

# Emulating atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 2: Applications

M. Meinshausen<sup>1</sup>, T. M. L. Wigley<sup>2</sup>, and S. C. B. Raper<sup>3</sup>

<sup>1</sup>Earth System Analysis, Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

<sup>2</sup>National Center for Atmospheric Research (NCAR), Boulder, CO, USA

<sup>3</sup>Manchester Metropolitan University (MMU), Manchester, UK

Received: 20 December 2007 – Published in Atmos. Chem. Phys. Discuss.: 27 March 2008

Revised: 20 January 2011 – Accepted: 2 February 2011 – Published: 16 February 2011

**Abstract.** Intercomparisons of coupled atmosphere-ocean general circulation models (AOGCMs) and carbon cycle models are important for galvanizing our current scientific knowledge to project future climate. Interpreting such inter-comparisons faces major challenges, not least because different models have been forced with different sets of forcing agents. Here, we show how an emulation approach with MAGICC6 can address such problems. In a companion paper (Meinshausen et al., 2011a), we show how the lower complexity carbon cycle-climate model MAGICC6 can be calibrated to emulate, with considerable accuracy, globally aggregated characteristics of these more complex models. Building on that, we examine here the Coupled Model Intercomparison Project's Phase 3 results (CMIP3). If forcing agents missed by individual AOGCMs in CMIP3 are considered, this reduces ensemble average temperature change from pre-industrial times to 2100 under SRES A1B by 0.4 °C. Differences in the results from the 1980 to 1999 base period (as reported in IPCC AR4) to 2100 are negligible, however, although there are some differences in the trajectories over the 21st century. In a second part of this study, we consider the new RCP scenarios that are to be investigated under the forthcoming CMIP5 intercomparison for the IPCC Fifth Assessment Report. For the highest scenario, RCP8.5, relative to pre-industrial levels, we project a median warming of around 4.6 °C by 2100 and more than 7 °C by 2300. For the lowest RCP scenario, RCP3-PD, the corresponding warming is around 1.5 °C by 2100, decreasing to around 1.1 °C by 2300 based on our AOGCM and

carbon cycle model emulations. Implied cumulative CO<sub>2</sub> emissions over the 21st century for RCP8.5 and RCP3-PD are 1881 GtC (1697 to 2034 GtC, 80% uncertainty range) and 381 GtC (334 to 488 GtC), when prescribing CO<sub>2</sub> concentrations and accounting for uncertainty in the carbon cycle. Lastly, we assess the reasons why a previous MAGICC version (4.2) used in IPCC AR4 gave roughly 10% larger warmings over the 21st century compared to the CMIP3 average. We find that forcing differences and the use of slightly too high climate sensitivities inferred from idealized high-forcing runs were the major reasons for this difference.

## 1 Introduction

In our companion paper, we summarized the uses and advantages of simple climate models. MAGICC6 was documented and the methods used to calibrate MAGICC against CMIP3 AOGCMs and C<sup>4</sup>MIP carbon cycle models were described (Meinshausen et al., 2011a) (henceforth MRW).

This part 2 applies the calibrated MAGICC6 model to the interpretation of AOGCM and carbon cycle intercomparison exercises. We determine the possible effect that incomplete forcing series could have on the CMIP3 temperature evolutions. For the forthcoming CMIP5 intercomparison, projections are presented for the four Representative Concentration Pathways (RCPs).

Tests in MRW showed that the calibrated MAGICC model can have emulation skill both within the temperature range over which calibrations were performed as well as outside. Specifically, we found a close fit to the SRES A2 scenario, which was not used for calibration. Independent tests suggest



Correspondence to: M. Meinshausen  
(malte.meinshausen@pik-potsdam.de)

that emulations with MAGICC can successfully reproduce the deep mitigation scenario results from HadCM3 (Lowe et al., 2009).

These tests are important because the very high RCP8.5 scenario beyond 2100 as well as the low RCP3-PD scenario go outside the MAGICC6 calibration range. The forthcoming CMIP5 intercomparison data will allow us to further verify the ability of this simple model to emulate global average and large-scale diagnostics over a wide range of scenarios.

This paper is structured as follows: Sect. 2 “Methods” briefly describes the MAGICC6 parameter settings, radiative forcing assumptions and the experimental setup to emulate both SRES and RCP scenarios. Section 3 “The effect of incomplete forcings in CMIP3 results” estimates the effect of quantitatively different or incomplete forcing assumptions on the CMIP3 ensemble results. Section 4 “Design and analysis of the RCPs” provides projections for the Representative Concentration Pathways – used in the CMIP5 intercomparison exercise. Section 5 “Analysis of AR4 results” assesses the reasons why temperature projections obtained by MAGICC4.2 (as used in the IPCC AR4) were roughly 10% warmer than CMIP3 ensemble means. Finally, Section 6 “Discussion” addresses volcanic forcing assumptions and the interpretation of multi-model ensembles, while Sect. 7 concludes.

## 2 Methods

### 2.1 Parameter sets and forcing assumptions for CMIP3 and C<sup>4</sup>MIP emulations

MAGICC6 has an updated carbon cycle routine, modified indirect aerosol forcing parameterizations, a larger set of calibration parameters, and enhanced flexibility compared with earlier versions of MAGICC. The emulations used here are the result of the “calibration III” procedure described in MRW and applied to the CMIP3 AOGCM data for 19 models. This involved adjusting eight climate response parameters to reproduce hemispheric land and ocean temperature timeseries as well as ocean heat uptake. Carbon cycle results from C<sup>4</sup>MIP models were used to calibrate MAGICC6 emulations for atmospheric CO<sub>2</sub> concentrations, carbon pools and carbon fluxes (in total 7 variables, each for an uncoupled and coupled model setup with over 200-yr long time series – as shown in Fig. 14 of MRW). When using both AOGCM and carbon cycle model emulations together, we give priority to the climate model parameters derived from the CMIP3 AOGCM comparisons. With 10 carbon cycle parameter sets (Table B4 in MRW) for each of the 19 AOGCM parameter sets (Table B3 in MRW), there are 190 emulations for any one modeled SRES emissions scenario.

To determine the effect of the use of different forcings by individual models in CMIP3, we are interested in the effect on temperature change when a complete set of forcings for

future scenarios is applied, rather than the AOGCM-specific subsets of forcings (see Table 2 in MRW). The common forcing timeseries we use (see Fig. 1 in the Supplement in MRW) match the point forcing estimates for individual forcing agents in year 2005 provided by IPCC AR4 WG1 Table 2.12 (Forster et al., 2007). For consistency, we compared non-efficacy adjusted forcings to the AR4 Table 2.12 values.

As discussed by Collins et al. (2006), radiative forcing parameterizations in AOGCMs can result in substantial deviations from more accurate line-by-line radiative transfer schemes. These deviations are, for example, apparent in the reported  $\Delta Q_{2\times}$  values, with values ranging from 3.09 to 4.06 Wm<sup>-2</sup> across the AR4 AOGCMs (see first column of Table B3 in MRW). To remove these differences we used a central estimate of 3.71 Wm<sup>-2</sup> (Myhre et al., 1998) consistent with line-by-line models (Collins et al., 2006), although it is recognized that there is some uncertainty in the “true” CO<sub>2</sub> forcing – in particular when considering indirect forcing effects that are traditionally subsumed in the definition of feedbacks. As discussed in MRW (Sect. 6.2 therein), we do not take into account fast forcing adjustments for CO<sub>2</sub>, but apply radiative forcings according to the standard definition used in IPCC AR4.

Finally, for our projections we use CO<sub>2</sub> concentrations that are generated internally from emissions using each of the ten carbon cycle model calibrations in combination with each AOGCM calibration. This takes into account the effect that a warmer AOGCM is likely to see higher CO<sub>2</sub> concentrations because of carbon cycle feedbacks. For the calibration of MAGICC model parameters (see MRW), we used the prescribed CO<sub>2</sub> concentrations (Bern-CC reference case specified in IPCC TAR), which were likely used by most CMIP3 AOGCMs.

### 2.2 Setup for projecting temperatures under the RCPs

For projections under the RCP scenarios we use forcing time series described in Meinshausen et al. (2011b) (see <http://www.pik-potsdam.de/~mmalte/rcps/>). For most CMIP5 experiments (Taylor et al., 2009), the models will be forced with specified GHG concentrations (Meinshausen et al., 2011b), together with tropospheric ozone fields, 4-D aerosol fields (Lamarque et al., 2011) and land use change patterns (Hurt et al., 2011) to (at least partly) ensure better model-to-model and scenario-to-scenario comparability. The recommended default GHG concentrations have been generated with MAGICC6 using a default parameter setting to match the median of the AOGCM calibrations presented here and in MRW. For the carbon cycle, emulations of the Bern-CC model have been employed in Meinshausen et al. (2011b). Going beyond the standard CMIP5 time horizon of 2300, we extended here our analysis until 2500 (cf. Wigley et al., 2009). RCP extensions between 2300 and 2500 follow the same guiding principles as the RCPs extensions in 2300, i.e., constant emissions for RCP3-PD and constant

concentrations for the higher three RCPs (data up to 2500 provided here <http://www.pik-potsdam.de/~mmalte/rcps/>).

As an application of the calibrated MAGICC6 model, we give temperature projections, with uncertainties, for the RCPs. We employ the 19 individual AOGCM emulations using calibration “III” as described in MRW. Carbon cycle emulations do not influence the temperature outcomes of these projections because CO<sub>2</sub> concentrations are prescribed. However, as part of the CMIP5 exercise, it is planned to retrieve the implied emissions for those models that include a coupled carbon cycle model. Thus, in a second experiment, we calculate the implied fossil CO<sub>2</sub> emissions for the prescribed RCP concentrations. Specifically, we combine 19 CMIP3 and 9 C<sup>4</sup>MIP calibration sets as an ensemble of 171 parameter combinations to estimate implied emissions. We leave out one of the 10 carbon cycle emulations (IPSL) for two reasons: because the IPSL calibration is least able to reproduce the IPSL carbon cycle dynamics towards the end of the 21st century compared to all other C<sup>4</sup>MIP emulations, although the fit is still rather close (see MRW); and because, when comparing long-term ocean carbon uptake to the results from Orr (2002), our IPSL emulation performed least well over multi-century timescales. We therefore place low confidence on the ability of the IPSL calibration to simulate plausible long-term carbon cycle behavior beyond 2100.

