Supplement: Evaluating Climate Emulation: Unit Testing of Simple Climate Models

Adria K. Schwarber[1], Steven J. Smith[1,2], Corinne A. Hartin[2], Benjamin Aaron Vega-Westhoff[3], Ryan Sriver[3]

[1]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742

[2]Joint Global Change Research Institute, 5825 University Research Ct, College Park, MD 20740

[3]Department of Atmospheric Sciences, University of Illinois Urbana-Champaign, Champaign, IL 61820

Correspondence to: Adria K. Schwarber (adria.schwarber@gmail.com)

Contents

S1 Supplementary Method

We note that unit testing in software refer to a specific method of comparing output from the smallest portion of code, called a unit (i.e., function), to known outputs (Clune and Rood 2011). Here, we use this term in a similar way as van Vuuren *et al.* (van Vuuren et al. 2011), where MAGICC 6.0 was used as the reference output to compare several human-Earth system models. We conduct our unit test with comparable inputs and compare model-generated outputs from several SCMs.

We conduct perturbations of three contrasting chemical species: carbon dioxide (CO_2), methane (CH_4), and black carbon (BC). We begin with CO_2 because this well-mixed greenhouse gas is the largest contributor to anthropogenic forcing changes (Myhre et al. 2013). Methane is also of interest because it is a shorter-lived greenhouse gas, with chemical interactions with itself and other species (Cicerone, R.J.; Oremland 1988). Finally, we use BC perturbations to represent aerosols more generally because we are interested in model responses to a short-lived climate forcers (Bond et al. 2013; Harmsen et al. 2015). SCM representations of other aerosols species are similar so we do not conduct impulse tests of other species.

The SCMs we use are readily comparable because they read in similar emissions files. Background trajectory emissions are taken from the published Representative Concentration

Pathway (RCP) 4.5 scenario (Thomson et al. 2011) database, which means that all calculations in the main paper are conducted relative to a changing CO_2 concentration background, unless otherwise noted. SCMs are often used to project global mean temperature over various future scenarios, so this is the most relevant type of background on which to test these models. Conducting these experiment with a constant CO_2 background, as previously used in the literature (Joos et al. 2013), requires inverse modeling of the individual models to produce constant CO_2 concentration emissions files. Our methodology is easier to implement as a regular unit test. To this end, we provide comparable input emission files used in this paper.

In many SCMs, forcing over historical periods is explicitly calibrated to a model base year, so it is not possible to conduct perturbations during these time periods. Therefore, our perturbations are conducted in 2015 to avoid the model base years of our SCMs. In the main paper, we show some model responses out to 2300, the end of the MAGICC model runs, equal to 285 years after the perturbation. Additional results are in the Supplement (SI8).

We run reference scenarios in the SCMs, followed by each perturbation case described below. For each experiment (see below) we report the response, which is obtained by subtracting the reference from the perturbation results. For instance, the CO_2 concentration response is obtained as follows:

$$CO_2Concentration_{response}(t) = CO_2Concentration_{perturabtion}(t) - CO_2Concentration_{reference}(t) \quad (1)$$

We conducted the following impulse tests:

a. Concentration impulse (CO_2).

These SCMs can be used in a mode where CO_2 concentrations are exogenously specified. We carry out this experiment by instantaneously increasing CO_2 concentration by 200 ppm in 2015. After 2015, CO_2 concentrations return to the baseline levels following the published RCP4.5 scenario. Note, we do not conduct separate forcing impulse experiments because this is functionally equivalent to a concentration impulse. In this experiment, we are only interested in the dynamics of the models' temperature response. This experiment eliminates the added uncertainty in the emissions to concentrations calculation and complicating factors from carbon cycle feedbacks.

b. Emissions impulse (BC, CH_4, CO_2).

For this experiment all models were run with an emissions input. We carry out this experiment by increasing individual emissions (BC, CH_4, or CO_2) in one year. Following that year, the emissions return to the RCP4.5 pathway for all subsequent years. In this experiment CO_2 concentrations are allowed to vary as determined by each model. We find our perturbation values by doubling the 2015 value for each chemical species equal to a 9.2 PgC pulse of CO_2, a 329 Tg pulse of CH_4, and a 7981 Gg pulse of BC. We also perturb CO_2 emissions in 2010, 2020, 2030, 2040, 2050 to understand changes in model responses over time and see very small difference in the model response (SI5). We compare results from three comprehensive SCMs to two IR models, AR5-IR and FAIR model (Millar et al. 2017; Myhre et al. 2013) (SI2).

We also compared results to several ESMs and EMICs by carrying out a 100 GtC CO_2 impulse, following Joos *et al*. (Joos et al. 2013) (SI11). This is approximately 10x the CO_2 perturbation pulse described above.

Finally, we conduct a 4xBC emissions step experiment. We compare the SCM temperature responses with the response of a complex climate model used by Sand *et al.* (Sand et al. 2016) (SI12).

c. Step increase in CO_2 concentration (instantaneous $4\times CO_2$ concentration experiment).

Similar to comparison (a), in this experiment, CO_2 concentrations are prescribed. We have CO_2 concentrations follow a pre-industrial pathway (278.0516 ppmv in 1765) until 2014. The CO_2 concentration is quadrupled (4x) in 2015, and maintained at this level until 2300. This follows experimental protocol used in the CMIP5 experimental design (Taylor, Stouffer, and Meehl 2012).

We compare these results to drift-corrected (Gupta et al. 2013) global mean temperature results from 15 complex climate models from the CMIP5 archive. We drift-correct the CMIP5 global mean temperature time series by subtracting the slope of the linear fit from the full time series for each individual model.

We ran Hector v2.0 with few changes to the default configuration file settings. We changed two model time steps in Hector v2.0: (1) the carbon-cycle-solver.cpp time step from dt(0.3) to dt(0.1) and (2) the ocean_component.hpp OCEAN_MIN_TIMESTEP from 0.3 to 0.01 to allow for the carbon cycle, in particular the ocean carbon cycle to accurately integrate across the sharp gradient introduced by these experiments. In experiments where we constrained the CO_2 concentration, these changes significantly increase the model run time for this scenario.

Additionally, we used an equilibrium climate sensitivity (ECS) value of 3°C in the SCMs, with the exception of the idealized SCMs, FAIR and AR5-IR, where these parameters cannot be set by the user (see Table S9).

S2 Model Design

We conduct unit tests within three comprehensive SCMs and two stylized SCMs. The three comprehensive SCMs have structural differences worth noting. Hector v2.0, has explicit ocean carbon chemistry in four boxes, where ocean carbon uptake is a non-linear function of the solubility of carbon. MAGICC 5.3 BC-OC and 6.0 have differential hemispheric forcing over land and ocean, thereby calculating temperature over each box. Important characteristics of the carbon and climate components of each model are shown in Table S1.

Table S1 Main carbon cycle and climate characteristics of SCMs and IRFs

Model	Model description	Carbon cycle	Climate component
Hector v2.0 (C. A. Hartin et al. 2015; Corinne A Hartin et al. 2016; Kriegler 2005a)	mechanistic climate carbon-cycle model	One-pool atmosphere, three-pool land, and four-pool ocean	Global Energy balance model, with ocean heat diffusion
MAGICC 5.3 BC-OC (Raper and Cubasch 1996; S. J. Smith and Bond 2014; Wigley and Raper 1992)	mechanistic climate carbon-cycle model	One-pool atmosphere, three-pool land, and one-pool ocean	4-box Energy balance model, with ocean heat upwelling diffusion
MAGICC 6.0 (Meinshausen, Raper, and Wigley 2011)	mechanistic climate carbon cycle model	One-pool atmosphere, three-pool land, and one-pool ocean	4-boxEnergy balance model, with ocean heat upwelling diffusion
AR5-IR (Myhre et al. 2013)	Impulse-response function	Impulse-response function	Equilibrium temperature as a function of RF
FAIR v1.0 (Millar et al. 2017)	Impulse-response function	Four timescale impulse-response function with state-dependence of the CO_2 airborne fraction	Equilibrium temperature as a function of RF; IRF with two timescales

Some SCMs also include representations of aerosol dynamics, though the model representations differ. As mentioned in the main paper, unlike Hector v2.0, both version of MAGICC have differential hemispheric forcing over land and ocean. AR5-IR represents BC forcing response as a simple exponential, similar to the response from greenhouse gas forcing. FAIR v1.0, used here, represents the relationship between CO_2-only emissions, concentrations, and temperature. Other versions of FAIR include non-CO_2 forcing, such as BC.

S2.1 Model Settings

Here we discuss the model settings used in our experiments, noting any changes made to the default settings. The three comprehensive SCMs were run with the same ocean diffusivity value and ECS value, unless otherwise noted.

S2.2 AR5-IR

The IPCC AR5 (Myhre et al. 2013) describes a multi-gas impulse function using a multi-gas equivalence metric, Absolute Global Temperature Potential (AGTP), to compare temperature changes at a chosen time in response to a unit pulse of emissions *i*. AGTP is found via a convolution of the fraction of the species *i* remaining in the atmosphere after an emissions pulse and the climate response to a unit forcing $R_T(t) = \sum_{j=1}^{M} \frac{c_j}{d_j} \exp(-\frac{t}{d_j})$ (1).

$$AGTP_i(H) = \int_0^H RF_i(t) R_T(H-t)dt \qquad (2)$$

and $RF_i(t) = A_i R_i(t)$, (3)

where for most species $R_i(t) = \exp(-\frac{t}{\tau_i})$, (4)

and for CO_2 $R_{CO_2}(t) = a_0 + \sum_{i=1}^{N} a_i \exp(-\frac{t}{\tau_i})$, (5)

and A_i is the radiative efficiency yielding, the general equation:

$$AGTP_i(H) = A_i \sum_{j=1}^{2} \frac{\tau c_j}{\tau - d_j} \left(\exp\left(\frac{-H}{\tau}\right) - \exp\left(\frac{-H}{d_j}\right) \right) \qquad (6)$$

AGTP can then be used to calculate global mean temperature change from any given emission scenario using,

$$\Delta T = \sum \int_{o}^{t} E_i(s) AGTP_i(t-s) ds \quad (7)$$

where E_i are the emissions of a species, t is the time horizon, and s is the time of emissions(Myhre et al. 2013). For this paper, AR5-IR was recoded in R and is available for download with the Supplementaty Materials.

S2.3 FAIR

The FAIR v1.0 model is a modified version of the AR5-IR carbon cycle component, updated to include the state-dependence of the CO_2 airborne fraction to reproduce the relationship between CO_2-only emissions, concentrations, and temperature over the historical period. Millar *et al.*(Millar et al. 2017) began with the impulse response functions used for calculation of multi-gas equivalence metrics in IPCC-AR5(Myhre et al. 2013) and extended the CO_2 IRF by coupling the carbon-cycle to the thermal response and to cumulative carbon uptake by terrestrial and marine sinks. FAIR is available for download at https://github.com/OMS-NetZero/FAIR.

Here, we use the first iteration of FAIR, but we note that two new versions have recently been published, FAIR v1.1 and FAIR v1.3. FAIR v1.3 extends the original version to, "calculate non-CO2 greenhouse gas concentrations from emissions, aerosol forcing from aerosol precursor emissions, tropospheric and stratospheric ozone forcing from the emissions of precursors, and forcings from black carbon on snow, stratospheric methane oxidation to water vapour, contrails and land use change (C. J. Smith et al. n.d.)."

S2.4 MAGICC 5.3 BC-OC

MAGICC 5.3 BC-OC is a version of MAGICC 5.3 developed in conjunction with the Global Change Assessment Model (GCAM). MAGICC 5.3 used here is available in GCAM version 4.4, available for download at https://github.com/JGCRI/gcam-core/releases. The major change in this version of MAGICC was the addition of explicit BC and OC(S. J. Smith and Bond 2014). To enable MAGICC 5.3 within GCAM, the climate model must be set to <Value name = "climate">../input/climate/magicc.xml</Value> within the configuration file. We ran this model with all its default configuration settings, unless otherwise noted in the text.

S2.5 MAGICC 6.0

MAGICC 6.0 was run with all the default settings. For the main experiments, the climate sensitivity was set to 3.0°C to match the default setting of MAGICC 5.3 BC-OC and Hector v2.0, unless otherwise noted. The MAGICC 6.0 executable is available for free download here: http://www.magicc.org/.

S2.6 Hector v2.0 Settings

We use a new version of Hector (v2.0), an open-source, object-oriented, simple global climate carbon-cycle model (C. A. Hartin et al. 2015). The model can found at: http://github.com/JGCRI/hector. In the version used here (Hector v2.0), Hector v1.0 is coupled to a 1-D diffusive heat and energy balance model (DOECLIM: Diffusion Ocean Energy balance CLIMate model). DOECLIM is well documented and has been widely used in climate uncertainty studies (Bakker et al. 2017; Kriegler 2005b; Urban et al. 2014). DOECLIM includes three tunable parameters: climate sensitivity, ocean vertical heat diffusivity, and a scaling factor for aerosol forcing (Garner, Reed, and Keller 2016). Using default values for these parameters, we find that the new coupled model (Hector v2.0) exhibits improved vertical ocean structure and heat uptake, as well as surface temperature response to radiative forcing, compared to earlier versions of Hector.

S3 CMIP5 Model Data

The CMIP5 model data used to produce Figure 4, Figure S12, and Figure S22 is described here. Climate model output from 15 models was obtained from the CMIP5 data archive (http://cmip-pcmdi.llnl.gov/cmip5/data_portal.html) and the World Date Center for Climate site (http://cerawww.dkrz.de/WDCC/ui/Index.jsp). The long-term drift is removed from the CMIP5 model data by subtracting the linear trend from the corresponding pre-industrial control run (Gupta et al. 2013). Table S2 provides the CMIP5 modeling centre name and the model name from Figure 4.

Table S2 CMIP5 and SCM model information

Centre(s)	Model name
Beijing Climate Center (BCC) **China**	BCC-CSM1.1
Canadian Centre for Climate Modelling and Analysis (CCCma) **Canada**	CanESM2
Centre National de Recherches Météorologiques, Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CNRM-CERFACS) **France**	CNRM-CM5-2
Institut Pierre Simon Laplace (IPSL) **France**	IPSL-CM5A-LR
	IPSL-CM5A-MR
	IPSL-CM5B-LR
Institute of Atmospheric Physics, Chinese Academy of Sciences (LASG-CESS) **China**	FGOALS-g2
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (MIROC) **Japan**	MIROC-ESM
	MIROC5
Max Planck Institute for Meteorology (MPI-M) **Germany**	MPI-ESM-MR
	MPI-ESM-P
NASA/GISS (Goddard Institute for Space Studies; NASA-GISS) **USA**	GISS-E2-H
	GISS-E2-R
Geophysical Fluid Dynamics Laboratory (NCAR; NSF-DOE-NCAR) **USA**	GFDL-CM3
	GFDL-ESM2G

S4 Sensitivity Experiments in MAGICC 5.3

We conduct two sensitivity experiments to illustrate there is little impact of these choices on the model responses: (1) perturb CO_2 emissions in different years and (2) perturb CO_2 emissions at different levels in 2015.

S4.1 Impact of Changes to the Years of Emission Impulses

We test CO_2 emissions perturbations in different years from the default 2015 used in the main text. **Error! Not a valid bookmark self-reference.** shows the global mean temperature response ormalized by the 2010 global mean temperature response from a CO_2 emissions pulse in MAGICC 5.3. We found a maximum of 0.028°C/PgC difference in the response in MAGICC 5.3 and, therefore, carried out the remainder of the experiment in 2015, avoiding model base years.

