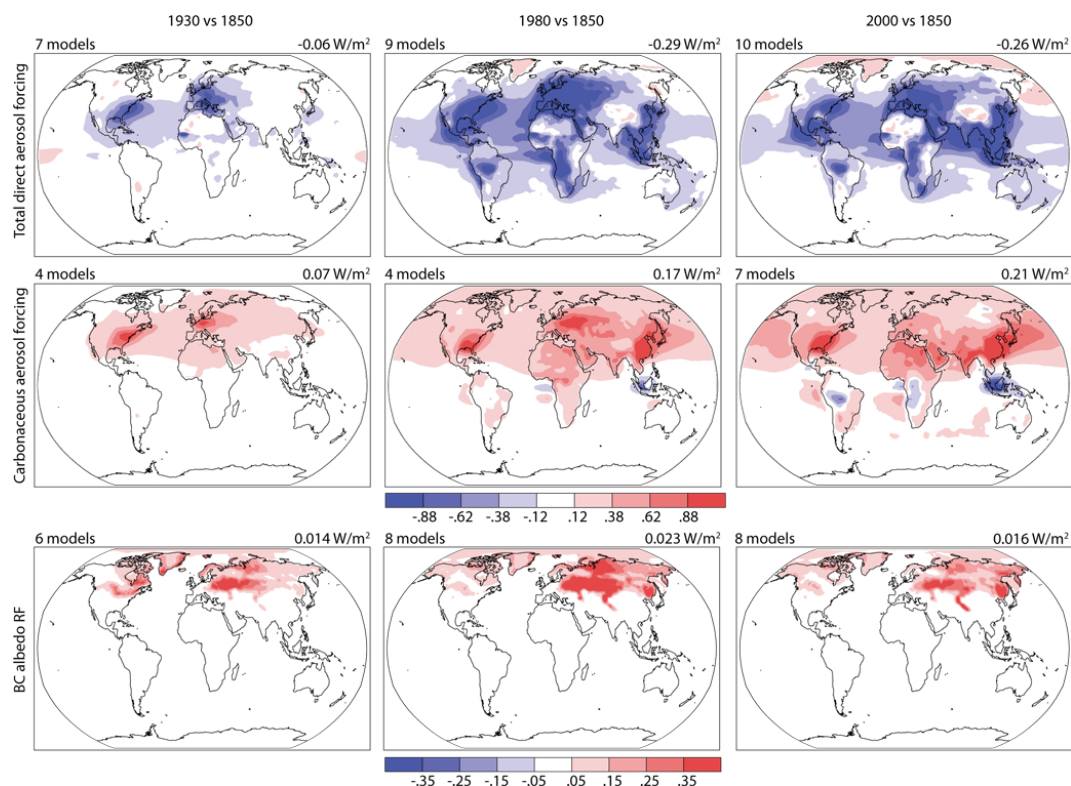
Fig. 12. Multi-model mean all aerosol RF, carbonaceous aerosol RF and BC albedo RF for 1930, 1980 and 2000 (all relative to 1850; W m^{-2}). The mean 2000 carbonaceous forcing in the 4 models with results at earlier times is also 0.21 W m^{-2}. Mean total 2000 direct forcing in the 7 models with results for 1930 and in the 9 models with results for 1980 is also -0.26 W m^{-2}.

Manchuria. Arctic forcing is also substantial, though not as large as at lower latitudes where sunlight is more plentiful.

We also examine the distribution of forcing through time, focusing on the total and carbonaceous aerosol (carbonaceous aerosol RF is simpler to display than the three separate components, and the OA ff+bf and BB forcings are comparatively small). RF in 1930 was primarily concentrated over Europe and eastern North America for both cases (Fig. 12). Magnitudes increased substantially in those areas from 1930 to 1980, while large forcings also appeared over east Asia, and for the total, over parts of Africa, Latin America and southeast Asia. From 1980 to 2000, negative all aerosol RF increased substantially over south and southeast Asia, where increased sulfate outweighed increased BC (Fig. 13). These more nearly offset one another over East Asia, leading to small trends in all aerosol RF there. Over Europe and North America, however, forcing is positive due to declining sulfate, which over Europe outweighs a large reduction in fossil+biofuel BC RF during this period. Tropical trends are attributable to changes in biomass burning (they are not from fossil+biofuel BC RF; Fig. 13), which increased in Indonesia and decreased in western Africa.

From 1930 to 1980 BC albedo forcing declines over North America while increasing markedly over Eurasia (Fig. 12). Trends are especially large over western Russia, an area

downwind of European BC sources and one with substantial biomass burning. During 1980 to 2000, BC albedo forcing decreases in most of Russia, while increasingly greatly over the Karakoram and Manchuria.

The all aerosol 2000 to 2030 RF is similar in many respects to the 1980 to 2000 RF (Fig. 14). Under RCP2.6 or RCP8.5 negative all aerosol RF over Europe and North America continues to decline, though more so under RCP2.6. Likewise negative all aerosol RF continues to increase over South Asia, especially under RCP8.5. As with the recent past, trends in scattering and absorbing aerosols are more nearly balanced over east Asia, leading to modest all aerosol RF there. By 2100, aerosol forcing declines sharply in magnitude virtually everywhere (regions with negative 1850 to 2000 forcing show positive RF, and vice-versa). Carbonaceous aerosol RF similarly declines in the RCPs. As with the global mean, differences between scenarios are larger at 2030 than 2100.

5.2 Aerosol effective radiative forcing

In addition to direct aerosol RF, we calculate the "effective radiative forcing" (ERF), defined here as the top-of-the-atmosphere (TOA) net energy flux change with ocean conditions held fixed but all other processes allowed to respond
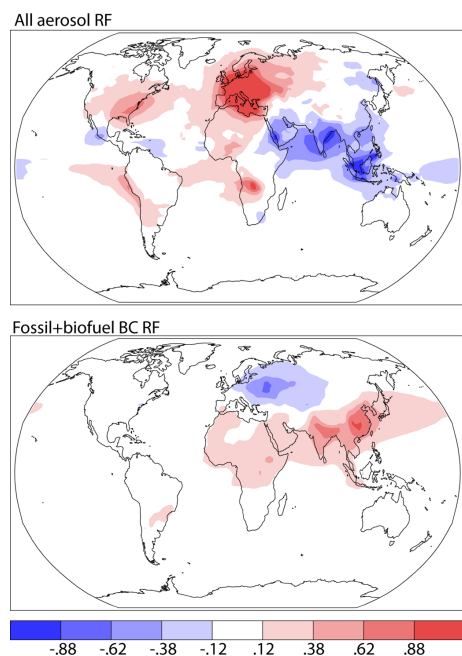
All aerosol RF

Fossil+biofuel BC RF

Fig. 13. Multi-model mean difference in all aerosol RF and in fossil+biofuel BC RF between 2000 and 1980 for models with results available at both times.

to the aerosol changes. Along with the direct RF, ERF thus includes aerosol indirect effects on clouds via microphysics (affecting cloud albedo and cloud lifetime) as well as responses of water vapor, lapse rate and clouds to aerosol thermodynamic impacts (including so-called "semi-direct" effects). In half the models, BC-induced albedo changes are also included (GISS-E2-R, GISS-E2-R-TOMAS, MIROC-CHEM and NCAR-CAM5.1). ERF is in general a better indicator of the eventual climate response than RF (Hansen et al., 2005; Lohmann et al., 2010). Note that since land temperatures are allowed to adjust, the ERF include a small portion of response which lowers its magnitude by ∼ 10 % or less (Hansen et al., 2005; Andrews et al., 2012). It is relatively straightforward to account for this bias at the global scale, but it is not clear how to do so at the regional scale and hence we do not remove this bias.

ERF values come from comparing the 1850 simulations against additional simulations that (1) used a different year's short-lived species emissions, (2) maintained 1850 climate and WMGHG concentrations, and (3) included interactions with radiation and clouds for aerosols but not for ozone. Thus aerosols were the only changes influencing radiation. CMIP5 included a very similar pair of simulations with fixed ocean boundary conditions and 1850 to 2000 aerosol concentration changes (although using aerosols from historical simulations includes the influence of climate change on aerosols). We evaluate aerosol ERF from the CMIP5 simulations for three models: LMD and HadGEM2 (ACCMIP results not available), and CSIRO (BC albedo did not affect radiation in AC-
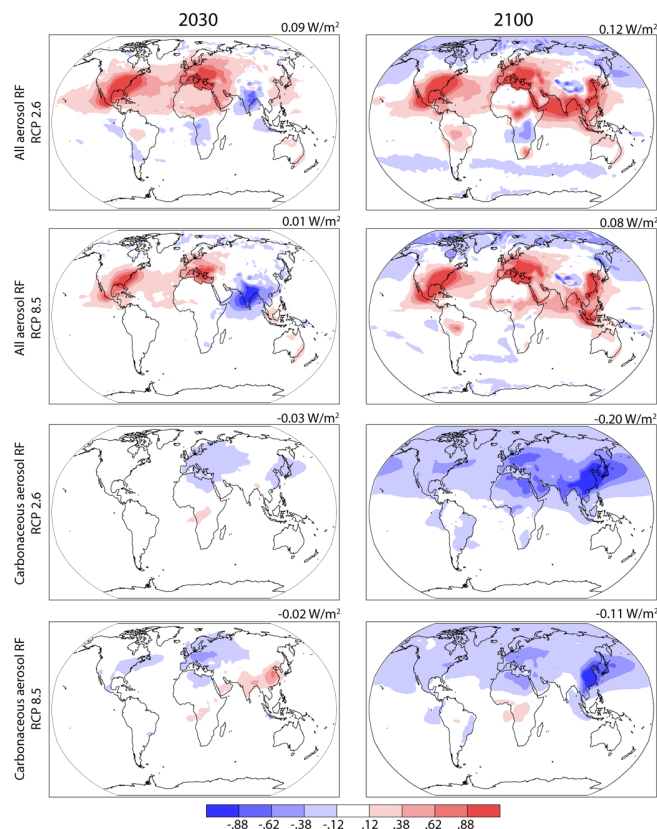
Fig. 14. All aerosol RF and carbonaceous aerosol RF for 2030 and 2100 relative to 2000 under RCP2.6 and RCP8.5.

CMIP runs but did in CMIP5 runs). In the NCAR-CAM5.1 simulations, background climate and WMGHG conditions were fixed at 2000 rather than 1850, and ozone precursor emissions did not change (Ghan et al., 2013).

