

Research



**Cite this article:** Gregory JM, Andrews T, Good P. 2015 The inconstancy of the transient climate response parameter under increasing CO<sub>2</sub>. *Phil. Trans. R. Soc. A* **373**: 20140417. <http://dx.doi.org/10.1098/rsta.2014.0417>

Accepted: 6 May 2015

One contribution of 12 to a discussion meeting issue ‘Feedbacks on climate in the Earth system’.

**Subject Areas:**

atmospheric science, climatology, oceanography

**Keywords:**

climate sensitivity, ocean heat uptake, radiative forcing, climate change, climate modelling

**Author for correspondence:**

J. M. Gregory  
e-mail: [j.m.gregory@reading.ac.uk](mailto:j.m.gregory@reading.ac.uk)

# The inconstancy of the transient climate response parameter under increasing CO<sub>2</sub>

J. M. Gregory<sup>1,2</sup>, T. Andrews<sup>2</sup> and P. Good<sup>2</sup>

<sup>1</sup>NCAS-Climate, University of Reading, Reading, UK

<sup>2</sup>Met Office Hadley Centre, FitzRoy Road, Exeter, UK

In the Coupled Model Intercomparison Project Phase 5 (CMIP5), the model-mean increase in global mean surface air temperature  $T$  under the 1pctCO<sub>2</sub> scenario (atmospheric CO<sub>2</sub> increasing at 1% yr<sup>-1</sup>) during the second doubling of CO<sub>2</sub> is 40% larger than the transient climate response (TCR), i.e. the increase in  $T$  during the first doubling. We identify four possible contributory effects. First, the surface climate system loses heat less readily into the ocean beneath as the latter warms. The model spread in the thermal coupling between the upper and deep ocean largely explains the model spread in ocean heat uptake efficiency. Second, CO<sub>2</sub> radiative forcing may rise more rapidly than logarithmically with CO<sub>2</sub> concentration. Third, the climate feedback parameter may decline as the CO<sub>2</sub> concentration rises. With CMIP5 data, we cannot distinguish the second and third possibilities. Fourth, the climate feedback parameter declines as time passes or  $T$  rises; in 1pctCO<sub>2</sub>, this effect is less important than the others. We find that  $T$  projected for the end of the twenty-first century correlates more highly with  $T$  at the time of quadrupled CO<sub>2</sub> in 1pctCO<sub>2</sub> than with the TCR, and we suggest that the TCR may be underestimated from observed climate change.

## 1. Introduction

The transient climate response (TCR) is defined [1] as the global mean surface air temperature change  $T$  (with respect to the unperturbed state) under the 1pctCO<sub>2</sub> scenario, in which the atmospheric CO<sub>2</sub>

© 2015 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

concentration increases at  $1\% \text{ yr}^{-1}$ , at the time when it reaches twice its initial value (after 70 years,  $1.01^{70} = 2$ ). The TCR is regarded as a benchmark for the magnitude of climate change to be expected from an atmosphere–ocean general circulation model (AOGCM) in response to increasing radiative forcing, it being assumed that many aspects of climate change scale with  $T$  [2]. The spread of TCR in the AOGCMs of the Coupled Model Intercomparison Project Phase 5 (CMIP5) is  $1.2\text{--}2.4 \text{ K}$  (90% confidence interval) [3].

The 1pctCO<sub>2</sub> forcing scenario is of plausible magnitude for the twenty-first century but idealized. To make a link with more realistic scenarios, Gregory & Forster [4] proposed a simple model of global mean climate change, in which  $F(t) = \rho T(t)$ , where  $F$  is the effective radiative forcing (henceforth ‘forcing’ for brevity),  $t$  is time and  $\rho$  is the climate resistance ( $\text{W m}^{-2} \text{ K}^{-1}$ ). In this model,  $T(t)$  is instantaneously related to  $F(t)$ , and the TCR is  $F_{2\times}/\rho$ , where  $F_{2\times}$  is the forcing due to doubling the CO<sub>2</sub> concentration. We call this the ‘zero-layer model’ [5], because finite heat capacity does not appear in it.

In the zero-layer model, we assume that all the heat added to the climate system is stored in the ocean, so that the net downward radiative flux  $N$  at the top of the atmosphere equals the rate of ocean heat uptake. An observational assessment of heat uptake by the climate system during 1971–2010 indicates that 93% of it is accounted for by increased ocean heat content, with most of the remainder being latent heat associated with loss of land ice [6]. CMIP5 AOGCMs do not simulate most changes in the mass of land ice. Consequently, in these models,  $N$  is a very good approximation to the rate of ocean heat uptake, for time scales longer than about a year (fig. 13.8 of [7]; [8]), while on shorter time scales the heat capacity of land and atmosphere cannot be neglected.

We can write  $\rho = \alpha + \kappa$ , where  $\alpha$  is the climate feedback parameter and  $\kappa$  is the ocean heat uptake efficiency [9,10]. These quantities, respectively, measure the tendency of the surface climate system, as it warms up, to lose more heat to space (by net outgoing radiation, at a rate  $R = \alpha T$ ) and into the ocean (at a rate  $N = \kappa T$ ). Thus  $F = R + N$ . The temperature  $T$  is a skin temperature, a property of the surface climate system, including a shallow upper-ocean layer, which is regarded as having negligible heat capacity. The deep ocean is regarded as a heat sink with infinite heat capacity—it absorbs heat but does not warm up.

The zero-layer model predicts  $T = F/\rho$  for time-dependent forcing. This is formally analogous to the prediction of the steady-state  $T = F/\alpha$  for constant forcing. The equilibrium climate sensitivity (ECS), defined as  $T(=F_{2\times}/\alpha)$  for constant  $2 \times \text{CO}_2$  concentration, is a benchmark originally introduced for climate models with ‘slab’ or ‘mixed-layer’ oceans, which cannot be used for projecting time-dependent climate change, because they do not model ocean heat uptake realistically. The ocean heat uptake efficiency does not affect ECS because there is no ocean heat uptake at equilibrium, and the forcing is balanced only by increased radiation to space. Since  $\rho = \alpha + \kappa > \alpha$ , TCR is less than ECS, by about one-third for the 1pctCO<sub>2</sub> scenario in CMIP5 [11], because ocean heat uptake mitigates the rate of warming.

We recommend the term ‘transient climate response parameter’ (TCRP) for  $T/F = 1/\rho$  ( $\text{KW}^{-1} \text{ m}^2$ ), the increase in global mean surface air temperature per unit increase in forcing (in any scenario, not necessarily 1pctCO<sub>2</sub>). We use ‘parameter’ in naming the TCRP to denote the same relationship to the TCR as exists between the climate sensitivity parameter  $1/\alpha$  and the ECS, the corresponding quantities for equilibrium climate change. Held *et al.* [12] propose ‘transient climate sensitivity’ for  $1/\rho$ , but Merlis *et al.* [13] use that phrase for  $F_{2\times}/\rho$ , and this could be confusing; moreover, the mention of ‘sensitivity’ could cause confusion with ECS.

The zero-layer model is a helpful interpretative picture, but it is accurate only if the TCRP  $1/\rho$  is constant. Gregory & Forster [4] showed that the TCRP is nearly constant for the HadCM3 AOGCM under anthropogenic historical and Special Report on Emission Scenarios (SRES) forcing, and for the AOGCMs of the Coupled Model Intercomparison Project Phase 3 (CMIP3) under 1pctCO<sub>2</sub> up to  $2 \times \text{CO}_2$ , but that it cannot be constant in scenarios where the forcing stabilizes, to which the ECS eventually applies. Gregory & Forster [4] also noted that, in the 1pctCO<sub>2</sub> experiments, the increase in  $T$  for the second doubling of CO<sub>2</sub>, from  $2 \times \text{CO}_2$  to  $4 \times \text{CO}_2$ , is greater than the TCR, i.e.  $T$  under the first doubling. The forcing increase on doubling CO<sub>2</sub>

from any level is always  $F_{2\times}$ , if we make the usual assumption (to which we return in §5) that  $\text{CO}_2$  forcing depends logarithmically on  $\text{CO}_2$  concentration  $C$  according to  $F = F_{2\times} \ln(C/C_0)/\ln 2$  [14–16], where  $C_0$  is a reference level. Hence, the larger  $T$  increase for the second doubling indicates that the TCRP increases ( $\rho$  decreases) as the experiment proceeds. This behaviour is shown by both CMIP3 and CMIP5 AOGCMs.