### 2.3 Setup for analyzing previous MAGICC4.2 emulation results

For comparison and historical context, the MAGICC4.2 emulations presented in IPCC AR4 are analyzed here to try to understand why the mean emulations were approximately 10% warmer over the 21st century compared to the CMIP3 AOGCM data. For replicating and analyzing the emulation results presented in the IPCC AR4, we use MAGICC4.2 (which is the version used in the AR4). Note that the calibration of MAGICC4.2 was based solely on the idealized CO<sub>2</sub>-only experiments (Sect. 5). The MAGICC6 “calibration I” procedure (presented in MRW) is equivalent to the calibration employed for MAGICC4.2 for the AR4 (see Table 1 in MRW). Both calibration methods yield very similar climate sensitivity estimates for all AOGCMs (see Table 4 in MRW, columns 4 and 5). One exception is the HadGEM1 model, for which this study benefited from additional 1pctto4x data. In general, MAGICC4.2 was calibrated to match both 1pctto2x and 1pctto4x experiments, but in the case of HadGEM1, data for the 1pctto4x has not been available at the time of calibrating MAGICC4.2. This leads to a 0.5 °C larger climate sensitivity estimate in the MAGICC6 calibration. In addition to the improved model structure, MAGICC6 calibrations benefited from complete hemispheric land/ocean AOGCM data sets allowing a more accurate determination of the land/ocean warming characteristics for all models. If only the data used for calibrating MAG-

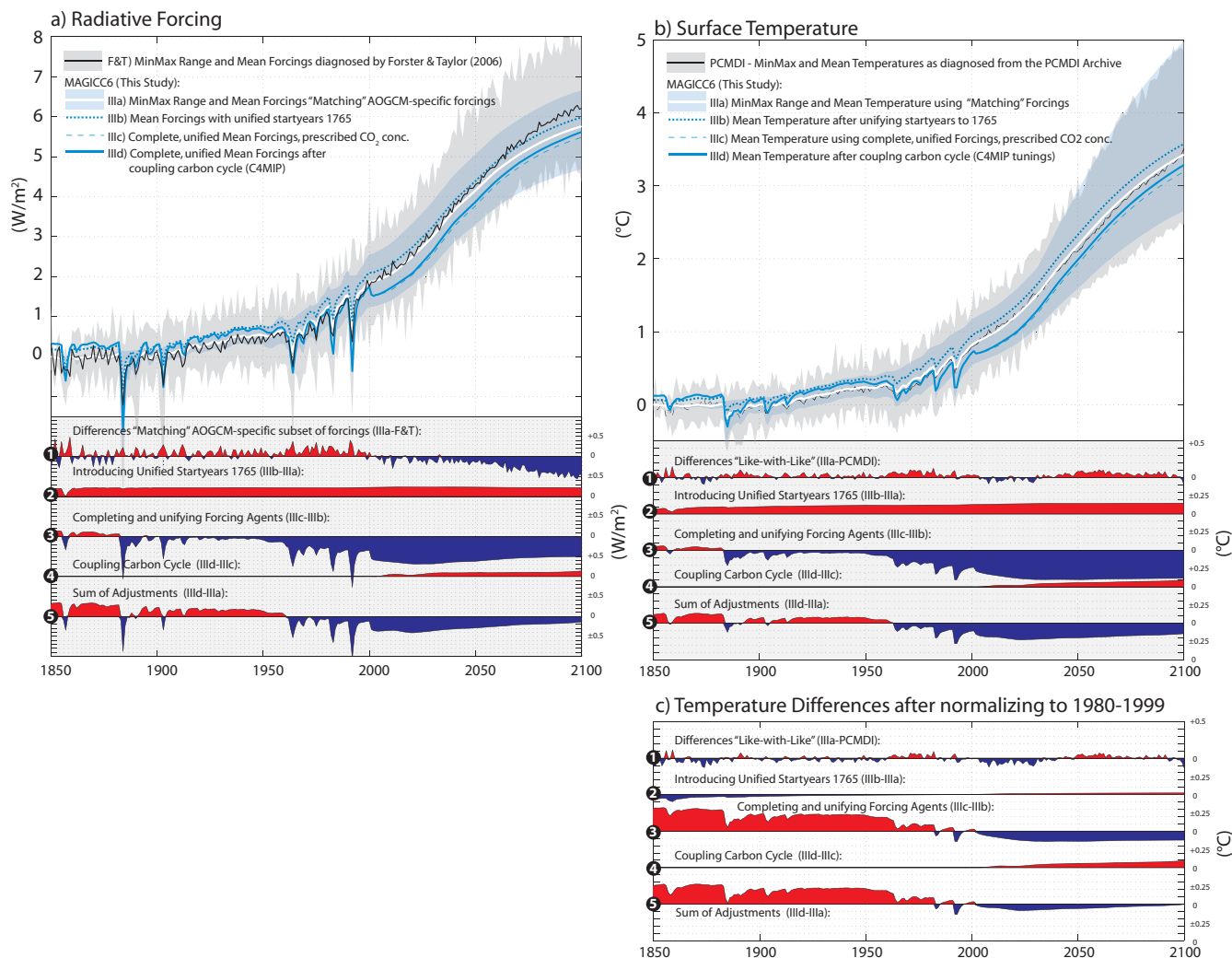
ICC4.2 are used in calibrating version 6, then the agreement in retrieved climate sensitivities is even closer, providing an important check on the credibility of the model’s structure in earlier and simpler versions of MAGICC.

## 3 The effect of incomplete forcings on CMIP3 results

In this section we analyze the CMIP3 intercomparison exercise results using MAGICC6. We begin by comparing our AOGCM-specific forcing sets with independently estimated time series from Forster and Taylor (2006), hereafter called F&T (Sect. 3.1). We then estimate the effects on global-mean temperature change that arise because: the AOGCM experiments branched off the control runs in different years (Sect. 3.2); because incomplete sets of forcing agents were used (Sect. 3.3); and because CO<sub>2</sub> concentrations were prescribed rather than prescribing emissions in coupled carbon cycle models (Sect. 3.4). As a representative example, we consider results for the A1B scenario (Sect. 3.5). The last subsection (Sect. 3.6) provides a comprehensive and unified set of projections for the three illustrative SRES scenarios used in CMIP3, based on the full forcing timeseries and using the full range of both C<sup>4</sup>MIP carbon cycle as well as CMIP3 climate model emulations.

### 3.1 Forcing comparisons

The starting point here is the set of AOGCM-specific, efficacy-adjusted forcings that we used in calibrating MAGICC parameters (see Table 2 in MRW) – referred to below as “matching” forcings. We compare these with the forcings diagnosed by F&T, which, in the absence of forcing information from individual models, are the only other available model-specific forcing estimates. Comparing the across-model means of these two forcing data sets reveals a close match up to the middle of the 21st century (see first sub-panel indicated with black circled “1” in Fig. 1a). Thereafter, our “matching” forcings are lower than the diagnosed F&T forcings, probably due to an overestimation of the effective forcings in F&T towards the second half of the 21st century. This could arise because F&T assumed constant (rather than increasing) effective climate sensitivities, estimated from the first 70 years of the idealized forcing scenarios. Two factors support this hypothesis. Firstly, in the idealized scenarios, for which the standard forcing at tropopause level after stratospheric adjustment is better known, the diagnosis of effective climate sensitivities (see equation 3 in MRW) reveals higher climate sensitivities for higher forcing levels for a number of models - as shown in Fig. 1 in MRW for CCSM3 and ECHAM5/MPI-OM. Secondly, continuing the analysis by F&T beyond 2100 suggests increasing diagnosed forcings for some models, inconsistent with the fact that forcings in these runs were set constant after 2100 (Meehl et al., 2005a).



**Fig. 1.** Net radiative forcing (panel **a**) and global-mean surface temperatures (**b**) for the SRES A1B scenario according to MAGICC6 (calibration III) as presented in this study. For comparison, the radiative forcing applied is compared to the means across all 19 archived CMIP3 AOGCMs, as diagnosed by Forster and Taylor (2006) (F&T, see panel **a**). For temperature, the MAGICC6 emulations are compared to the means diagnosed from the matching set of 19 CMIP3 AOGCMs relative to their respective start years. Differences relative to 1980–1899 are also shown (panel **c**). See Table 3 in the companion paper MRW and text for discussion of the different forcing adjustments and temperature effects (black circled numbers). The roman number III denotes the calibration method, while the small latin letters a, b, c, and d denote the forcing assumptions, ranging from (**a**) AOGCM-specific forcing subsets, (**b**) forcings relative to a unified 1765 starting year, (**c**) complete and unified forcings and (**d**) full forcings including  $\text{CO}_2$  concentrations with coupled carbon cycles.

The emulated temperature perturbations (mean across all emulations) in MAGICC6 using the “matching” forcings are within  $0.1^{\circ}\text{C}$  of the AOGCM mean throughout the whole emulation period (see black circled “1” in Fig. 1b). If we had used the F&T forcings (which are higher in the second half of the 21st century) we would have overestimated the AOGCMs temperature response. Relative to the base period 1980–1999, the difference in projected warming for 2090–2099 between the MAGICC6 emulations and the mean AOGCM result averaged over all SRES scenarios considered is  $0.04^{\circ}\text{C}$  or 2% less (cf. column AOGCM and IIIa for “Period 2” in Table 3 in MRW).

### 3.2 The effect of different starting years

One of the difficulties in interpreting the AOGCM results for the past is that modelling groups assumed different starting years in which the 20th century simulations (20c3m) diverge from the pre-industrial control runs. Unifying these starting years to 1765 shifts the forcing to higher values. This is because the forcing increments for  $\text{CO}_2$  and other considered forcing agents between 1765 and the starting years, e.g. 1850 or 1900, are now taken into account (see “2” in Fig. 1a). Had all AOGCMs started their simulations in 1765, their temperature projections for the 21st century could be expected to

be approximately  $0.1\text{ }^{\circ}\text{C}$  warmer relative to 1765 (see “2” in Fig. 1b), although this effect almost vanishes when taking differences of future projections from the 1980–1999 base period (see “2” in Fig. 1c).