The ACCMIP models' mean and approximate 5–95 % confidence interval (1.65-σ) for ERF is -1.2 ± 0.5 W m^{-2}. The spatial pattern of 1850 to 2000 aerosol ERF is broadly similar to the aerosol RF pattern (Fig. 15). ERF is relatively stronger over outflow regions, however. This is likely because anthropogenic aerosols have an enhanced effect on clouds in remote areas where there are few natural cloud condensation nuclei and high humidity. The ERF is positive in several regions, including the Sahara, parts of the Himalayas/Karakoram, and over both polar regions. Over much of the Arctic Ocean, values are more than 0.5 W m^{-2}. Arctic ERF is especially large in boreal spring, with values exceeding 0.75 W m^{-2} over large areas of the Arctic Ocean and Greenland (Fig. 16). This is attributable to both the more positive RF over highly reflective surfaces discussed previously and to the greater influence of clouds on longwave versus shortwave radiation at high latitudes. These results suggest that aerosols may have played a greater role in rapid Arctic climate change than generally appreciated.
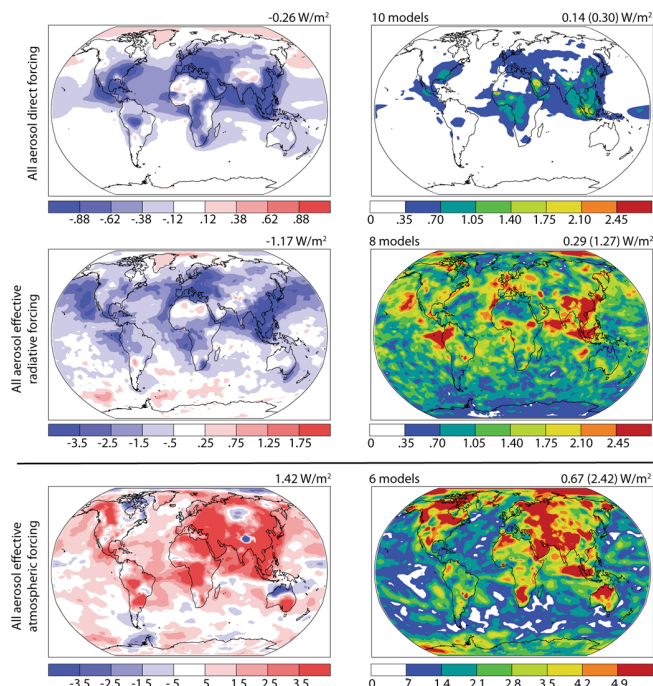
Fig. 15. Industrial era multi-model mean (left column) and standard deviation (right column) of direct aerosol RF, aerosol ERF, and atmospheric forcing (all in W m^{-2}). Numerical values as in Fig. 11.
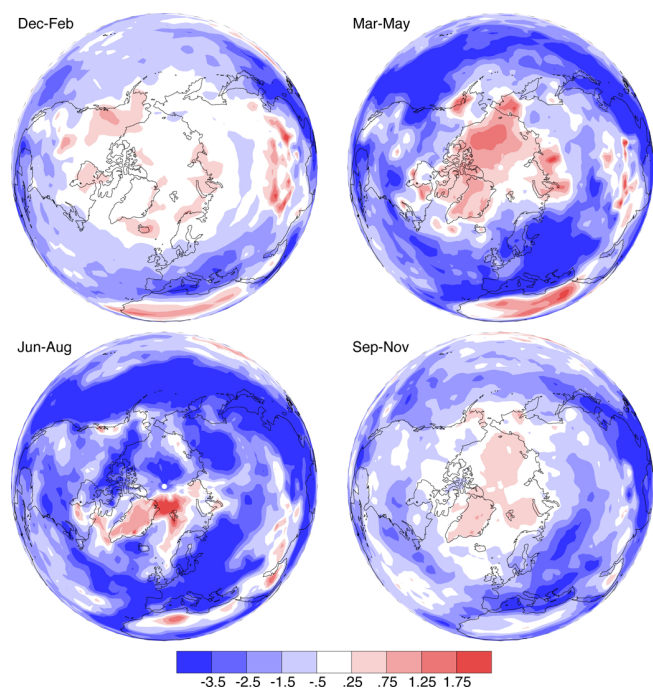

Fig. 16. Multi-model mean Northern Hemisphere aerosol ERF for 2000 relative to 1850 for the indicated seasons.
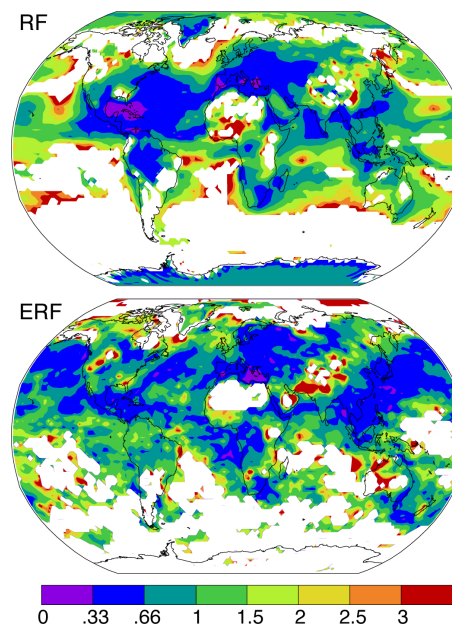

Fig. 17. The ratio of the standard deviation across models to the multi-model mean value for RF and ERF. Absolute values of the 1850 to 2000 forcings are used in these ratios. Areas with forcing less than 1/4 of the global mean are blank to avoid division by small values. Both analyses use the same 8 models.

Variability across models is large in many locations for ERF. The global average of the standard deviation at all points is 1.27 W m^{-2}, far larger than the standard deviation of any individual aerosol component's RF. The standard deviation across each model's global mean ERF is only 0.29 W m^{-2}, however, indicating that models produce fairly similar total aerosol ERFs but with forcing locations shifted between models. The global mean ERF standard deviation is 25 % of the multi-model mean, less than the comparable ratio for all aerosol RF (50 %) or RF by component (35–40 % for sulfate and BCff+bf, 80 % or more for others). The noisy ERF structure indicates that calculating local ERF values may be difficult for small forcings, however, whereas these can be easily isolated in the RF methodology that is not influenced by meterological variability. Over regions with substantial ERF, the relative standard deviation of the ERF is no larger than that of RF (Fig. 17). Over some areas, such as parts of east Asia, it is actually smaller.

Effective atmospheric forcing by aerosols (defined as TOA minus surface ERF) shows strong absorption of energy where BCff+bf and biomass burning RF are large. There are also indications of dynamics changes, including shifts over North America and Australia. The 2000 versus 1850 reduction in surface shortwave flux due to aerosol (including their effects on clouds, water vapor, etc.) is 2.50 ± 0.81 W m^{-2}, in excellent agreement with the most recent IPCC assessment of −2.3 ± 1.0 W m^{-2} (Denman et al., 2007).
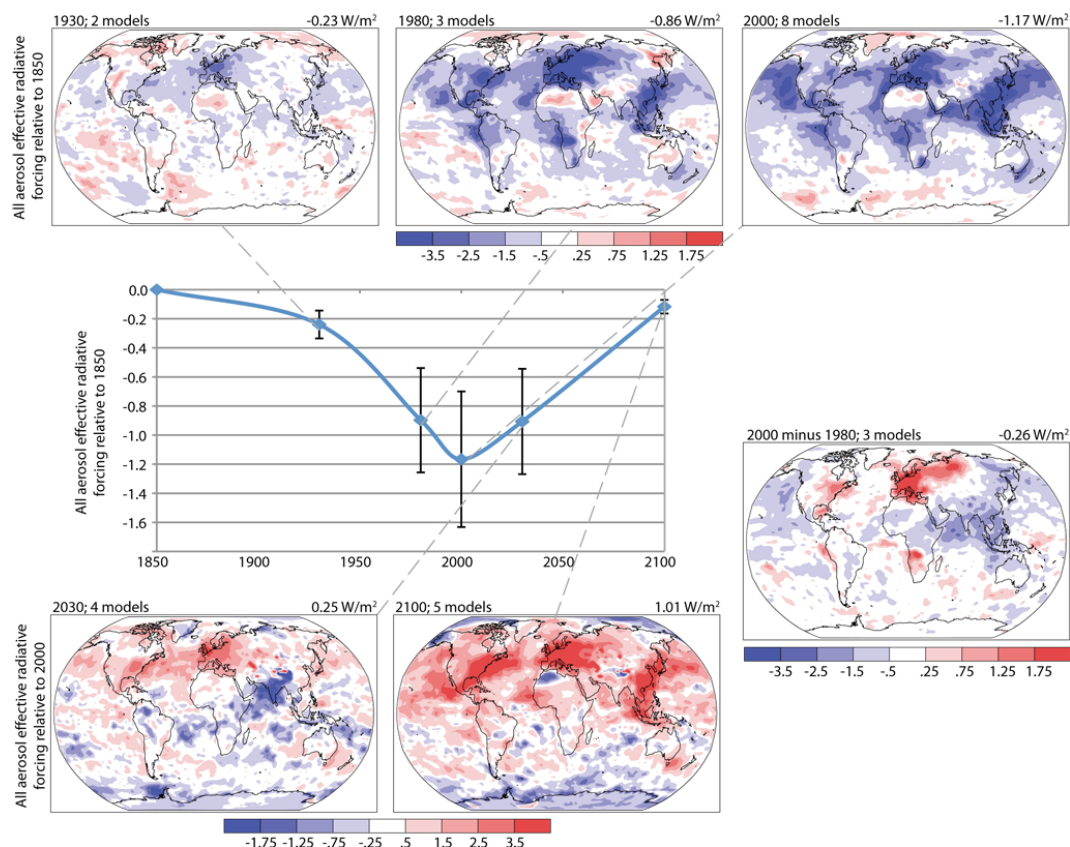
Fig. 18. Spatial patterns of all aerosol ERF for 1930, 1980 and 2000 relative to 1850, for 2030 and 2100 under RCP8.5 relative to 2000 and for 2000 relative to 1980. The timeseries shows global mean values all relative to 1850. All ERF values other than at 2000 are based on fractional changes relative to 2000 in models with data available at both times, and uncertainty is assigned as the relative uncertainty at 2000 (1.65 standard deviation as 5–95 % confidence interval).

Global mean ERF minus RF is $-0.90 \, \mathrm{W \, m^{-2}}$, providing a rough estimate of aerosol indirect effects on clouds (though it includes additional responses such as water vapor and lapse rate adjustments). A large number of model studies constrained by satellite data summarized in Lohmann et al. (2010) found ERF ranging from approximately -0.6 to $-1.6 \, \mathrm{W \, m^{-2}}$, with roughly 1/3 direct and 2/3 indirect. These values are in good agreement with our results. Most inverse methods based on observed temperature changes, ocean heat uptake, and cacluated non-aerosol forcings produce ERF best estimates of -0.8 to $-1.6 \, \mathrm{W \, m^{-2}}$ (Murphy et al., 2009; Shindell and Faluvegi, 2009; Church et al., 2011; Hansen et al., 2011), though some find smaller values (e.g. Libardoni and Forest, 2011). Hence the ACCMIP models' ERF is consistent with most prior studies, though weaker values have also been reported.

As the most ERF results were available for 2000, we evaluate global mean ERF at other times based on fractional differences relative to 2000 in models that diagnosed ERF at both times. The temporal evolution of aerosol ERF does not closely follow the temporal evolution of all aerosol RF (Fig. 18). Through 1980, ERF follows scattering aerosol RF,

which increases by 377 % from 1930 to 1980 while ERF increases by ~ 375 % (all aerosol RF increases by more than 500 %). ERF continues to increase from 1980 to 2000, however, while both all aerosol RF and scattering aerosol RF tend to decrease. Aerosol ERF became more negative in all three models that calculated aerosol ERF differences between 1980 and 2000, and did not closely track RF. This suggests that, unsurprisingly, ERF may be quite sensitive to background aerosol loading, the geographic location of the aerosols, and the mixture of aerosol types. In particular, increases in negative ERF from 1980 to 2000 are large over East Asia and the Pacific outflow regions, with large values extending all the way to the eastern North Pacific (Fig. 18), while increases in negative all aerosol RF are fairly small and localized there during this time (Fig. 13). Thus the global mean RF trends are dominated by decreasing aerosol over Europe and North America, but the ERF is dominated by increasing negative forcing over and downwind of Asia. Thus suggests that recent aerosol ERF has been more strongly influenced by recent increases in Asian emissions than by coincident European or North American decreases due to the former being upwind of the large, comparatively pristine Pacific

Ocean. Though results are only available from three models, this is consistent with the relatively stronger ERF vs. RF response in outflow regions seen in the analysis of all models. Analysis of ocean heat uptake and the planetary energy budget suggests that aerosol ERF indeed became more negative in the late 1990s and early 2000s relative to the 1980s and early 1990s (Church et al., 2011).

Diagnoses of future ERF are only possible for RCP8.5 at 2030 and 2100. We analyze flux differences between the full future simulations and runs with 2000 emissions and future climate (i.e. ocean conditions are at future values in both cases). When ozone changed in the RCP8.5 simulations, we subtract ozone forcing to get aerosol ERF. ERF is positive for 2030 relative to 2000 over Europe and the US and negative over South Asia and the Himalayas (Fig. 18). This pattern is similar to the 2000 versus 1980 ERF, except that ERF ceases becoming more negative over and downwind of east Asia. The global mean ERF trend therefore changes direction during this time. By 2100, the ERF has become positive nearly everywhere relative to 2000. The spatial pattern closely resembles the inverse of the 1980 or 2000 ERF (relative to 1850), indicating that most of the historical aerosol forcing has been removed. The primary exceptions are the negative tropical African and South American ERFs related to biomass burning.

Global mean ERF becomes less negative from 2000 to 2030 under RCP8.5, and by 2100 nearly recovers to its 1850 value. Unlike RF, future ERF becomes less negative in all the models, including the GFDL-AM3 and GISS-E2-R models that showed steady or increasingly negative future aerosol RF. We expect similar 2100 ERF under the other RCPs since they all remove most anthropogenic aerosol and aerosol precursor emissions, except for ammonia (Lamarque et al., 2013). There are differences in the timing of reductions, however, making it difficult to infer ERF at earlier times for the other RCPs.