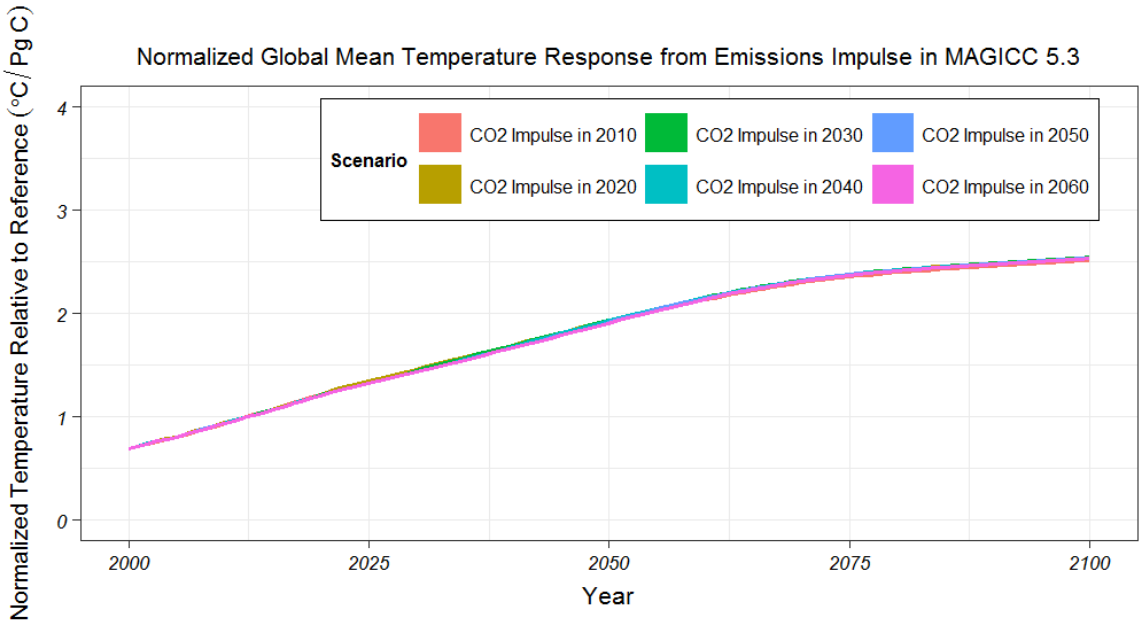

Figure S1 *Normalized global mean temperature response from CO_2 emissions impulses in MAGICC 5.3 carried out in different years.*

S4.2 Impact of Emissions Pulses Size on Temperature Response

In the main text we carried out annual emissions perturbations equivalent to doubling the value in 2015 to avoid model base years. Figure S2 shows the global mean temperature response normalized by the perturbation size for different CO_2 perturbation sizes in 2015 in MAGICC 5.3. We found there was a maximum difference of 0.0015°C/PgC, and thus we continued our experiments using only one perturbation value.

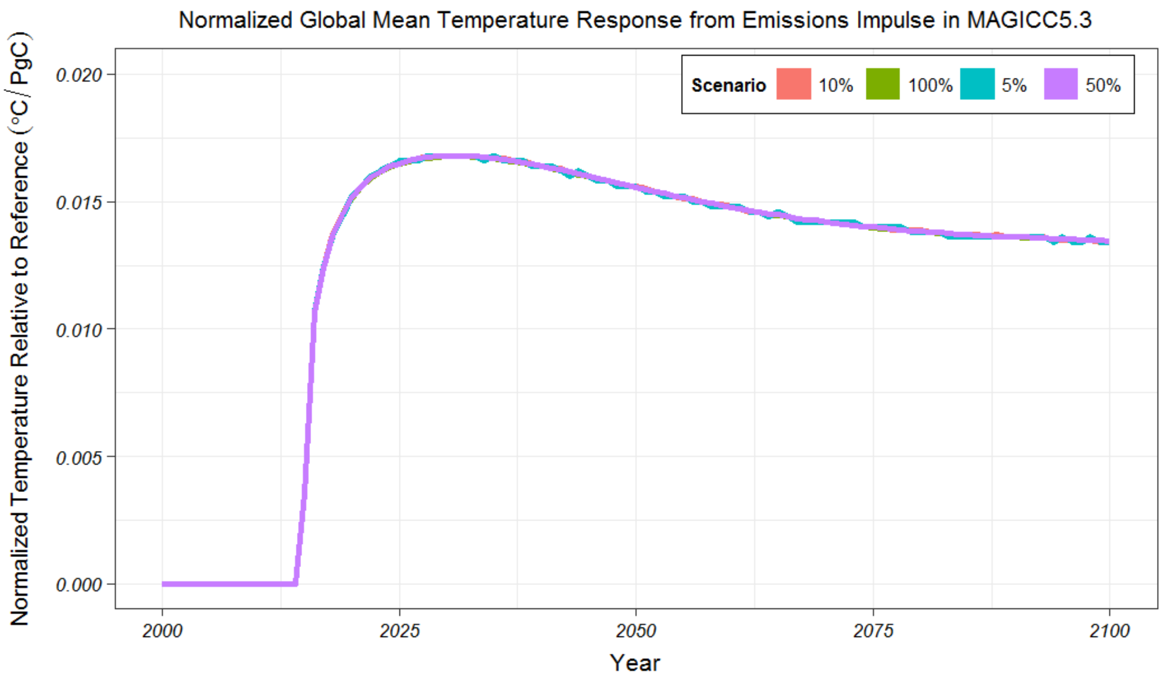

Figure S2 *Normalized global mean temperature response from different sized CO_2 emissions impulses in MAGICC 5.3 in 2015.*

S5 Adjusted Total Forcing Response

We found that MAGICC 5.3, MAGICC 6.0, and Hector v2.0 respond similarly to a CO_2 concentration impulse, with differences in the forcing and temperature responses arising from the treatment of time within each model. Hector v2.0, for example, reads in end-of-year emissions and carries out calculations of concentration, forcing, and temperature using that same classification of time. MAGICC 5.3 and MAGICC 6.0 read in end-of-year emissions and calculate concentration, forcing, and temperature at mid-year values, and successively reports temperature at the end-of year. This change in the timing effects the impulse response by distributing the pulse over more time periods. Here, we offer an adjustment for the forcing response to a CO_2 concentration impulse.

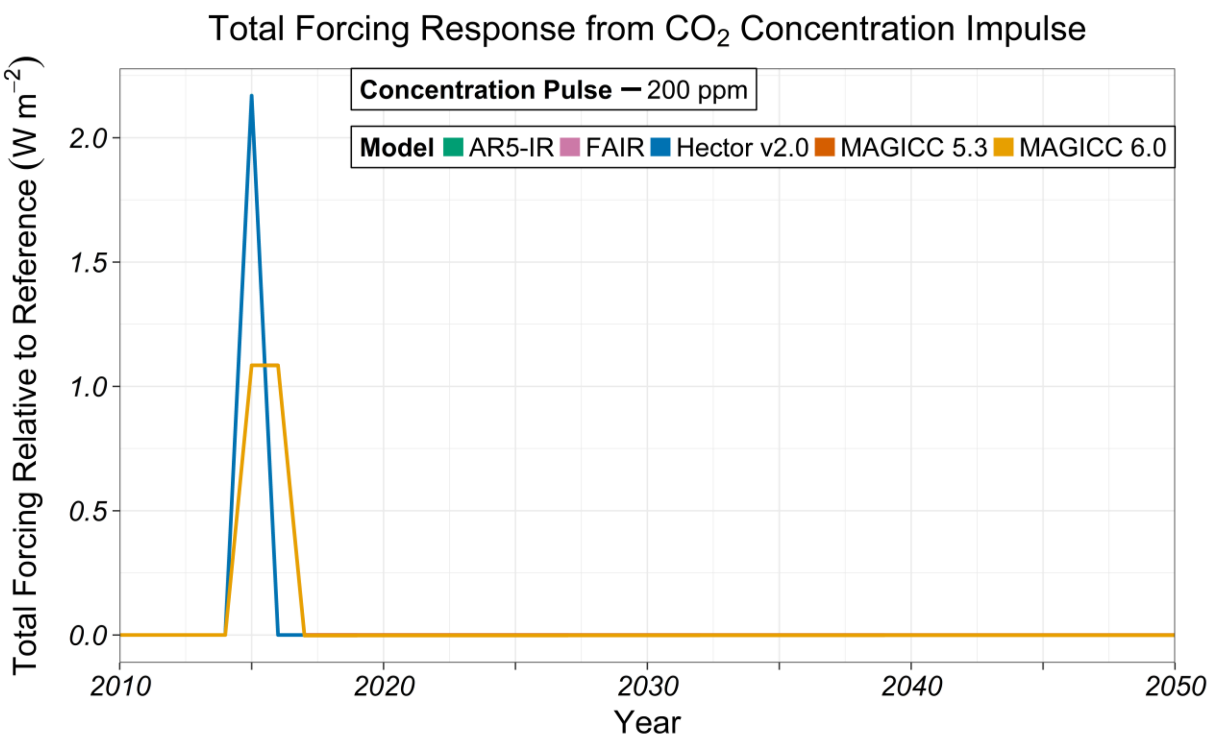

Figure S3 *Total forcing response from a CO_2 concentration impulse in SCMs. All three SCMs have a collinear response (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR –pink).*

Due to the differences in model treatment of time, we offer a correction to the forcing in two of the SCMs. MAGICC 5.3 and MAGICC 6.0 calculate forcing in mid-year, while Hector v2.0 reports forcing at the end of a year. The result is a broadened impulse response peak in both versions of MAGICC, compared to Hector v2.0. The total forcing response from both version of MAGICC, however, can be adjusted with the following equation:

$$\boldsymbol{F_i = (2xf_i) - f_{i-1}} \tag{8}$$

where $\mathrm{F_i}$ is the adjusted forcing, $\mathrm{f_i}$ is the unadjusted forcing at the current time step, and $\mathrm{f_{i-1}}$ is the unadjusted forcing at the previous time step.

Figure S4 shows the total forcing response adjusted from mid-year reporting, to end-year reporting using equation (SI. Eqn. 8). We can also apply this adjustment to the BC impulse, however, the MAGICC 6.0 distribution is larger in this case.

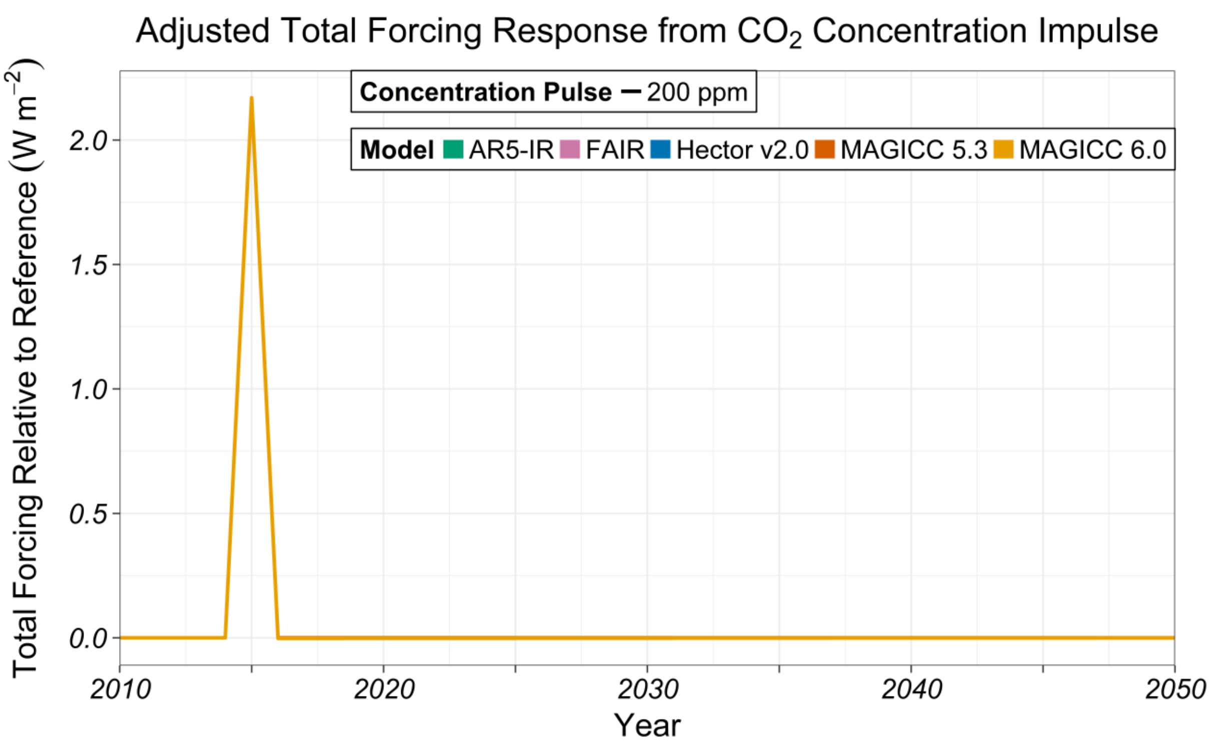

Figure S4 *Total forcing response from a CO_2 concentration impulse in SCMs. All three SCMs have a collinear response (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR –pink).*

S6 Total Forcing Response from BC Emissions Impulse

We see in Figure S5 that the model responses to a pulse of BC have similar patterns of instantaneous behavior seen in Fig. 1 from the CO_2 concentration pulse. In general, the models behave similarly in response to a BC pulse; Hector v2.0 and AR5-IR have a collinear response, while MAGICC 6.0 distributes the BC emissions pulse over 3 years.

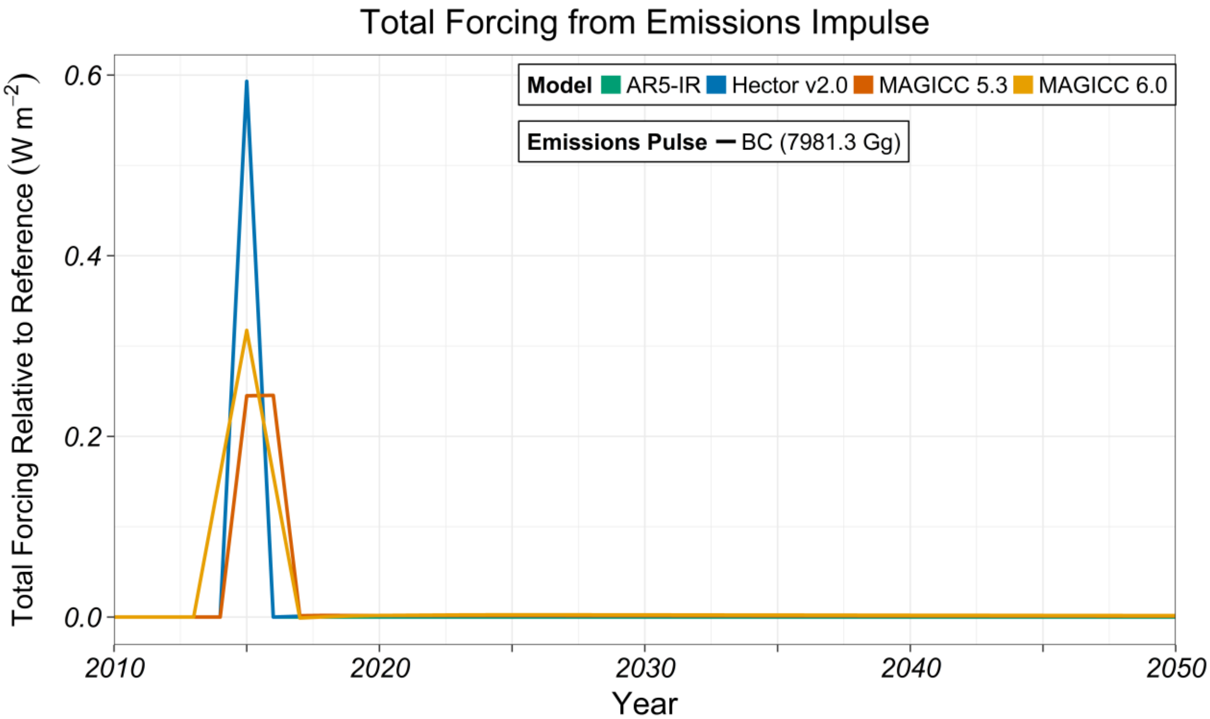

Figure S5 *Total forcing response from a BC emissions perturbation in SCMs (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR-green). AR5-IR and Hector v2.0 are collinear.*

S7 CO_2 Concentration Responses from Emissions Impulses

Figure S6 shows the CO_2 concentration responses from a BC and CH_4 emissions pulse. Every model response shows an eventual CO_2 concentration increases from a BC impulse; a secondary effect from the temperature increase. From a CH_4 and BC emissions pulse, the CO_2 concentration response is stronger in MAGICC 6.0, followed by MAGICC 5.3 and Hector v2.0. MAGICC 6.0, however, shows an initial decrease in CO_2 concentration response from the BC pulse.