### 3.3 The effect of using complete and unified forcings

Here we address the issue, what if the overall best-estimate forcings are used in each MAGICC emulation, instead of “matching” model-specific forcings? Specifically, if an AOGCM had left out a specific forcing agent (e.g. indirect aerosol effects or tropospheric ozone), we now run the emulations with these forcings included (see Table 2 in MRW). We also unify the  $\text{CO}_2$  forcing ( $\Delta Q_{2\times} = 3.71\text{ Wm}^{-2}$ ) and adjust the historical volcanic forcing to a zero mean. These changes have a significant effect on the applied forcing (see “3” in Fig. 1a) and, hence, on temperature projections.

Relative to the starting years of the emulations, when a unified and complete forcing set is used, the temperatures drop by around  $0.4\text{ }^{\circ}\text{C}$  for much of the 21st century (see “3” in Fig. 1b). However, when taking differences relative to 1980–1999, the MAGICC 21st century temperatures only cool by  $0.1\text{ }^{\circ}\text{C}$  on average (see “3” in Fig. 1c). The main reason for these differences is the neglect in some models of indirect aerosol forcing. As the relative contribution of indirect aerosol forcing is greater in the period to 1999 than thereafter, it is not surprising that its influence is greater in the historical period.

### 3.4 Coupling the carbon cycle

Running the AOGCM emulations coupled with the carbon cycle calibrations for the future overcomes the inconsistency that warmer AOGCMs used the same prescribed  $\text{CO}_2$  concentrations as colder AOGCMs, and so do not account for carbon cycle feedbacks. By combining each CMIP3 AOGCM with each of the C<sup>4</sup>MIP emulations, the coupled AOGCM carbon cycle emulations yield, on average,  $0.2\text{ Wm}^{-2}$  additional forcing (see “4” in Fig. 1a), resulting in an additional warming of roughly  $0.1\text{ }^{\circ}\text{C}$  by the end of the 21st century (see “4” in Fig. 1b and c).

### 3.5 The net effect of harmonized forcing assumptions

When averaged across models, the sum of all forcing adjustments for most of the 20th century is up to  $0.2\text{ Wm}^{-2}$  (up to 1960) larger than the unadjusted forcing (see “5” in Fig. 1a). The time series is punctuated by negative spikes reflecting the unaccounted part of volcanic coolings in the past. After 1960, the sum of forcing adjustments is negative, and is dominated by volcanic forcing spikes. Consequently, the direct AOGCM results are likely to show a larger warming trend over the 2nd half of the 20th century compared to what they would have been with common and complete forcing assumptions, in part because of missing or low volcanic

forcing assumptions in some models. After 2000, there is no volcanic forcing record, so the volcano spikes no longer appear. Instead, this forcing component is assumed to drop from the year 2000 no-forcing level to its pre-2000 mean volcanic forcing level by about  $0.2\text{ Wm}^{-2}$ . Thus, future projections under our unified forcing series assumptions include an expected mean cooling effect due to volcanic aerosols eruptions. We prefer this assumption over alternative assumptions such as that there will be no future volcanic eruptions, or a repetition of an historical period with volcanic forcing spikes.

In the above, we concentrated on A1B. The deviation of the emulated temperatures to AOGCM data are similarly small across all SRES scenarios, with a maximum deviation of 2.2% for SRES B1 at the end of the 21st century (cf. the columns “AOGCM” and “IIIa” for “Period 2” in Table 3 in MRW). Note, however, that differences for the “year 2000 concentration stabilization” (COMMIT) scenarios are larger (mean  $-0.1\text{ }^{\circ}\text{C}$ ) (Wigley, 2005; Meehl et al., 2005b). The “COMMIT” scenarios are challenging to emulate, as different AOGCMs show a wide range of 20th century temperature evolutions with, for some models, relatively strong short-term variability. Because there is a wide range of forcing levels that are held constant after year 2000, the inter-model differences in unrealized warming at the end of the 20th century manifest themselves in a rather wide spread of temperatures under the “COMMIT” scenario ( $0.17\text{ }^{\circ}\text{C}$  to  $1.0\text{ }^{\circ}\text{C}$  by 2090–2099 relative to 1980–1999).

In case of the SRES A2 scenario, not all AOGCMs provided integrations: in particular not the MIROC 3.2 (hires) AOGCM, the “warmest” among all CMIP3 AOGCMs. The SRES A2 mean across AOGCM emulations over the 21st century is  $0.09\text{ }^{\circ}\text{C}$  warmer when all AOGCMs are averaged, compared to the case where only the limited set of models is averaged that ran this scenario under CMIP3. Additionally, a general feature of the forcing harmonization is that the emulated warming up to the end of the 20th century is generally reduced, while the 21st century warming is relatively unaffected. This is primarily because a number of AOGCMs left out indirect aerosol forcing effects (e.g. CCSM3, CAOGCM3.1(T47), CSIRO-MK3.0, FGOALS-g1.0, GFDL-CM2.0 etc. – see Table 2 in MRW), which leads to an overestimate of historical forcing, but little relative change in future forcing. In the case of SRES A2, the  $0.09\text{ }^{\circ}\text{C}$  additional warming noted above due to the inclusion of all AOGCMs is overcompensated by additional cooling due to aerosols, resulting in a net downward adjustment of  $0.1\text{ }^{\circ}\text{C}$  for the 21st century (cf. columns IIIa with the AOGCM specific subsets of forcings and column IIIId with complete forcings for all AOGCM emulations in Table 3 in MRW). Furthermore, an effect of similar magnitude compared to the forcing adjustments is due to the choice of the reporting period. While in IPCC TAR, results were stated for 2100, IPCC AR4 choose the 2090–2099 period, which results in a lowering of the stated temperatures by up to  $0.25\text{ }^{\circ}\text{C}$

under the SRES A2 scenario (cf. “Period 2” and “Period 3” in Table 3 in MRW).

### 3.6 Projections for CMIP3 SRES scenarios

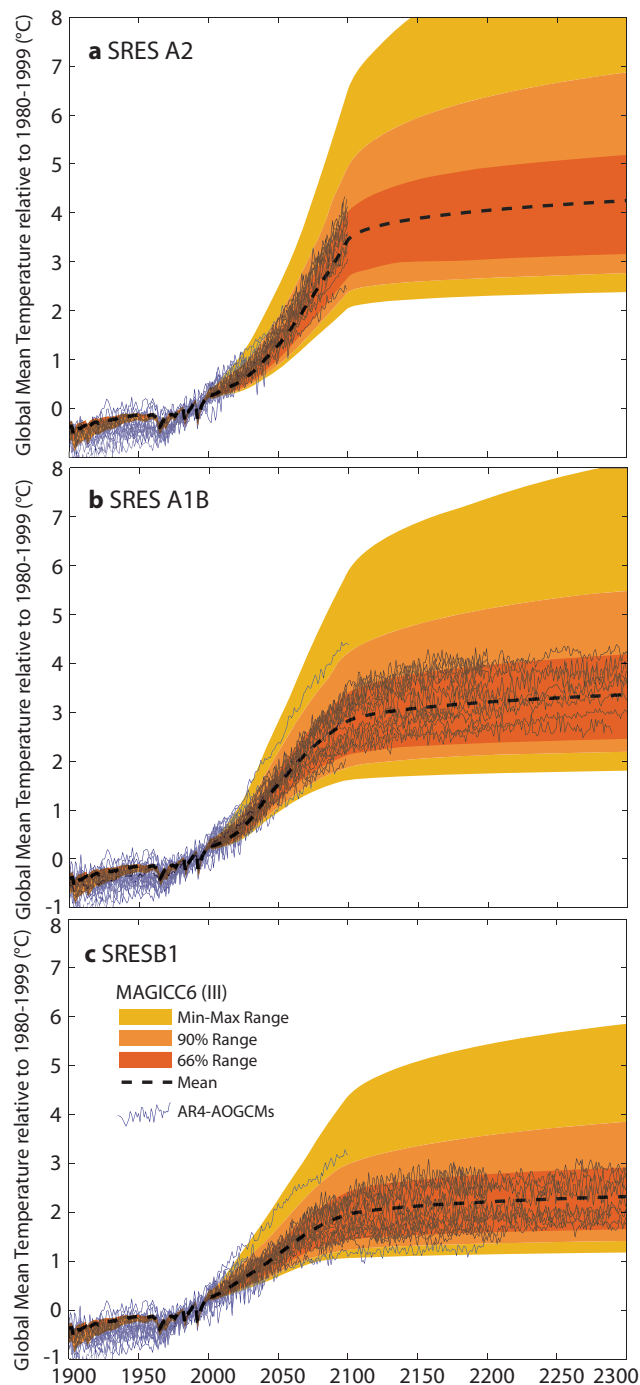
Here, we present the full set of SRES temperature projections combining the effects of uncertainties in both the carbon cycle and GCM emulations. With 19 AOGCM and 10 carbon cycle emulations, this makes a total of 190 cases (given that this section investigates projections only up to 2100, we included the IPSL emulation, which we exclude for projections beyond 2100). Making the debatable assumption that all models are random and equally likely representations of the real world, Bayesian credible regions can be easily inferred: the lowest 9 and the highest 9, for example, effectively define the 90% credible region. One obvious limitation is that what we have here is not the full range of parameter possibilities but a more restricted “ensemble of opportunity”. Comparing the 21st century temperature evolutions of the full set of 190 emulations with the original AOGCM data shows good agreement, despite the numerous forcing adjustments we have made to obtain a unified and complete set of forcings (see Fig. 2). The 90% C.I. for temperature changes relative to 1980 to 1999 is  $-31\%$  to  $+43\%$ , when averaging the results for 2050, 2075 and 2100 across the SRES scenarios (see Table 1). The right-skewed nature of the distribution, i.e., the fact that the upper bound of the 90th percentile range is further from the mean than the lower bound, is consistent with the expert judgment of  $-40\%$  to  $+60\%$  for a likely (66% CI) confidence range provided in IPCC AR4 based on multiple lines of evidence (Meehl et al., 2007; Knutti et al., 2008). Whatever the true uncertainty range is, it is clear that the ensemble of opportunity presented here must underestimate this range.