Understanding the contribution of specific aerosol types to ERF is important for attribution of historical changes to particular emissions and especially for assessing the impact of potential future emissions pathways or emissions mitigation policies. Little information on the ERF attributable to specific aerosol types is available, however. Chuang et al. (2002) reported that cloud forcing due to sulfate and carbonaceous aerosols followed the burden fairly closely (within 15 %), which would imply a slightly larger historical indirect effect from sulfate. Jacobson (2002) suggested that the indirect effect of sulfate was roughly double that of carbonaceous aerosols, however. Among the ACCMIP models, MIROC-CHEM and GISS-E2-R performed simulations to isolate the contribution of individual aerosol components to the 2000 versus 1850 ERF (see Appendix E). In the MIROC-CHEM model, nearly all the indirect forcing is attributable to sulfate (93 %), with only small negative forcings ($\sim$ 10 % W m^{-2}) from BC and OC (Takemura, 2012). In the GISS-E2-R model, sulfate contributes a large nega-

tive cloud forcing (-0.45 ± 0.03 W m^{-2}) while OC causes a much smaller cloud forcing of -0.17 ± 0.03 W m^{-2} and BC causes a positive cloud forcing (0.23 ± 0.03 W m^{-2}). The total cloud forcing is -1.13 W m^{-2}, suggesting that additional components (nitrate and SOA) are important in that model or the sum is highly non-linear. Hence there seems to be consistency in attributing the largest share of aerosol indirect forcing to sulfate, presumably owing to its greater solubility, but the relative values for individual components span a wide range in extant studies, and the sign of the BC-induced indirect effect is not clear (see Appendix E).

5.3 Comparison of forcing with AOD and estimated CMIP5 model forcing

As calculation of RF adds some computational expense, and calculation of ERF adds a great deal, forcing is often not diagnosed. We therefore test how well forcing can be estimated based on more readily available AOD changes. AOD changes are highly correlated ($r^2 = 0.95$) with RF calculated in the ACCMIP models (Fig. 19). RF per unit AOD change, which we call Normalized Radiative Forcing (NRF), is -7.6 W m^{-2} (-8.9 to -6.6 at the 95 % confidence interval (CI)). The correlation is higher ($r^2 = 0.88$) for the larger 1850 to 2000 RF than for the 1980 to 2000 RF ($r^2 = 0.72$). Hence AOD may provide a reasonable indicator of RF for large AOD changes, but does not appear to be as reliable for smaller changes.

AOD changes are a fairly poor indicator of ERF, however, with an r^2 correlation of 0.48 in the ACCMIP models. The ERF/dAOD ratio is -30 W m^{-2} with a very large 95 % CI spanning -660 to -15. Note that both the RF and ERF analyses include cases in which changes in global mean AOD of near zero produce small but substantial forcings. Aerosol burden changes are generally not well correlated with RF, but analysis of burdens provides insight into the relative importance of emissions and lifetime changes in driving aerosol forcing (see Appendix F). If we use the RF/dAOD ratio calculated here to adjust model forcings for the bias relative to satellite observations, assuming all the bias can be attributed to trend underestimates, the multi-model mean would increase by -0.11 W m^{-2}. While this would produce an RF value (-0.37 W m^{-2}) in fairly good agreement with our best estimate, some of the bias relative to observations may be systematic over time.

AOD trends are available from a large number of CMIP5 models. Using the NRF of -7.6 W m^{-2} derived from ACCMIP, we estimate the direct aerosol RF based on CMIP5 AODs (Fig. 19; Appendix G). The estimated aerosol RFs in the subset of ACCMIP models with CMIP5 AOD data is -0.31 W m^{-2}, a value larger than the true forcing diagnosed in those models (-0.24 W m^{-2}). This suggests that if anything, the AOD-based RF estimates may be biased high. The estimated aerosol RFs in the non-ACCMIP CMIP5 models have a mean of -0.21 W m^{-2}, and most of the CMIP5
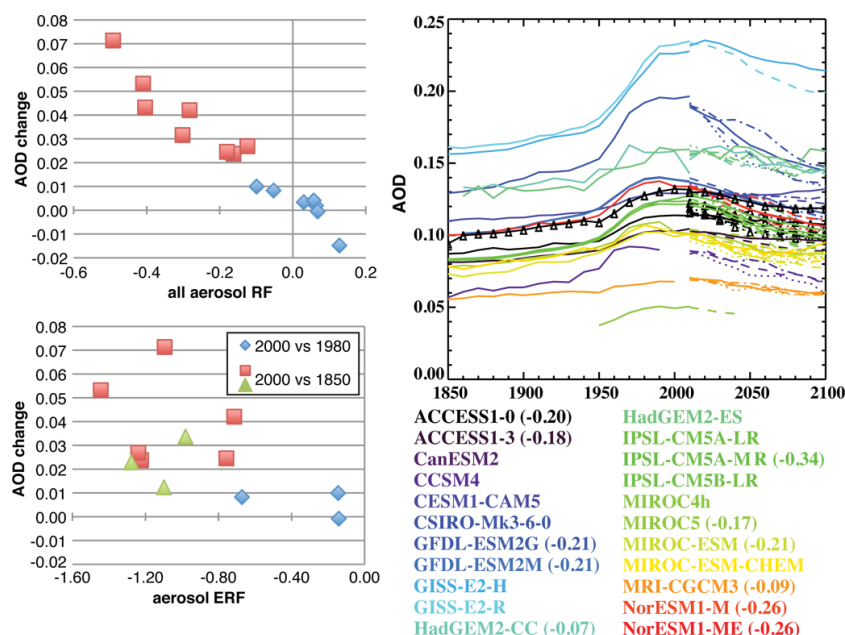
Fig. 19. AOD change versus RF (top left) and AOD change versus ERF (bottom left) from ACCMIP models (squares and diamonds) for the indicated periods. 2000 relative to 1850 ERF versus AOD changes are also shown for three additional CMIP5 models (green triangles). AOD versus time in available CMIP5 simulations (right), with estimated 2000–1850 aerosol RF given for each non-ACCMIP model for which AOD data was available. The open triangles indicate the multi-model mean. RCP2.6 is dotted, RCP4.5 dashed, RCP6.0 dash-dot, and RCP8.5 dash-dot-dot.

models not in ACCMIP have estimated aerosol RF less than the range encompassed by our best estimate ($-0.42 \pm 0.09\,\mathrm{W\,m^{-2}}$). Across both projects, only a few models have aerosol RF within the range of our best estimate, with all the others too small. Note that in many cases different model versions from the same institution have similar aerosols, but not always.

The present-day modeled AOD shows an enormous range for CMIP5 models, and the bias relative to AeroNet is strongly correlated with the AOD change ($r^2 = 0.84$). This suggests that our screening by skill in capturing present-day AOD provided a useful constraint on RF because present-day AOD is highly correlated with long-term aerosol changes. The CMIP5 multi-model mean AOD, excluding the two GISS models with much larger all-sky than clear-sky AODs, is biased 13 % low versus AeroNet. As the CMIP5 multi-model mean estimated forcing is roughly half the best estimate presented here, the AOD changes may be too small by an even larger factor.

As described previously, some models performed simulations under CMIP5 that allow aerosol ERF to be diagnosed for 2000 relative to 1850. Analysis of ERF was performed for six additional models (Appendix G). AOD changes from 1850 to 2000 were available for three of those models (MRI-CGCM3, MIROC5, and NorESM1-M), allowing those models to be added to the AOD change versus ERF comparison (Fig. 19). Consistent with the ACCMIP models, there

is no clear relationship, and the correlation across both sets of models decreases to $r^2 = 0.34$.

6 Total anthropogenic composition forcing

ACCMIP characterized radiative forcing from ozone as well as from aerosols. Ozone RF was calculated offline using the NCAR Community Climate System Model 4 radiative transfer model (RTM) and allowing stratospheric temperatures to adjust (Conley et al., 2012). We compute net longwave and shortwave all-sky flux at the tropopause (based on a climatology of tropopause pressure from the NCAR/NCEP reanalyses) varying only the ozone distribution. Results presented here are for ozone changes throughout the atmosphere. Detailed analyses of ACCMIP tropospheric ozone RF in Stevenson et al. (2013) show that use of a different RTM yields values 10 % higher, providing a rough estimate of uncertainty associated with the RTM. Young et al. (2013) show that the ACCMIP models generally capture observed 1980 to 2000 total ozone column trends relatively well. Here we analyze the ozone RF from most of the same models providing AC-CMIP aerosol simulations (Table 7), including all but two of the CMIP5 subset of ACCMIP models (Table 1). The 1850 to 2000 ozone forcing is $0.33 \pm 0.10\,\mathrm{W\,m^{-2}}$. Comparison of the radiative impact of present-day tropospheric ozone in these six models versus that of ozone observations from the Tropospheric Emission Spectrometer shows that these models have

global mean biases of $0.035 \pm 0.044 \, \text{W m}^{-2}$ (instantaneous longwave forcing), much smaller than the industrial-era RF (Bowman et al., 2012). Stevenson et al. (2013) also show that their ozone forcing results are only very weakly sensitive to changes in the models included (analyzing sets ranging from 4 to 17 models). While present-day biases and very recent trends do not necessarily constrain the long-term behavior of ozone, to the best of our ability to evaluate them these models appear to produce realistic ozone and ozone trends.

As variation between models in the RF due to WMGHGs is small, we examined this in only two models: NCAR-CAM3.5 and GISS-E2-R for 2000 relative to 1850. The spatial patterns of RF are indeed quite similar, and hence we take the mean as representative of the geographic distribution of RF from WMGHGs for all times, and scale the values uniformly to match the global mean RF due to WMGHGs prescribed in the historical period and under the RCPs. Uncertainty is estimated to be 10 % of the RF (Forster et al., 2007).

WMGHG forcing is relatively homogeneous, with slightly greater values at subtropical latitudes where clouds are less prevalent (Fig. 20). WMGHG forcing increases continually with time in the past and in the future under RCP8.5 (in RCP2.6 WMGHG forcing decreases after ~ 2050). Ozone forcing is positive between $\sim 45°$ S and $90°$ N, but is negative over and near Antarctica for the 1980 to 2030 timeslices due to the Antarctic ozone hole, which was not yet present in 1930 and has recovered by 2100. Positive ozone forcing maximizes in the subtropics similarly to the WMGHGs, and increases from 1850 through 2000. Ozone forcing continues to rise under RCP8.5 while it decreases to nearly its 1930 value by 2100 under RCP2.6. Relative uncertainties for ozone are substantially larger than for WMGHGs, but are smaller than those for aerosols (Fig. 11).

We then create composite fields of WMGHG, ozone and aerosol forcing, using ERF for aerosols and RF for WMGHGs and ozone. Though clearly it would be preferable to use the same metric for all agents, studies to date suggest that for both WMGHG and ozone ERF and RF values are probably within 5 % (Hansen et al., 2005; Andrews and Forster, 2008; Lohmann et al., 2010). Though aerosol ERF was only calculated for RCP8.5, as discussed previously we can regard the 2100 RCP8.5 aerosol ERF as a reasonable approximation to 2100 aerosol ERF under the other scenarios as well.

Total anthropogenic composition forcing relative to 1850 shows positive global mean values throughout the historical period, but distinct regional differences (Fig. 21). The net forcing is strongly negative over many industrialized areas in 1980 as negative aerosol ERF outweighs positive GHG (WMGHG+ozone) forcing. WMGHG forcing rises sharply in the late 20th century while negative aerosol forcing over Europe and North America declines, so that by 2000 the net forcing is near zero over Europe and positive over North America, while remaining strongly negative only over east

Table 8. Global mean annual average anthropogenic composition forcings and uncertainties.