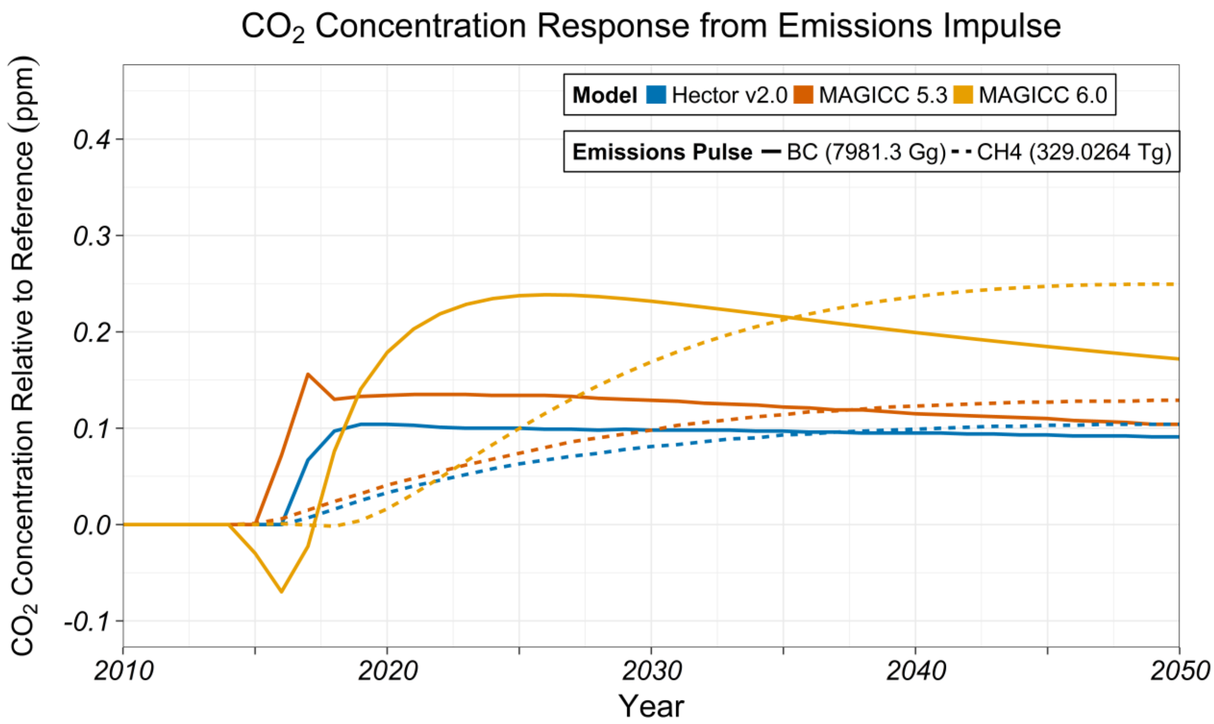

Figure S6 *CO_2 concentration response from CH_4 and BC emissions perturbation (B) in SCMs (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue) illustrating the carbon-cycle feedbacks present in each model.*

Figure S6 also shows that CH_4 emission perturbations impact CO_2 concentration within both versions of MAGICC. The discrepancy between the MAGICC and Hector responses is partly due to CH_4 oxidation in MAGICC 5.3. The MAGICC 6.0 response is larger in Figure S6 because of the temporal distribution of the pulse, however the general shape of the response is similar to the other two SCMs.

AR5-IR is notably absent from Figure S6 because, in this IRF, the CO_2 concentration is not affected by rising temperature or CO_2 accumulation from BC or CH_4 emissions perturbations (Millar et al. 2017). Similarly, the FAIR model (Millar et al. 2017) is absent from Figure S6. The CO_2 concentration response to a CO_2 emissions impulse in FAIR can be seen in Figure S8.

The CH_4 chemistry components in Hector v2.0 and MAGICC 5.3 BC-OC are nearly identical, accounting for the similarities between these two SCMs responses (C. A. Hartin et al. 2015). MAGICC 5.3, however, includes CH_4 oxidation to CO_2, which might account for this response difference. To test this, Figure S7 shows the CO_2 concentration response from emissions impulse in SCMs. MAGICC 5.3 is shown with and without CH_4 oxidation included for a clearer comparison the Hector v2.0 response. With the CH_4 oxidation turned off, the MAGICC 5.3 BC-OC response is similar to Hector v2.0 with only a slight difference after 2025.

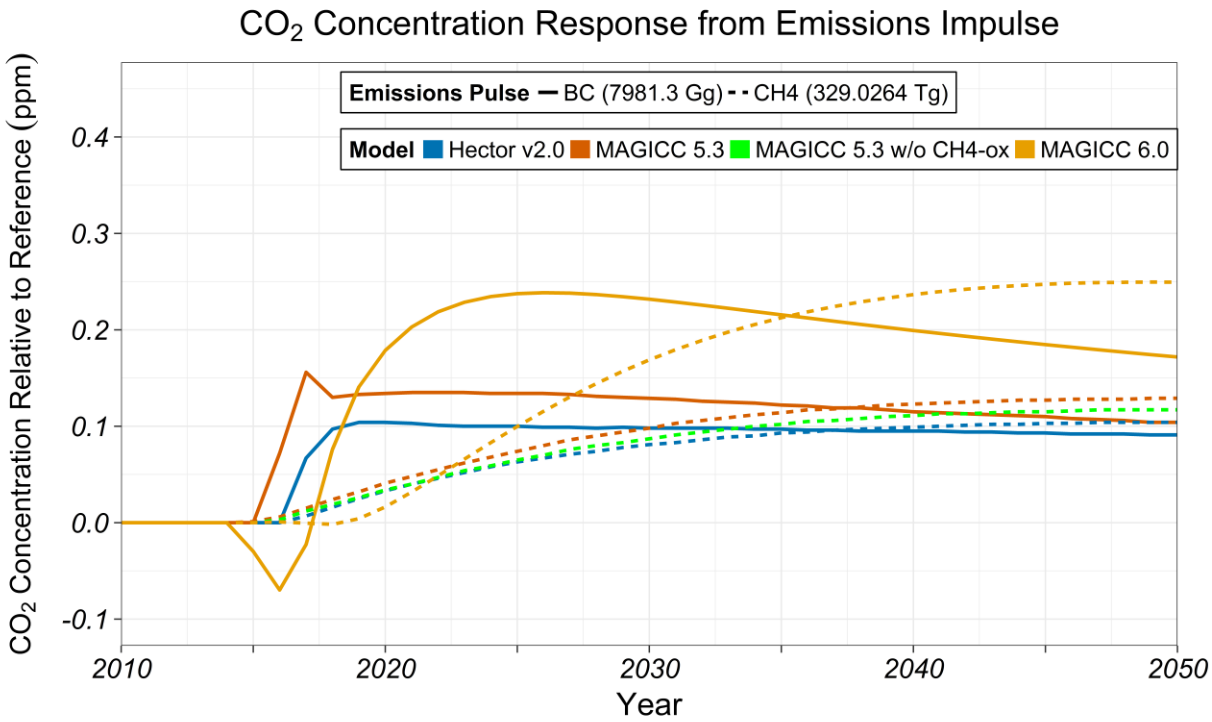

Figure S7 *CO_2 concentration response from emissions impulse in SCMs. MAGICC 5.3 is shown with and without CH_4 oxidation included (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue)..*

385

S8 Model Responses out to 2300

Figure S8 - Figure S12 show the CO_2 concentration response, total forcing response, and global mean temperature response from an emissions impulse, respectively, to the end of the model period equal to 2300.

S8.1 CO_2 Concentration Response to a CO_2 Emissions Pulse

Figure S8 shows the CO_2 concentration response from a CO_2 emissions pulse in the SCMs out to 2300. We see that the SCMs respond similarly to this perturbation, with the exception of the stylized SCM, FAIR, which has a weaker response.

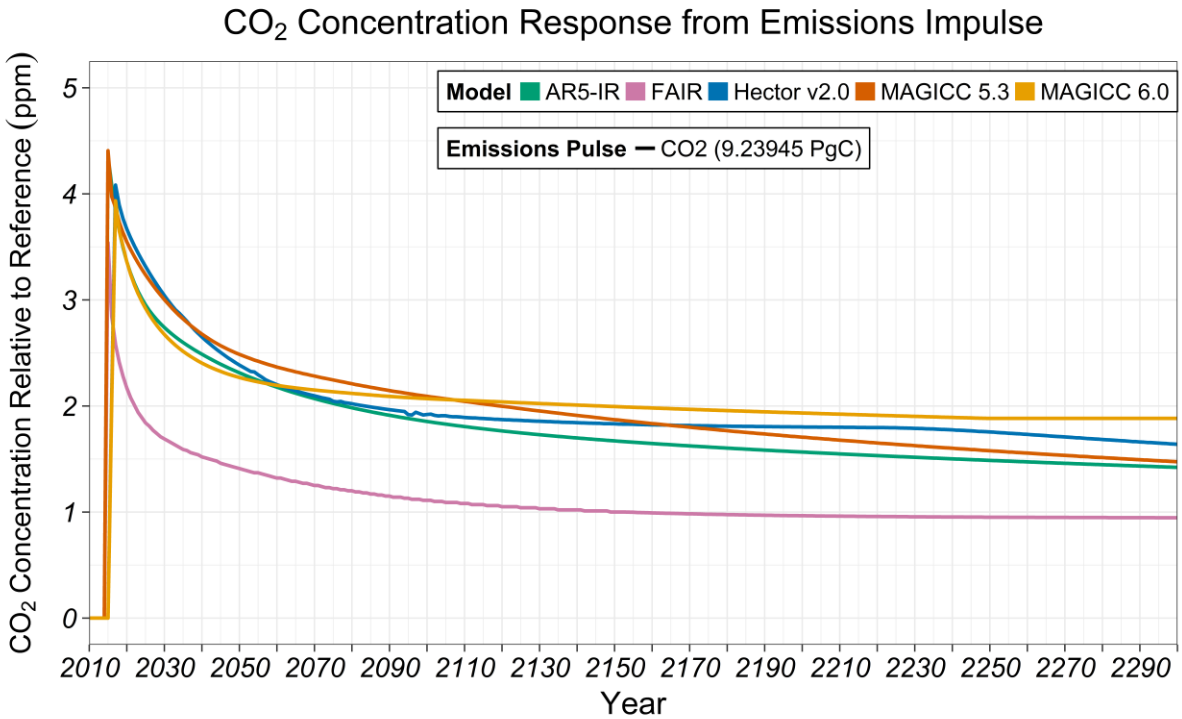

Figure S8 *Carbon dioxide concentration response from a CO_2 emissions pulse in SCMs (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR - pink).*

S8.2 CH_4 Concentration Response from CH_4 Emissions Pulse

Figure S9 shows the CH_4 concentration response from a CH_4 emissions pulse in the comprehensive SCMs out to 2300. The stylized SCMs do not report CH_4 concentrations. We see that the comprehensive SCMs behave similarly in their response to this perturbation, especially after 2050 when the response tends towards 0 ppb.

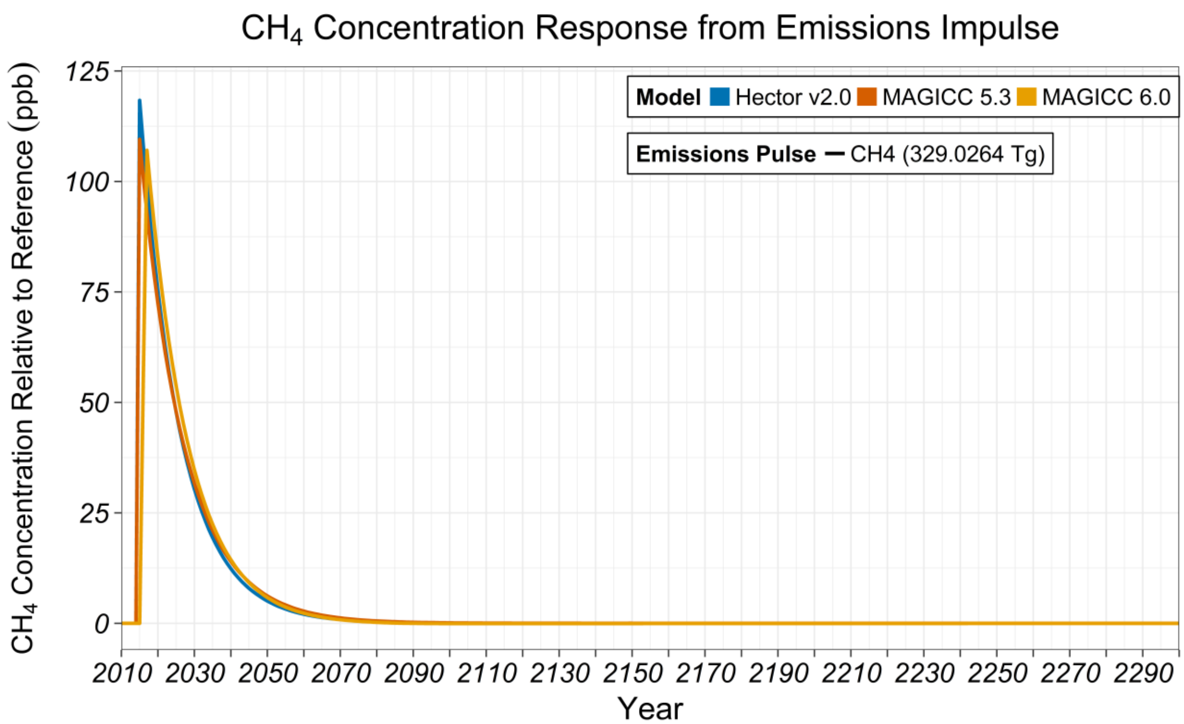

Figure S9 *Methane concentration response from a CH_4 emissions pulse in SCMs out to 2300 (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue).*

S8.3 CO_2 Concentration Response to a BC or CH_4 Emissions Pulse

Figure S10 shows the CO_2 concentration response from a CH_4 and BC emissions perturbations in the SCMs out to 2300. We see that the SCMs behave differently across the entire time series. Hector v2.0 changes state after 2225, a feature being investigated by the modeling team who originally calibrated the model out to 2300.