## 4 Design and analysis of the RCPs

In this section, we illustrate three applications of the calibrated MAGICC6 model relevant to the forthcoming CMIP5 experiments and the creation of the RCP scenarios (Moss et al., 2010).

### 4.1 Calculation of recommended RCP GHG concentrations

MAGICC6 has been used to create the harmonized and recommended greenhouse gas concentration series to be used by AOGCMs and Earth System Models (ESMs, i.e., AOGCMs with coupled carbon cycle models) taking part in the CMIP5 experiment. We produced these default GHG concentrations from the original scenario emission data (van Vuuren et al., 2007; Riahi et al., 2007; Clarke et al., 2007; Fujino et al., 2006) by assuming a median default parameter setting within the range of C<sup>4</sup>MIP and CMIP3 calibrations. See Meinshausen et al. (2011b) for a detailed description and



**Fig. 2.** Global-mean surface temperatures relative to 1980–1999 for the original AOGCMs (thin lines) and for the emulations using 10 C<sup>4</sup>MIP carbon cycle emulations and 19 CMIP3 AOGCM emulations for the three SRES scenario B1, A1B and A2 using complete, unified forcings (calibration III, as described in MRW). The shaded areas denote the ranges of the 190 cross-combinations of emulations, specifically the ranges in which the middle 66% (dark), 90% (medium) and all (bright patches) of the 190 emulations are located.

**Table 1.** Distributions of temperature projections for 2050, 2075 and 2100 for the six illustrative SRES scenarios, for the “Year 2000 concentration stabilization” (COMMIT) experiment and the four RCP scenarios – relative to the 1980–1999 mean. To convert warming relative to 1980–1999 as provided here into warming relative to a proxy for pre-industrial levels (i.e., the average over the first 70 years of the observational record, 1850–1919), 0.5 °C should be added (see HadCRUt3v as described in Brohan et al., 2006). The distribution percentiles in the header row denote the cumulative density of occurrence of the cross-combination between calibrations to CMIP3 AOGCMs and C<sup>4</sup>MIP carbon cycle models. The row denoted “Dev. From Mean” provides the average deviations from the mean across all six SRES scenarios and the three time slices.

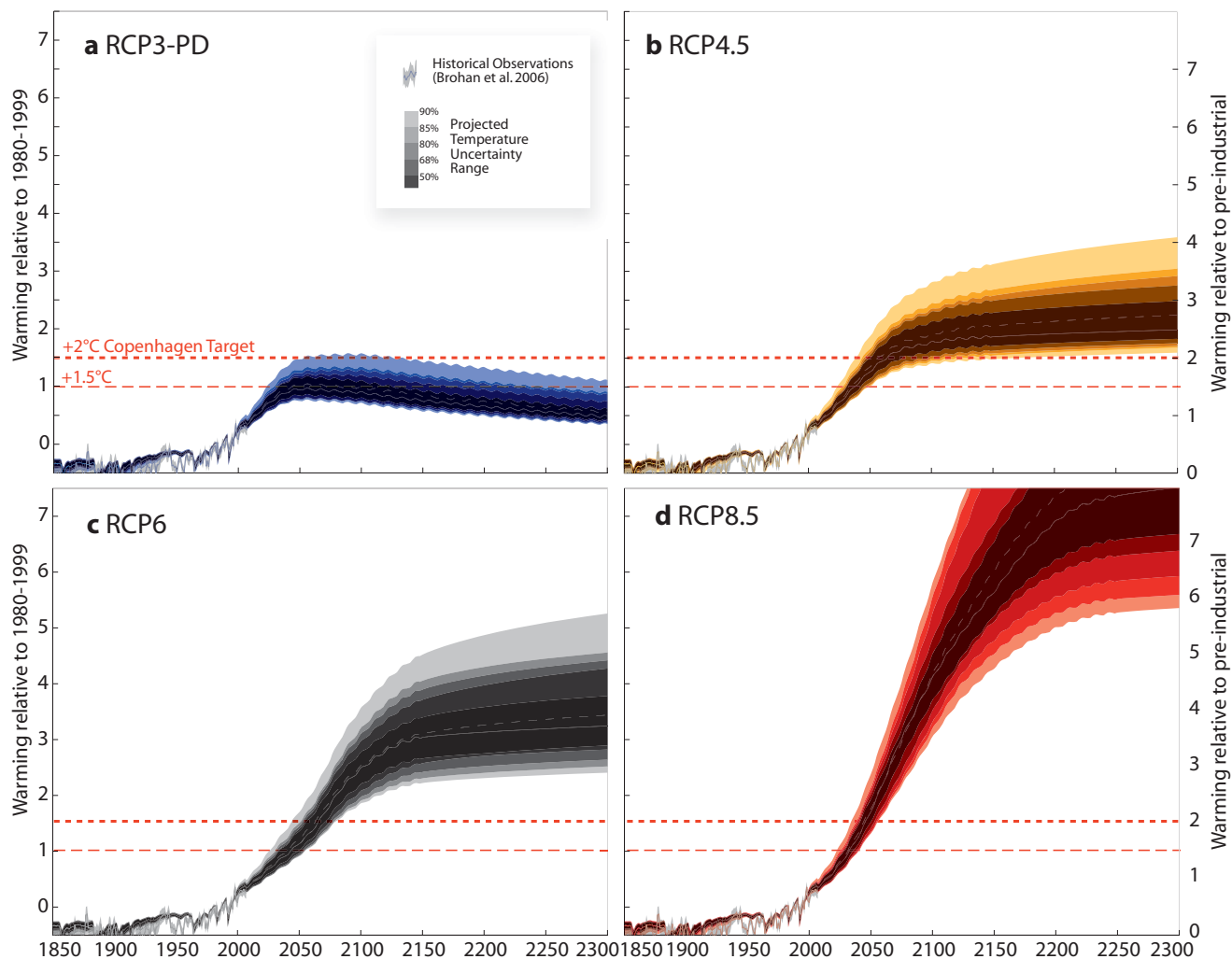
Scenario	Years	MIN	5%	16.7%	MEDIAN	MEAN	83.3%	95%	MAX
SRES A1B	2050	1.00	1.10	1.23	1.51	1.57	1.85	2.18	2.77
	2075	1.42	1.64	1.83	2.27	2.33	2.77	3.27	4.40
	2100	1.65	1.94	2.18	2.78	2.87	3.44	4.14	5.89
SRES A1FI	2050	1.11	1.22	1.36	1.67	1.71	1.99	2.35	2.94
	2075	1.87	2.14	2.39	2.97	3.02	3.51	4.28	5.32
	2100	2.43	2.85	3.26	3.96	4.09	4.79	5.88	7.74
SRES A1T	2050	1.02	1.11	1.23	1.52	1.56	1.84	2.22	2.78
	2075	1.33	1.50	1.69	2.07	2.16	2.59	3.12	4.20
	2100	1.37	1.64	1.85	2.32	2.45	2.99	3.67	5.31
SRES A2	2050	0.82	0.92	1.02	1.27	1.32	1.56	1.85	2.42
	2075	1.44	1.63	1.83	2.28	2.33	2.74	3.22	4.21
	2100	2.07	2.42	2.70	3.37	3.45	4.05	4.88	6.43
SRES B1	2050	0.68	0.77	0.86	1.07	1.12	1.34	1.63	2.16
	2075	0.97	1.07	1.25	1.57	1.63	1.99	2.38	3.21
	2100	1.06	1.25	1.46	1.87	1.96	2.38	2.86	4.23
SRES B2	2050	0.81	0.90	1.00	1.23	1.28	1.52	1.87	2.38
	2075	1.14	1.26	1.44	1.77	1.85	2.22	2.71	3.60
	2100	1.42	1.64	1.85	2.34	2.43	2.91	3.54	4.97
Dev. From Mean		−39%	−31%	−22%	−3%	0%	19%	43%	91%
COMMIT	2050	0.36	0.36	0.39	0.47	0.51	0.59	0.93	0.93
	2075	0.38	0.38	0.42	0.50	0.55	0.64	1.03	1.03
	2100	0.39	0.39	0.44	0.51	0.57	0.67	1.10	1.10
RCP3-PD	2050	0.73	0.75	0.79	0.94	1.01	1.17	1.47	1.65
	2075	0.71	0.73	0.77	0.92	1.00	1.18	1.54	1.74
	2100	0.69	0.72	0.76	0.91	1.00	1.18	1.58	1.80
RCP45	2050	1.00	1.01	1.07	1.25	1.34	1.55	1.87	2.08
	2075	1.26	1.29	1.40	1.63	1.74	2.05	2.48	2.77
	2100	1.36	1.41	1.53	1.78	1.92	2.30	2.81	3.16
RCP6	2050	0.90	0.91	0.96	1.13	1.21	1.39	1.70	1.89
	2075	1.40	1.42	1.54	1.82	1.90	2.21	2.63	2.91
	2100	1.81	1.86	2.08	2.49	2.54	2.96	3.56	3.90
RCP85	2050	1.31	1.32	1.43	1.71	1.76	2.00	2.38	2.61
	2075	2.12	2.17	2.44	3.02	2.95	3.36	4.03	4.31
	2100	2.92	3.05	3.54	4.12	4.19	4.78	5.81	6.05

<http://www.pik-potsdam.de/~mmalte/rcps/> for the available datasets.