The aim of this paper is to explain the inconstancy of TCRP in CMIP5 AOGCMs under  $\text{CO}_2$  forcing, which has implications for their projections under representative concentration pathway (RCP) scenarios. This leads us to review several of the issues investigated by Long & Collins [17] from a different perspective. We use data from  $\text{CO}_2$  and RCP experiments with the AOGCMs listed in table 1, and to remove model climate drift we subtract the state of the parallel pre-industrial control run (piControl, with constant atmospheric composition), as a function of time. We begin by describing two simple models which are helpful in interpretation of the AOGCM results.

## 2. Linearity and the step model

Linear systems theory assumes that the response of the system depends linearly on forcing, so that the response to a sum of forcings equals the sum of responses to individual forcings. Thus, if  $X_i(t)$  is the response in some quantity of the system to forcing  $F_i(t)$ , the response to  $F(t) = \sum_i F_i(t)$  is  $X(t) = \sum_i X_i(t)$ . This implies that forcing  $kF$  gives response  $kX$  for any constant scaling factor  $k$ .

By applying this theory to the climate system, the ‘step model’ [18] estimates the response  $X(t)$  of any climate variable to a forcing scenario  $F(t)$  according to

$$X(t) = \sum_{t'=1}^t X_s(t-t'+1) \frac{F(t') - F(t'-1)}{F_s}, \quad (2.1)$$

where the sum is over years, and  $X_s(t)$  is the response after  $t$  years to a ‘step forcing’  $F_s$ , which is instantaneously switched on at  $t = 0$  and held constant indefinitely.

The step model thus regards the response to any time-dependent forcing scenario as the sum of responses to a succession of  $t$  annual forcing increments  $F(t) - F(t-1)$ , with  $F(0) = 0$ , and these individual responses are estimated by scaling the response to  $F_s$ . The response to the step forcing is used rather like a Green function (which is the response to a  $\delta$ -function forcing). The idea is illustrated schematically by Good *et al.* [19] (their fig. 1b), who show that it works well for global mean surface air temperature change and ocean heat uptake simulated in CMIP5 1pctCO2 and RCP scenarios, using the step response from the CMIP5 abrupt4xCO2 experiments, in which the  $\text{CO}_2$  concentration is quadrupled at  $t = 0$  and subsequently held constant, i.e.  $F_s = F_{4\times}$ , the forcing due to  $4 \times \text{CO}_2$ .

Linearity as defined above and by Good *et al.* [20] is different from the property that the quantities of the system vary together in linear relationships as a function of state. Long & Collins [17] call that property ‘linearity in time’, because change in the state of the system is usually a matter of time. If the system is not linear, the  $X$ – $Y$  relationship will not generally be linear, because  $X$  and  $Y$  will depend in different nonlinear ways on the forcing. If the system is linear, the step model will predict any quantities  $X$  and  $Y$  correctly and individually, even if they are not linearly related in time, but the form of the relationship between them will be a function of the time dependence of the system. For example, the increase in ocean heat content  $S = \int N \, dt$ , so the relationship in time between  $N$  and  $S$  is not generally linear; one possible time dependence of the system is  $N \propto t \Rightarrow S \propto t^2 \propto N^2$ . However, if the system is linear, meaning the step model predicts  $N$  correctly, it must also predict  $S$  correctly, regardless of the nonlinearity in time.

## 3. Two-layer model

The two-layer model [5,12,21–23] consists of the upper and the deep ocean, like the zero-layer model (§1), but the layers have finite non-zero heat capacities  $C_u$  and  $C_d$ ,  $C_u \ll C_d$ . Their

**Table 1.** Parameters of the 16 CMIP5 AOGCMs used in this study. The last two lines of the table show ensemble mean and standard deviation. The TCR is the transient climate response evaluated as the time-mean  $T$  with respect to control for years 61–80 in the 1pctCO2 scenario, and  $T_{140}$  is for year 140 in the same scenario, estimated from a linear regression against time for years 121–140. The parameters  $\alpha$ ,  $\kappa$  and  $\rho$  for years 61–80 and 121–140 of 1pctCO2 are evaluated as the ratios of the time-means  $(F - N)/T$ ,  $N/T$ ,  $F/T$ , using the logarithmic formula for  $F(\text{CO}_2)$  with the  $2 \times \text{CO}_2$  forcing  $F_{2\times}$  obtained by regression of  $N$  against  $T$  in years 1–20 of the abrupt4xCO2 scenario. ECS is the effective climate sensitivity estimated as  $F_{2\times}/\alpha$ , using the TCR  $\alpha$ . For RCP4.5 and RCP8.5,  $N$  and  $T$  are the time-means with respect to control for years 2081–2100, and  $\kappa = N/T$ . TCR,  $T_{140}$ , RCP  $T$  and ECS are in  $\text{K}$ ,  $F_{2\times}$  in  $\text{W m}^{-2}$ ,  $\alpha$ ,  $\kappa$  and  $\rho$  in  $\text{W m}^{-2} \text{K}^{-1}$ .

AOGCM	$F_{2\times}$	TCR	$T_{140}$	1pctCO2 years 61–80				1pctCO2 years 121–140				RCP8.5 2081–2100				RCP4.5 2081–2100			
				$\alpha$	$\kappa$	$\rho$		$\alpha$	$\kappa$	$\rho$		$N$	$T$	$\kappa$		$N$	$T$	$\kappa$	
				TCR			$\frac{T_{140}}{T}$												
ACCESS1.0	3.44	1.98	4.59	2.32	1.00	0.74	1.74	0.99	0.51	1.51	2.56	4.46	5.31	0.57	1.37	2.63	2.63	0.52	3.45
ACCESS1.3	3.42	1.65	4.02	2.43	1.22	0.85	2.07	1.14	0.62	1.76	2.65	4.25	4.25	0.62	1.33	2.37	2.37	0.56	2.80
CanESM2	4.11	2.41	5.24	2.17	1.14	0.56	1.70	1.14	0.45	1.59	2.51	5.31	5.31	0.47	1.07	3.10	3.10	0.35	3.60
CNRM-CM5	3.56	2.08	4.52	2.17	1.13	0.59	1.71	1.08	0.49	1.57	2.12	4.24	4.24	0.50	0.98	2.64	2.64	0.37	3.16
CSIRO-Mk3.6.0	3.30	1.78	4.51	2.53	1.12	0.74	1.85	1.02	0.56	1.57	2.70	4.55	4.55	0.59	1.33	2.73	2.73	0.49	2.96
GFDL-CM3	3.60	1.95	4.79	2.45	1.12	0.72	1.84	1.03	0.49	1.52	2.72	4.83	4.83	0.56	1.37	2.87	2.87	0.48	3.20
HadGEM2-ES	3.26	2.50	5.44	2.17	0.75	0.55	1.30	0.76	0.44	1.20	2.50	4.84	4.84	0.52	1.15	2.70	2.70	0.43	4.32
INM-CM4	3.23	1.29	3.01	2.34	1.63	0.88	2.51	1.55	0.63	2.18	2.27	3.37	3.37	0.67	1.11	1.98	1.98	0.56	1.98
IPSL-CM5A-LR	3.43	2.04	5.21	2.56	0.99	0.69	1.68	0.88	0.50	1.38	2.88	5.57	5.57	0.52	1.24	3.22	3.22	0.39	3.46
IPSL-CM5A-MR	3.44	2.03	5.11	2.52	1.01	0.69	1.70	0.87	0.55	1.41	2.96	5.29	5.29	0.56	1.28	3.17	3.17	0.40	3.40
MIROC-ESM	4.59	2.16	5.56	2.58	1.32	0.80	2.13	1.10	0.61	1.71	3.22	5.27	5.27	0.61	1.37	3.03	3.03	0.45	3.47

(Continued.)