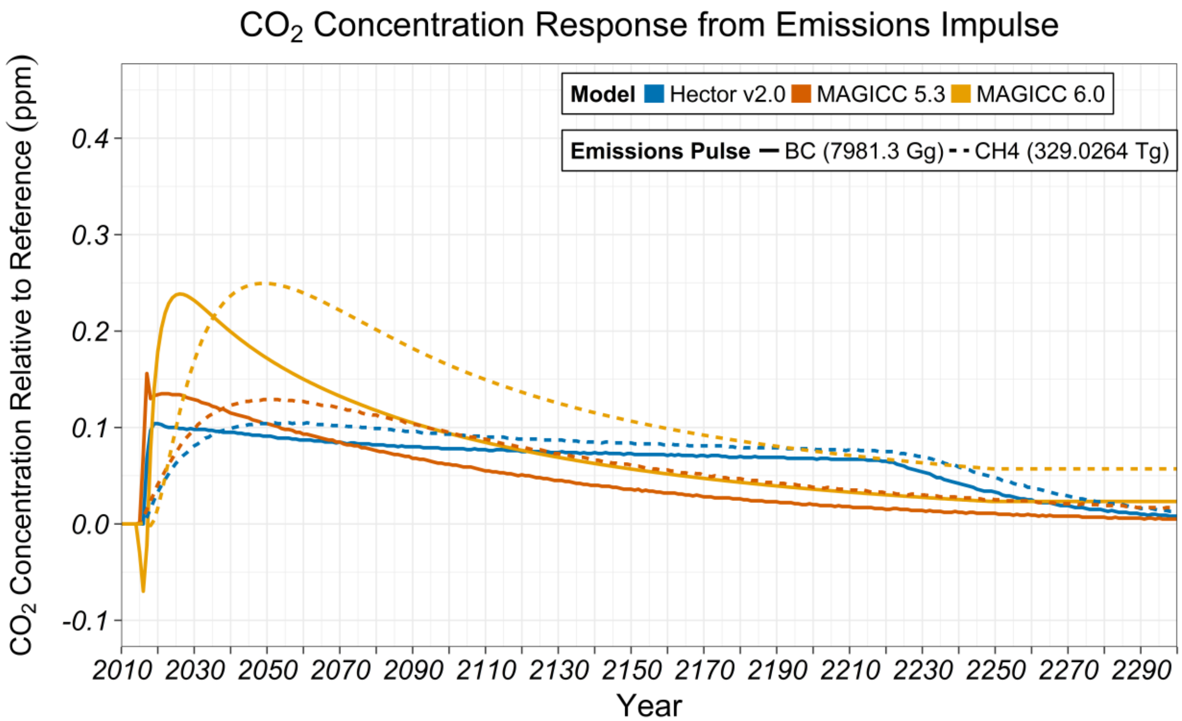

Figure S10 *CO_2 concentration response from emissions perturbations in SCMs out to 2300 (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue).*

S8.4 Total Forcing Response to a CO_2 or CH_4 Emissions Pulse

We report the total forcing response from the models, rather than the individual species' forcing responses for comparability. This has little impact on the results because, in the case of the non-CO_2 species, the total forcing is dominated by the CO_2 response, which is removed by subtracting the reference case.

Figure S11 shows the total forcing response from a CH_4 and CO_2 emissions perturbations in the SCMs out to 2300. FAIR does not report total forcing.

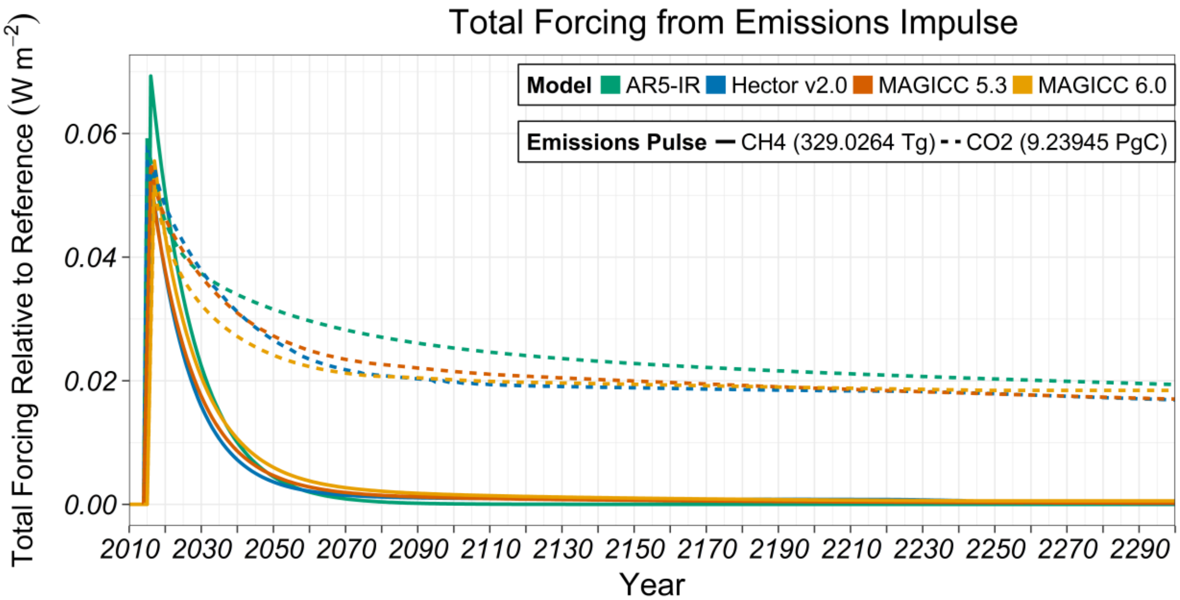

Figure S11 *Total forcing response from emissions perturbations in SCMs out to 2300 (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green).*

S8.5 Global Mean Temperature Response to a CH_4 or CO_2 Emissions Pulse

Figure S12 shows the temperature response from a CH_4 and CO_2 emissions perturbations in the SCMs out to 2300. We see that most of the SCM responses differ slightly immediately following the perturbation, but converge over time. AR5-IR has a stronger response than the other SCMs immediately following the perturbation. More details are included in the main paper.

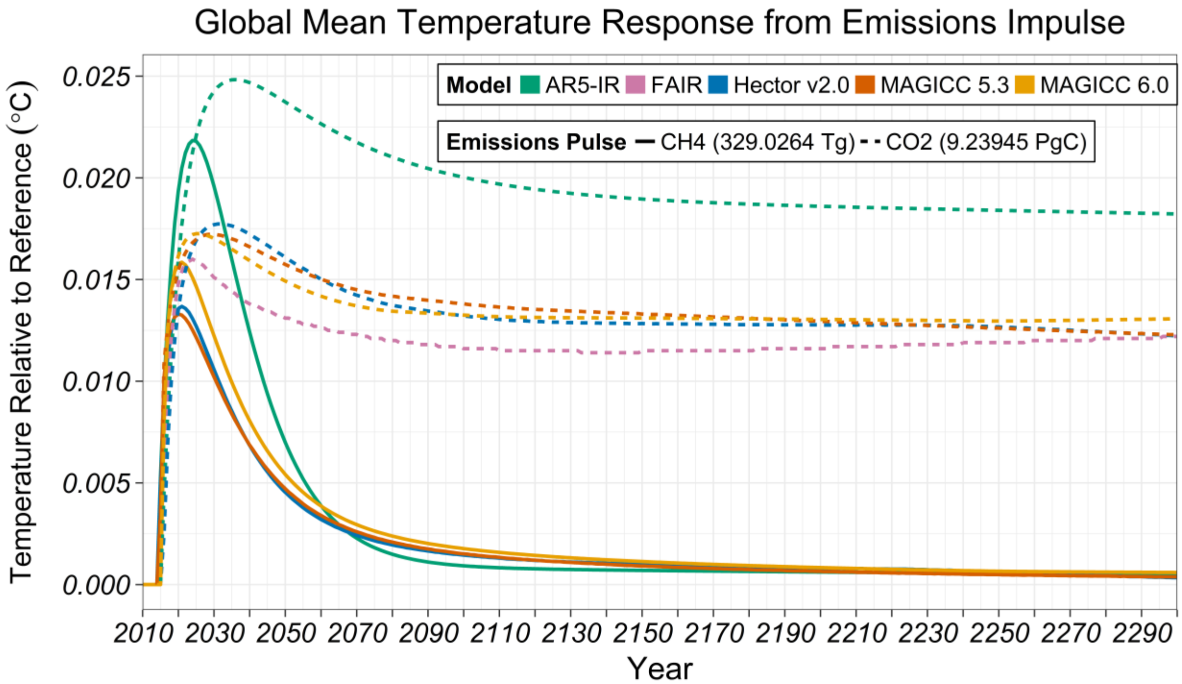

Figure S12 *Global mean temperature response from emissions perturbations in SCMs out to 2300 (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR - pink).*

S9 Time Integrated Responses

Figure S13 – Figure S18 shows the integrated forcing and temperature response for the full suite of experiments to the end of the model period. The data tables in this section provide numerical data (rounded to three significant figures) supporting the integrated forcing or temperature response figures. The data tables also include percent differences found using the following formula:

$$Percent\ Difference_{i,t} = \left(\frac{Model\ response_{i,t} - Average\ Comprehensive\ Model\ Response_t}{Average\ Comprehensive\ Model\ Response_t}\right) \times 100 \qquad (9)$$

where *t* is the time horizon and *i* is the individual model. A positive percent difference indicates that the model response is stronger than the average comprehensive model response, while a negative value indicates the model response was weaker than the average comprehensive model response.

S9.1 Time Integrated Responses from a CO_2 Concentration Impulse

Figure S13 shows the time integrated total forcing response from a CO_2 concentration impulse.

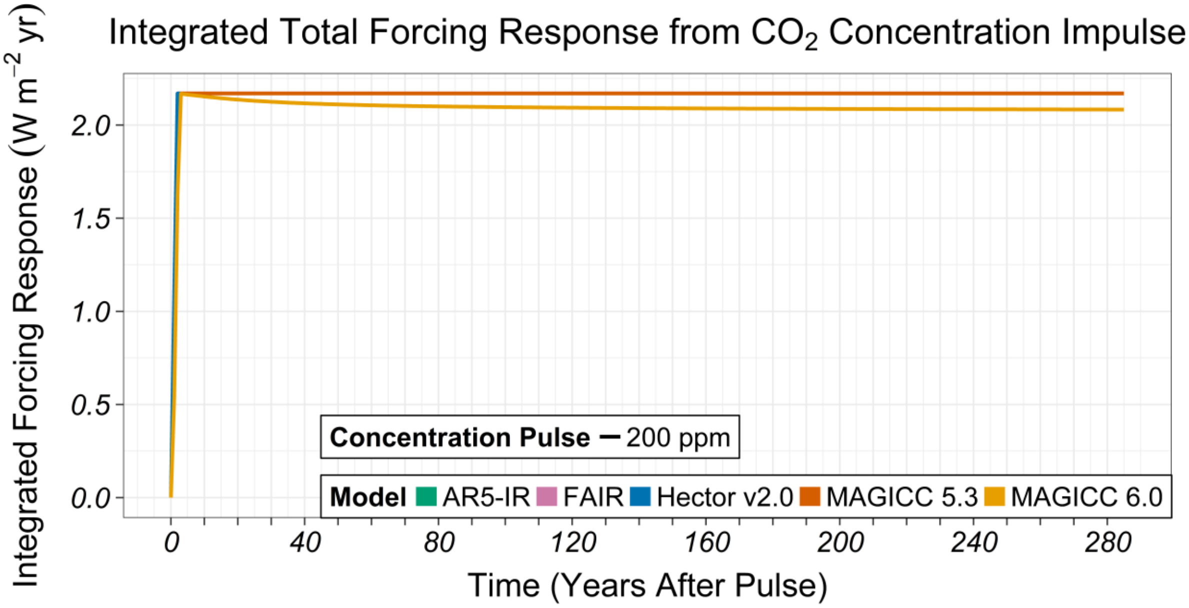

Figure S13 *Time integrated forcing response from a CO_2 concentration impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR - pink).*

Figure S14 shows the time integrated global mean temperature response from a CO_2 concentration impulse to the end of the model period. We see that the comprehensive SCMs respond similarly, while AR5-IR has a stronger response and FAIR, a slightly weaker response. The associated values time integrated temperature responses are in Table S3.

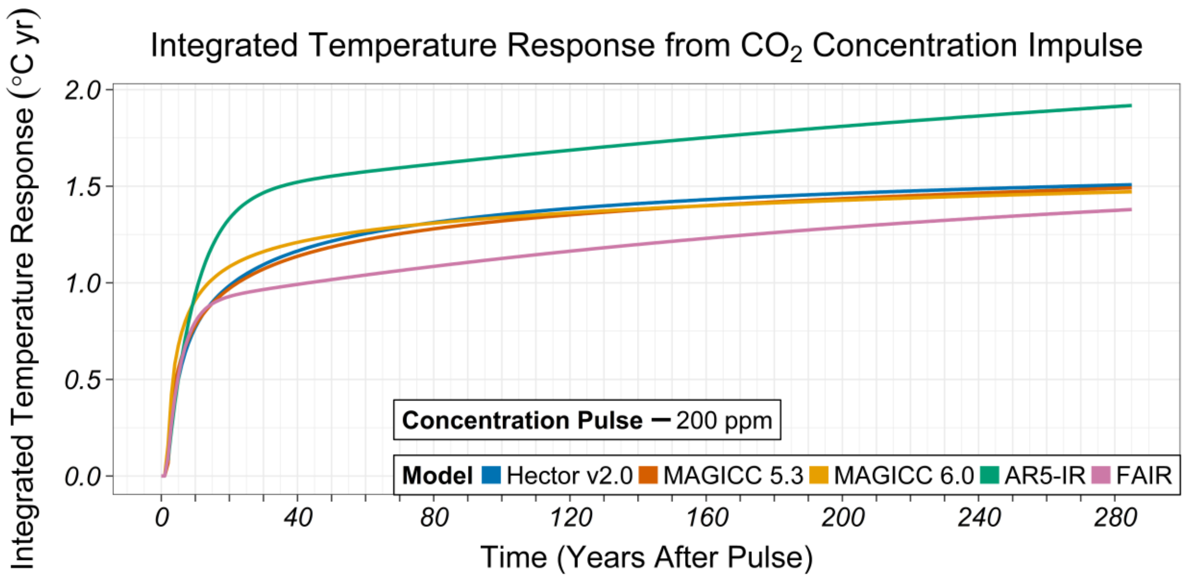

Figure S14 *Time integrated temperature response from a CO_2 concentration impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR - pink).*

Table S3 Integrated Temperature Responses from a CO_2 Concentration Impulse in the SCMs

Time After Pulse	Integrated Temperature Response (°Cyr)						Percent Difference from Comprehensive SCMs Average (%)				
	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	FAIR	AR5-IR	Average of Comprehensive SCMs	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	FAIR	AR5-IR
10	0.85	0.98	0.85	0.87	0.24	0.90	-4.79	9.59	-4.79	-3.57	-73.5
20	1.00	1.11	1.02	0.94	1.10	1.04	-4.25	6.29	-2.04	-9.80	5.33
50	1.20	1.25	1.22	1.02	1.39	1.22	-2.07	2.26	-0.19	-16.6	13.7
100	1.32	1.34	1.35	1.13	1.65	1.34	-1.25	0.25	1.00	-15.5	23.4
150	1.39	1.39	1.42	1.13	1.74	1.40	-0.71	-0.71	1.43	-19.3	24.3
285	1.46	1.47	1.51	1.38	1.92	1.48	-1.31	-0.63	1.94	-6.71	29.8

S9.2 Time Integrated Responses from a CO_2 Emissions Impulse

Figure S15 and Figure S16 shows the integrated forcing (Table S4) and temperature response (Table S5) for the CO_2 emissions impulse experiment to the end of the model period, respectively. The numerical data is shows in the Table S4 and Table S5.