#### 4.2 Estimation of temperature evolutions under the RCPs

We present here temperature projections for the four RCPs and their extensions to 2500 based on our CMIP3 calibrations (Fig. 3). We show that, relative to pre-industrial, the high-end RCP8.5 scenario leads to an increase in global mean surface temperature of 4.6 °C (3.6 to 6.3 °C 90% cred-

ible interval) by 2100. A century later by 2200, the RCP8.5’s median temperature projection is in excess of 7 °C. RCP6, which is similar to the SRES A1B scenario, has a temperature increase between 2 °C and 4 °C by 2100 (Fig. 3c). The warming under the medium-low RCP4.5 scenario exceeds 2 °C relative to the pre-industrial level, with a median warming of 2.5 °C and a 90% credible interval between approximately 2.1 and 4 °C by 2300 (Fig. 3b). The lowest RCP3-“Peak&Decline” (RCP3-PD) pathway leads to a maximum of global-mean temperatures shortly after 2050 with a



**Fig. 3.** Global-mean surface temperatures for the Representative Concentration Pathways using prescribed concentrations (Meinshausen et al., 2011b) and 19 CMIP3 AOGCM emulations. Future short-term oscillations are due to the assumed solar forcing (Lean and Rind, 2009). Historical temperature observations through to 2005 including 90% uncertainty ranges are also shown (Brohan et al., 2006). For illustrative purposes, the 2 °C and 1.5 °C levels above pre-industrial temperatures are indicated by red dashed lines, corresponding to temperature limits adopted and put forward for review in the Copenhagen Accord.

median estimate slightly above 1.5 °C and a range between 1.3 to 2.0 °C (90% CI) with a slow continuous decline of roughly 0.2 °C per century thereafter (see Table 1, which provides warmings relative to the 1980–1999 period, cf. Schewe et al., 2011).

It should be noted that the CMIP3 calibration of MAGICC6 used only monotonically increasing or constant GHG concentrations spanning a limited range. Two of the RCP cases, therefore, lie outside the calibration range, and the extension to 2500 also extends the projections into no-analogue territory. For example, in the high-end RCP8.5 scenario, we are faced with a CO<sub>2</sub> concentration that is seven times higher than pre-industrial levels – in addition to substantial non-CO<sub>2</sub> forcings. The highest scenarios investigated in CMIP3 were the idealized quadrupling CO<sub>2</sub>-only experiment and the

SRES A2 experiment (which reached roughly a tripling of the pre-industrial CO<sub>2</sub> concentration in 2100). The temperature projections under RCP8.5 must therefore be interpreted cautiously: they are qualitatively robust, but must involve unknown quantitative uncertainties at high temperatures. In Fig. 3d we have therefore cut the temperature projections off above 8 °C relative to pre-industrial levels. In spite of the manifest uncertainties, we can be at least 90% confident that RCP8.5 temperatures will exceed global-mean warming of 5 °C by 2150 relative to pre-industrial levels (see Fig. 3d).

#### 4.3 Estimation of emissions under the RCPs

As part of the CMIP5 exercise, ESMs will be used to infer CO<sub>2</sub> emissions for the prescribed concentration pathways.



Here we emulate this experiment, starting from the recommended default CO<sub>2</sub> concentrations that we generated previously (Meinshausen et al., 2011b) and which are going to be prescribed in ESMs.

There is an element of unavoidable circularity here. The RCPs were defined initially as radiative forcing scenarios. For each RCP, an Integrated Assessment model was used, via a multi-gas optimization procedure (see e.g. Clarke et al., 2009, 2007), to determine the multi-gas emissions that, for that particular integrated assessment model, would lead to the prescribed forcing trajectory. We could, therefore consider those emissions to be the basic data that describe the RCPs. If other models were used to calculate forcings for those emissions, there would be a range of corresponding forcing trajectories.

For CMIP5, however, what is required for each RCP is a single set of concentration data. This is the only way to ensure a set of like-with-like comparisons between the various AOGCMs participating in this exercise. To determine those concentrations MAGICC6 default parameter settings for the AOGCM emulations and Bern parameter settings for the carbon cycle were used. For the CMIP5 exercise, therefore, it is these concentrations that should be considered as the basic RCP-defining data.

Using inverse methods it is now possible to determine the emissions that correspond to the RCP concentrations (similarly to Wigley et al., 2009) – and this is one of the CMIP5 tasks set for those models that include an interactive carbon cycle. Of course, each model used for the inverse exercise will give different results – but the results, forming an “ensemble of opportunity”, will not span the full uncertainty range of emissions corresponding to the prescribed concentrations. To obtain insights into a fuller emissions uncertainty range, we use MAGICC6 with the set of 171 combinations of AOGCM and carbon cycle model parameterizations (one of the ten carbon cycle models could not be used for this exercise, see above).

The medians of the inverse CO<sub>2</sub> emissions trajectories for the four RCPs are (as would be expected) similar to the harmonized CO<sub>2</sub> emissions that were initially used to derive the CO<sub>2</sub> concentrations, indicating that the Bern C<sup>4</sup>MIP emulation (which was chosen as default) does indeed present a middle-of-the range model within the set of C<sup>4</sup>MIP emulations (Fig. 4). Noteworthy are the large fluctuations of implied inverse emissions between 1950 and 2000. In part this is an artifact of inverse modelling, where results can be very sensitive to small variability in the rate of change of the driver concentrations. This especially affects the times before 2005, when the driver concentrations are observed atmospheric values. Decadal and annual fluctuations in atmospheric concentrations, which might partly result from internal natural variability, can hence lead to relatively large fluctuations in the derived inverse emissions. Interestingly, the sharp drop in inverse CO<sub>2</sub> emissions around World War II is not matched

by the available inventory data (Marland et al., 2006) shown here for historical emissions (Fig. 4).

The multiple carbon cycle emulations re-confirm one central point: namely that, irrespective of the ultimate CO<sub>2</sub> stabilization level, emissions will have to return to near-zero levels in the long-term (cf. Matthews and Caldeira, 2008). As shown, the largest uncertainties in emissions occur under the highest scenario RCP8.5 (Fig. 4d). See Table 2 for the range of cumulative CO<sub>2</sub> emissions for each RCP.

## 5 Analysis of AR4 results (obtained with MAGICC4.2)

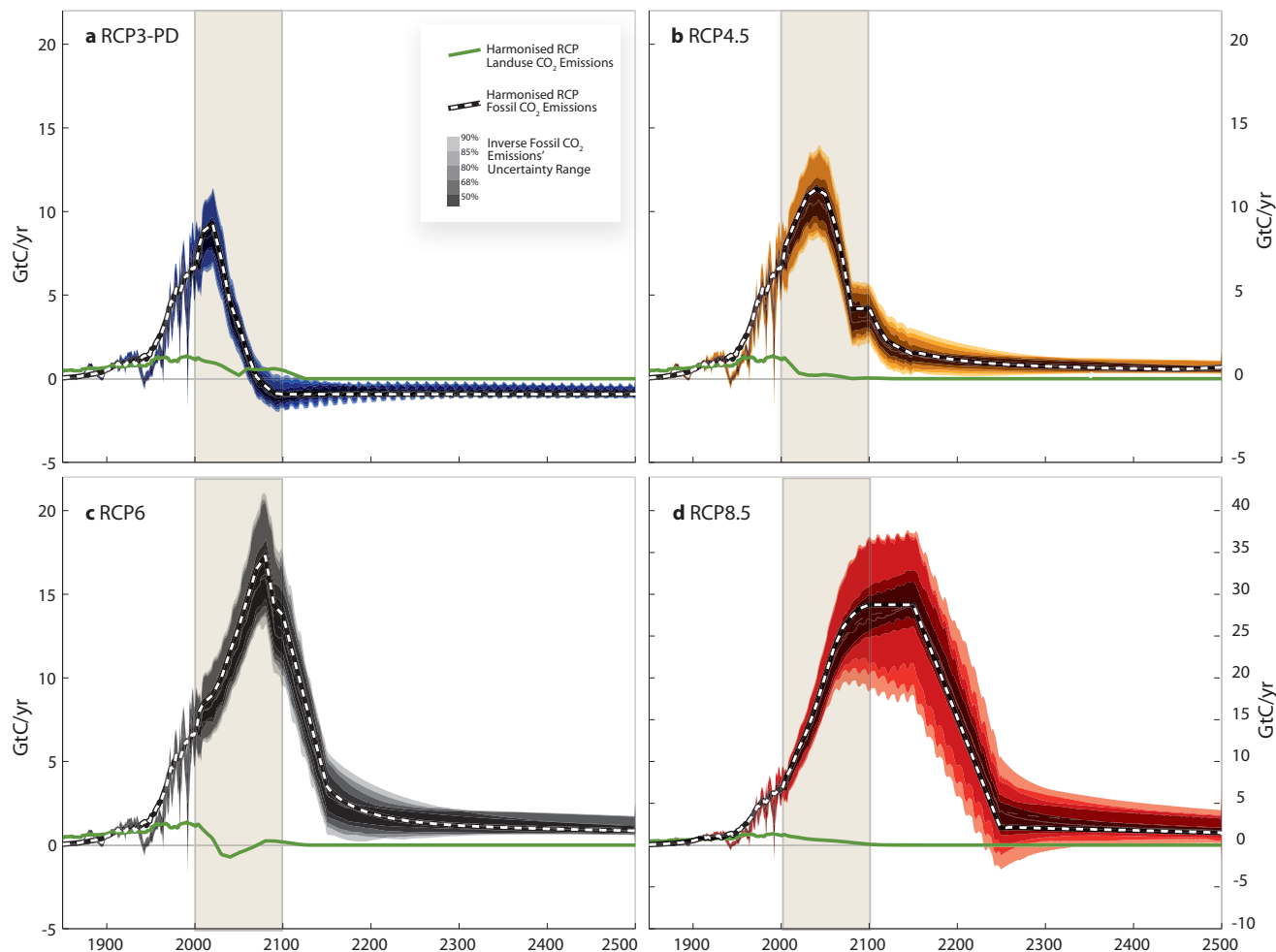
In an Appendix on sea level rise in IPCC AR4, Sect. 10.A.1 (p. 844), it is stated that “Under the SRES scenarios [...] the ensemble average of the tuned versions of MAGICC [4.2] gives about 10% greater temperature rise [...] over the 21st century (2090 to 2099 minus 1980 to 1999) than the average of the corresponding AOGCMs. The MAGICC radiative forcing is close to that of the AOGCMs (as estimated for A1B by Forster and Taylor, 2006), so the mismatch suggests there may be structural limitations on the accurate emulations of AOGCMs by the SCM”.