**Table 1.** (Continued.)

AOGCM	$F_{2\times}$	TCR	$T_{140}$	1pctCO2 years 61–80			1pctCO2 years 121–140			RCP8.5 2081–2100			RCP4.5 2081–2100				
				$T_{140}$ $\frac{\Delta T}{\Delta t}$	$\alpha$	$\kappa$	$\rho$	$\alpha$	$\kappa$	$\rho$	$N$	$T$	$\kappa$	$N$	$T$	$\kappa$	
MIROC5	4.47	1.51	3.74	2.48	2.11	0.85	2.96	1.74	0.71	2.46	2.73	3.74	0.73	1.32	2.10	0.63	2.12
MPI-ESM-LR	4.48	2.06	5.03	2.45	1.45	0.72	2.18	1.31	0.49	1.80	2.35	4.41	0.53	1.12	2.41	0.46	3.08
MPI-ESM-MIR	4.52	2.04	4.81	2.36	1.54	0.68	2.22	1.39	0.52	1.92	2.36	4.38	0.54	0.99	2.55	0.39	2.94
MRI-CGCM3	3.55	1.56	4.03	2.59	1.62	0.66	2.28	1.37	0.44	1.81	1.91	3.66	0.52	0.82	2.08	0.40	2.19
NorESM1-M	3.78	1.39	3.61	2.61	1.79	0.94	2.73	1.50	0.67	2.17	2.66	3.76	0.71	1.24	2.18	0.57	2.11
mean	3.76	1.90	4.58	2.42	1.31	0.73	2.04	1.18	0.54	1.72	2.57	4.50	0.58	1.19	2.61	0.46	3.02
SD	0.50	0.35	0.72	0.15	0.35	0.11	0.43	0.27	0.08	0.33	0.32	0.66	0.08	0.17	0.40	0.08	0.65

temperatures are  $T_u$  and  $T_d$  relative to initial equilibrium. The upper-layer temperature change  $T_u$  is identified with the global mean surface air temperature change  $T$ . The upper ocean participates in the surface energy budget and experiences the forcing, while the deep ocean receives a heat flux  $H = \gamma(T_u - T_d)$  from the upper ocean, with  $\gamma$  ( $\text{W m}^{-2} \text{K}^{-1}$ ) being a constant. Thus

$$C_u \frac{dT_u}{dt} = F - \alpha T_u - \gamma(T_u - T_d),$$

$$C_d \frac{dT_d}{dt} = \gamma(T_u - T_d)$$

and

$$N = C_u \frac{dT_u}{dt} + C_d \frac{dT_d}{dt} = F - \alpha T_u.$$

Geoffroy *et al.* [22] fitted the two-layer model to CMIP5 results for abrupt4xCO<sub>2</sub>. It is useful to consider how the model behaves qualitatively under this scenario, which clearly separates the various time scales of response to forcing, because it is used by the step model to predict the response to other scenarios. If a constant positive  $F$  is imposed at  $t = 0$ , as in abrupt4xCO<sub>2</sub>,  $T_d \ll T_u$  while  $t$  is small, because the deep ocean has much larger heat capacity and warms up more slowly than the upper ocean. During this phase,  $T_d$  can be neglected in  $\gamma(T_u - T_d)$ , and

$$C_u \frac{dT_u}{dt} \simeq F - \alpha T_u - \gamma T_u = F - \rho T_u \quad (3.1)$$

if we identify  $\kappa$  with  $\gamma$ , so that  $\rho = \alpha + \gamma$ . Geoffroy *et al.* [22] call this form the ‘one-layer model’ and we will refer to it as the ‘upper-layer model’. In this approximation, the upper layer has finite and the deep layer infinite heat capacity. The solution of equation (3.1) is  $T_u = (F/\rho)(1 - e^{-\rho t/C_u})$ . At  $t = 0$ ,  $T_u = 0$  and  $dT_u/dt = F/C_u$ , but, as  $T_u$  rises,  $dT_u/dt \rightarrow 0$ , on the time scale  $C_u/\rho$  characteristic of the upper ocean, which is 3.9 years using the CMIP5 ensemble mean parameters of Geoffroy *et al.* (tables 3 and 4 of [22]).

After 10–20 years,  $dT_u/dt \ll F/C_u$ , and the upper-ocean thermal inertia can thereafter be neglected, because it stores only a small fraction of the heat uptake, which goes mainly into the deep ocean. By setting  $C_u = 0$  in equation (3.1), we obtain the zero-layer model  $F = \rho T_u$ , which is valid while  $T_d \simeq 0 \Rightarrow N = H = \gamma T_u$  (or  $N = \kappa T_u$  if we identify  $\kappa$  with  $\gamma$ ).

For larger  $t$ , the zero-layer model is inaccurate because the deep ocean warms up, i.e.  $T_d > 0$ . However,  $C_u$  remains negligible, so

$$F - \alpha T_u = N = H = C_d \frac{dT_d}{dt} = \gamma(T_u - T_d). \quad (3.2)$$

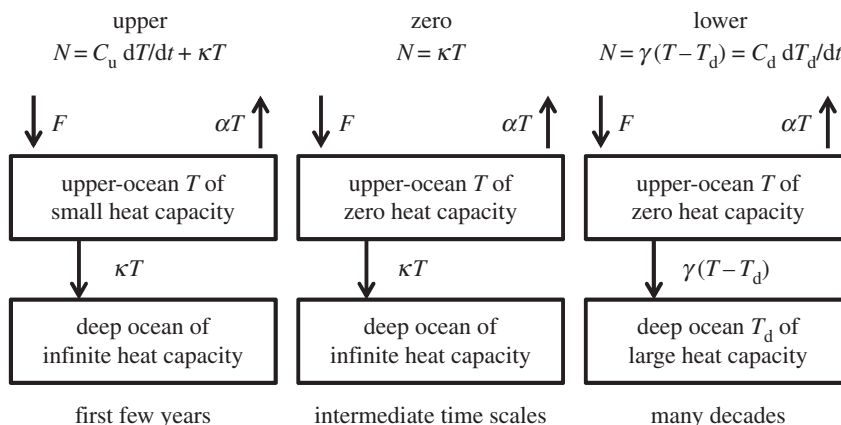
In this approximation, which we call the ‘deep-layer model’, the upper layer has zero and the deep layer finite heat capacity. The solution of equation (3.2) is

$$T_u = \frac{F}{\alpha} \left( 1 - \left( \frac{\gamma}{\rho} \right) e^{-t/\tau_d} \right) \quad \text{and} \quad T_d = \frac{F}{\alpha} (1 - e^{-t/\tau_d}) \quad \tau_d \equiv \frac{\rho C_d}{\alpha \gamma}.$$

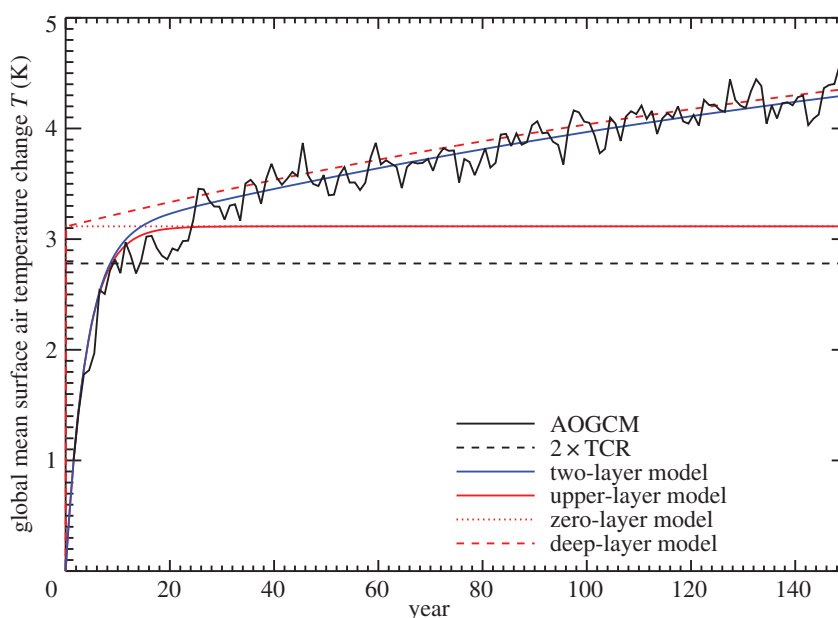
The formula given by Geoffroy *et al.* [22] for the ‘slow’ time scale of warming is approximated by our  $\tau_d$  if  $C_d \gg C_u$ . Using their CMIP5 ensemble-mean parameters,  $\tau_d = 239$  years.

The upper-, zero- and deep-layer approximations to the two-layer model are shown schematically in figure 1. As an example, we compare  $T$  simulated by the NorESM1-M AOGCM for abrupt4xCO<sub>2</sub> with the prediction of the two-layer model fitted to this AOGCM by Geoffroy *et al.* [22] (figure 2, also shown in their fig. 2). The two-layer model is close to the forced response of  $T$  in the AOGCM, although slightly overestimating in the first 20 years. It does not generate unforced variability.

The upper-layer model prediction of  $T$  is indistinguishable from the two-layer one for about the first 10 years, but then levels off at  $F_{4\times}/\rho = 3.1 \text{ K}$ , consistent with equation (3.1). The zero-layer model predicts constant  $T$  at this same level from  $t = 0$ . These two models cannot predict the continuing slow rise of  $T$  due to the warming of the deep ocean, which reduces  $H$ . Their  $T$  at long time scales would be twice the TCR ( $=F_{2\times}/\rho$ ) if the two-layer model were exact.