	WMGHG RF	Ozone RF	Aerosol ERF	Net
1930	0.58 ± 0.06	0.09 ± 0.05	-0.24 ± 0.10	0.44 ± 0.12
1980	1.56 ± 0.16	0.30 ± 0.17	-0.90 ± 0.36	1.00 ± 0.42
2000	2.30 ± 0.23	0.33 ± 0.18	-1.17 ± 0.47	1.51 ± 0.55
2030 RCP8.5	3.64 ± 0.36	0.43 ± 0.20	-0.91 ± 0.36	3.20 ± 0.55
2100 RCP2.6	2.83 ± 0.28	0.14 ± 0.12	$-0.12 \pm 0.10^*$	2.86 ± 0.32
2100 RCP4.5	4.33 ± 0.43	0.23 ± 0.15	$-0.12 \pm 0.10^*$	4.44 ± 0.46
2100 RCP6.0	5.60 ± 0.56	0.25 ± 0.09	$-0.12 \pm 0.10^*$	5.74 ± 0.58
2100 RCP8.5	8.27 ± 0.83	0.55 ± 0.30	-0.12 ± 0.05	8.71 ± 0.88

All values are relative to 1850. Uncertainties are 5–95 % confidence intervals, assigned as 10 % for WMGHG RF (Forster et al., 2007), as 1.65 times the standard deviation across models for ozone RF, and as 40 % for aerosol ERF, which is 1.65 times the year 2000 standard deviation.
* For 2100 ERF, calculated ERF under RCP8.5 was used for all scenarios, with uncertainty doubled for other scenarios to account for potential differences relative to RCP8.5.

and southeast Asia. The negative aerosol forcing over central Africa and northwestern South America visible in 1980 is still present in 2000, but the increased WMGHG forcing balances it by that time.

Net anthropogenic composition forcing remains negative over a shrinking portion of southeast Asia in 2030 under RCP8.5, and remains small over parts of Africa and South America with substantial biomass burning, and over Antarctica. Nearly everywhere else forcing exceeds $2 \, \text{W m}^{-2}$, and in many areas is greater than $4 \, \text{W m}^{-2}$. By 2100 under RCP8.5, WMGHG forcing is so large and aerosol forcing so small that there is relatively little spatial variation in the net forcing, which is mostly between 6 and $10 \, \text{W m}^{-2}$. Forcing at 2100 is much lower under the RCP2.6 scenario, but is again dominated by WMGHGs and so is relatively uniform.

Early in the 20th century, forcing from aerosols and ozone largely offset one another, so that the net anthropogenic composition forcing follows the WMGHG forcing (Fig. 21). Aerosol ERF grows increasingly negative through 2000, masking a considerable portion of the GHG forcing. In the 21st century, aerosol masking is reduced and the net forcing again approaches the WMGHG forcing. Uncertainties are heavily influenced by aerosols, so that relative uncertainties maximize in 1980, and are far larger in 1980, 2000 and 2030 than in 2100 (Table 8). As WMGHGs dominate 2100 forcing, the net value is close to the RCP targets for all RCPs (Table 8).

In an earlier generation of climate models (those used in the IPCC Third Assessment Report), there was a distinct correlation between the magnitude of negative aerosol forcing and climate sensitivity (Kiehl, 2007). A similar analysis for the CMIP5 subset of ACCMIP models shows that in this generation of models there is now an anti-correlation between historical aerosol RF and equilibrium climate sensitivity (ECS; taken from Andrews et al., 2012; Bitz et al., 2012; Gettelman et al., 2012) and GISS-E2-R simulations) (Fig. 22). However, ECS is not particularly correlated with
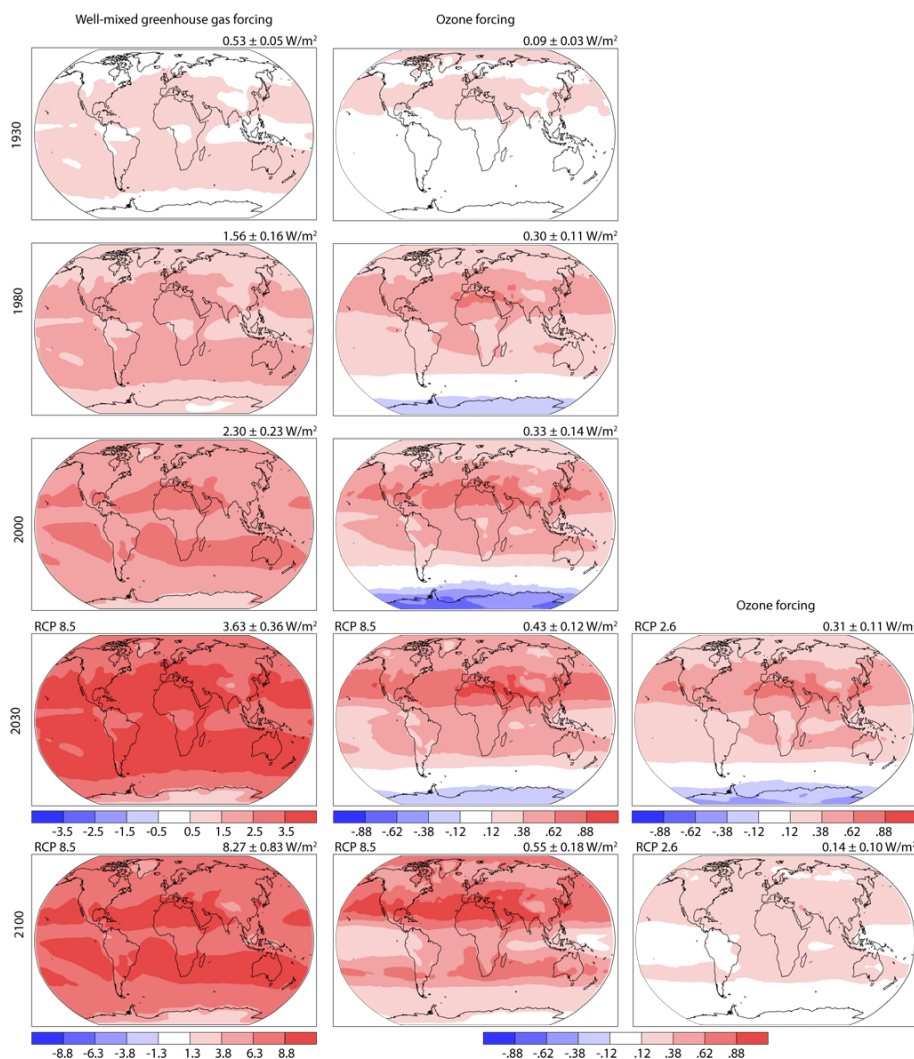
Fig. 20. Maps of WMGHG and ozone RF (W m^{-2}) for various historical and future times/scenarios. Note change in scale for WMGHGs in 2100.

aerosol ERF, which is the more relevant quantity, nor with total forcing. Similar conclusions come from comparing forcing with transient climate response. We examine modeled change in comparison with both the Climate Research Unit (CRU) and GISS historical temperature datasets (Brohan et al., 2006; Hansen et al., 2006). As the latter begins only in 1880, we use GISS 1880s to 1996–2005 plus CRU 1850s to 1880s in that case. Many models have difficulty in capturing the observed global mean historical warming (Fig. 22). We hypothesize that representations of aerosol-cloud interactions have become so complex that the emergent aerosol ERF cannot be readily predicted or adjusted, and hence models must accept limitations in their historical simulations to maintain their most realistic representation of aerosol-cloud physical processes.

There is a clear relationship between the historical response and the imposed forcing across the ACCMIP and CMIP5 models (Fig. 22). The regression through the mean of all models is 0.57 C response per W m^{-2} forcing (0.53 C per W m^{-2} without the three models that do not include aerosol indirect effects: top three sets of points in the lower panel), though there is substantial variation from this fit due to variations in modeled climate sensitivity and ocean heat uptake. Although uncertainties in both those terms are substantial, the results suggest that given the temperature response per unit forcing in the current generation of ACCMIP/CMIP5 models, aerosol ERF from about −0.8 to −1.5 W m^{-2} (along with ozone RF of ∼ 0.3 W m^{-2}) is consistent with the observed historical warming. Account for the land temperature adjustment incorporated into these ERF estimates implies a bias-corrected ERF of about −0.9 to −1.6 W m^{-2}. An ERF in this range is in good agreement with most of the inverse calculations and satellite-constrained model studies discussed previously (Sect. 5.2).
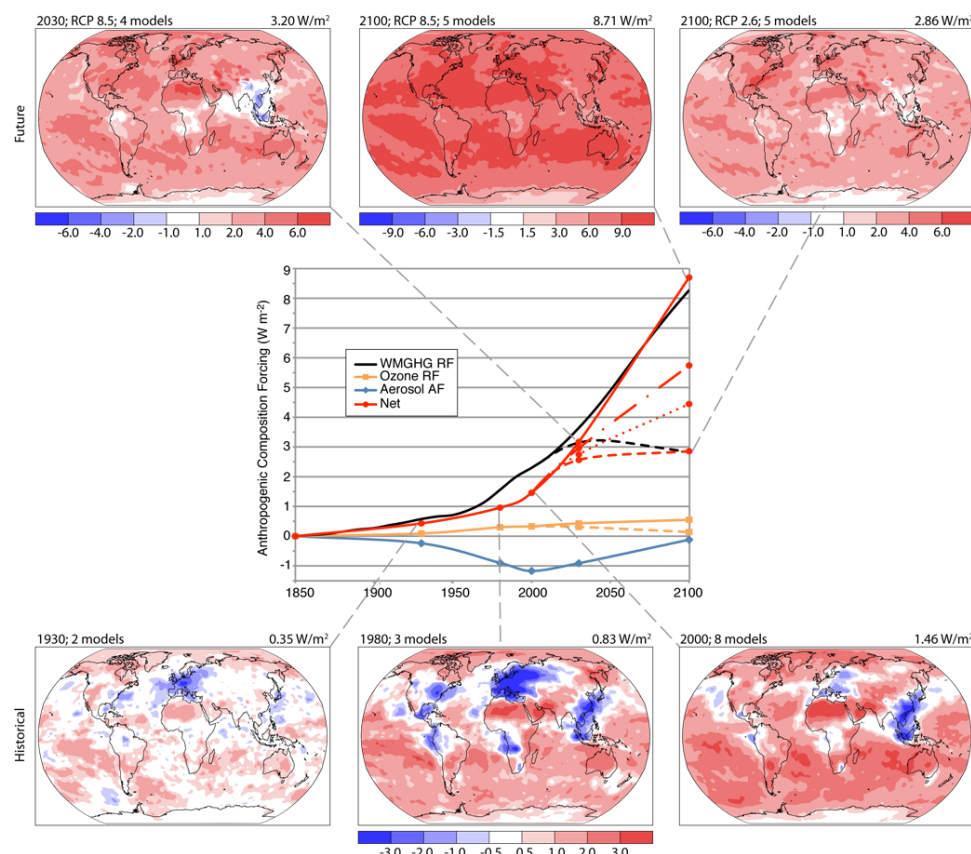
Fig. 21. Maps of total anthropogenic composition forcing and timeseries of global mean by component. Top row: maps of 2030 and 2100; bottom row: maps of 1930, 1980 and 2000 (note change in scale relative to top row); center panel: timeseries of global mean forcing with symbols indicating the times at which ACCMIP simulations were performed (solid lines are RCP8.5, dashed lines RCP2.6, dash-dot RCP6.0, and dotted RCP4.5). RCPs 2.6, 4.5 and 6.0 net forcings at 2100 are approximate values using aerosol ERF projected for RCP8.5. Uncertainty ranges for the global means are given in Table 8.

7 Conclusions

We have evaluated the ACCMIP aerosol models against observations of the AOD climatology and trends. The models generally capture the observed magnitude of present day AOD within 30 %, though with a tendency to be biased low, and represent much of the spatial and seasonal structure (correlations typically 0.5 to 0.7). Fine-mode AOD and AAOD show less agreement with observations, with large underestimates for AAOD. Analysis of AOD trends suggests that most models realistically reproduce changes over North America and Europe during the last few decades, while results for Asia are more ambiguous as the observations are highly sensitive to the satellite analysis methodology. Many of the ACCMIP models do not include nitrate and SOA, which is a primary cause of the general low bias in aerosols, though there are also clear underestimates of AOD in biomass burning regions. CMIP5 models also appear to have an overall low bias in AOD, which again is partially attributable to the lack of nitrate and SOA in many models, and most ACCMIP and CMIP5 models likely underestimate aerosol RF. Further-

more, there is evidence from one model that forcing by SOA induced by land-use changes may be large, and from two models that nitrate may become the largest aerosol RF component in the latter part of the 21st century, highlighting the need for more models to represent these processes.