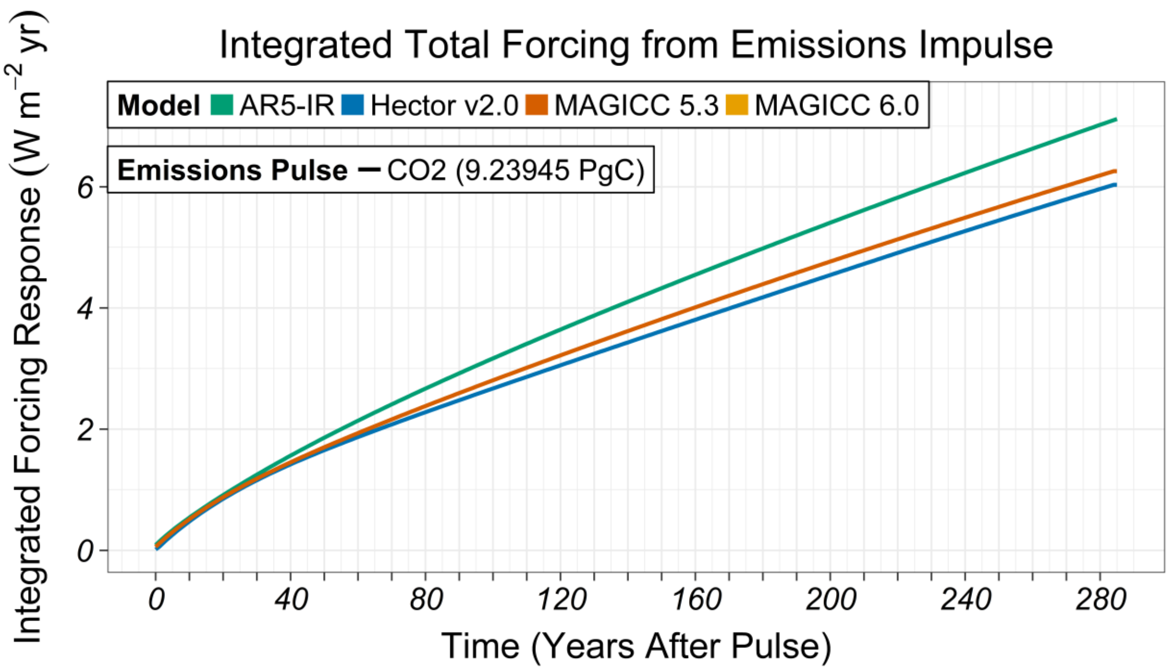

Figure S15 *Time integrated total forcing response from a CO_2 emissions impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green).*

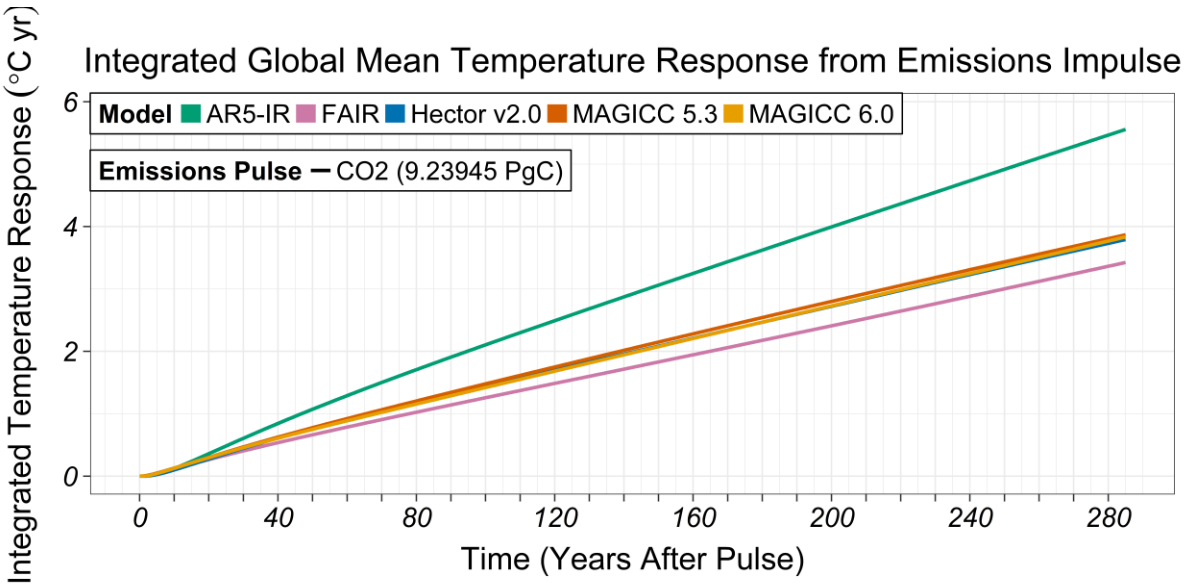

Figure S16 *Time integrated temperature response from a CO_2 emissions impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR - pink).*

Table S4 Integrated Forcing Responses from a CO_2 Emissions Impulse in the SCMs

Time After Pulse	Integrated Forcing Response ($Wm^{-2}yr$)					Percent Difference from Comprehensive SCMs Average (%)			
	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	AR5-IR	Average of Comprehensive SCMs	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	AR5-IR
10	0.51	0.43	0.48	0.54	0.47	8.67	-9.51	0.85	14.38
20	0.88	0.75	0.85	0.91	0.82	6.39	-9.38	2.99	10.63
50	1.70	1.48	1.65	1.86	1.61	5.63	-8.28	2.65	15.38
100	2.81	2.50	2.67	3.17	2.66	5.52	-5.96	0.44	19.13
150	3.82	3.47	3.62	4.32	3.63	4.97	-4.52	-0.45	18.87
285	6.26	5.97	6.03	7.12	6.09	2.79	-1.89	-0.90	16.98

Table S5 Integrated Temperature Responses from a CO_2 Emissions Impulse in the SCMs

Time After Pulse	Integrated Temperature Response (°Cyr)						Percent Difference from Comprehensive SCMs Average (%)				
	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	FAIR	AR5-IR	Average of Comprehensive SCMs	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	FAIR	AR5-IR
10	0.16	0.16	0.13	0.14	0.18	0.15	6.81	5.49	-12.31	-5.71	20.66
20	0.33	0.33	0.31	0.29	0.42	0.32	2.99	1.44	-4.43	-9.38	29.90
50	0.81	0.78	0.79	0.69	1.00	0.79	1.86	-1.69	-0.17	-13.07	26.15
100	1.50	1.45	1.46	1.27	1.93	1.47	2.16	-1.59	-0.57	-13.51	31.44
150	2.17	2.10	2.10	1.85	2.86	2.12	2.20	-1.10	-1.10	-12.87	34.69
285	3.87	3.85	3.80	3.44	5.87	3.84	0.83	0.17	-1.00	-10.38	52.93

S9.3 Time Integrated Responses from a CH_4 Emissions Impulse

Figure S17 and Figure S18 shows the integrated forcing (Table S6) and temperature response (Table S7) for the CH_4 emissions impulse experiment to the end of the model period. The numerical data in Table S6 and Table S7.

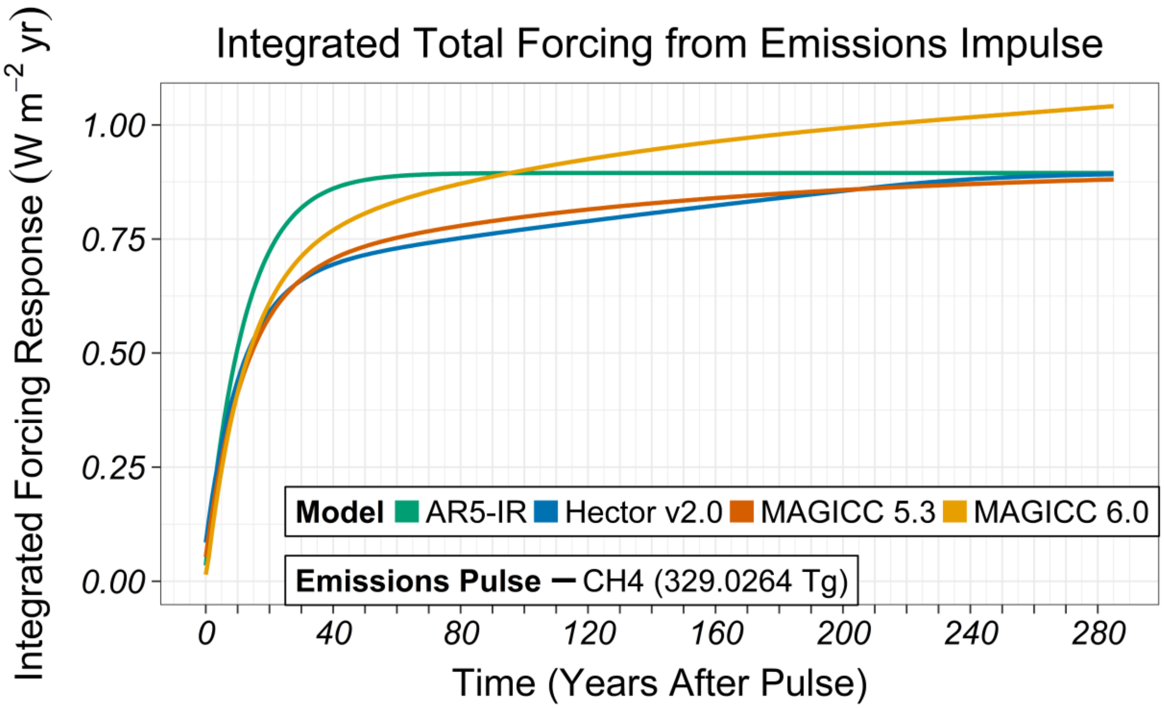

Figure S17 *Time integrated total forcing response from a CH_4 emissions impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green).*

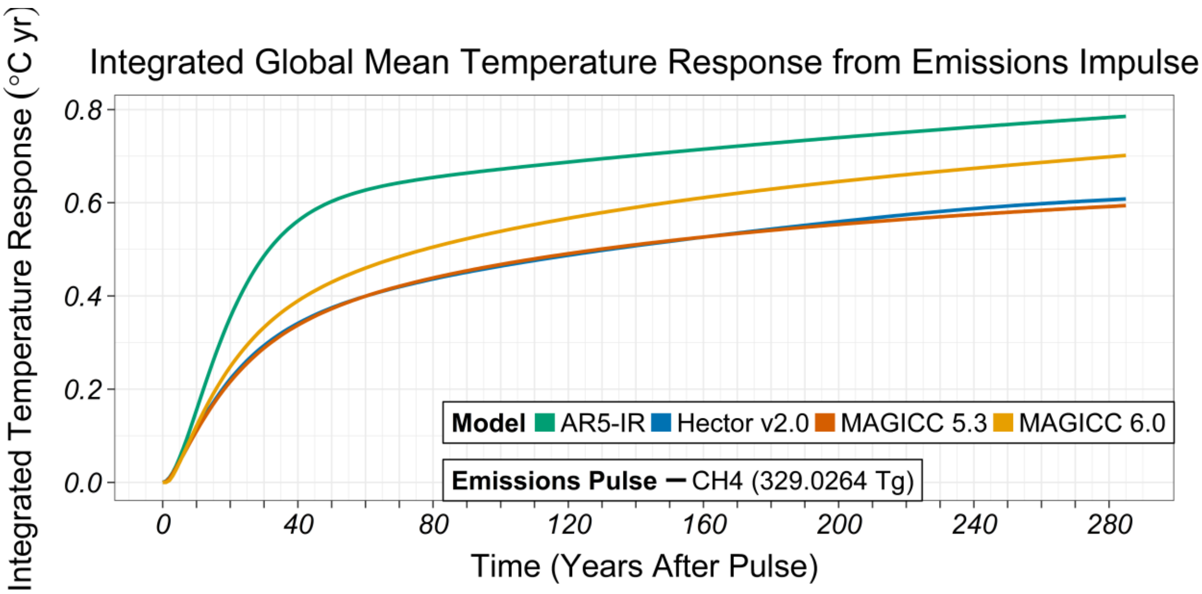

Figure S18 *Time integrated temperature response from a CO_2 emissions impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR - pink).*

Table S6 Integrated Forcing Responses from a CH_4 Emissions Impulse in the SCMs

Time After Pulse	Integrated Forcing Response ($Wm^{-2}yr$)					Percent Difference from Comprehensive SCMs Average (%)			
	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	AR5-IR	Average of Comprehensive SCMs	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	AR5-IR
10	0.41	0.41	0.44	0.51	0.42	-2.14	-2.14	4.28	21.1
20	0.58	0.61	0.59	0.72	0.59	-2.31	2.76	-0.45	21.9
50	0.73	0.81	0.72	0.88	0.75	-2.44	7.28	-4.84	16.9
100	0.80	0.90	0.77	0.89	0.82	-3.04	9.36	-6.32	8.63
150	0.83	0.95	0.82	0.89	0.87	-3.88	9.95	-6.07	3.03
285	0.88	1.04	0.89	0.89	0.94	-6.01	10.9	-4.94	-4.62

Table S7 Integrated Temperature Responses from a CH_4 Emissions Impulse in the SCMs

Time After Pulse	Integrated Temperature Response (°Cyr)					Percent Difference from Comprehensive SCMs Average (%)			
	MAGICC 5.3 BC-OC	**MAGICC 6.0**	**Hector v2.0**	**AR5-IR**	**Average of Comprehensive SCMs**	**MAGICC 5.3 BC-OC**	**MAGICC 6.0**	**Hector v2.0**	**AR5-IR**
10	0.13	0.15	0.14	0.16	0.14	-4.26	7.10	-2.83	17.2
20	0.23	0.27	0.24	0.36	0.25	-5.56	8.68	-3.12	47.4
50	0.38	0.44	0.38	0.45	0.40	-5.03	9.55	-4.52	12.3
100	0.47	0.54	0.47	0.58	0.49	-4.54	9.88	-5.35	17.4
150	0.52	0.60	0.52	0.70	0.55	-4.99	10.2	-5.17	28.1
285	0.60	0.70	0.61	0.85	0.64	-6.20	10.5	-4.31	33.5

S9.4 Time Integrated Responses from a BC Emissions Impulse

Figure S19 and Figure S20 shows the integrated forcing and temperature response for the BC emissions impulse experiment to the end of the model period, respectively. We used FAIR v1.0, which only represented the response from CO_2 emissions. An updated version, FAIR v1.3, was recently released and includes non-CO_2 forcing. SI. Table 8 shows the integrated temperature response data.