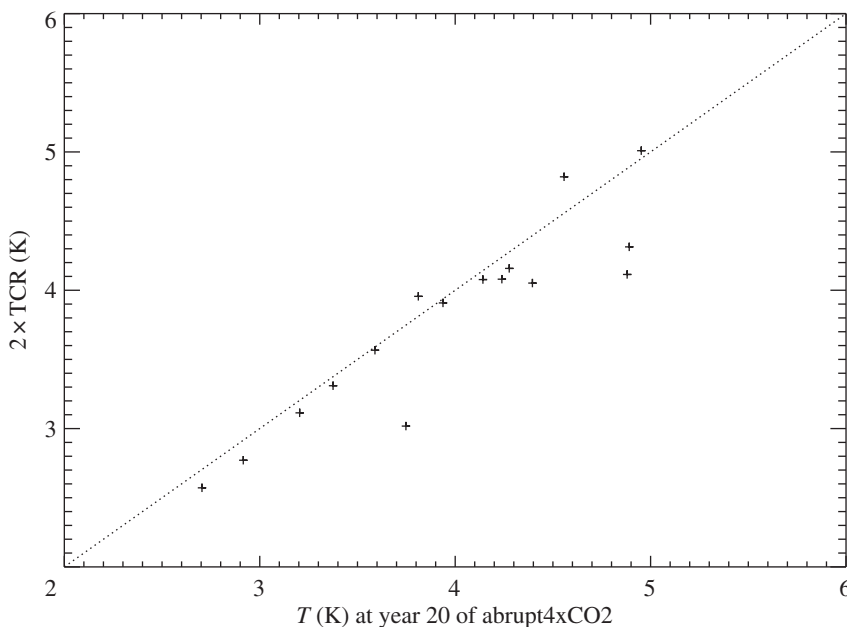
**Figure 1.** Schematic of the upper-, zero- and deep-layer models of ocean heat uptake, which are three approximations to the two-layer model, applicable on the time scales shown in the bottom row. In all versions of the model, the global energy balance is  $F = N + \alpha T$ . Global mean surface air temperature change  $T$  is identified with the upper-layer temperature change.



**Figure 2.** Global mean surface air temperature change  $T$  in the NorESM1-M AOGCM abrupt4xCO<sub>2</sub> simulation compared with predictions by the zero-, one- and two-layer models.

Like the zero-layer model, the deep-layer model predicts  $T = F_{4\times}/\rho$  at  $t = 0$ , because  $C_u = 0 \Rightarrow F = \rho T_u$  when  $T_d = 0$ . Unlike in the zero-layer model,  $T_d$  can rise in the deep-layer model, and  $T_u$  comes quite close to the AOGCM and two-layer model after about 20 years. Because the upper layer takes up heat in the initial phase of the two-layer model, but has no heat capacity in the deep-layer model,  $T_d$  is greater at all times in the latter. At large  $t$ , when  $C_u dT_u/dt$  is negligible in both models,  $0 = F - \alpha T_u - \gamma(T_u - T_d) \Rightarrow T_u = (F + \gamma T_d)/\rho$ . The greater  $T_d$  in the deep-layer model thus accounts for its small overestimate of  $T_u$  at large  $t$ . As  $t \rightarrow \infty$ ,  $T_u, T_d \rightarrow F/\alpha$  and  $H \rightarrow 0$  in both deep- and two-layer models.





**Figure 3.** Comparison of the TCR with the global mean surface air temperature change  $T$  after 20 years of abrupt4xCO<sub>2</sub> in the ensemble of 16 CMIP5 AOGCMs of table 1. The dotted line is  $2 \times \text{TCR} = T$ .

In summary, the upper-layer model is accurate for about 20 years under abrupt4xCO<sub>2</sub>, and the deep-layer model thereafter. The zero-layer model is obviously inadequate because it predicts constant  $T$ . However, its prediction is quite accurate around  $t = 20$  years, when  $F = \rho T$  is a good approximation, as can be seen for the example of NorESM1-M (figure 2), in which the zero-layer  $T$  is within 0.5 K of the two-layer model and the AOGCM for an intermediate phase spanning roughly years 10–60. In the set of 16 CMIP5 AOGCMs of table 1, the correlation between  $2 \times \text{TCR}$  and  $T$  at year 20 in abrupt4xCO<sub>2</sub> is 0.91 (figure 3). On average, the former is 5% smaller. The differences have the same causes as the inconstancy of TCRP under 1pctCO<sub>2</sub>, which we address in the following sections.

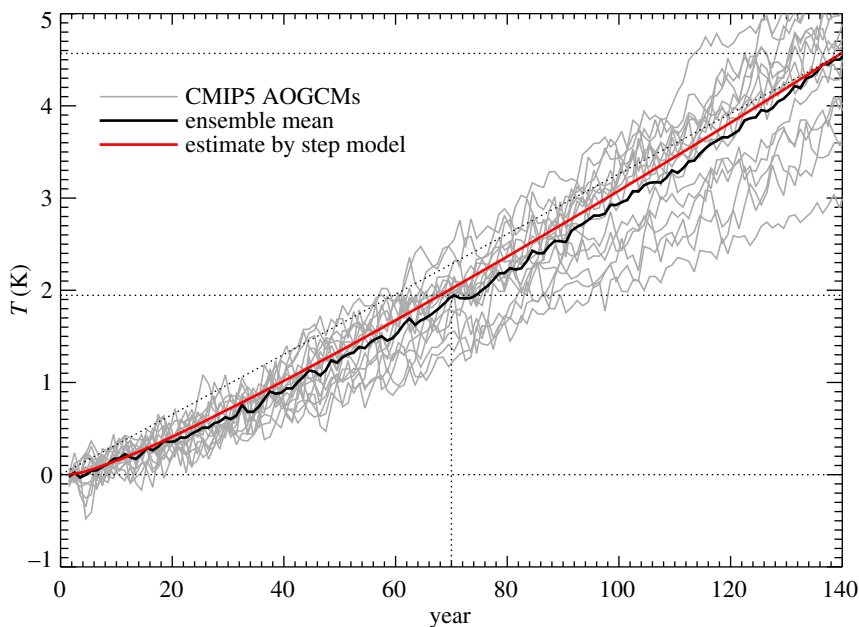
## 4. Climate response to increasing CO<sub>2</sub>

The inconstancy of the TCRP under increasing CO<sub>2</sub> can be conveniently quantified by the ratio  $T_{140} \div \text{TCR}$ , where  $T_{140} = T(t = 140 \text{ years})$ , at the time of  $4 \times \text{CO}_2$ . This ratio would be 2 if TCRP were constant. In our ensemble of 16 CMIP5 AOGCMs, the ratio lies in the range 2.1–2.6 and has a mean of 2.4 (table 1), i.e. 20% larger than expected from constant TCRP.

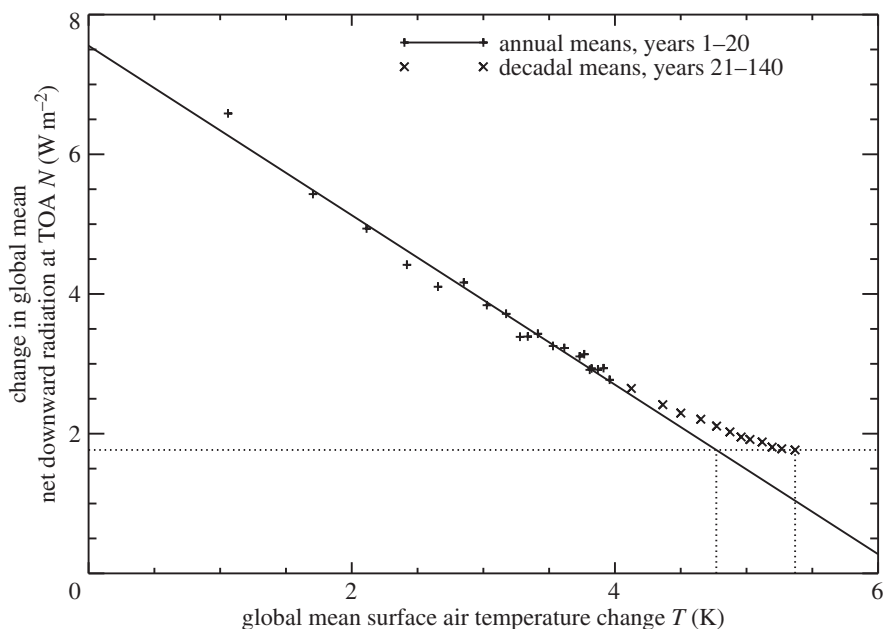
A constant geometric growth rate  $p$  of CO<sub>2</sub> concentration, following  $C = C_0 p^t$ , with  $p = 1.01$  for 1pctCO<sub>2</sub>, gives  $F = F_{2 \times} \ln(C/C_0) / \ln 2 = F_{2 \times} \ln p^t / \ln 2 = t F_{2 \times} \ln p / \ln 2$ , i.e.  $F \propto t$  and  $dF/dt = F_{2 \times} \ln p / \ln 2$  is a constant. Constant TCRP therefore predicts  $T = F/\rho \propto t$  under 1pctCO<sub>2</sub>, but in the AOGCMs  $T$  rises at an accelerating rate (figure 4; [17]).