We have used the ERF metric to characterize aerosol forcing, and shown that the relative variation in ERF across models is approximately equal to, or even less than, that in aerosol RF at both global and regional scales. As the majority of aerosol forcing is indirect, RF alone provides a very incomplete portrayal of aerosol impacts. A disadvantage of ERF is that it cannot be diagnosed in transient simulations with multiple forcings the way RF can. Trends in these metrics are in opposite directions for some periods. In particular, during 1980 to 2000, aerosol RF becomes less negative while aerosol ERF continues its historical trend becoming more negative. The ERF trends appear to be largely driven by increases in Asian emissions of both sulfate precursors and BC that largely offset one another's RF while sulfate appears to have a stronger impact on clouds than BC. This is consistent with the large ERF seen downwing of east Asia and
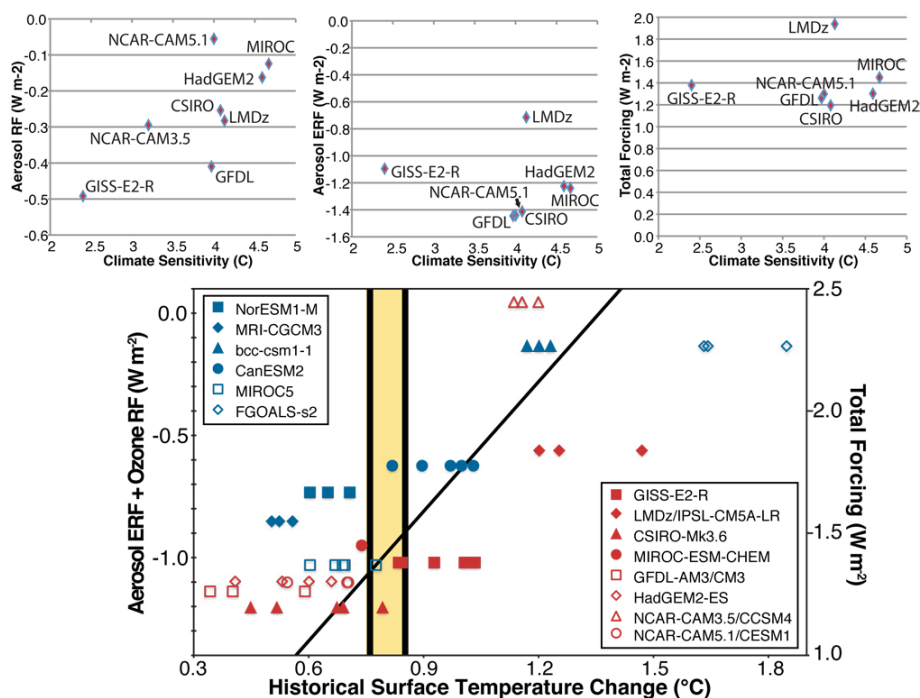
Fig. 22. Relationship between 1850 to 2000 forcings and ECS (top row) or historical warming (bottom panel) in the indicated models. Total forcing includes RF values of 2.30, −0.17, 0.07 and 0.10 W m^{-2} for WMGHGs, land-use, stratospheric water, and solar irradiance, respectively. In the lower panel, temperature changes are differences between 1850–1859 and 1996–2005 from observations (CRU left vertical line, GISS right vertical line) and models. For CMIP5 models (Table G2), the ACCMIP mean ozone forcing of 0.35 W m^{-2} is used. Points show temperatures from individual realizations. The black line is the regression through the model means.

with sulfate being a more abundant and soluble aerosol than BC, and with model results showing that the ERF/RF ratio is greater for sulfate than for BC (or OC). However, much more work is needed to characterize ERF by emitted species and region. Given current uncertainties, ERF cannot be adjusted to account for missing components, further emphasizing the need for complete representations of aerosol types.

Our results suggest that while pollution controls in North American and Europe have reduced aerosol forcing, increases in Asian emissions have more than compensated so that globally there has not yet been an unmasking of WMGHG forcing via aerosol reductions. Instead, the continued increase in negative aerosol ERF may have contributed to relative slow rates of global warming during recent years.

The models show a large reduction in negative aerosol ERF at 2030 and 2100. With the pollution controls envisioned under the RCPs, the combined aerosol ERF and ozone RF becomes 5 % or less of the total anthropogenic composition forcing in the latter part of the century. This would lead to an almost total unmasking of WMGHG forcing. Under the RCP2.6 scenario, in fact, the projected change in aerosol ERF from 2000 to 2100 is actually larger than the increased WMGHG forcing. Such a complete reduction in pollutant emissions may be overly optimistic, however, as current legislation certainly does not set the world on such a track (Shindell et al., 2012; Pozzer et al., 2012).

Appendix A

Clear-sky versus all-sky AOD

All-sky conditions determine forcing but observations are usually restricted to clear-sky conditions. Differences between all-sky and clear-sky AOD were only reported for a few models, however. These differences are especially large for the GISS-E2-R model, while they are fairly minor for the other two models reporting both values (CSIRO-Mk3.6 and GISS-E2-R-TOMAS). For the GISS-E2-R model, large positive biases versus observations were found when using all-sky AODs, which are replaced by small negative biases using clear-sky values. The incorporation of large aerosol water uptake in cloudy regions, which causes strong non-linearities in optical properties at very high relative humidity (RH) values, seems to have a large influence on the all-sky values in this model. In support of this hypothesis, we note that GISS-E2-R all-sky values are much larger than clear-sky values for sulfate, moderately larger for sea-salt and nitrate, and quite similar for other components, and hence follow the relative solubility of the different species. GISS-E2-R calculates clear-sky AOD by including only AOD values calculated in model locations where clouds are not present, rather than performing a global calculation with clouds removed from the model. This technique is more comparable to the sampling
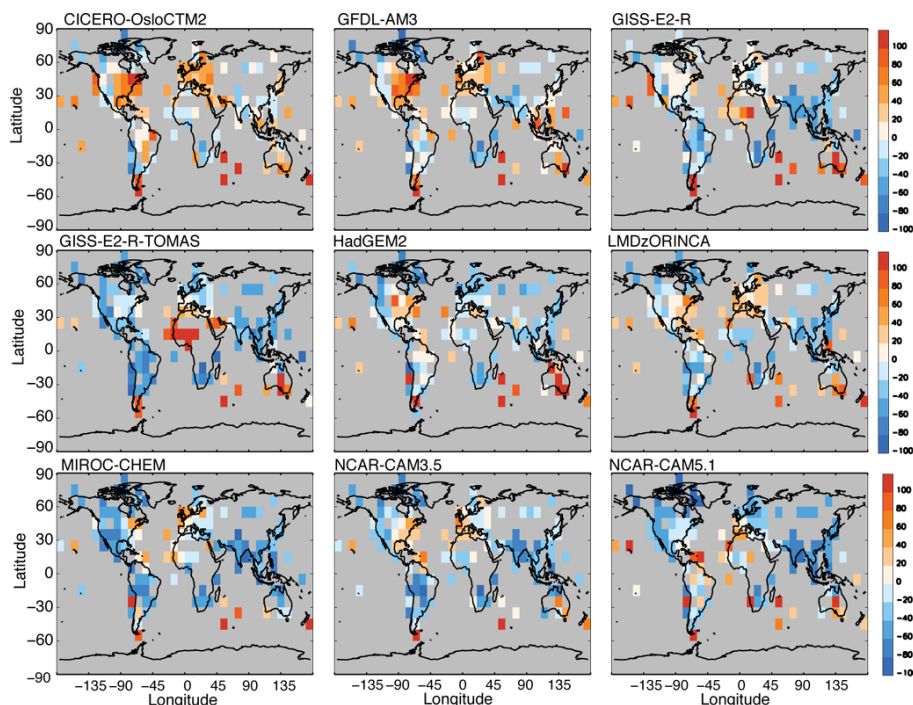
Fig. B1. Bias (%) in models for present-day annual average AOD relative to AeroNet for locations with AeroNet observations. Clear-sky AOD is used for the GISS-E2-R model.

of the satellites, but leads to spatial and temporal differences in sampling compared with all-sky calculations in addition to the role of the clouds themselves. It appears that in GISS-E2-R, there is in general substantially more AOD when and where clouds are present. In contrast, for CSIRO-Mk3.6, AOD is less where clouds are present. Several competing factors are at work, including increased wet removal rates and cloud scavenging in cloudy areas leading to lower AOD but increased in-cloud oxidation rates and RH in cloudy areas leading to higher AOD, so it is perhaps not surprising that clear-sky versus all-sky differences can be of either sign.

For other models, HadGEM2 reports that its AOD is clear-sky, while MIROC-CHEM and CICERO-OsloCTM2 report all-sky but note clear-sky is very similar in their models. In addition to the issue of sampling that could bias RH if all-sky rather than only clear-sky areas are sampled, the RH used in calculation of water uptake is also important and diverges between models. For example, LMDzORINCA uses the clear-sky humidity to compute aerosol growth, while NCAR-CAM5.1 and MIROC-CHEM report that all-sky grid cell mean RH is used. Concerted efforts to diagnose model behavior for the relevant parameters, including how aerosol mass, cloud scavenging and RH are apportioned between clear and cloudy portions of model grid boxes, will be required to better understand the divergence across models in the all-sky/clear-sky AOD ratios.

Appendix B

Regional comparison of modeled AOD with AeroNet

As discussed in the main text, the models show regional biases that are consistent across models in some areas. In particular, Fig. 4 shows that virtually all models have large negative biases over east Asia. The only two models that do not greatly underestimate AOD over east Asia have large positive biases over Europe and North America, where the other models match AeroNet relatively well. Hence the biases appear systematic across models. Additional spatial information can be seen in Fig. B1, which shows that, as noted in the main text, biases over remote island stations in the central Pacific and Indian Oceans, over the southern tip of South America, and over Australia and New Zealand are nearly always positive, while biases over the Russian and especially the Western Hemisphere Arctic are typically negative.

Appendix C

Regional comparison of modeled AAOD with observations

We quantitatively evaluate AAOD on a regional basis against both OMI and AeroNet observations. We calculate the ratio of regionally averaged modeled to observed AAOD. The ratio values using OMI data are calculated including only
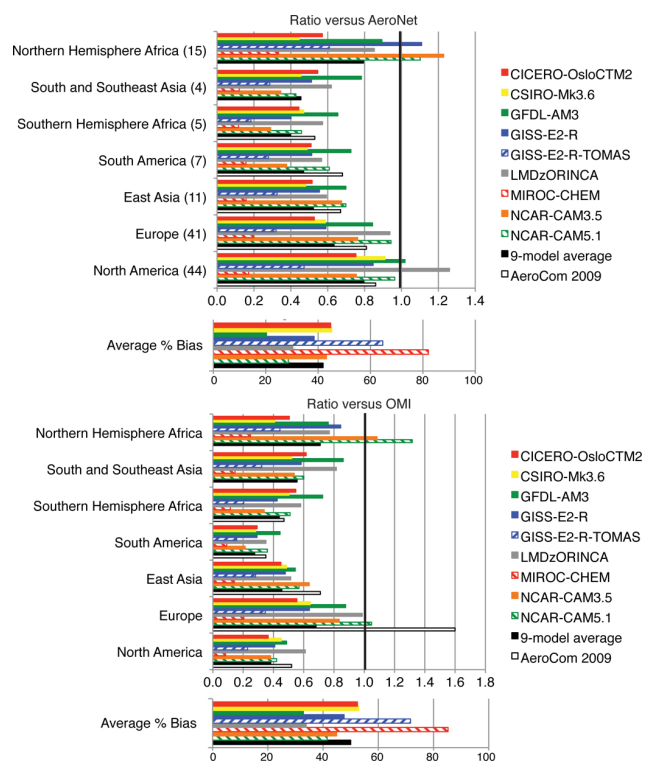
Fig. C1. Ratio of regional average model to retrieved AeroNet (top) and OMI land-area (bottom) clear-sky AAOD at 550 nm for the AC-CMIP models. Number of measurement sites is given for AeroNet for each region. The AeroNet data are for 1996–2006, v2 level 2, with annual averages for each year used if more than eight months were present, and monthly averages required more than 10 days of measurements. The values at 550 nm were determined using the 0.44 and 0.87 μm Angstrom parameters. Regions are shown in Fig. 6 and are defined as North America (130 to 70° W; 20 to 55° N), Europe (15° W to 45° E; 30 to 70° N), East Asia (100 to 160° E; 30 to 70° N), South America (85 to 40° W; 34 to 2° S), Southern Hemisphere Africa (20° W to 45° E; 34 to 2° S), South and Southeast Asia (60 to 110° E; 10 to 30° N) and Northern Hemisphere Africa (20° W to 60° E; 0 to 30° N; which includes the Arabian Peninsula). Also shown is the average of 14 AeroCom models from the analysis of Koch et al. (2009) (except for NH Africa and S/SE Asia, which were not calculated for AeroCom). Average % biases shown at the bottom of each analysis are average absolute values over the 7 regions for a given ACCMIP model.