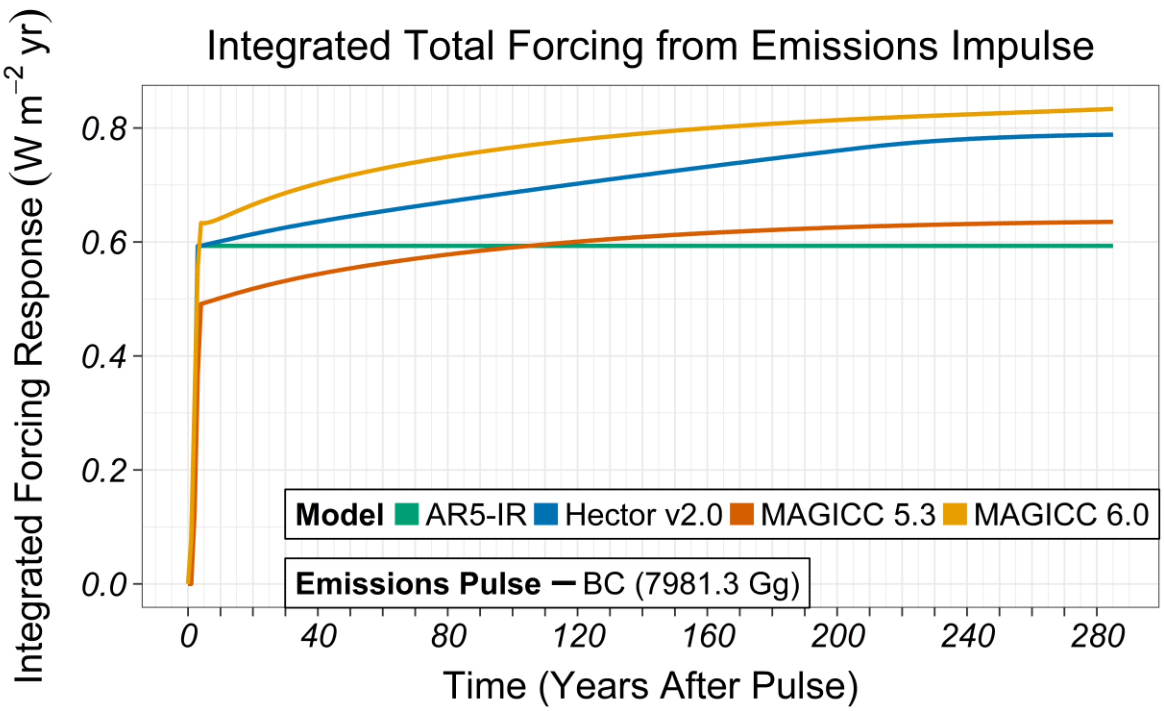

Figure S19 *Time integrated total forcing response from a BC emissions impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green).*

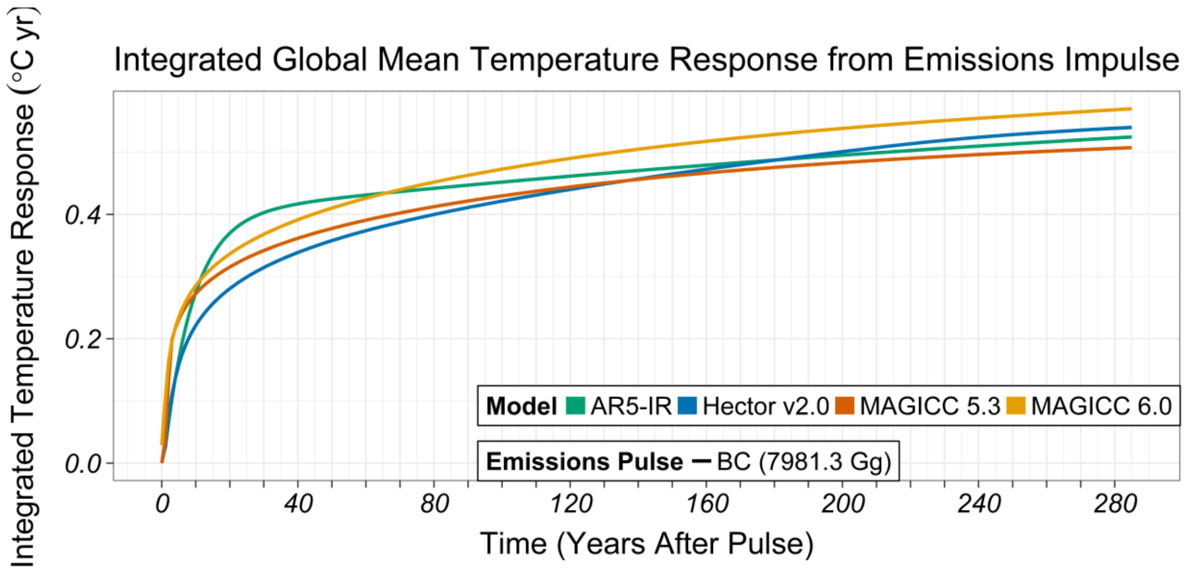

Figure S20 *Time integrated temperature response from a BC emissions impulse for the SCMs to the end of the model period (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR - pink).*

We see that Hector v2.0, which does not differentiate BC forcing over land and ocean and has a 9% weaker response 20 years after the pulse. MAGICC 6.0 diverges from the MAGICC 5.3 temperature response 20 years after the pulse. AR5-IR represents the temperature response from a BC perturbation as a simple exponential decay analogous to the greenhouse gas IRF, leading to a much stronger integrated temperature response (20%) 20 years after the pulse.

Table S8 Integrated Temperature Responses from a BC Emissions Impulse in the SCMs

Time After Pulse	Integrated Temperature Response (°Cyr)					Percent Difference from Comprehensive SCMs Average (%)			
	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	AR5-IR	Average of Comprehensive SCMs	MAGICC 5.3 BC-OC	MAGICC 6.0	Hector v2.0	AR5-IR
10	0.28	0.30	0.24	0.30	0.27	3.91	9.22	-13.1	11.0
20	0.32	0.34	0.29	0.38	0.32	1.13	8.12	-9.25	19.3
50	0.38	0.41	0.36	0.43	0.38	-1.22	7.43	-6.21	10.7
100	0.43	0.47	0.42	0.45	0.44	-2.68	7.12	-4.44	2.22
150	0.46	0.51	0.47	0.48	0.48	-3.80	6.76	-2.96	-0.92
285	0.51	0.57	0.54	0.53	0.54	-5.90	5.73	0.17	-2.56

S10 Temporal Response of SCMs Compared to CMIP5

Here we compare the 20-year moving average at time t=30, t= 50, t=70, t=100, and t=130 in the CMIP5 models and SCMs to show the temporal response of temperature. Hector v2.0 and MAGICC 5.3 have a faster response than the other SCMs and the majority of the complex models to an abrupt 4xCO_2 concentration step.

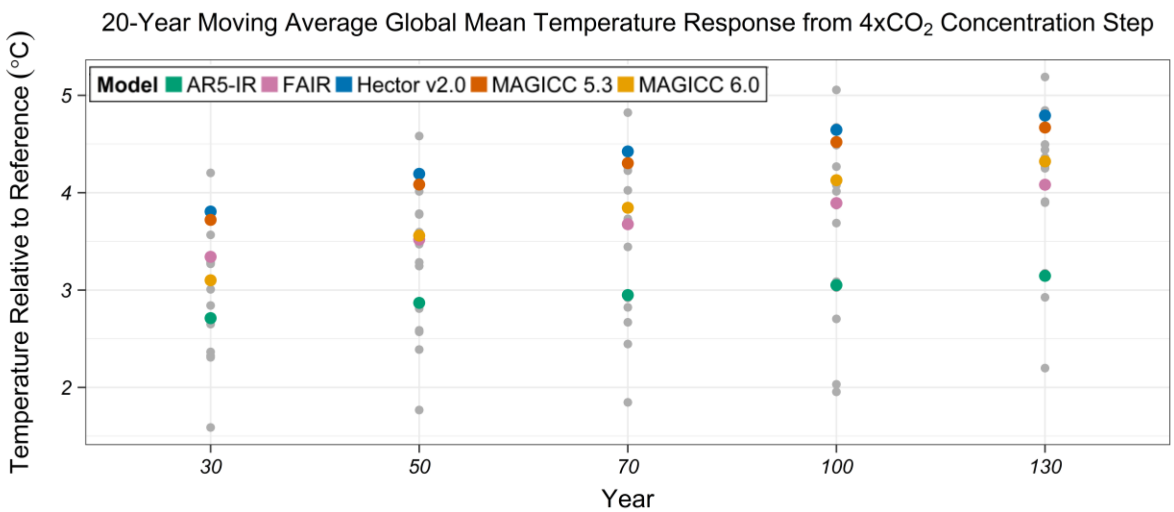

Figure S21 *20-Year moving average centered at year shown of the global mean temperature response from 4xCO_2 concentration step in CMIP5 models (grey) and* SCMs *(MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, FAIR – pink, AR5-IR –green).*

Table S9 shows the ECS values and the realized warming fraction (RWF) for the CMIP5 data and SCMs used to produce Figure 5. The RWF reveals that the SCMs used in this study generally warm faster than the more complex models in CMIP5.

Table S9 *CMIP5 and SCM model information with ECS and RWF*

Centre(s)	Model name	ESC (°C)	RWF (%) $LN(2)\ x\ \frac{Average\ of\ the\ last\ 40\ years}{ECS}$
Beijing Climate Center (BCC) **China**	BCC-CSM1.1	2.8	83.4
Canadian Centre for Climate Modelling and Analysis (CCCma) **Canada**	CanESM2	3.7	77.0
Centre National de Recherches Météorologiques, Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CNRM-CERFACS) **France**	CNRM-CM5-2	3.3	69.7
Institut Pierre Simon Laplace (IPSL) **France**	IPSL-CM5A-LR	4.1	69.8
	IPSL-CM5A-MR	NA	NA
	IPSL-CM5B-LR	3.6	72.9
Institute of Atmospheric Physics, Chinese Academy of Sciences (LASG-CESS) **China**	FGOALS-g2	NA	NA
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and	MIROC-ESM	4.7	65.1
	MIROC5	2.7	68.6

Japan Agency for Marine-Earth Science and Technology (MIROC) **Japan**			
Max Planck Institute for Meteorology (MPI-M) **Germany**	MPI-ESM-MR	NA	NA
	MPI-ESM-P	3.5	71.7
NASA/GISS (Goddard Institute for Space Studies; NASA-GISS) **USA**	GISS-E2-H	2.3	70.2
	GISS-E2-R	2.1	62.3
Geophysical Fluid Dynamics Laboratory (NCAR; NSF-DOE-NCAR) **USA**	GFDL-CM3	4.0	70.5
	GFDL-ESM2G	2.4	82.9
Raper et al. 1996; Wigley and Raper 2002; Smith and Bond 2014	MAGICC 5.3 BC-OC	3.0*	82.0
Meinshausen et al. 2011	MAGICC 6.0	3.0*	83.8
Hartin et al. 2015 Hartin et al. 2016	Hector v2.0	3.0*	90.3
Millar et al. 2017	FAIR	2.75	86.2
Myhre et al. 2013	AR5-IR	2.7	66.8

*Unless otherwise noted.

Note: NA denotes models that have not reported an ESC value from Table 9.5 in IPCC AR5(Flato et al. 2013).

S10.1 Changing Equilibrium Climate Sensitivity Values in SCMs with Comparison to CMIP5

Here we reproduce Figure 5 from the main paper using different ECS values in Hector v2.0, MAGCC 5.3, and MAGIC 6.0. We run each of these SCMs with a climate sensitivity values of 2.1°C, the same as GISS-E2-R, and 4.7°C, the same as MIROC-ESM. These two model values were selected because they represent the largest range of climate sensitivity values in the model data used here. Figure S22 shows the global mean temperature response from 4xCO_2 concentration step in CMIP5 models and SCMs. The SCMs were run with two different ECS values. Figure S22a shows the SCM response with an ECS value of 2.1°C and Figure S22b shows the SCM responses with an ECS value of 4.7°C.

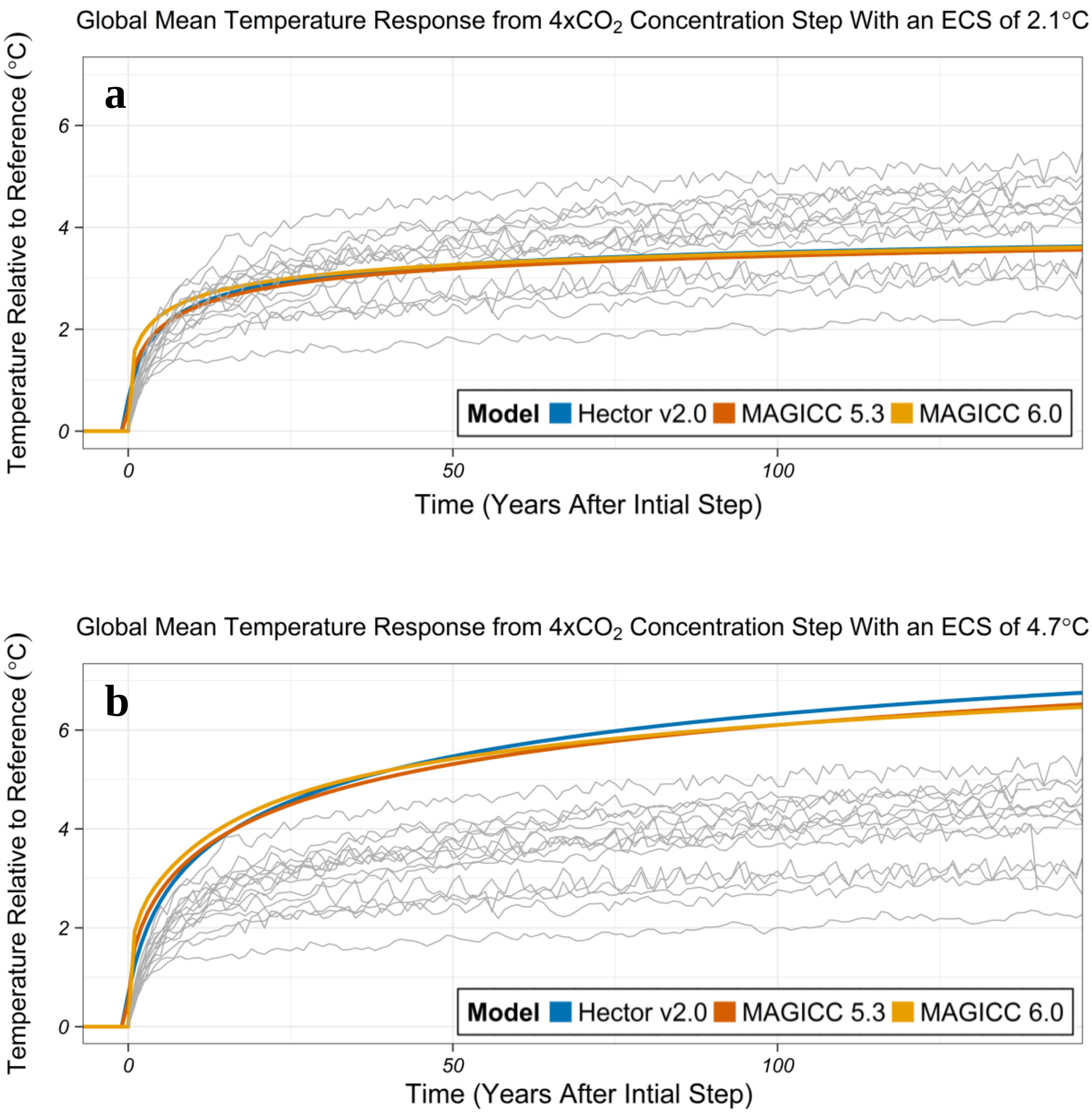

Figure S22 *Global mean temperature response from 4xCO2 concentration step in CMIP5 models and SCMs, as in Fig. 5, with the SCMs run with two different ECS values. Fig. 22a shows the SCM response with an ECS value of 2.1°C, and Fig. 22b shows the SCM responses with an ECS value of 4.7°C (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue)*

S11 Comparison to Previous Impulse Reponses Work by Joos *et al.*(Joos et al. 2013)

We conducted the same perturbation experiment done by Joos et al. (Joos et al. 2013) with our three comprehensive SCMs and two stylized SCMs, however, we do not conduct this against a constant CO_2 concentration background. Instead, we use the RCP 4.5 scenario and add a 100GtC CO_2 pulse in 2015. It is useful to note that MAGICC 6.0 was used both in this study and by Joos *et al.* The versions used in each study differ slightly. Joos *et al.* used MAGICC model version 6.3 run in 171 different parameter settings that emulate 19 AOGCMs and 9 coupled climate-carbon cycle models. MAGICC 6.0 used in this study was set at the default setting using the AOGCM multi-model mean.

Table S10 shows the time-integrated airborne fraction at chosen time horizons from the 100 GtC pulse of CO_2 emissions. The Table S10 results are graphically represented in SI. Fig. 23. These results are largely discussed in the main paper.