Under any constant forcing,  $N = F - \alpha T$  is a straight line if  $\alpha$  is constant [24]. In most CMIP5 abrupt4xCO<sub>2</sub> experiments, the  $N$ – $T$  relationship is not exactly linear [25]; there is a weak concave-upward curvature (figure 5) indicating a decrease in  $\alpha$  [26–29]. Due to this effect, the ensemble-mean  $T$  after 140 years of constant  $4 \times \text{CO}_2$  is about 10% larger than estimated from the  $N$  of that time using the linear regression of  $N$  against  $T$  from the first 20 years (as shown in figure 5 for the time-mean of years 131–140). The time-mean  $F$  during 140 years of 1pctCO<sub>2</sub> is half of that in abrupt4xCO<sub>2</sub>, i.e. equal to  $2 \times \text{CO}_2$  on average, and has less time to take effect because it rises gradually, so we might expect the effect of this nonlinearity on 1pctCO<sub>2</sub> to be smaller, perhaps

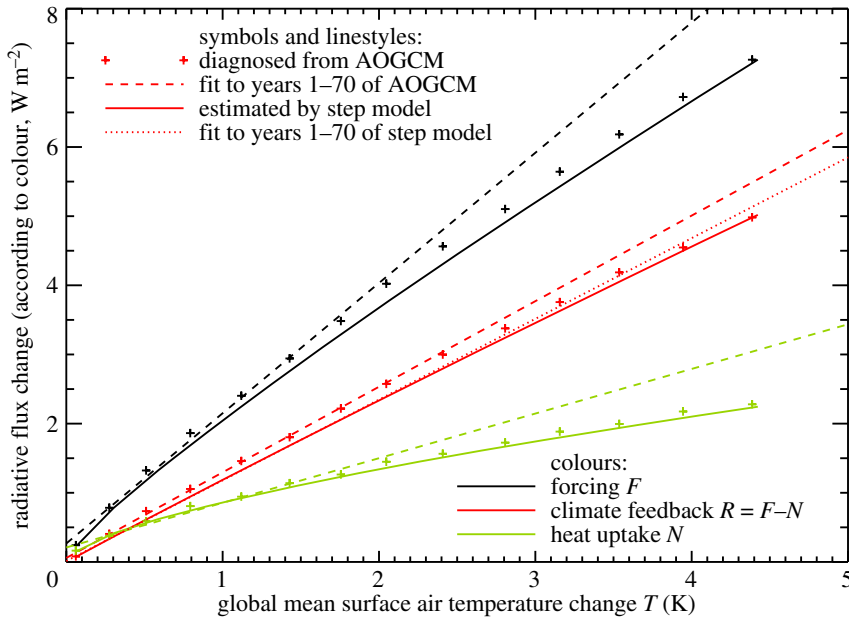




**Figure 4.** Global mean surface air temperature change  $T$  in CMIP5 1pctCO<sub>2</sub> experiments, their ensemble mean and the ensemble mean of simulations by the step model using the abrupt4xCO<sub>2</sub> response of each AOGCM. The ensemble-mean warming caused by the first and second doublings of CO<sub>2</sub> can be read from the graph using the horizontal and vertical dotted lines. The diagonal dotted line indicates a constant  $dT/dt$ .



**Figure 5.** Relationship between annual means of the change in global mean net downward radiative flux  $N$  at the top of the atmosphere and the change in surface air temperature  $T$  in the ensemble mean of CMIP5 abrupt4xCO<sub>2</sub> experiments. The solid line is a linear regression of  $N$  against  $T$  for years 1–20. The dotted lines indicate the time-mean  $N$  and  $T$  for years 131–140 and the estimate of  $T$  from the linear regression.



**Figure 6.** The relationship between the change in the energy budget of the climate system (radiative forcing  $F$ , climate feedback  $R$  and net downward radiative flux  $N$  at the top of the atmosphere, which equals net heat uptake) and the change in surface air temperature  $T$  in the ensemble mean of CMIP5 1pctCO<sub>2</sub> experiments. The plus symbols show decadal means diagnosed from the AOGCMs. The dashed lines show linear functions of  $T$ , obtained by regression of the AOGCM results for years 1–70. The solid lines show estimates from the step model using  $F(t)$  for 1pctCO<sub>2</sub> and the abrupt4xCO<sub>2</sub> response. The dotted line shows a linear function of step-model  $T$ , obtained by regression of the step-model estimate of  $R$  for years 1–70.

about 5%. Although it is in the right sense, and must contribute, it is too small to explain the inconstancy of TCRP by 20% on the model mean under 1pctCO<sub>2</sub>.

As a test case, we take the NorESM1-M AOGCM under 1pctCO<sub>2</sub>, for which  $T_{140}/\text{TCR}$  is particularly large (table 1). Geoffroy *et al.* [22] fit the two-layer model to NorESM1-M assuming a linear  $N(T)$  relationship in abrupt4xCO<sub>2</sub>, while Geoffroy *et al.* [27] fit a modified version of the two-layer model, making  $\alpha$  depend on  $N$  following the formulation of Held *et al.* [12] for the efficacy of ocean heat uptake, in order to reproduce the AOGCM nonlinear  $N(T)$  relationship. Their two versions of the two-layer model give indistinguishable results for NorESM1-M  $T(t)$  under 1pctCO<sub>2</sub> (their fig. 4), and the same is true for the other AOGCMs they show, supporting our inference that the inconstancy of the climate feedback parameter under constant CO<sub>2</sub> is not the reason for the inconstancy of the TCRP under increasing CO<sub>2</sub>.

The zero-layer model assumes that  $F = \rho T$ . We calculate  $F(t)$  under 1pctCO<sub>2</sub> from the logarithmic formula (see also §5) with  $F_{2\times}$  in each AOGCM obtained by regressing  $N$  against  $T$  in its abrupt4xCO<sub>2</sub> experiment [19,25], in which the  $N$ -intercept is  $F = F_{4\times} = 2F_{2\times}$  [24]. We avoid most of the nonlinearity of the abrupt4xCO<sub>2</sub>  $N$ – $T$  relationship by using only the first 20 years in the regression (figure 5), though even during these years there may be some curvature. In the CMIP5 ensemble mean of 1pctCO<sub>2</sub> experiments, figure 6 shows that  $F \propto T$  holds fairly well up to  $2 \times \text{CO}_2$  (comparing the black plus symbols with the linear fit shown as a black dashed line), but it becomes inadequate thereafter as  $\rho$  decreases.

The same is true of the zero-layer model assumptions  $R = F - N = \alpha T$  and  $N = \kappa T$  with constant  $\alpha$  and  $\kappa$  (figure 6, red and green plus symbols and dashed lines). In the AOGCMs, the climate feedback parameter  $\alpha$  and the ocean heat uptake efficiency  $\kappa$  both tend to decrease (values for years 1–70 and years 121–140 are compared in table 1). This suggests that these two effects make distinct contributions to the inconstancy of the TCRP. We consider  $\alpha$  in §5 and  $\kappa$  in §6.

## 5. Climate feedback

The curvature of  $R = F - N$  versus  $T$  under 1pctCO<sub>2</sub>, indicating declining climate feedback parameter  $\alpha$ , is not reproduced by the step model, which predicts  $R \propto T$ , i.e. constant  $\alpha$  to a good approximation (figure 6, solid and dotted red lines), even though  $\alpha$  is not constant in the abrupt4xCO<sub>2</sub> experiment used by the step model (figure 5). This is a further demonstration that the decreasing  $\alpha$  under constant  $4 \times \text{CO}_2$  is not important for 1pctCO<sub>2</sub>.