land-area locations to be more comparable with AeroNet. For comparison, we show similar ratios calculated from an earlier AeroCom model intercomparison (Koch et al., 2009) (though that study did not include Northern Hemisphere Africa or South and Southeast Asia). Models underestimate AAOD in general compared with AeroNet (Fig. C1). This is especially the case for south and southeast Asia, South America, and Southern Hemisphere Africa, where the model average AAOD is less than half the observed, but also for east Asia. Note, however, that these four areas have very limited coverage of AeroNet sites. Model performance is gen-

erally better for North America, Europe and Northern Hemisphere Africa. In comparison with OMI land-area AAOD, the models show similar biases to AeroNet except over North America, where underestimations relative to OMI are much larger. In comparison with either set of observations, the GFDL-AM3 model shows the smallest overall underestimate of present-day AAOD, with the largest underestimates in GISS-E2-R-TOMAS and MIROC-CHEM. Biases in different regions tend to be systematic for a given model. For the four regions with the most pronounced low biases (south and southeast Asia, South America, Southern Hemisphere Africa and east Asia), every model examined is biased low in comparison with either AeroNet or OMI. High biases are seen most commonly for Northern Hemisphere Africa, where 2 of 9 models are high relative to OMI and 3 are high relative to AeroNet. The AAOD in this region, which includes the Sahara and Arabian deserts, is dominated by mineral dust aerosols rather than BC (which contributes less than 5 % of total BC+dust AAOD), suggesting that dust loading or absorption may be too high in these models (though biases are comparatively small). We note that calculations of the regionally averaged ratios of modeled to observed AAOD at individual locations always show larger values. This indicates that the model biases are most pronounced at locations with large AAOD.

Biases in the ACCMIP models are often fairly similar to those in the 2009 AeroCom analysis, though typically somewhat larger in comparison with AeroNet. Comparisons with OMI for Europe are very different, however, with the very large overestimate in the AeroCom models replaced by a smaller but still large underestimate in the ACCMIP models. Comparison of the models against OMI over the full land and ocean areas reduces the model biases in all regions. In particular, multi-model mean ratios increase by 0.06 for east Asia, 0.07 for North America, and 0.08 for south and southeast Asia. This accounts for a good portion of the difference for North America relative to the 2009 AeroCom analysis, which did not remove oceanic areas, though it accounts for only a small portion of the east Asia difference.

Appendix D

BC albedo forcing methodology and comparison with prior studies

BC albedo forcing was calculated in offline simulations conducted using prescribed meteorology from 1994–2000, with spinup from 1994–1995 and analysis (averaging) over 1996–2000. Black carbon and dust deposition fields from each ACCMIP model were prescribed with monthly resolution (annually-repeating), and linearly interpolated to the model timestep. The land simulations applied the NCAR Community Land Model 4 (Lawrence et al., 2011), using bias-corrected atmospheric forcing data from (Qian et al., 2006),

and run at 1.9×2.5 degree resolution. A sensitivity test run at 0.9×1.25 degree resolution showed global mean values within 1 % of those obtained at the coarser resolution. The sea-ice temperature, wind, specific humidity, and surface pressure forcing data come from NCEP, radiation data are from GISS, and precipitation data from the GCGCS blended product. The land snow treatments of aerosol processes and radiative transfer are described by Flanner et al. (2007) and Lawrence et al. (2011), and the new sea-ice aerosol and radiation treatments are described by Holland et al. (2012). The snow and sea-ice fields generated with these offline configurations agree better with observed conditions during this time period than those simulated with coupled land-ocean-atmosphere simulations, but the precipitation and aerosol deposition fluxes are less compatible with each other than in coupled aerosol-climate simulations. The influence of this incompatibility on simulated surface snow BC concentrations and radiative forcing is somewhat mitigated by the use of temporally-smoothed monthly aerosol deposition fields. Additional analyses of the BC deposition fields, including extensive comparisons with both recent snowpack measurements and historical trends from ice-cores, as well as further discussion of the BC albedo forcing can be found in Lee et al. (2013).

The BC albedo forcing estimates reported here are smaller than those reported in previous studies (e.g. Flanner et al., 2007, 2009) because of the offline configuration that was applied, which produces less snow cover (and hence less area over which the forcing can operate) in the Tibetan Plateau and other parts of Asia. Because the forcing was quantified using snow and ice states which are representative of 1996–2000, and likely diminished relative to previous periods, actual BC snow forcings in 1850 may have been slightly greater (Lawrence et al., 2012). Additionally, BC albedo forcing is sensitive to the methodology of the calculation, with values reported from calculations internal to three of the ACCMIP models showing substantial variations from these offline results in their magnitude, though their time-dependence is similar (Lee et al., 2013).

Appendix E

Calculation of indirect forcing attributable to specific aerosol components

The indirect aerosol forcing attributable to specific aerosol types was analyzed in only two ACCMIP models as these experiments are very computationally expensive. In the GISS-E2-R model, with all other conditions set at year 2000 values, sulfate, BC and OA were individually removed and their direct RF and ERF were diagnosed in 50-yr atmosphere simulations with fixed ocean conditions. At the same time, a cloud forcing diagnostic was saved that calculates the flux perturbation due to the model's clouds relative to zero clouds ev-

Table E1. Direct RF, cloud forcing and ERF caused by individual aerosol components and all aerosols in the GISS-E2-R and MIROC-CHEM models (W m^{-2}).

GISS-E2-R	Direct RF	Cloud forcing	ERF
Sulfate	-0.37 ± 0.03	-0.45 ± 0.03	-0.61 ± 0.05
OC	-0.14 ± 0.03	-0.17 ± 0.03	-0.25 ± 0.05
BC	0.24 ± 0.03	0.23 ± 0.03	0.46 ± 0.04
Total	-0.49 ± 0.03	-1.13 ± 0.03	-1.10 ± 0.05
MIROC-CHEM			
BCff+bf			-0.12 ± 0.06
OAff+bf			-0.16 ± 0.06
BB			-0.14 ± 0.06
Sulfate			-1.68 ± 0.06
Total			-1.80 ± 0.06

erywhere, so that the difference between the 2000 control run cloud forcing and the cloud forcing in the aerosol perturbation experiments yields the cloud forcing due to aerosols. In the MIROC-CHEM model, transient diagnostics of the indirect effect were obtained via simulations with only a single aerosol type varying in time. Forcings are calculated as the difference between the last 20 yr of the simulation (1991–2010) and the first 20 yr (1851–1870).

The results of these calculations are presented in Table D1. The GISS simulations reveal that for OA and BC, cloud forcing is equivalent in magnitude to RF, and ERF is statistically equivalent to a linear sum of the direct RF and the cloud forcing (though the mean ERF estimated for OA is ~ 20 % less than the direct plus cloud RF). In contrast, for sulfate the cloud forcing is larger than the direct RF, and the ERF is clearly less (~ 25–30 %) than the direct plus cloud forcings. This suggests that the greater solubility of sulfate causes it to have an enhanced cloud forcing relative to OA by more efficiently serving as cloud condensation nuclei, and that the overall effect of BC on clouds is dominated by thermodynamic effects of local heating rather than microphysical effects in this model. In the all aerosol case, cloud forcing is much greater than direct RF, which is likely due to the logarithmic dependence of cloud droplet number concentration on nucleation sites (Gultepe and Isaac, 1999) leading to a greater response at the low aerosol numbers reached when all aerosols are removed simultaneously. The all aerosol ERF is, like sulfate (and perhaps OA), approximately 25–30 % less than the sum of the direct and cloud forcings. This suggests that other rapid responses, such as adjustment of the temperature lapse rate and water vapor concentration, compensate for some of the cloud response.

In contrast, in the MIROC-CHEM simulations, BC causes a negative indirect forcing, but forcing by sulfate strongly dominates the total aerosol indirect forcing. The sum of all aerosol forcings is again weaker than the sum of individual components, consistent with the non-linear response at low aerosol number densities discussed above.

Table F1. Emissions or removal and burdens for sulfate and BC.

	SO_4 removal (Tg yr^{-1})	BC emissions (Tg yr^{-1})	SO_4 burden (Tg)	BC burden (Tg)	SO_4 lifetime (days)	BC lifetime (days)
1850						
CICERO-OsloCTM2	61	3.1	1.4	0.07	8.6	8.0
GFDL-AM3	35	3.1	0.7	0.05	7.4	6.2
GISS-E2-R	74	4.0	0.7	0.05	3.5	5.0
GISS-E2-R-TOMAS	81	3.1	1.3	0.07	5.9	8.0
HadGEM2	60	3.1	0.6	0.10	3.8	12.3
MIROC-CHEM	62	3.1	1.1	0.04	6.4	4.4
NCAR-CAM3.5	58	3.1	0.6	0.05	3.7	5.9
NCAR-CAM5.1		3.1		0.03		4.0
Average	61	3.2	0.9	0.06	5.6	6.8
Std dev	13	0.4	0.3	0.02	2.0	2.7
2000						
CICERO-OsloCTM2	132	7.8	2.8	0.17	7.6	7.9
GFDL-AM3	104	7.7	2.3	0.13	7.9	6.2
GISS-E2-R	153	8.8	1.3	0.14	3.0	5.7
GISS-E2-R-TOMAS	167	7.8	2.4	0.17	5.2	8.3
HadGEM2	164	7.7	1.5	0.32	3.4	15.2
MIROC-CHEM	198	7.8	2.2	0.11	4.1	5.2
NCAR-CAM3.5	163	7.7	1.7	0.13	3.9	5.9
NCAR-CAM5.1		7.8		0.08		3.9
Average	155	7.9	2.0	0.16	5.0	7.4
Std dev	27	0.4	0.5	0.07	2.0	3.4
2000–1850						
CICERO-OsloCTM2	71	4.7	1.3	0.10	−0.9	−0.1
GFDL-AM3	69	4.7	1.5	0.08	0.5	−0.1
GISS-E2-R	80	4.8	0.5	0.08	−0.5	0.7
GISS-E2-R-TOMAS	85	5.7	1.1	0.13	−0.7	0.4
HadGEM2	104	4.6	0.9	0.16	−0.5	0.3
MIROC-CHEM	136	4.7	1.1	0.07	−2.3	0.8
NCAR-CAM3.5	105	4.7	1.1	0.08	0.2	0.0
NCAR-CAM5.1		4.7		0.05		−0.1
2000–1850 mean total	93	4.8	1.1	0.09	−0.6	0.2
std dev total	24	0.4	0.3	0.04	0.9	0.4
2000–1850 mean % change	157	152	130	165	−10	4
std dev % change	46	17	56	21	14	8

The sign of the indirect impact of BC differs in these two models. There is a large range seen in the literature, which shows values ranging from about -0.35 W m^{-2} to $+0.3$ W m^{-2} just for the effect of BC on mixed-phase or ice clouds (Penner et al., 2009; Liu et al., 2009). In addition, while many climate models find a substantial negative BC indirect forcing, observationally-constrained estimates often indicate this forcing is positive (Ruckstuhl et al., 2010; Kaufman and Koren, 2006). This has led recent assessments to conclude that the most likely range for BC's indirect forcing is -0.4 to $+0.4$ W m^{-2} (UNEP WMO, 2011; Shindell et al., 2012). Hence reliable quantification of aerosol ERF due to

individual components in general, and due to BC in particular, remains a substantial challenge for the community.

Appendix F

Analysis of sulfate and BC burden and lifetime changes

Examination of changes in sulfate removal and BC emissions shows large variations across models for sulfate removal (Table F1), implying substantial differences in sulfur emissions (as these are balanced) probably due to varying dimethyl sulfide emission responses to climate change. In contrast, BC

Table G1. Aerosol AOD and estimated RF from CMIP5 models.