In this section, we examine this statement and show that differences in the forcing assumptions of individual AOGCMs, rather than structural limitations, are a key factor in explaining this discrepancy. MAGICC4.2 was calibrated to individual AOGCMs using only results from idealized CO<sub>2</sub>-only scenarios (1% per year increases to 2× and 4× CO<sub>2</sub>). If MAGICC4.2 is successfully calibrated in this way, but fails to reproduce AOGCM results for multi-gas scenarios, it is important to understand why, since it would imply that MAGICC4.2 should, perhaps, not be used outside its calibration region.

There are a number of reasons why MAGICC4.2 might give temperature projections for the SRES scenarios that differ from the AOGCM results. It could be the case that the model-specific climate sensitivities optimal for fitting the idealized CO<sub>2</sub>-only runs are too high for emulating the AOGCMs’ behaviour for SRES scenarios. Alternatively, the extra warming could be due to forcing differences between our best estimate full forcing and the specific forcing used by individual AOGCMs. The warmer MAGICC response in this latter case would represent a likely correction to the AOGCM temperature predictions as these did not account for all forcings. As discussed below, it turns out that the answer is likely to be a combination of the two effects.

### 5.1 Comparing forcings

SRES projections with the calibrated MAGICC4.2 model were calculated by assuming central forcing estimates for all individual major forcings listed in Table 2.12 of Forster et al. (2007). The MAGICC4.2 central forcing estimate must differ from the average of the AOGCM-forcing series, simply



**Fig. 4.** Inverse fossil CO<sub>2</sub> emissions calculated for prescribed default RCP concentrations. The uncertainty ranges (colored shaded areas) are based on 171 model emulations, combining 19 CMIP3 AOGCM and 9 C<sup>4</sup>MIP carbon cycle model emulations. Harmonized global RCP land-use CO<sub>2</sub> emissions are shown in green, with scaled historical land-use CO<sub>2</sub> emissions in the past based on Houghton and Hackler (2002). Harmonized RCP fossil CO<sub>2</sub> emissions, used to produce default RCP CO<sub>2</sub> concentrations using the same model (MAGICC6) are shown as white-dashed lines (Meinshausen et al., 2011b). Shaded vertical areas from 2000 to 2100 denote the time-span of RCPs, with extensions following thereafter. Due to the focus of the C<sup>4</sup>MIP intercomparison on only one relatively high scenario (SRES A2) up to 2100, the extension of the calibrated C<sup>4</sup>MIP carbon cycle emulations for deeper scenarios and longer time frames is uncertain. Note that the vertical axis in panel (d) is a factor two higher compared to the other panels.

because some of the AOGCMs did not take indirect aerosol effects into account. As a check on the full forcings used by MAGICC4.2, we compare them to the forcings diagnosed by F&T, noting that these forcings might be overestimated in the second half in the 21st century because of the constant sensitivity assumption employed by these authors.

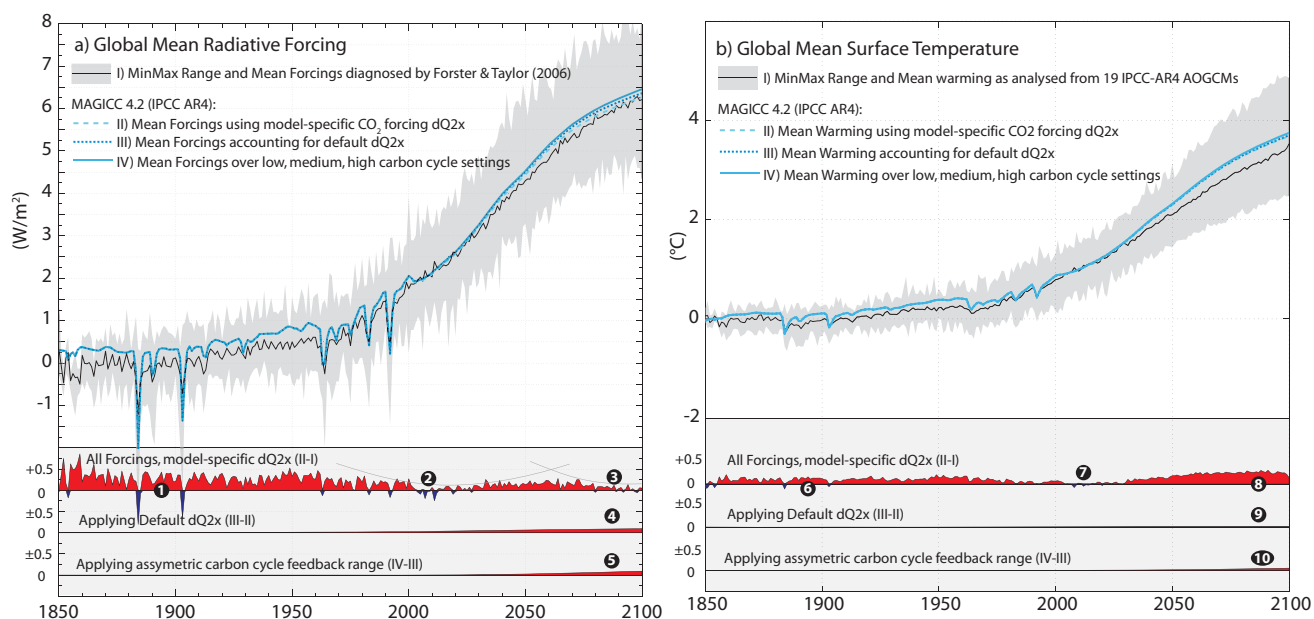
The most obvious difference is a higher forcing in MAGICC4.2 from 1850 to 1970, starting to diminish thereafter. There are three reasons for this: Firstly, MAGICC4.2 applied forcings from 1765 for all models, while many AOGCMs branched off the pre-industrial control runs later. Secondly, MAGICC4.2 applied volcanic forcings differently from the way these were applied in the AOGCMs (i.e., those AOGCMs that included volcanic forcing – not all did).

Thirdly, a number of models ignored some important forcings, such as those due to indirect aerosol effects and/or stratospheric ozone changes.

More specifically, forcing differences start to diminish after around 1970 and are small by the beginning of the 21st century (“2” in Fig. 5a). This is primarily due to the more pronounced negative forcing from aerosols in MAGICC4.2 compared to the average of the AOGCMs, of which only 9 models included indirect aerosol effects. Furthermore, in the year 2000, the MAGICC4.2 forcing has a step downward to match the historical mean volcanic forcing. Towards the end of the 21st century, applied forcings in MAGICC4.2 increase again above the diagnosed F&T AOGCM forcings, partially due to adjustments of the CO<sub>2</sub> forcing strength (“4”

**Table 2.** Cumulative fossil CO<sub>2</sub> emissions retrieved by prescribing recommended RCP concentrations (Meinshausen et al., 2011b) and using nine C<sup>4</sup>MIP carbon cycle emulations to inversely retrieve fossil CO<sub>2</sub> emissions. Harmonized land-use CO<sub>2</sub> RCP emissions (last column) were prescribed.

Scenario	Time Horizon	Fossil CO <sub>2</sub> (GtC)		land-use CO <sub>2</sub> (GtC)
		Median	80% Range	Median
RCP3-PD	2000–2049	322	(288–370)	41
	2000–2099	314	(267–421)	67
	2000–2199	222	(177–337)	74
RCP4.5	2000–2049	461	(418–510)	25
	2000–2099	752	(675–880)	30
	2000–2199	894	(800–1136)	31
RCP6	2000–2049	453	(412–501)	5
	2000–2099	1162	(1053–1296)	6
	2000–2199	1607	(1474–1977)	8
RCP8.5	2000–2049	598	(542–648)	43
	2000–2099	1821	(1637–1974)	60
	2000–2199	4283	(4097–4846)	61



**Fig. 5.** Net radiative forcing (panel a) and global-mean surface temperatures (b) for the SRES A1B scenario using MAGICC4.2 as presented in IPCC AR4. The standard radiative forcing used by MAGICC4.2 is compared to the means across all 19 CMIP3 AOGCMs, as diagnosed by Forster and Taylor (2006) (panel a), resulting in an apparently relatively close agreement (see text). In contrast, the AOGCM-specific subsets of forcings used in calibrating MAGICC6, as shown in Fig. 1, are lower towards the end of the 21st century. For temperature, the MAGICC4.2 emulations are shown compared to the means diagnosed from the matching set of 19 CMIP3 AOGCMs. See text for discussion (black circled numbers). As presented in IPCC AR4 (see Fig. 10.26 in Meehl et al., 2007), the MAGICC 4.2 temperature data shown here are given relative to a 21-year mean around 1990 of 0.52 °C above 1861–1890 (Brohan et al., 2006). The AOGCM temperature perturbations are shown relative to their control runs as a means to removing control run drift. The mean across AOGCMs relative to their control runs agrees well with the observational data around 1990, although there is a significant spread across AOGCMs (grey shading).

in Fig. 5a) to a central 3.71 Wm<sup>-2</sup> estimate for doubled CO<sub>2</sub> forcing, and partially due to the application of the low and high carbon cycle feedback estimates, which cause on average an increase of applied forcings (“5” in Fig. 5a).

Both the forcings in MAGICC4.2 and the diagnosed F&T forcings are likely to be higher than the actually effective forcings in the AOGCMs towards the end of the 21st century (see “3” in Fig. 5a). On the one hand, we expect the

MAGICC forcing to show a larger increase between 1980–1999 and the end of the 21st century due to the reduced masking of the warming trend due to reducing aerosol emissions – given that some aerosol effects are not included in many AOGCM forcings and therefore the reduction of the masking, if any, is less in these models. In MAGICC4.2, the fact that the indirect aerosol effect was modeled solely as a function of SO<sub>x</sub> aerosols contributed to this strong reduction in the aerosol masking, as SRES SO<sub>x</sub> emissions are assumed to decline faster than nitrate or other aerosol emissions. On the other hand, the diagnosed F&T forcings are probably overestimates due to the increasing effective climate sensitivity in some AOGCMs, as detailed above (see Sect. 3). F&T assumed constant climate sensitivities.