Table S10 Time-integrated Airborne Fraction from a 100 GtC CO_2 Emissions Impulse in SCMs Compared to Results from Table 4 in Joos *et al.* (Joos et al. 2013)

Time Horizon	**20 yr**	**50 yr**	**100 yr**
NCAR CSM1.4	13.8	27.8	46.6
HadGEM2-ES	14.7	30.9	53.3
MPI-ESM	14.5	29.2	48.8
Bern3D-LPJ (reference)	15.4	34.3	61.9
Bern3D-LPJ ensemble	15.1 (14.0-16.0)	32.7 (28.9-36.0)	57.6 (48.9-65.6)
Bern2.5D-LPJ	13.9	29.7	51.1
CLIMBER2-LPJ	13.0	26.8	49.2
DCESS	14.6	31.8	56.3
GENIE ensemble	13.6 (10.9-17.6)	28.9 (21.7-41.4)	50.5 (38.3-77.9)
LOVECLIM	13.5	27.9	45.3
MESMO	15.1	33.6	61.1
UVic2.9	13.7	29.5	53.0
ACC2	13.7	27.9	46.5
Bern-SAR	14.0	29.0	48.9
TOTEM2	16.9	38.3	66.6
MAGICC 6.0 ensemble	14.0 (12.0-16.1)	29.6 (23.6-35.7)	51.8 (40.0-64.2)
Multi-model mean	14.3 ± 1.8	30.2 ± 5.7	52.4 ± 11.3
Hector v2.0	16.2	34.0	58.3
MAGICC 5.3	16.0	33.4	58.3
MAGICC 6.0	15.3	32.2	57.9
AR5-IR	15.0	31.0	53.1
FAIR	14.6	32.6	61.6

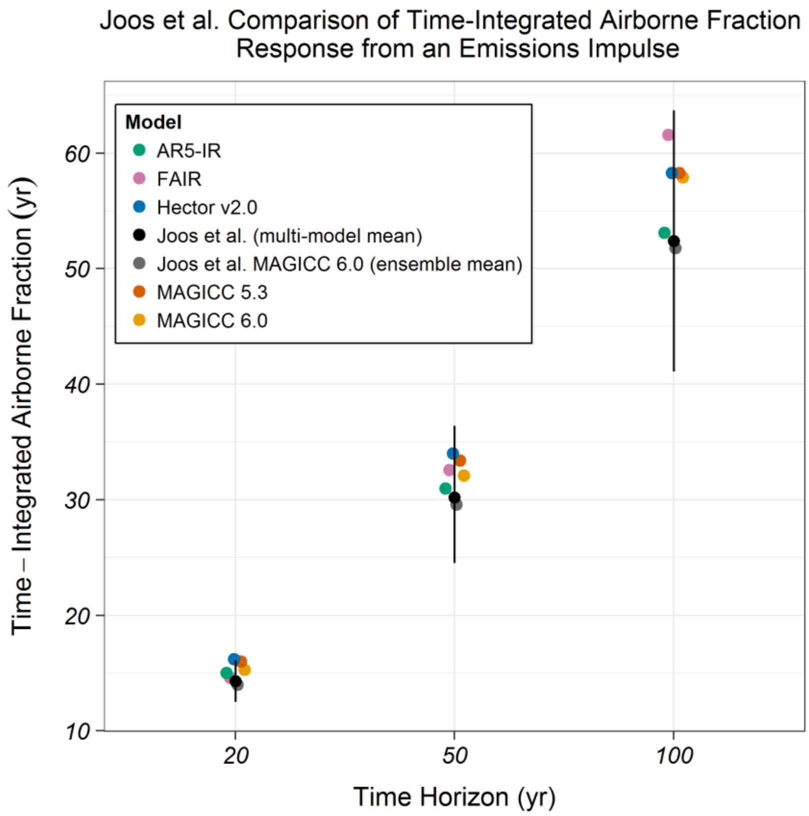

Figure S23 *Time-integrated airborne fraction from a 100GtC CO_2 emissions impulse in SCMs compared to Joos et al. This is not a direct comparison because we did not perform this experiment with a constant CO_2 concentration background, as done by Joos et al. The colored points represent the time-integrated airborne fraction in the SCMs used in this study, following Joos et al., and the Joos et al. MAGICC 6.0 ensemble mean. The black point is the Joos et al. multi-model mean and the vertical black line represents the range of the Joos et al. model results. (Joos et al. MAGICC 6.0 ensemble mean –grey, MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR –pink).*

We also indirectly compare the temperature response of the comprehensive SCMs and more complex models in Joos *et al*. MAGICC 6.0 was used both here and by Joos *et al*., and we find similar responses with ≤ 1 °C yr difference from Joos *et al.* at each reported period. Though the other two comprehensive SCMs were not used by Joos *et al.*, their similar responses to our MAGICC 6.0 allow us to make a larger conclusion, as done in the main paper. Using this logic, we are able to validate our SCM responses from a finite pulse, without conducting this experiment in ESMs or EMICs, directly. We find that the comprehensive SCM responses are generally less varied, close to the Joos *et al.* ensemble mean 20 years after the pulse, and below most Joos *et al.* model responses 50 and 100 years after the pulse (see Figure S24).

Table S11 Time-integrated temperature response from a 100 GtC CO_2 Emissions Impulse in SCMs Compared to Results from Table 7 in Joos *et al.*

Time Horizon	**20 yr**	**50 yr**	**100 yr**
NCAR CSM1.4	2.53	7.36	10.6
HadGEM2-ES	4.24	12.4	30.3
MPI-ESM	3.83	8.84	19.1
Bern3D-LPJ (reference)	4.11	12.1	24.5
Bern3D-LPJ ensemble	3.20 (2.1-4.6)	8.61 (5.1-13.5)	17.3 (9.5-29.3)
Bern2.5D-LPJ	3.15	8.40	17.1
CLIMBER2-LPJ	3.05	7.96	16.5
DCESS	3.38	9.96	20.6
GENIE ensemble	3.77	10.54	21.6
LOVECLIM	0.22	3.46	7.83
MESMO	4.41	12.5	26.0
UVic2.9	3.40	9.17	18.5
ACC2	3.99	10.55	20.0
Bern-SAR	n/a	n/a	n/a
TOTEM2	n/a	n/a	n/a
MAGICC 6.0 ensemble	3.64 (2.7-4.7)	8.96 (6.6-12.7)	17.2 (12-26)
Multi-model mean	3.29 ± 2.03	9.13 ± 4.45	18.7 ± 11.1
Hector v2.0	3.05	8.20	15.54
MAGICC 5.3	3.13	8.19	15.73
MAGICC 6.0	3.39	8.28	15.54

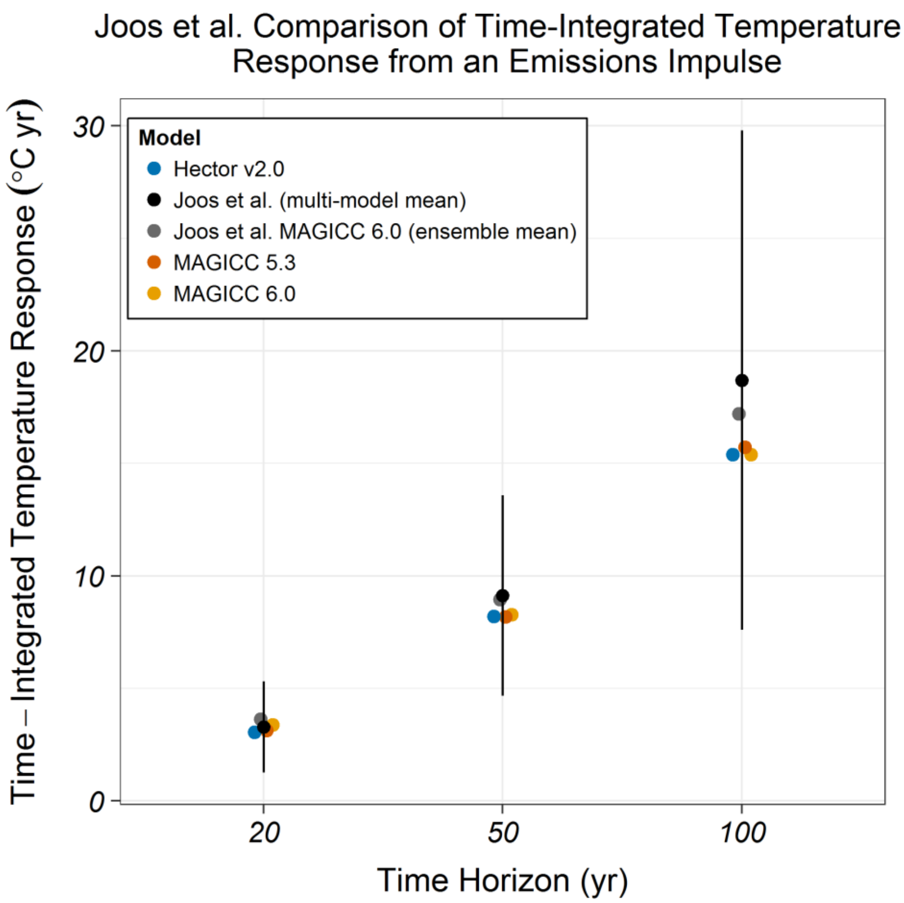

Figure S24 *Time-integrated temperature response from a 100GtC CO_2 emissions impulse in SCMs compared to Joos et al. This is not a direct comparison because we did not perform this experiment with a constant CO_2 concentration background, as done by Joos et al. The colored points represent the time-integrated temperature response in the SCMs used in this study, following Joos et al., and the Joos et al. MAGICC 6.0 ensemble mean. The black point is the Joos et al. multi-model mean and the vertical black line represents the range of the Joos et al. model results. (Joos et al. MAGICC 6.0 ensemble mean –grey, MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue).*

We compare the comprehensive SCM responses from the 100GtC CO_2 pulse to our earlier experiment using a ~10GtC CO_2 pulse. We find that the relative behavior of the comprehensive SCMs in the 100 GtC CO_2 impulse is similar to the response pattern from the smaller pulse experiment (see Figure 3a and Figure S25). The MAGICC 6.0 temperature response pattern is consistent with our prior experiments, where we see an initially stronger response (10 years following the perturbation) compared to the other comprehensive SCMs. Due to the initial oscillatory behavior in complex model responses (see Figure 2a in Joos *et al.*(Joos et al. 2013)), it is difficult to compare SCM responses to complex models on these short time scale.

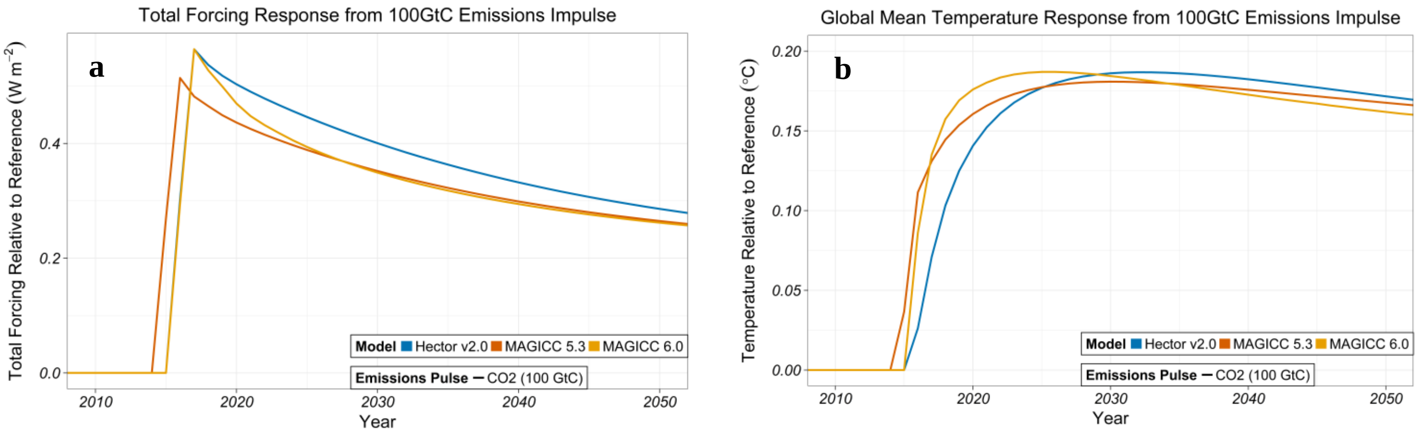

Figure S25 *Total forcing response (a) and global mean temperature response (b) from a 100GtC CO2 emissions impulse in the SCMs (MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue).*

S12 Investigating Temperature Response from BC Step Experiment

We investigate SCM responses to a black carbon (BC) emissions step by quadrupling (4x) the values in 2015. We choose two of the SCMs, Hector v2.0 and MAGICC 5.3, as examples and compare the temperature response to Figure 1 in Sand *et al* (Sand et al. 2016). Sand *et al.* finds that after applying a 25x BC emissions step to NorESM1-M, a complex climate model, the temperature response levels off after it reaches 1.2K after less than 10 years. Sand *et al.* applies a large BC step to increase the signal in the complex model, while we apply a smaller step in the SCMs. We find that the SCM responses to a BC emissions step continue to increase 10 years after the perturbation, suggesting that the SCMs fail to capture aerosol dynamics.

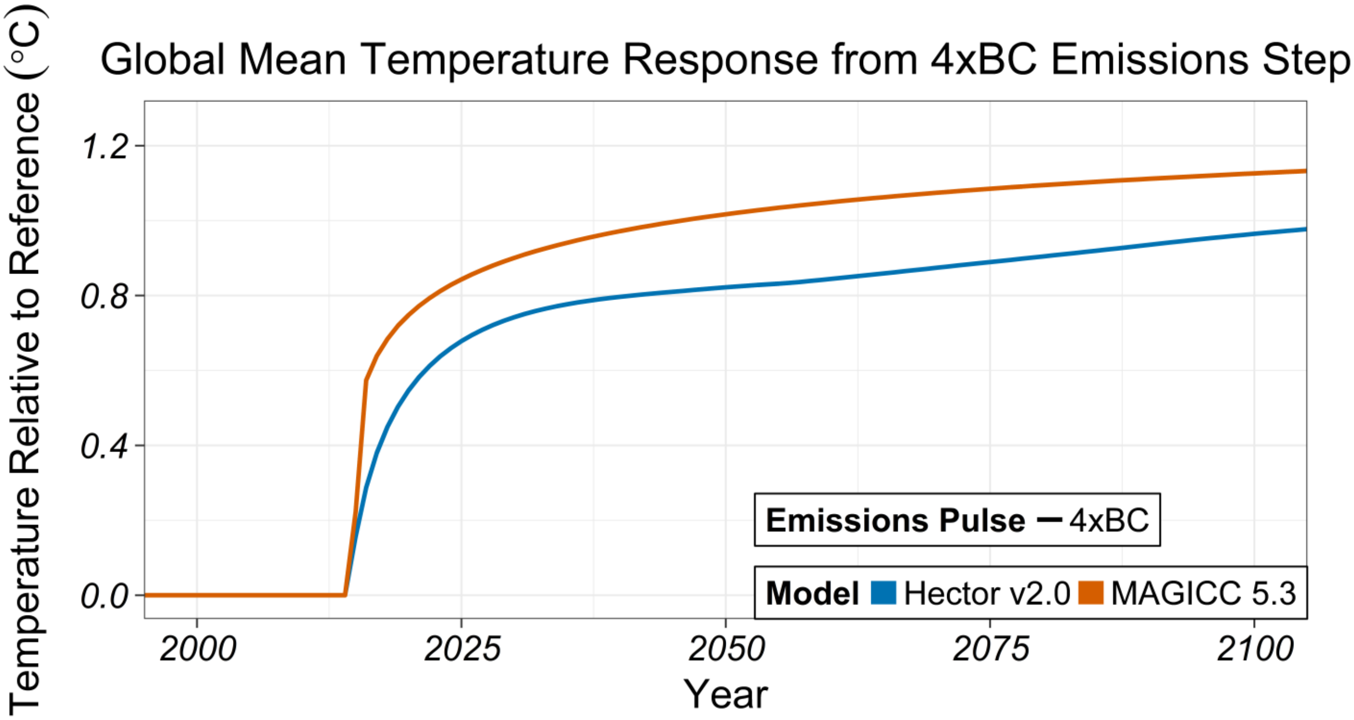

Figure S26 *Global mean temperature response from a 4xBC emissions step in the SCMs (MAGICC 5.3 BC-OC – red, Hector v2.0 – blue).*

S13 Summary of SCM Performance in Table 1

We provide a summary of SCM performance for each of our recommended unit tests in Table 1 in the main paper. Here, we describe the performance scale used in Table 1.