Model	Correlation vs. AeroNet	Bias vs. AeroNet	Estimated RF (NRF × dAOD)	Calculated RF (ACCMIP)
ACCESS1-0	0.62	−31	−0.20	
ACCESS1-3	0.47	−40	−0.18	
BNU-ESM	0.57	3		
CCSM4	0.59	−1	−0.25	−0.29
CESM1-CAM5	0.52	−23	−0.15	−0.06
CSIRO-Mk3-6-0	0.60	56	−0.51	−0.25
GFDL-CM3	0.51	9		−0.41
GFDL-ESM2G	0.53	−3	−0.27	
GFDL-ESM2M	0.51	−3	−0.27	
GISS-E2-H	0.56	78*	−0.54	
GISS-E2-R	0.54	79*	−0.56	−0.49
HadGEM2-CC	0.56	−11	−0.07	
HadGEM2-ES	0.57	−12	−0.25	−0.16
IPSL-CM5A-LR	0.55	−10	−0.31	−0.28
IPSL-CM5A-MR	0.56	−12	−0.34	
IPSL-CM5B-LR	0.55	−11		
MIROC4h	0.43	−42		
MIROC5	0.52	−20	−0.17	
MIROC-ESM	0.46	−27	−0.21	
MIROC-ESM-CHEM	0.45	−25	−0.17	−0.12
MRI-CGCM3	0.61	−55	−0.09	
NorESM1-M	0.65	1	−0.26	
NorESM1-ME	0.64	1	−0.26	
Mean across models	0.55	−13*	−0.27	

The Normalized RF used here is $-7.6\,\mathrm{W\,m^{-2}}$ per unit AOD change.
* We calculate mean bias excluding the GISS models as those models have much higher all-sky than clear-sky AOD (see Sect. 3.1).

emissions or removal changes are quite consistent across models. Both sulfate and BC burden changes show comparable spread, however, as do changes in lifetimes. Sulfate lifetimes typically decrease while BC lifetimes typically increase, however. The decrease in sulfate lifetime is consistent with increases in precipitation seen in most models (Lamarque et al., 2013). Changes in lifetime reflect both the influence of climate change and the shifting spatial distribution of emissions (as lifetimes vary regionally).

One reason that AOD changes may not provide a better indicator of forcing, especially when AOD changes are small, is because AOD changes include both cooling and warming agents. To see if these might be separated, we also examined correlations between the 2000 versus 1850 RF from sulfate and the change in sulfate burden and between BC ff+bf forcing and BC burden changes. Correlations between sulfate burden changes and sulfate RF are fairly weak at $r^2 = 0.44$, while correlations between BC burden changes and BC forcing are near zero. This suggests that variations in the optical properties of the aerosols are too large to allow burden changes alone to provide a good indicator of RF, as in prior studies (Schulz et al., 2006).

Appendix G

AOD in the CMIP5 models

Time varying AOD is available from many CMIP5 models. We compare the present-day AOD with AeroNet following the methods described in Sect. 3.2. We also evaluate the decadal mean AOD change between 1850 and 2000, and from that derive an estimated aerosol RF using the normalized RF (NRF) calculated from the ACCMIP models (see Sect. 5.3). As discussed in the main text, this estimate produces a larger mean forcing than that diagnosed directly in the ACCMIP models. This may stem from different experimental setups, such as the coupled ocean, and imperfections in the NRF-based estimates. Results are presented in Table G1.

Analysis of ERF from six additional CMIP5 models was possible based on the simulations and methodology described in Sect. 5.2. Results are presented in Table G2. Note that FGOALS-s2 and bcc-csm1-1 do not include aerosol indirect effects.

Table G2. Additional ERF values from CMIP5 models.

model	ERF
bcc-csm1-1	−0.38
CanESM2	−0.87
FGOALS-s2	−0.38
MIROC5	−1.28
MRI-CGCM3	−1.10
NorESM1-M	−0.98

Note that bcc-csm1-1 and FGOALS-s2 do not include aerosol indirect effects on clouds.

References

Acker, J. G. and Leptoukh, G.: Online Analysis Enhances Use of NASA Earth Science Data, Eos, Trans. AGU, 88, 14–17, 2007.

Andrews, T. and Forster, P. M.: CO_2 forcing induces semi-direct effects with consequences for climate feedback interpretations, Geophys. Res. Lett., 35, L04802, doi:10.1029/2007gl032273, 2008.

Andrews, T., Gregory, J., Webb, M., and Taylor, K.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophys. Res. Lett., 39, L09712, doi:10.1029/2012GL051607, 2012.

Bellouin, N., Jones, A., Haywood, J., and Christopher, S. A.: Updated estimate of aerosol direct radiative forcing from satellite observations and comparison against the Hadley Centre climate model, J. Geophys. Res., 113, D10205, doi:10.1029/2007JD009385, 2008.

Bellouin, N., Rae, J., Jones, A., Johnson, C., Haywood, J., and Boucher, O.: Aerosol forcing in the Climate Model Intercomparison Project (CMIP5) simulations by HadGEM2-ES and the role of ammonium nitrate, J. Geophys. Res.-Atmos., 116, D20206, doi:10.1029/2011JD016074, 2011.

Bitz, C., Shell, K., Gent, P., Bailey, D., Danabasoglu, G., Armour, K., Holland, M., and Kiehl, J.: Climate Sensitivity of the Community Climate System Model, Version 4, J. Climate, 25, 3053–3070, doi:10.1175/JCLI-D-11-00290.1, 2012.

Boer, G. J. and Yu, B.: Climate sensitivity and response, Clim. Dynam., 20, 415–429, doi:10.1007/s00382-002-0283-3, 2003.

Bowman, K., Shindell, D., Worden, H., Lamarque, J. F., Young, P. J., Stevenson, D., Qu, Z., de la Torre, M., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Horowitz, L. W., Josse, B., Lee, Y. H., MacKenzie, I., Myhre, G., Nagashima, T., Naik, V., Plummer, D., Rumbold, S., Skeie, R., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., Zeng, G., Kulawik, S., and Worden, J.: Observational constraints on ozone radiative forcing from the Atmospheric Chemistry Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys. Discuss., 12, 23603–23644, doi:10.5194/acpd-12-23603-2012, 2012.

Brohan, P., Kennedy, J., Harris, I., Tett, S., and Jones, P.: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, J. Geophys. Res.-Atmos., 111, D12106, doi:10.1029/2005JD006548, 2006.

Chan, P. K., Zhao, X., and Heidinger, A. K.: Long-term Aerosol Climate Data Record (CDR) Derived from Operational AVHRR Satellite Observations, Dataset Papers in Geosciences, in press, 2013.

Chuang, C. C., Penner, J. E., Prospero, J. M., Grant, K. E., Rau, G. H., and Kawamoto, K.: Cloud susceptibility and the first aerosol indirect forcing: Sensitivity to black carbon and aerosol concentrations, J. Geophys. Res., 107, 4564, doi: 10.1029/2000JD000215, 2002.

Church, J., White, N., Konikow, L., Domingues, C., Cogley, J., Rignot, E., Gregory, J., van den Broeke, M., Monaghan, A., and Velicogna, I.: Revisiting the Earth's sea-level and energy budgets from 1961 to 2008, Geophys. Res. Lett., 38, L18601, doi:10.1029/2011GL048794, 2011.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch,

S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev., 4, 1051–1075, doi:10.5194/gmd-4-1051-2011, 2011.

Conley, A. J., Lamarque, J.-F., Vitt, F., Collins, W. D., and Kiehl, J.: PORT, a CESM tool for the diagnosis of radiative forcing, Geosci. Model Dev. Discuss., 5, 2687–2704, doi:10.5194/gmdd-5-2687-2012, 2012.

Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., Dias, P. L. d. S., Wofsy, S. C., and Zhang, X.: Couplings Between Changes in the Climate System and Biogeochemistry, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Donner, L., Wyman, B., Hemler, R., Horowitz, L., Ming, Y., Zhao, M., Golaz, J., Ginoux, P., Lin, S., Schwarzkopf, M., Austin, J., Alaka, G., Cooke, W., Delworth, T., Freidenreich, S., Gordon, C., Griffies, S., Held, I., Hurlin, W., Klein, S., Knutson, T., Langenhorst, A., Lee, H., Lin, Y., Magi, B., Malyshev, S., Milly, P., Naik, V., Nath, M., Pincus, R., Ploshay, J., Ramaswamy, V., Seman, C., Shevliakova, E., Sirutis, J., Stern, W., Stouffer, R., Wilson, R., Winton, M., Wittenberg, A., and Zeng, F.: The Dynamical Core, Physical Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL Global Coupled Model CM3, J. Climate, 24, 3484–3519, doi:10.1175/2011JCLI3955.1, 2011.

Flanner, M. G., Zender, C. S., Randerson, J. T., and Rasch, P. J.: Present-day climate forcing and response from black carbon in snow, J. Geophys. Res., 112, D11202, doi:10.1029/2006JD008003, 2007.

Flanner, M. G., Zender, C. S., Hess, P. G., Mahowald, N. M., Painter, T. H., Ramanathan, V., and Rasch, P. J.: Springtime warming and reduced snow cover from carbonaceous particles, Atmos. Chem. Phys., 9, 2481–2497, doi:10.5194/acp-9-2481-2009, 2009.

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.

Geogdzhayev, I., Mishchenko, M., Terez, E., Terez, G., and Gushchin, G.: Regional advanced very high resolution radiometer-derived climatology of aerosol optical thickness and size, J. Geophys. Res.-Atmos., 110, D23205, doi:10.1029/2005JD006170, 2005.

Gettelman, A., Kay, J., and Shell, K.: The Evolution of Climate Sensitivity and Climate Feedbacks in the Community Atmosphere Model, J. Climate, 25, 1453–1469, doi:10.1175/JCLI-D-11-00197.1, 2012.

Ghan, S. J., Liu, X., Easter, R. C., Rasch, P., and Yoon, J.-H.: Toward a Minimal Representation of Aerosol Direct, Semi-Direct and Indirect Effects: Comparative Decomposition, J. Climate, in press, 2013.

Gultepe, I. and Isaac, G. A.: Scale effects on averaging cloud droplet and aerosol number concentrations: observations and models, J. Climate, 12, 1268–1279, 1999.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N. Y., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R. L., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M., and Zhang, S.: Efficacy of Climate Forcings, J. Geophys. Res., 110, D18104, doi:10.1029/2005JD005776, 2005.

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., and Medina-Elizade, M.: Global temperature change, P. Natl. Acad. Sci., 103, 14288–14293, 2006.

Hansen, J., Sato, M., Kharecha, P., and von Schuckmann, K.: Earth's energy imbalance and implications, Atmos. Chem. Phys., 11, 13421–13449, doi:10.5194/acp-11-13421-2011, 2011.

Holland, M., Bailey, D., Briegleb, B., Light, B., and Hunke, E.: Improved Sea Ice Shortwave Radiation Physics in CCSM4: The Impact of Melt Ponds and Aerosols on Arctic Sea Ice, J. Climate, 25, 1413–1430, doi:10.1175/JCLI-D-11-00078.1, 2012.

Hsu, N. C., Tsay, S.-C., King, M., and Herman, J. R.: Aerosol properties over bright-reflecting source regions, IEEE T. Geosci. Remote Sens., 42, 557–569, 2004.

Jacobson, M. Z.: Control of fossil-fuel particulate black carbon and organic matter, possibly the most effective method of slowing global warming, J. Geophys. Res., 107, 4410, doi:10.1029/2001JD001376, 2002.

Kahn, R. A., Nelson, D. L., Garay, M. J., Levy, R. C., Bull, M. A., Diner, D. J., Martonchik, J. V., Paradise, S. R., Hansen, E. G., and Remer, L. A.: MISR Aerosol Product Attributes and Statistical Comparisons With MODIS, Geosci. Remote Sens., 47, 4095–4114, 2009.

Kaufman, Y. and Koren, I.: Smoke and pollution aerosol effect on cloud cover, Science, 313, 655–658, doi:10.1126/science.1126232, 2006.

Kiehl, J.: Twentieth century climate model response and climate sensitivity, Geophys. Res. Lett., 34, L22710, doi:10.1029/2007GL031383, 2007.