In summary, the key point is that F&T estimates likely overestimate actual AOGCM forcings towards the end of the 21st century (relative to the 1980–1999 base period). Thus, the more comprehensive forcings applied in MAGICC4.2, although matching F&T forcings relatively closely, might actually be higher than those effective in the individual AOGCMs. As stated above, the more comprehensive forcing series used in MAGICC4.2, in many cases, is an improvement on the AOGCM forcings because many AOGCMs did not include all forcings: in particular indirect aerosol effects that lead to a net warming between 1980–1999 and the end of the century.

## 5.2 Comparing temperatures

We now consider the temperature consequences of these forcing differences. While MAGICC4.2 emulations show, in line with the forcing differences discussed above, a higher warming initially in the 20th century, the warming rate is lower than in the AOGCMs after approximately 1970 until the beginning of the 21st century. As noted above, the mean warming from 1980–1999 to 2090–2099 exhibited by the calibrated MAGICC4.2 emulations exceed the mean AOGCM warming (by  $\approx 10\%$ ).

A base period of 1900–1970 would have reduced the future temperature differences by about half, although MAGICC4.2 emulations would have exhibited cooler temperatures around the year 2000. Thus, some of the additional warming in MAGICC4.2 emulations relative to the CMIP3 AOGCMs and their 1980–1999 base period is due to the reduced aerosol masking effect, or in other words, due to the fact that not all AOGCMs included all aerosol forcings.

Some additional warming in the MAGICC4.2 emulations could be caused by climate sensitivity estimates made in the calibration of MAGICC4.2 that are optimal to explain the higher-forcing idealized scenarios but are too high for the multi-forcing agent SRES runs. This is because the MAGICC4.2 calibration attempted to find a single compromise climate sensitivity that emulated the rather high-forcing part of the idealized 1pctto4x scenarios as well as the lower 1pctto2x scenario, even though some AOGCMs exhibit increased ef-

fective climate sensitivities. The average climate sensitivity estimated for the AOGCMs decreases by approximately 2.5% if the calibration employs both the idealized and the SRES scenarios, rather than only the idealized scenarios (see difference between calibration I and II in Table 4 in MRW).

The correction of the CO<sub>2</sub> forcing strength, using a default  $\Delta Q_{2\times}$  parameter of 3.71 Wm<sup>-2</sup> has a very small influence on the mean temperature evolution (see “9” in Fig. 5b). A slight additional warming is noticeable due to the inclusion of the uncertainty in the carbon cycle feedbacks, averaging across the low, mid and high carbon cycle feedback settings, that were applied for IPCC AR4.

In summary, a major reason for MAGICC4.2 results being warmer than the average AOGCM projection is the difference between the AOGCM-specific incomplete forcing series and the full forcing series applied in MAGICC4.2. To some extent, this difference therefore represent a correction to the AOGCM results. The accuracy of this correction is, of course, limited by the realism of the applied forcings in MAGICC4.2, in particular for aerosol-induced forcings. On top of that, the calibrated climate sensitivities in MAGICC4.2 were probably, on average, higher than appropriate for the low-forcing part of the SRES scenarios. The successful emulations of AOGCMs using MAGICC6 in the present study (see Fig. 4 in MRW) show that there are no inherent structural limitations in simple models that might lead to problems in their ability to accurately emulate AOGCMs.

## 6 Discussion

### 6.1 Volcanic forcing assumptions

The following paragraph highlights one of the forcing assumptions that must be made in order to carry out future temperature projections, namely assumptions regarding volcanic forcing. Judging from their temperature evolutions, for those CMIP3 AOGCMs that included the effects of historical volcanic eruptions (see Table 2 in MRW), volcanic forcing was applied as a negative forcing only, i.e., the control run assumed zero stratospheric volcanic aerosol concentrations. The effect of this is to cause a long-term cooling trend after the runs branch off the pre-industrial control simulations (cf. Gregory, 2010). As stated above, this long-term cooling trend is spurious.

Volcanic forcing is not known beyond the present, so this leads to the question, what should be assumed for future volcanic forcing? (The same question applies to future solar forcing – see below.) As far as we can determine, almost all those models that included historical volcanic forcing assumed that the forcing (which was essentially zero at the end of the 20th century simulations) remained at zero. As there will be volcanic eruptions in the future, although we do not know when or how large these will be, we do know that their mean forcing (as it was in the 20th century) will be negative.

An alternative (and, we claim, more realistic) assumption would be to assume a constant negative forcing equal to the long-term (e.g. 20th century) mean. In fact, there are three possible constant-forcing assumptions for the future: zero forcing; continued forcing at the level that prevailed in the recent history; and forcing at the long-term mean historical forcing (or the mean over some representative period).

In our simulations, we use the long-term mean assumption for volcanic forcing, using the average over the 20th century. Furthermore, we set this mean to zero, in order to avoid the spurious cooling trend noted above. This assumption is consistent with assuming that the climate system was on average in equilibrium with this negative forcing in pre-industrial times. An equivalent approach is often used in HadCM3 simulations (J. Lowe, personal communication, 2007; see as well Fig. 1 in Stott et al., 2000). Note, however, that the HadCM3 runs stored in the PCMDI CMIP3 archive did not include volcanic forcings (in contrast to the information provided in Table 10.1 of Meehl et al., 2007). For solar forcing, we set the future level at the mean over the last 11 years, which is very close to the value in the year 2000.

## 6.2 Interpretation of multi-model ensembles

The assumption that each of the AOGCMs or each of the C<sup>4</sup>MIP models should be given equal weight, is certainly a simplification as it does not account for the different skills of these models and their structural dependence. For example, many AOGCMs share (to a varying degree) model components, such as the MOM ocean code (Bryan, 1969). To illustrate the problem, consider the hypothetical case where two AOGCMs are absolutely identical, but are submitted to an intercomparison exercise under different names by different modeling groups. Should that particular model then carry twice the weight in the ensemble average? Obviously not (for a discussion, see Tebaldi and Knutti, 2007; Knutti, 2010; Santer et al., 2009). Such an “equal likelihood” assumption affects both the uncertainty range and the multi-model ensemble means, with the latter often portrayed as “best estimates” (Meehl et al., 2007). In the absence of appropriate weights that would capture the individual models’ projection skill and interdependence, a second-best approach seems to continue the tradition of reporting unweighted ensemble means. There are two advantages in giving the multi-model average results. Apart from characterizing the overall performance of a range of models, there is strong evidence that ensemble means tend to outperform individual models for various performance metrics (see e.g. Tebaldi and Knutti, 2007, and references therein). If all models were equally likely to be “correct”, then the resulting ensemble of 171 emulations employed here could be interpreted as a probability distribution. In fact, the distributions spanned by the 171 emulations simply denote distributions of occurrences or “ensembles of opportunity” – a measure of uncertainty

that arises from inter-model differences. These “ensembles of opportunity” are a collection of best estimates made by each modeling group rather than an attempt to explore the extremes of the uncertainty range. “Ensembles of opportunity” are therefore likely to under-estimate the actual uncertainty.

The 90% ranges spanned by the emulations for CMIP3 SRES scenarios (−31% to +43%, cf. Table 1) are narrower compared to the “likely” IPCC AR4 range (−40% to +60%). This is consistent with the fact that the emulation results do not account for forcing uncertainties. Furthermore, the fact that the CMIP3 AOGCM and C<sup>4</sup>MIP carbon cycle sets do not span the complete range of plausible climate and carbon cycle responses supports the larger uncertainty ranges provided by IPCC AR4 (cf. Knutti et al., 2008). For example, independent estimates of climate sensitivity uncertainties (Meehl et al., 2007; Hegerl et al., 2007) find wider uncertainty ranges than the purely AOGCM-based range used here – as one would expect given that what we have here is a limited “ensemble of opportunity”.

## 7 Conclusions

We showed that MAGICC can successfully emulate global and hemispheric-average temperatures of AOGCMs, as well as key quantities of carbon cycle models. The difference between the calibrated MAGICC6 model and the mean of AOGCMs across all SRES scenarios is only 0.04 °C in global mean temperatures over the 21st century – when compared on the basis of AOGCM-specific forcing subsets. Given this high emulation skill, MAGICC can and does assist in both the diagnosis and design of AOGCM and ESM intercomparison exercises.

We provided examples of MAGICC6 applications regarding the diagnosis of previous intercomparison exercises, in particular CMIP3 and C<sup>4</sup>MIP. Direct AOGCM results are disparate because different models used different combinations and magnitudes of the suite of important forcing agents. This leads to ambiguities in determining the reasons for differences in AOGCM temperature projections: how much is due to different climate feedbacks and inertia, and how much is simply an expression of different forcing assumptions? MAGICC6, by using common forcings, can at least partly answer this question. Our results suggest that the mean of AOGCM projections for the SRES A1B scenario as reported in IPCC AR4 would have been 0.1 °C cooler over the 21st century, if all models had taken into account all forcing agents, primarily the indirect aerosol effects. This cooling is approximately offset by the warming that could have resulted from coupling carbon cycle models, so that reported IPCC AR4 ranges, with a baseline of 1980–1999, seem in the end unaffected from these adjustments. Until 1960, however, our results suggest that the mean AOGCM results are slightly cooler than if all forcings and common starting years in 1765 would have been applied. Between 1950 and 2000,

the direct AOGCM results are likely to show a larger warming trend compared to what they would have been with common and complete forcing assumptions, in part because of missing or low volcanic forcing assumptions in some models (see Fig. 1b and c). Volcanic forcings pose an additional complication for interpreting AOGCM results in the early 20th century runs. As the control run in all CMIP3 experiments did not include stratospheric volcanic aerosol loadings comparable to actual pre-industrial volcanic activity, an initial spurious cooling drift is embodied in the CMIP3 20th century AOGCM runs. A key conclusion is hence that in future AOGCM intercomparison exercises the effective forcing fields in each model should be diagnosed as far as possible. The separation between climate response uncertainties, forcing uncertainties and emissions uncertainties is impossible in the absence of such forcing diagnostics.