The inaccuracy of the step model can also be seen as a function of time (figure 4, and fig. 2a of [19]). The error is largest in the middle decades of the 140-year experiment, and decreases subsequently. In the case of the NorESM1-M AOGCM, the step model gives particularly large errors in  $T(t)$  under 1pctCO<sub>2</sub> (black and red lines in figure 7a). Using the parameters appropriate for NorESM1-M assuming a linear  $N(T)$  relationship for abrupt4xCO<sub>2</sub> (tables 3 and 4 of [22]), we see that the two-layer model behaves very similarly to the step model (figure 7a, solid green and red lines, which are almost coincident). The inaccuracy of the step model means that the response to 1pctCO<sub>2</sub> is not linear in the sense of linear systems theory (§2). We consider two possibilities to explain the nonlinear behaviour.

One possibility is that the logarithmic formula for  $F(\text{CO}_2)$  is not accurate. The forcing caused by successive doublings of CO<sub>2</sub> is shown to increase with CO<sub>2</sub> in the GISS model E atmospheric general circulation model (AGCM) (table 1 of [30]), the BMRC AGCM [31] and the NCAR CCSM3 AOGCM (table 4 of [32]; fig. 2a of [33]). The HadCM3 AOGCM shows the opposite behaviour:  $\frac{1}{2}F_{4\times} = 3.86 \pm 0.07 \text{ W m}^{-2}$  is 10% less than  $F_{2\times} = 4.28 \pm 0.16 \text{ W m}^{-2}$ .

In the results of Colman & McAvaney [31] (their fig. 1), the forcing caused by doubling CO<sub>2</sub> concentration from  $C$  to  $2C$  increases roughly linearly with  $\ln C$  for  $1\text{--}4 \times \text{CO}_2$ . This suggests a modification of the usually assumed logarithmic dependence to

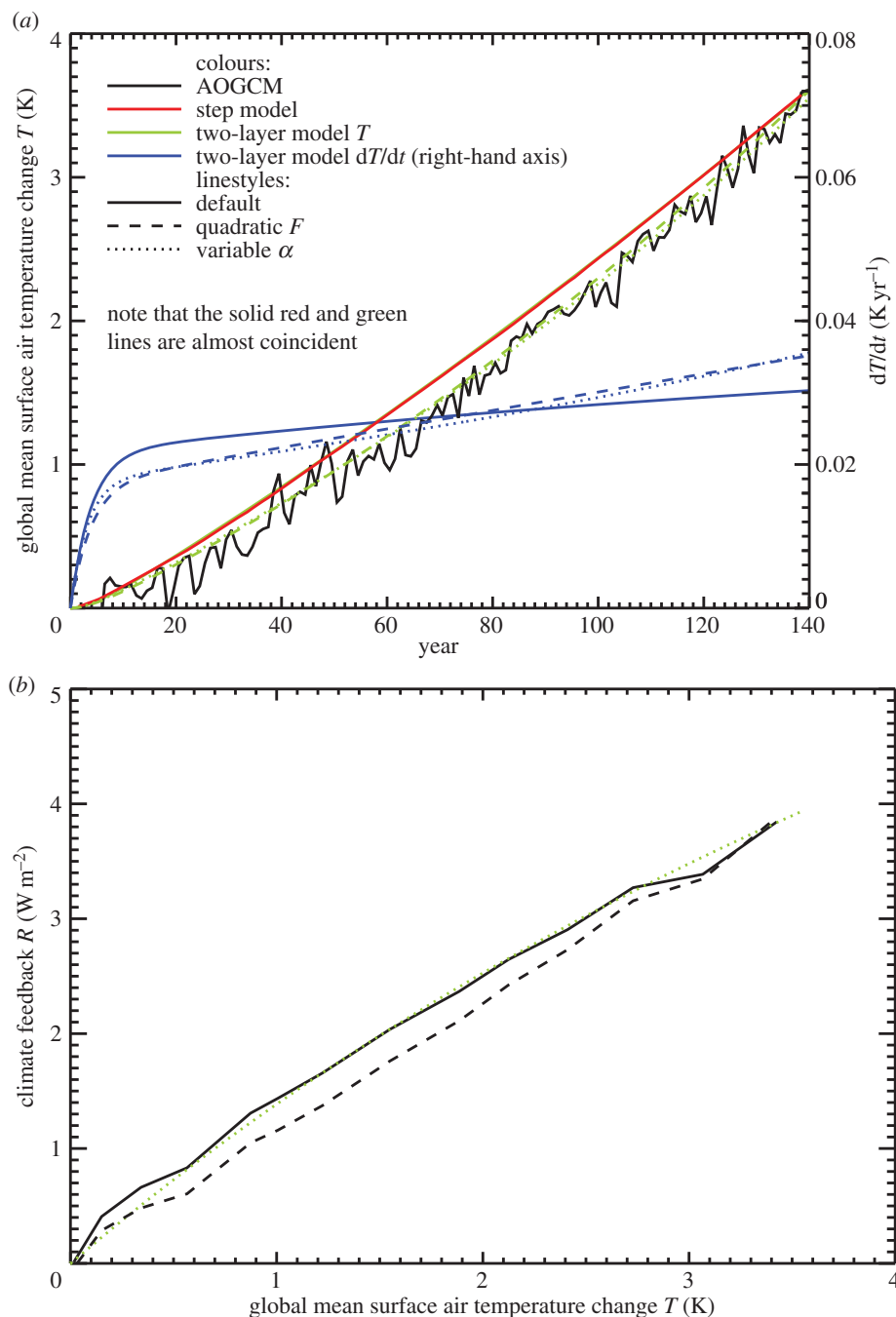
$$F = F_{4\times} \left( A \ln \frac{C}{C_0} + B \left( \ln \frac{C}{C_0} \right)^2 \right),$$

where  $A$  and  $B$  are constants, and  $C_0$  is the  $1 \times \text{CO}_2$  concentration, for which  $F(C_0) = 0$ . Considering results of radiative transfer calculations from a line-by-line code, Byrne & Goldblatt [34] propose this same form as an approximate expression for CO<sub>2</sub> forcing. With this form,  $\delta F \propto (A + 2B \ln C) \delta \ln C$ , consistent with Colman and McAvaney. If  $B = 0$ , the usual purely logarithmic formula  $F = AF_{4\times} \ln C/C_0$  is recovered, where  $A = 1/\ln 4$ .

The quadratic formula can be fitted to  $F(C)$ , requiring  $F(C_0) = 0$ , by linear regression of  $F/\ln c = F_{4\times}(A + B \ln c)$  against  $\ln c$ , where  $c \equiv C/C_0$ . The results of Hansen *et al.* [30] for  $F(C)$  with  $c$  in the range 1–16 are thus estimated with a RMS error of  $0.12 \text{ W m}^{-2}$  using  $A = 0.66$  and  $B = 0.051$ , and those of Jonko *et al.* [32] for effective radiative forcing with  $c = 2, 4, 8$  with an error of  $0.038 \text{ W m}^{-2}$  using  $A = 0.58$  and  $B = 0.109$ . The expression given by Byrne & Goldblatt [34] is equivalent to  $F_{4\times} = 8.12 \text{ W m}^{-2}$ ,  $A = 0.66$ ,  $B = 0.048$ .

In the 1pctCO<sub>2</sub> scenario,  $\ln C$  increases linearly with time, so  $F(t) = F_{4\times}(A't + B't^2)$ , with  $A' = A(\ln 2/70 \text{ yr})$  and  $B' = B(\ln 2/70 \text{ yr})^2$ . Using this quadratic form for  $F$  as a function of time with  $F_{4\times} = 6.2 \text{ W m}^{-2}$  for NorESM1-M [22,25] largely eliminates the  $T$  error in the two-layer model (the dashed green line in figure 7a follows the trajectory of the black line). This is simply because the quadratic  $F$  initially increases more slowly than the linear  $F$ , and later more quickly, and  $T$  behaves likewise (compare the dashed and solid blue lines in figure 7a). Using the quadratic  $F$  to compute  $R = F - N$  with the AOGCM  $N$  and  $T$  straightens the  $R$ – $T$  relationship in the AOGCM (compare the solid and dashed black lines in figure 7b). This explanation of the error thus allows  $\alpha$  to be constant; its apparent inconstancy in figure 6 would be due to the incorrect estimate of  $F(t)$ .