Using results from Joos *et al.* (Joos et al. 2013), we found that the MAGICC 6.0 temperature response to a 100GtC CO_2 emissions impulse was similar to more complex models. In the main text, therefore, we use MAGICC 6.0 as a reference to understand the other comprehensive SCM responses. We report that the comprehensive SCM carbon cycle representations, including from MAGICC 5.3 and Hector v2.0, generally capture more complex model responses. For unit tests where we cannot directly compare the responses to complex model, therefore, we use the comprehensive SCM average.

We developed the performance scale in Table 1 generally using the time-integrated temperature response percent difference from the comprehensive SCM average. We set the scale based on the range in percent differences found in our analysis: • : 0-10% difference, •• : 10-20% difference, and ••• : 20-30% difference from the comprehensive SCM average (see S9).

For example, we assign the comprehensive SCM responses to a CO_2 concentration impulse a three (•••) because the responses are within 10% of the comprehensive SCM average. The idealized SCMs, FAIR v1.0 and AR5-IR, have greater differences and are given a two (••) and a one (•), respectively.

Under the $4xCO_2$ concentration step experiment, we can compare the SCM response to more complex models from CMIP5. We assign MAGICC 6.0 a three (•••) because it appears to respond more reasonably under stronger forcing conditions than the other SCMs. We assign Hector v2.0, MAGICC 5.3, and FAIR a two (••) because these SCMs have initially quicker responses to an abrupt $4xCO_2$ concentration increase compared to the ESMs. We assign AR5-IR a one (•) because it has a slower response to an abrupt $4xCO_2$ concentration increase and is insensitive to changing background concentrations.

For CH_4 emissions impulses, we use the difference from the comprehensive SCM average to rate the responses. Unlike the 100GtC CO_2 and $4xCO_2$ step experiments, we cannot compare the SCM responses to more complex models, therefore, we are more lenient in our performance assignment against the comprehensive SCM average. CH_4 is a well-mixed GHG and, therefore, we expect that the climate system response to CH_4 concentration perturbations will be similar to that for CO_2. However, it would be useful to evaluate in more complex models if the simple representation of chemistry in the comprehensive SCMs adequately represents the time evolution of CH_4 concentrations in response to a change in emissions.

Finally, we assign ratings to the SCM responses to aerosols. We do not explicitly conduct aerosol experiments other than BC because the responses of the SCMs to other aerosols will be similar to their response to BC. We do not have a definitive reference for the time-dependent response to aerosol forcing perturbations. Instead, we rate the SCMs using the difference from the average of both MAGICC models, which both differentiate aerosol forcing between land and ocean, which results in a faster overall climate response to aerosols as compared to greenhouse gases (Shindell 2014). In the case of BC, we note that all SCM response ratings should be reduced from the values shown because they do not accurately represent the temporal response to a BC step found in an ESM (see S12). A more definitive evaluation of climate system responses to aerosol perturbations would be useful. This would require additional GCM simulations to step emission changes for various aerosol species and/or forcing mechanisms.

S14 Supplementary Data

Other supplementary materials for this manuscript include the following:

Dataset S1 (separate file)

Simple climate model responses from 4xBC emissions step.

Dataset S2 (separate file)

Simple climate model responses from $4xCO_2$ concentration step with 2.3 ocean diffusion and an ECS = 3 °C.

Dataset S3 (separate file)

Simple climate model responses from a 100PgC CO_2 emissions impulse experiment.

Dataset S4 (separate file)

Simple climate model responses from a CH_4 emissions impulse experiment.

Dataset S5 (separate file)

Simple climate model responses from a BC emissions impulse experiment.

Dataset S6 (separate file)

Simple climate model responses from CO_2 concentration impulse experiment.

Dataset S7 (separate file)

Simple climate model responses from CO_2 emissions impulse experiment.

Dataset S8 (separate file)

AR5-IR code to produce responses to BC emissions impulse.

Dataset S9 (separate file)

AR5-IR code to produce responses to CH_4 emissions impulse.

Dataset S10 (separate file)

AR5-IR code to produce responses to CO_2 emissions impulse.

Dataset S11 (separate file)

AR5-IR code to produce responses to 100PgC CO_2 emissions impulse for comparison to Joos *et al.*

Dataset S12 (separate file)

AR5-IR code to produce responses to CO_2 concentration step.

Dataset S13 (separate file)

FAIR CO_2 concentration impulse experiment input file.

Dataset S14 (separate file)

FAIR 4xCO_2 concentration step experiment input file.

Dataset S15 (separate file)

FAIR CO_2 emissions impulse experiment input file.

Dataset S16 (separate file)

FAIR 100Pg CO_2 emissions impulse experiment input file.

Dataset S17 (separate file)

FAIR CO_2 emissions impulse experiment reference input file.

Dataset S18 (separate file)

Hector v2.0 CO_2 concentration impulse experiment input file.

Dataset S19 (separate file)

Hector v2.0 CO_2 concentration impulse experiment reference input file.

Dataset S20 (separate file)

Hector v2.0 4xCO_2 concentration step experiment reference input file.

Dataset S21 (separate file)

Hector v2.0 4xCO_2 concentration step experiment input file.

Dataset S22 (separate file)

Hector v2.0 BC emissions impulse experiment input file.

Dataset S23 (separate file)

Hector v2.0 BC emissions step experiment input file.

Dataset S24 (separate file)

Hector v2.0 CH_4 emissions impulse experiment input file.

Dataset S25 (separate file)

Hector v2.0 CO_2 emissions impulse experiment input file.

Dataset S26 (separate file)

Hector v2.0 100Pg CO_2 emissions impulse experiment input file.

Dataset S27 (separate file)

Hector v2.0 emissions impulse experiment reference input file.

Dataset S28 (separate file)

Hector v2.0 emissions step experiment reference input file.

Dataset S29 (separate file)

MAGICC5.3 CO_2 concentration impulse experiment reference input file.

Dataset S30 (separate file)

MAGICC5.3 CO_2 concentration impulse experiment input file.

Dataset S31 (separate file)

MAGICC5.3 $4xCO_2$ concentration step experiment input file.

Dataset S32 (separate file)

MAGICC5.3 $4xCO_2$ concentration step experiment reference input file.

Dataset S33 (separate file)

MAGICC5.3 BC emissions impulse experiment input file.

Dataset S34 (separate file)

MAGICC5.3 BC emissions step experiment input file.

Dataset S35 (separate file)

MAGICC5.3 CH_4 emissions impulse experiment input file.

Dataset S36 (separate file)

MAGICC5.3 1% CO_2 emissions impulse experiment in 2010 input file.

Dataset S37 (separate file)

MAGICC5.3 1.01% CO_2 emissions impulse experiment in 2010 input file.

Dataset S38 (separate file)

MAGICC5.3 5% CO_2 emissions impulse experiment in 2010 input file.

Dataset S39 (separate file)

MAGICC5.3 10% CO_2 emissions impulse experiment in 2010 input file.

Dataset S40 (separate file)

MAGICC5.3 50% CO_2 emissions impulse experiment in 2010 input file.

Dataset S41 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2010 input file.

Dataset S42 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2015 input file.

Dataset S43 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2020 input file.

Dataset S44 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2030 input file.

Dataset S45 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2040 input file.

Dataset S46 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2050 input file.

Dataset S47 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2060 input file.

Dataset S48 (separate file)

MAGICC5.3 100% CO_2 emissions impulse experiment in 2070 input file.

Dataset S49 (separate file)

MAGICC5.3 100PgC CO_2 emissions impulse experiment in 2015 input file.

Dataset S50 (separate file)

MAGICC5.3 CO_2 emissions impulse experiment reference input file.

Dataset S51 (separate file)

MAGICC5.3 CO_2 emissions step experiment reference input file.

Dataset S52 (separate file)

MAGICC56.0 4xCO_2 concentration impulse experiment input file.

Dataset S53 (separate file)

MAGICC56.0 4xCO_2 concentration impulse experiment reference input file.

Dataset S54 (separate file)

MAGICC56.0 4xCO_2 concentration step experiment input file.

Dataset S55 (separate file)

MAGICC56.0 4xCO_2 concentration step experiment reference input file.

Dataset S56 (separate file)

MAGICC6.0 BC emissions impulse experiment input file.

Dataset S57 (separate file)

MAGICC6.0 CH_4 emissions impulse experiment input file.

Dataset S58 (separate file)

MAGICC6.0 100% CO_2 emissions impulse experiment input file.

Dataset S59 (separate file)

MAGICC6.0 100PgC CO_2 emissions impulse experiment input file.

Dataset S60 (separate file)

MAGICC6.0 emissions impulse experiment reference input file.

S15 References

Bakker, Alexander M. R., Tony E. Wong, Kelsey L. Ruckert, and Klaus Keller. 2017. "Sea-Level Projections Representing the Deeply Uncertain Contribution of the West Antarctic Ice Sheet." *Scientific Reports* 7(1): 3880. http://www.nature.com/articles/s41598-017-04134-5 (November 27, 2017).

Bond, T. C. et al. 2013. "Bounding the Role of Black Carbon in the Climate System: A Scientific Assessment." *Journal of Geophysical Research Atmospheres* 118(11): 5380–5552.

Cicerone, R.J.; Oremland, R.S. 1988. "GLOBAL BIOGEOCHEMICAL Methane (CH4) Is the Most Abundant Organic Gas in Measurement Data Has Been Reviewed by Ehhalt [1974] and Wofsy [1976] and Will Not Be Repeated Here . Since the Earth ' s Energy Balance (See The Quasi-Steady State (Defined In." 2(4): 299–327.

Clune, Thomas L, and R. B. Rood. 2011. "Software Testing and Verification in Climate Model Development." *IEEE Software* 28(6): 49–55. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5999647.

Flato, Gregory et al. 2013. "Evaluation of Climate Models." *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*: 741–866. http://elib.dlr.de/95697/1/Alet-Eyring-WG1AR5_Chapter09_FINAL.pdf (October 17, 2017).

Garner, Gregory, Patrick Reed, and Klaus Keller. 2016. "Climate Risk Management Requires Explicit Representation of Societal Trade-Offs." *Climatic Change* 134(4): 713–23. http://link.springer.com/10.1007/s10584-016-1607-3 (November 27, 2017).

Gupta, Alexander Sen, Nicolas C. Jourdain, Jaclyn N. Brown, and Didier Monselesan. 2013. "Climate Drift in the CMIP5 Models." *Journal of Climate* 26(21): 8597–8615.

Harmsen, Mathijs J H M et al. 2015. "How Well Do Integrated Assessment Models Represent Non-CO 2 Radiative Forcing?" *Climatic Change* 133: 565–82. https://link.springer.com/content/pdf/10.1007%2Fs10584-015-1485-0.pdf (December 19, 2017).

Hartin, C. A. et al. 2015. "A Simple Object-Oriented and Open-Source Model for Scientific and Policy Analyses of the Global Climate System - Hector v1.0." *Geoscientific Model Development* 8(4): 939–55.

Hartin, Corinne A, Benjamin Bond-Lamberty, Pralit Patel, and Anupriya Mundra. 2016. "Ocean Acidification over the next Three Centuries Using a Simple Global Climate Carbon-Cycle Model: Projections and Sensitivities." *Biogeosciences* 13: 4329–42. www.biogeosciences.net/13/4329/2016/ (October 17, 2017).

Joos, F. et al. 2013. "Carbon Dioxide and Climate Impulse Response Functions for the Computation of Greenhouse Gas Metrics: A Multi-Model Analysis." *Atmospheric Chemistry and Physics* 13(5): 2793–2825.

Kriegler, Elmar. 2005a. “Imprecise Probability Analysis for Integrated Assessment of Climate Change.” *Time*: 258. https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/index/index/docId/497 (October 29, 2017).

———. 2005b. “Imprecise Probability Analysis for Integrated Assessment of Climate Change.” *Time*: 258. https://www.pik-potsdam.de/members/edenh/theses/PhDKriegler.pdf (November 27, 2017).

Meinshausen, M., S. C B Raper, and T. M L Wigley. 2011. “Emulating Coupled Atmosphere-Ocean and Carbon Cycle Models with a Simpler Model, MAGICC6 - Part 1: Model Description and Calibration.” *Atmospheric Chemistry and Physics* 11(4): 1417–56.

Millar, J. Richard, Zebedee R. Nicholls, Pierre Friedlingstein, and Myles R. Allen. 2017. “A Modified Impulse-Response Representation of the Global near-Surface Air Temperature and Atmospheric Concentration Response to Carbon Dioxide Emissions.” *Atmospheric Chemistry and Physics* 17(11): 7213–28.

Myhre, Gunnar et al. 2013. “Anthropogenic and Natural Radiative Forcing.” *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*: 659–740.

Raper, S C B, and U Cubasch. 1996. “Emulation of the Results from a Coupled General Circulation Model Using a Simple Climate Model.” *Geophysical Research Letters* 23(10): 1107–10.

Sand, M. et al. 2016. “Response of Arctic Temperature to Changes in Emissions of Short-Lived Climate Forcers.” *Nature Climate Change* 6(3): 286–89.

Shindell, Drew T. 2014. “Inhomogeneous Forcing and Transient Climate Sensitivity.” *Nature Climate Change* 4(4): 18–21.

Smith, Christopher J et al. “FAIR v1.1: A Simple Emissions-Based Impulse Response and Carbon Cycle Model.” https://www.geosci-model-dev-discuss.net/gmd-2017-266/gmd-2017-266.pdf? (December 19, 2017).

Smith, S J, and T C Bond. 2014. “Two Hundred Fifty Years of Aerosols and Climate: The End of the Age of Aerosols.” *Atmospheric Chemistry and Physics* 14(2): 537–49. www.atmos-chem-phys.net/14/537/2014/ (October 12, 2017).

Taylor, Karl E., Ronald J. Stouffer, and Gerald A. Meehl. 2012. “An Overview of CMIP5 and the Experiment Design.” *Bulletin of the American Meteorological Society* 93(4): 485–98.

Thomson, Allison M. et al. 2011. “RCP4.5: A Pathway for Stabilization of Radiative Forcing by 2100.” *Climatic Change* 109(1): 77–94. http://link.springer.com/10.1007/s10584-011-0151-4 (October 12, 2017).

Urban, Nathan M. et al. 2014. “Historical and Future Learning about Climate Sensitivity.” *Geophysical Research Letters* 41(7): 2543–52. http://doi.wiley.com/10.1002/2014GL059484 (November 27, 2017).

van Vuuren, Detlef P. et al. 2011. “How Well Do Integrated Assessment Models Simulate Climate Change?” *Climatic Change* 104(2): 255–85.

Wigley, T. M. L., and S. C. B. Raper. 1992. “Implications for Climate and Sea Level of Revised IPCC Emissions Scenarios.” *Nature* 357(6376): 293–300. http://www.nature.com/articles/357293a0 (October 12, 2017).