Furthermore, we showed how MAGICC can assist in the design and planning of future intercomparison exercises. Pre-empting AOGCM results for the forthcoming CMIP5 results, we present here global mean temperature projections for the RCP scenarios. The highest RCP, RCP8.5, is projected to result in temperatures in excess of 7 °C by 2300, while the lowest RCP, RCP3-PD, is projected to peak slightly above 1.5 °C (1.3–2.0, 90% range), and then decrease by approximately 0.2 °C per century thereafter – due to negative emissions after 2070. While MAGICC6 has been used to assist the CMIP5 exercise by providing the RCP's GHG concentrations that will be prescribed in AOGCMs, our temperature projections will be an independent test of both (a) the emulation skill of MAGICC6 for new, i.e., non-calibrated, scenarios and (b) for the difference between the CMIP3 and CMIP5 generation of AOGCMs. Attributing differences between our projected RCP temperature ranges and the forthcoming range of CMIP5 projections to both effects (a) and (b) will then only be possible by ex-post analysis of CMIP5 results - taking into account the specific subsets of forcings that have actually been applied in individual AOGCM runs.

The future development of MAGICC will focus on the emulation of the next-generation of CMIP5 ESMs. In addition, future enhancements of MAGICC regarding gas-cycle parameterisations, ozone chemistry, indirect forcing effects and aerosol interactions can strengthen its role as a cross-disciplinary model for global change and impact assessments.

*Acknowledgements.* We would like to acknowledge the numerous collaborations during the preparation process of IPCC AR4 and beyond, without which this study would not have been possible. Specifically, we would like to thank J. Gregory, K. Taylor, P. Gleckler, B. Santer, J. Meehl, J. Arblaster, R. Lieberman, F. Joos, J. Daniel, G. Velders and R. Knutti. R. Knutti is as well deeply thanked for discussions and comments on an earlier draft of this manuscript, as are J. Lowe, T. Schneider von Deimling, R. Schofield, V. Brovkin, B. Hare and E. Kriegler. We acknowledge the modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI)

for collecting and archiving the model output of the third coupled model intercomparison (CMIP3), and the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) for organizing the model data analysis activity. The multi-model data archive is supported by the Office of Science, US Department of Energy.

Edited by: W. E. Asher

## References

- Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D.: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, 111, D12106, doi:10.1029/2005JD006548, 2006.
- Bryan, K.: A numerical method for the study of the circulation of the world ocean, *J. Comput. Phys.*, 4, 347–376, doi:10.1016/0021-9991(69)90004-7, 1969.
- Clarke, L., Edmonds, J., Jacoby, H., Pitcher, H., Reilly, J., and Richels, R.: Scenarios of Greenhouse Gas Emissions and Atmospheric Concentrations – US Climate Change Science Program Synthesis and Assessment Product 2.1a, Tech. rep., US Department of Energy, 2007.
- Clarke, L., Edmonds, J., Krey, V., Richels, R., Rose, S., and Tavoni, M.: International climate policy architectures: Overview of the EMF 22 International Scenarios, *Energy Economics*, 31, S64–S81, 2009.
- Collins, W. D., Ramaswamy, V., Schwarzkopf, M. D., Sun, Y., Portmann, R. W., Fu, Q., Casanova, S. E. B., Dufresne, J. L., Fillmore, D. W., Forster, P. M. D., Galin, V. Y., Gohar, L. K., Ingram, W. J., Kratz, D. P., Lefebvre, M. P., Li, J., Marquet, P., Oinas, V., Tsushima, Y., Uchiyama, T., and Zhong, W. Y.: Radiative forcing by well-mixed greenhouse gases: Estimates from climate models in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4), *J. Geophys. Res.*, 111, D14317, doi:10.1029/2005JD006713, 2006.
- Forster, P., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R., Fahey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R.: Chapter 2: Changes in Atmospheric Constituents and in Radiative Forcing, in: IPCC Fourth Assessment Report WG I, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, 129–234, 2007.
- Forster, P. M. D. and Taylor, K. E.: Climate forcings and climate sensitivities diagnosed from coupled climate model integrations, *J. Climate*, 19, 6181–6194, 2006.
- Fujino, J., Nair, R., Kainuma, M., Masui, T., and Matsuoka, Y.: Multi-gas mitigation analysis on stabilization scenarios using AIM global model, *Energy J.*, 343–353, 2006.
- Gregory, J. M.: Long-term effect of volcanic forcing on ocean heat content, *Geophys. Res. Lett.*, 37, L22701, doi:10.1029/2010GL045507, 2010.
- Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N., Luo, Y., Marengo Orsini, J. A., Nicholls, N., Penner, J. E., and Stott, P. A.: Understanding and Attributing Climate Change, 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*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tig-

- nor, M., and Miller, H., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007.
- Houghton, R. A. and Hackler, J. L.: Carbon Flux to the Atmosphere from Land-Use Changes, In *Trends: A Compendium of Data on Global Change*, 2002.
- Hurttt, G., Chini, L., Frolking, S., Betts, R., Edmonds, J., Feddema, J., Fisher, G., Goldewijk, K., Hibbard, K., Houghton, R., Janetos, A., Jones, C., Kinderman, G., Konoshita, T., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D., and Wang, Y.: Land Use Change and Earth System Dynamics, *Clim. Change*, submitted, 2011.
- Knutti, R.: The end of model democracy?, *Clim. Change*, 2010.
- Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meehl, G. A., Meinshausen, M., Murphy, J. M., Plattner, G.-K., Raper, S. C. B., Stocker, T. F., Stott, P. A., Teng, H., and Wigley, T. M. L.: A review of uncertainties in global temperature projections over the twenty-first century, *J. Climate*, 21, 2651–2663, 2008.
- Lamarque, J.-F., Kyle, G. P., Meinshausen, M., Riahi, K., Steven, J. S., van Vuuren, D. P., Conley, A. J., and Vitt, F.: Global and regional evolution of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways, *Clim. Change*, accepted.
- Lean, J. L. and Rind, D. H.: How will Earth's surface temperature change in future decades?, *Geophys. Res. Lett.*, 36, L15708, doi:10.1029/2009GL038932, 2009.
- Lowe, J. A., Huntingford, C., Raper, S. C. B., Jones, C. D., Lidicoat, S. K., and Gohar, L. K.: How difficult is it to recover from dangerous levels of global warming?, *Environ. Res. Lett.*, 4, 014012, 1–9, 2009.
- Marland, G., Boden, T. A., and Andres, R. J.: Global, Regional, and National Fossil Fuel CO<sub>2</sub> Emissions, In *Trends: A Compendium of Data on Global Change*, 2006.
- Matthews, H. D. and Caldeira, K.: Stabilizing climate requires near-zero emissions, *Geophys. Res. Lett.*, 35, L04705, doi:10.1029/2007GL032388, 2008.
- Meehl, G. A., Covey, C., McAvaney, B., Latif, M., and Stouffer, R. J.: Overview of coupled model intercomparison project, *B. Am. Meteorol. Soc.*, 86, 89–93, 2005a.
- Meehl, G. A., Washington, W. M., Collins, W. D., Arblaster, J. M., Hu, A. X., Buja, L. E., Strand, W. G., and Teng, H. Y.: How much more global warming and sea level rise?, *Science*, 307, 1769–1772, 2005b.
- Meehl, G. A., Stocker, T. F., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J., Noda, A., Raper, S. C. B., Watterson, I., Weaver, A., and Zhao, Z.-C.: Chapter 10: Global Climate Projections, in: *IPCC Fourth Assessment Report*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, 747–845, 2007.
- 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, 2011a.
- Meinshausen, M., Smith, S., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A. M., Velders, G. J. M., and van Vuuren, D.: The RCP Greenhouse Gas Concentrations and their Extension from 1765 to 2300, *Clim. Change*, submitted, 2011b.
- Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, *Nature*, 463, 747–756, doi:10.1038/nature08823, 2010.
- Myhre, G., Highwood, E. J., Shine, K. P., and Stordal, F.: New estimates of radiative forcing due to well mixed greenhouse gases, *Geophys. Res. Lett.*, 25, 2715–2718, 1998.
- Orr, J.: Global Ocean Storage of Anthropogenic Carbon (GOSAC). *EurOCMIP-2 Final Report*, Tech. rep., IPSL/CNRS, 2002.
- Riahi, K., Gruebler, A., and Nakicenovic, N.: Scenarios of long-term socio-economic and environmental development under climate stabilization., *Technol. Forecast. Soc. Change*, 74, 887–935, 2007.
- Santer, B. D., Taylor, K. E., Gleckler, P. J., Bonfils, C., Barnett, T. P., Pierce, D. W., Wigley, T. M. L., Mears, C., Wentz, F. J., Bruggemann, W., Gillett, N. P., Klein, S. A., Solomon, S., Stott, P. A., and Wehner, M. F.: Incorporating model quality information in climate change detection and attribution studies, *Proc. Natl. Acad. Sci.*, 106, 14778–14783, 2009.
- Schewe, J., Levermann, A., and Meinshausen, M.: Climate change under a scenario near 1.5°C of global warming: Monsoon intensification, ocean warming and steric sea level rise, *Earth Syst. Dynam. Disc.*, 1, 297–324, doi:10.5194/esdd-1-297-2010, 2010.
- Stott, P. A., Tett, S. F. B., Jones, G. S., Allen, M. R., Mitchell, J. F. B., and Jenkins, G. J.: External Control of 20th Century Temperature by Natural and Anthropogenic Forcings, *Science*, 290, 2133–2137, 2000.
- Taylor, K., Stouffer, R. J., and Meehl, G. A.: A summary of the CMIP5 Experiment Design, [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf), 2009.
- Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, *Philosoph. Trans. Roy. Soc. a-Math. Phys. Eng. Sci.*, 365, 2053–2075, 2007.
- van Vuuren, D., den Elzen, M., Lucas, P., Eickhout, B., Strengers, B., van Ruijven, B., Wonink, S., and van Houdt, R.: Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs, *Clim. Change*, 81, 119–159, 2007.
- Wigley, T. M. L.: The climate change commitment, *Science*, 307, 1766–1769, 2005.
- Wigley, T. M. L., Clarke, L. E., Edmonds, J. A., Jacoby, H. D., Paltsev, S., Pitcher, H., Reilly, J. M., Richels, R., Sarofim, M. C., and Smith, S. J.: Uncertainties in climate stabilization, *Clim. Change*, 97, 85–121, 2009.