The second possibility to explain the  $T$  error is that  $\alpha$  depends on time or on  $C$ —these two possibilities cannot be distinguished in 1pctCO<sub>2</sub>, since  $F \propto t$ . Jonko *et al.* [32] find that  $\alpha$  decreases for successive doublings of CO<sub>2</sub> in CCSM3. In the HadCM3 AOGCM,  $\alpha$  is about 20% larger for constant  $2 \times \text{CO}_2$  than  $4 \times \text{CO}_2$  ( $1.47 \pm 0.10$  and  $1.25 \pm 0.04 \text{ W m}^{-2} \text{ K}^{-1}$ , respectively, from ensemble means of seven integrations). Using the quadratic formula for  $F(C)$  fitted for HadCM3 with  $c = C/C_0 = 2$  and  $4$ , we can calculate  $\alpha = (F - N)/T$  during the 1pctCO<sub>2</sub> experiment. We



**Figure 7.** Response to 1pctCO<sub>2</sub> in the NorESM1-M AOGCM and the step and two-layer models. (a)  $T$  (left-hand axis) and  $dT/dt$  (right-hand axis, two-layer model only) as a function of time  $t$ . (b)  $R = F - N$  versus  $T$ . The solid black AOGCM curve for  $R$  is calculated assuming the logarithmic formula for  $F$ .

find that it declines linearly in time (not shown). Regressing  $\alpha(t)$  against  $t$ , we estimate that  $\alpha = 1.5 \text{ W m}^{-2} \text{ K}^{-1}$  and  $1.2 \text{ W m}^{-2} \text{ K}^{-1}$  at  $t = 70$  and  $140$  years, when  $c = 2$  or  $4$ , respectively, in agreement with the results from abrupt2xCO<sub>2</sub> and abrupt4xCO<sub>2</sub>. Extrapolating backward to the start of the experiment indicates  $\alpha = 1.8 \text{ W m}^{-2} \text{ K}^{-1}$  for  $c = 1$ . Thus,  $\alpha$  declines by about 40% during the 140 years of 1pctCO<sub>2</sub> in HadCM3.

For the sake of illustration, we therefore assume likewise in the two-layer model for NorESM1-M that  $\alpha$  for infinitesimal  $\text{CO}_2$  perturbations with respect to the control  $\text{CO}_2$  is 40% larger than  $\alpha$  for  $4 \times \text{CO}_2$  ( $1.11 \text{ W m}^{-2} \text{ K}^{-1}$ ), and that  $\alpha$  decreases linearly with time during the 1pct $\text{CO}_2$  scenario. With this form for  $\alpha$  (using  $F$  linear in time, as we did originally), we again largely eliminate the error in  $T$  (compare the dotted green and solid black lines in figure 7a). The explanation is somewhat different from that for the quadratic  $F$ . Initially,  $T = 0$ , so  $N = F$  for both constant and decreasing  $\alpha$  and  $T$  rises initially at the same rate (solid and dotted blue lines in figure 7b). As soon as  $T > 0$ , the larger  $\alpha$  reduces  $N = F - \alpha T$ , so the rate of warming with variable  $\alpha$  falls behind. Later on,  $\alpha$  becomes more similar in the two cases, but the case with variable  $\alpha$  by then has lower  $T$ , so  $N$  is larger, the rate of warming becomes greater than for constant  $\alpha$ , and the error in  $T$  decreases. Using this form for  $\alpha$ , the two-layer model reproduces the nonlinear  $R(T)$  relationship of the AOGCM (dotted green and solid black lines in figure 7b).

Thus, the error in the estimate of  $T$  for NorESM1-M under 1pct $\text{CO}_2$  by the step and two-layer models can be explained equally well by  $F$  increasing more rapidly than logarithmic with  $\text{CO}_2$ , or by  $\alpha$  decreasing during the experiment (as a function of  $\text{CO}_2$  concentration or of time). These are nonlinear effects in the sense of §2. (They were mitigated by Good *et al.* [19] in their use of the step model to predict differences in  $T$  between RCPs, rather than predicting  $T$  in individual RCPs.) Either effect could account for the deviation of the step model from the AOGCM in intermediate years of 1pct $\text{CO}_2$  in the CMIP5 ensemble (figure 4). Consequently either could also explain the apparent inconstancy of the climate feedback parameter (figure 6). The two are distinct physical explanations, but with the experiments currently available we cannot tell which of them is more important in explaining the CMIP5 behaviour. (They could be distinguished using abrupt2x $\text{CO}_2$  experiments; see §8.)

## 6. Ocean heat uptake

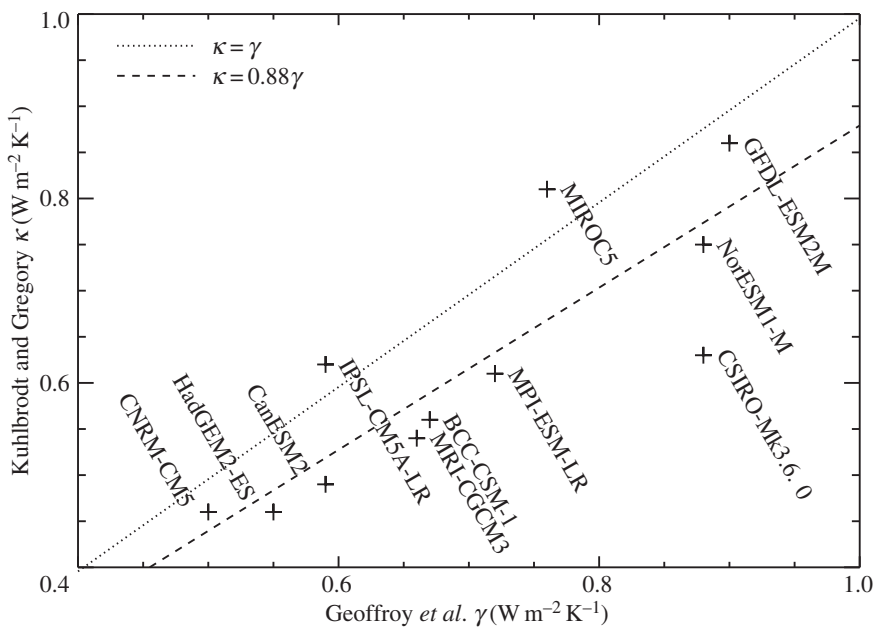
The curvature of the  $N$ - $T$  relationship under 1pct $\text{CO}_2$  indicates declining ocean heat uptake efficiency  $\kappa = N/T$  [35]. This behaviour is well reproduced by the step model (figure 6, solid and dotted green lines). It follows that the behaviour of ocean heat uptake efficiency under 1pct $\text{CO}_2$  must be entirely a consequence of effects that are also observed under abrupt4x $\text{CO}_2$ .

Therefore, the ocean heat uptake efficiency for 1pct $\text{CO}_2$  should be related to the zero-layer approximation of the two-layer model, which applies on the time scale of a few decades, as we showed above (§3). This can be tested for CMIP5 models by comparing the values of  $\kappa$  obtained from regression of  $N(t)$  against  $T(t)$  in 1pct $\text{CO}_2$  experiments [11] for years 1–70, during which constant  $\kappa$  is a reasonably good approximation (figure 6), with those obtained for  $\gamma$  by fitting the two-layer model to abrupt4x $\text{CO}_2$  experiments [22]. These two datasets and fitting procedures are entirely independent, and there is a highly significant correlation of 0.83 between  $\kappa$  and  $\gamma$  in the set of 11 AOGCMs for which both are available (figure 8).