Koch, D., Schulz, M., Kinne, S., McNaughton, C., Spackman, J. R., Balkanski, Y., Bauer, S., Berntsen, T., Bond, T. C., Boucher, O., Chin, M., Clarke, A., De Luca, N., Dentener, F., Diehl, T., Dubovik, O., Easter, R., Fahey, D. W., Feichter, J., Fillmore, D., Freitag, S., Ghan, S., Ginoux, P., Gong, S., Horowitz, L., Iversen, T., Kirkevåg, A., Klimont, Z., Kondo, Y., Krol, M., Liu, X., Miller, R., Montanaro, V., Moteki, N., Myhre, G., Penner, J. E., Perlwitz, J., Pitari, G., Reddy, S., Sahu, L., Sakamoto, H., Schuster, G., Schwarz, J. P., Seland, ø., Stier, P., Takegawa, N., Takemura, T., Textor, C., van Aardenne, J. A., and Zhao, Y.: Evaluation of black carbon estimations in global aerosol models, Atmos. Chem. Phys., 9, 9001–9026, doi:10.5194/acp-9-9001-2009, 2009.

Koch, D., Bauer, S. E., Del Genio, A., Faluvegi, G., McConnell, J. R., Menon, S., Miller, R. L., Rind, D., Ruedy, R., Schmidt, G. A., and Shindell, D.: Coupled Aerosol-Chemistry-Climate Twentieth-Century Transient Model Investigation: Trends in Short-Lived Species and Climate Responses, J. Climate, 24,

2693–2714, doi:10.1175/2011JCLI3582.1, 2011.

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

Lamarque, J.-F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J., and Tyndall, G. K.: CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model, Geosci. Model Dev., 5, 369–411, doi:10.5194/gmd-5-369-2012, 2012.

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.

Lawrence, D., Oleson, K., Flanner, M., Thornton, P., Swenson, S., Lawrence, P., Zeng, X., Yang, Z., Levis, S., Sakaguchi, K., Bonan, G., and Slater, A.: Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model, Journal of Advances in Modeling Earth Systems, 3, M03001, doi:10.1029/2011MS000045, 2011.

Lawrence, D., Oleson, K., Flanner, M., Fletcher, C., Lawrence, P., Levis, S., Swenson, S., and Bonan, G.: The CCSM4 Land Simulation, 1850–2005: Assessment of Surface Climate and New Capabilities, J. Climate, 25, 2240–2260, doi:10.1175/JCLI-D-11-00103.1, 2012.

Lee, Y. H. and Adams, P. J.: Evaluation of aerosol distributions in the GISS-TOMAS global aerosol microphysics model with remote sensing observations, Atmos. Chem. Phys., 10, 2129–2144, doi:10.5194/acp-10-2129-2010, 2010.

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.

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K.-H., Wang, M., Laszlo, I., Nakajima, T., and Maring, H.: Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective, Ann. Geophys., 27, 2755–2770, doi:10.5194/angeo-27-2755-2009, 2009.

Libardoni, A. and Forest, C.: Sensitivity of distributions of climate system properties to the surface temperature dataset, Geophys. Res. Lett., 38, D03204, doi:10.1029/2011GL049431, 2011.

Liu, X., Penner, J., and Wang, M.: Influence of anthropogenic sulfate and black carbon on upper tropospheric clouds in the NCAR CAM3 model coupled to the IMPACT global aerosol model, J. Geophys. Res.-Atmos., 114, D04204, doi:10.1029/2008JD010492, 2009.

Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A., Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A. M. L., Hess, P., Mahowald, N., Collins, W., Iacono, M. J., Bretherton, C. S., Flanner, M. G., and Mitchell, D.: Toward a minimal representation of aerosols in climate models: description and evaluation in the Community Atmosphere Model CAM5, Geosci. Model Dev., 5, 709–739, doi:10.5194/gmd-5-709-2012, 2012.

Lohmann, U., Rotstayn, L., Storelvmo, T., Jones, A., Menon, S., Quaas, J., Ekman, A. M. L., Koch, D., and Ruedy, R.: Total aerosol effect: radiative forcing or radiative flux perturbation?, Atmos. Chem. Phys., 10, 3235–3246, doi:10.5194/acp-10-3235-2010, 2010.

Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration, Atmos. Chem. Phys., 11, 1417–1456, doi:10.5194/acp-11-1417-2011, 2011.

Ming, Y. and Ramaswamy, V.: Nonlinear Climate and Hydrological Responses to Aerosol Effects, J. Climate, 22, 1329–1339, doi:10.1175/2008jcli2362.1, 2009.

Murphy, D. M., Solomon, S., Portmann, R. W., Rosenlof, K. H., Forster, P. M., and Wong, T.: An observationally based energy balance for the Earth since 1950, J. Geophys. Res., 114, D17107, doi:10.1029/2009JD012105, 2009.

Penner, J. E., Chen, Y., Wang, M., and Liu, X.: Possible influence of anthropogenic aerosols on cirrus clouds and anthropogenic forcing, Atmos. Chem. Phys., 9, 879–896, doi:10.5194/acp-9-879-2009, 2009.

Pozzer, A., Zimmermann, P., Doering, U. M., van Aardenne, J., Tost, H., Dentener, F., Janssens-Maenhout, G., and Lelieveld, J.: Effects of business-as-usual anthropogenic emissions on air quality, Atmos. Chem. Phys., 12, 6915–6937, doi:10.5194/acp-12-6915-2012, 2012.

Qian, T., Dai, A., Trenberth, K., and Oleson, K.: Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations, Journal of Hydrometeorology, 7, 953–975, doi:10.1175/JHM540.1, 2006.

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

Remer, L., Kaufman, Y., Tanre, D., Mattoo, S., Chu, D., Martins, J., Li, R., Ichoku, C., Levy, R., Kleidman, R., Eck, T., Vermote, E., and Holben, B.: The MODIS aerosol algorithm, products, and validation, J. Atmos. Sci., 62, 947–973, doi:10.1175/JAS3385.1, 2005.

Remer, L., Kleidman, R. G., Levy, R., Kaufman, Y., Tanre, D., Mattoo, Martins, J., Ichoku, C., Koren, I., Yu, H., and Holben, B.: Global aerosol climatology from the MODIS satellite sensors, J. Geophys. Res., 113, D14S07, doi:10.1029/2007JD009661, 2008.

Rotstayn, L. D. and Lohmann, U.: Tropical rainfall trends and the indirect aerosol effect, J. Climate, 15, 2103–2116, 2002.

Rotstayn, L. D., Jeffrey, S. J., Collier, M. A., Dravitzki, S. M., Hirst, A. C., Syktus, J. I., and Wong, K. K.: Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: a study using single-forcing climate simulations, Atmos. Chem. Phys., 12, 6377–6404, doi:10.5194/acp-12-6377-2012, 2012.

Ruckstuhl, C., Norris, J., and Philipona, R.: Is there evidence for an aerosol indirect effect during the recent aerosol optical depth decline in Europe?, J. Geophys. Res.-Atmos., 115, 4699, doi:10.1029/2009JD012867, 2010.

Schulz, M., Textor, C., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Dentener, F., Guibert, S., Isaksen, I. S. A., Iversen, T., Koch, D., Kirkevåg, A., Liu, X., Montanaro, V., Myhre, G., Penner, J. E., Pitari, G., Reddy, S., Seland, Ø., Stier, P., and Takemura, T.: Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations, Atmos. Chem. Phys., 6, 5225–5246, doi:10.5194/acp-6-5225-2006, 2006.

Schwarz, J. P., Spackman, J. R., Gao, R. S., Watts, L. A., Stier, P., Schulz, M., Davis, S. M., Wofsy, S. C., and Fahey, D. W.: Global-scale black carbon profiles observed in the remote atmosphere and compared to models, Geophys. Res. Lett., 37, 812, doi:10.1029/2010GL044372, 2010.

Shindell, D. and Faluvegi, G.: Climate response to regional radiative forcing during the 20th century, Nat. Geosci., 2, 294–300, 2009.

Shindell, D., Schulz, M., Ming, Y., Takemura, T., Faluvegi, G., and Ramaswamy, V.: Spatial scales of climate response to inhomogeneous radiative forcing, J. Geophys. Res., 115, D19110, doi:10.1029/2010JD014108, 2010.

Shindell, D., Kuylenstierna, J. C. I., Vignati, E., van Dingenen, R., Amann, M., Klimont, Z., Anenberg, S. C., Muller, N., Janssens-Maenhout, G., Raes, F., Schwartz, J., Faluvegi, G., Pozzoli, L., Kupiainen, K., Hoeglund-Isaksson, L., Emberson, L., Streets, D., Ramanathan, V., Hicks, K., Oanh, N. T. K., Milly, G., Williams, M., Demkine, V., and Fowler, D.: Simultaneously Mitigating Near-Term Climate Change and Improving Human Health and Food Security, Science, 335, 183–189, doi:10.1126/science.1210026, 2012.

Shindell, D. T., Pechony, O., Voulgarakis, A., Faluvegi, G., Nazarenko, L., Lamarque, J.-F., Bowman, K., Milly, G., Kovari, B., Ruedy, R., and Schmidt, G.: Interactive ozone and methane chemistry in GISS-E2 historical and future climate simulations, Atmos. Chem. Phys., 13, 2653–2689, doi:10.5194/acp-13-2653-2013, 2013.

Skeie, R. B., Berntsen, T. K., Myhre, G., Tanaka, K., Kvalevåg, M. M., and Hoyle, C. R.: Anthropogenic radiative forcing time series from pre-industrial times until 2010, Atmos. Chem. Phys., 11, 11827–11857, doi:10.5194/acp-11-11827-2011, 2011.

Stevenson, D. S., Young, P. J., Naik, V., Lamarque, J. F., Shindell, D. T., Voulgarakis, A., Skeie, R. B., Dalsøren, 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., and Wild, O.: Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., in press, 2013.

Szopa, S., Balkanski, Y., Schulz, M., Bekki, S., Cugnet, D., Fortems-Cheiney, A., Turquety, S., Cozic, A., Deandreis, C., Hauglustaine, D., Idelkadi, A., Lathiere, J., Lefevre, F., Marchand, M., Vuolo, R., Yan, N., and Dufresne, J. L., Aerosol and Ozone changes as forcing for Climate Evolution between 1850 and 2100, Clim. Dyn., in press, 2012.

Takemura, T.: Distributions and climate effects of atmospheric aerosols from the preindustrial era to 2100 along Representative Concentration Pathways (RCPs) simulated using the global aerosol model SPRINTARS, Atmos. Chem. Phys., 12, 11555–11572, doi:10.5194/acp-12-11555-2012, 2012.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Feichter, H., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I., Iversen, I., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: Analysis and quantification of the diversities of aerosol life cycles within AeroCom, Atmos. Chem. Phys., 6, 1777–1813, doi:10.5194/acp-6-1777-2006, 2006.

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P., Veefkind, P., and Levelt, P.: Aerosols and surface UV products from Ozone Monitoring Instrument observations: An overview, J. Geophys. Res.-Atmos., 112, D24S47, doi:10.1029/2007JD008809, 2007.

UNEP WMO (United Nations Environment Programme and World Meteorological Organization): Integrated Assessment of Black Carbon and Tropospheric Ozone, Nairobi, Kenya, 285 pp., 2011.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: an overview, Climatic Change, 109, 5–31, 2011.

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, doi:10.5194/gmd-4-845-2011, 2011.

Young, P. J., Archibald, A. T., Bowman, K. W., Lamarque, J.-F., Naik, V., Stevenson, D. S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R. M., Eyring, V., Faluvegi, G., Horowitz, L. W., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R. B., Shindell, D. T., Strode, S. A., Sudo, K., Szopa, S., and Zeng, G.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13, 2063–2090, doi:10.5194/acp-13-2063-2013, 2013.

Yu, H., Chin, M., Remer, L., Kleidman, R., Bellouin, N., Bian, H., and Diehl, T.: Variability of marine aerosol fine-mode

fraction and estimates of anthropogenic aerosol component over cloud-free oceans from the Moderate Resolution Imaging Spectroradiometer (MODIS), J. Geophys. Res.-Atmos., 114, D10206, doi:10.1029/2008JD010648, 2009.

Zhang, H., Shen, Z., Wei, X., Zhang, M., and Li, Z.: Comparison of optical properties of nitrate and sulfate aerosol and the direct radiative forcing due to nitrate in China, Atmos. Res., 113, 113–125, 2012.

Zhao, T., Laszlo, I., Guo, W., Heidinger, A., Cao, C., Jelenak, A., Tarpley, D., and Sullivan, J.: Study of long-term trend in aerosol optical thickness observed from operational AVHRR satellite instrument, J. Geophys. Res.-Atmos., 113, D07201, doi:10.1029/2007JD009061, 2008.