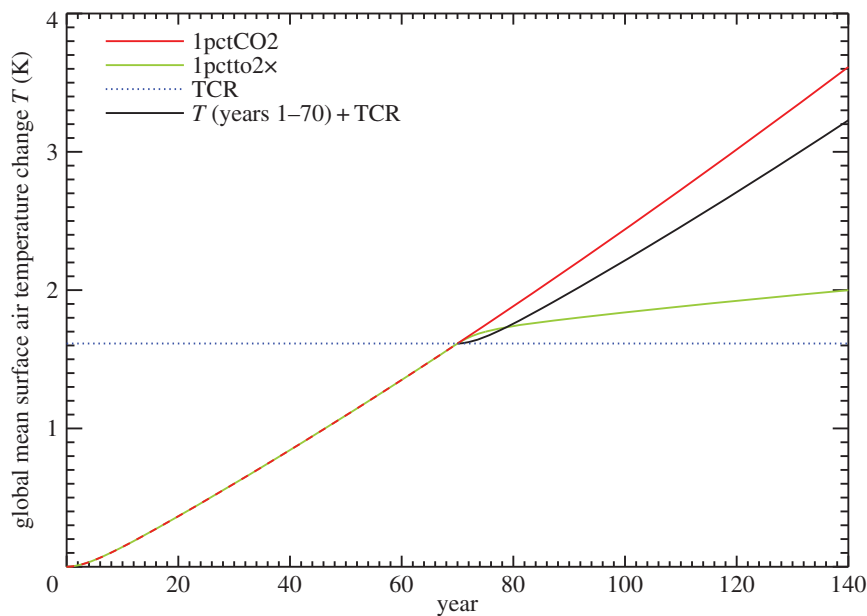
Because  $T_d \neq 0$ , the zero-layer approximation of  $N = H = \gamma(T_u - T_d)$  by  $\kappa T_u$  requires  $\kappa = N/T_u = \gamma(1 - T_d/T_u) < \gamma$ , as can be seen by comparison with the 1:1 line in figure 8. The correlation of  $\kappa$  with  $C_d$  of Geoffroy *et al.* [22] (their  $C_0$ ) is 0.63, which is significant at the 5% level. We suggest that this is because, for a given  $T_u$  and  $N$ , a larger  $C_d$  gives a smaller  $T_d$  and implies a larger  $\kappa$ . There is no significant correlation of  $\kappa$  with  $C_u$ . We conclude that the spread of ocean heat uptake efficiency across models is predominantly caused by the spread in the strength of the thermal coupling between upper and deep oceans.

The ‘cold start’ effect [36,37] is a consequence of the inconstancy of  $\kappa$ . It is the observation that under steadily increasing forcing, beginning from a steady state, the initial rate of warming is less than the rate after a few decades. A qualitative explanation of the effect can be given in terms of the zero-layer model  $T = F/\rho$ , in which the rate of warming  $dT/dt = (1/\rho) dF/dt$ , assuming the rate of change of  $\rho$  is relatively small. If  $\kappa$  declines, so does  $\rho$ , and hence  $dT/dt$  increases with time under a forcing increasing at a constant rate  $dF/dt = S$ , as in 1pct $\text{CO}_2$ .

This phenomenon is illustrated using results from the two-layer model (fitted to NorESM1-M, [22]) in figure 9, in which we compare  $T(t)$  during the first and second doublings of  $\text{CO}_2$  (black

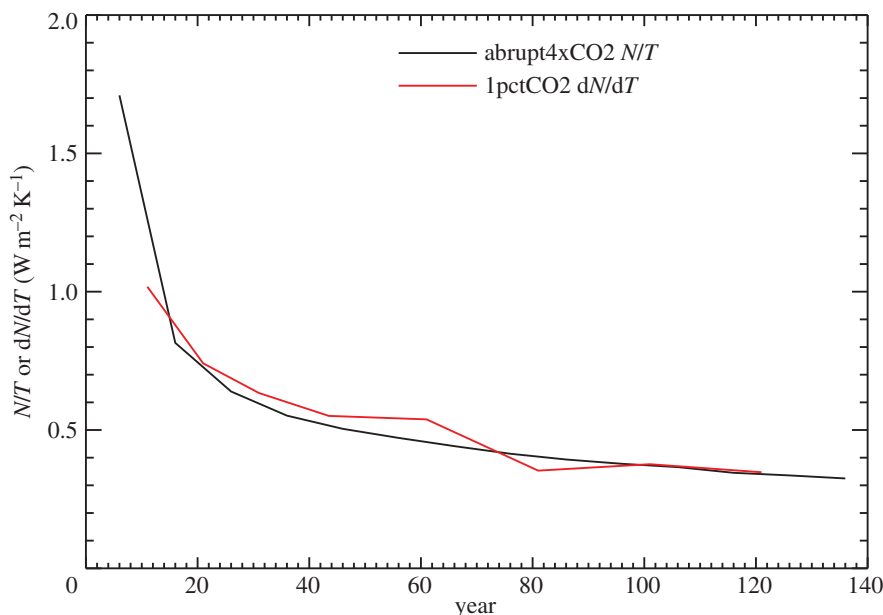


**Figure 8.** Relationship between the ocean heat uptake efficiency  $\kappa$  diagnosed from CMIP5 AOGCMs [11] and the thermal coupling  $\gamma$  in the two-layer model fitted to AOGCMs [22].



**Figure 9.** Comparison of global mean surface air temperature change  $T$  under the 1pctCO2 and 1pctto2x scenarios in the two-layer model fitted to the NorESM1-M AOGCM. The two scenarios are the same for years 1–70. The black line is  $T$  in years 1–70 translated rightward by 70 years and upward by the TCR, i.e.  $T$  at year 70.

and red lines). The cold start effect is the rapid separation of the lines in the first decade, but they continue to diverge thereafter as  $\kappa$  declines. Since the two-layer model is linear in its response to forcings,  $T$  for years 71–140 of 1pctCO2 (red line in figure 9, CO<sub>2</sub> increasing from  $2 \times$  CO<sub>2</sub> to  $4 \times$  CO<sub>2</sub>, following CO<sub>2</sub> increasing at  $1\% \text{ yr}^{-1}$  for 70 years) is the sum of  $T$  for years 1–70 of 1pctCO2 (CO<sub>2</sub> increasing at  $1\% \text{ yr}^{-1}$  up to  $2 \times$  CO<sub>2</sub>, the difference between the black and blue lines) and  $T$



**Figure 10.** Comparison of  $dN/dT$  as a function of time in 1pctCO2 with  $N/T$  as a function of time in abrupt4xCO2 for the CMIP5 ensemble mean.

for years 71–140 of the 1pctto2x scenario (constant  $2 \times \text{CO}_2$ , following  $\text{CO}_2$  increasing at  $1\% \text{ yr}^{-1}$  for 70 years, green line). This decomposition suggests the interpretation that the  $T$  increase for the second doubling of  $\text{CO}_2$  is greater than the TCR because of the ‘commitment’ to further warming accumulated during years 1–70 and realized during years 71–140 of 1pctto2x, during which  $T$  rises (green line) while  $F$  is constant at  $F_{2\times}$  and  $N = F - \alpha T$  is declining. This last statement implies that ocean heat uptake efficiency  $\kappa = N/T$  is decreasing. Thus, the interpretations of this paragraph and the preceding one are equivalent.

Under abrupt4xCO2,  $N$  decreases from  $F$  at  $t=0$  to zero in the final steady state, while  $T$  increases from 0 to  $F/\alpha$  (figure 5, noting that  $\alpha$  is not constant, as discussed in the previous section). Hence,  $\kappa = N/T$  declines continuously in abrupt4xCO2, from infinity to zero (figure 10). Given this behaviour, it is perhaps surprising that  $\kappa$  can be treated as constant for a few decades under 1pctCO2. A quantitative understanding of this can be gained from the step model.

The step response and the scenario response are in a special relationship in the case of constant  $dF/dt$ , because  $F(t) - F(t-1) = S$  for all  $t$ , and equation (2.1) becomes

$$X(t) = \frac{S}{F_s} \sum_{t'=1}^t X_s(t-t'+1) = \frac{S}{F_s} \sum_{t'=1}^t X_s(t').$$

Although we do not apply the step model in a continuous-time form, it helps the interpretation to regard this as

$$X(t) = \frac{S}{F_s} \int_0^t X_s(t') dt' \Rightarrow \frac{dX}{dt} = \frac{S}{F_s} X_s.$$

Thus, the *rate of change* of the 1pctCO2 response after  $t$  years is proportional to the *response itself* to the abrupt4xCO2 step after  $t$  years.

Hence

$$\frac{dT}{dt} = \frac{S}{F_s} T_s \quad \text{and} \quad \frac{dN}{dt} = \frac{S}{F_s} N_s. \quad (6.1)$$

Under abrupt4xCO2,  $T_s$  is initially zero and rapidly increases, but after the upper ocean time scale of 10–20 years it tends to level off, with comparatively slow warming thereafter (figures 2 and 5).



Therefore, under 1pctCO<sub>2</sub>, the *rate* of warming  $dT/dt$  is initially zero but rapidly increases on the upper ocean time scale. This is the cold start effect. Thereafter,  $dT/dt$  increases slowly under 1pctCO<sub>2</sub> because  $T_s$  increases slowly in response to step forcing.

Moreover,

$$\frac{dN}{dT} = \frac{dN}{dt} \frac{dt}{dT} = \frac{N_s}{T_s}, \quad (6.2)$$

i.e. the tangent slope of the  $N(T)$  relationship in 1pctCO<sub>2</sub> equals the slope of the line from the origin to  $(T_s, N_s)$  after the same time in abrupt4xCO<sub>2</sub> (figure 10). These quantities are initially large and decrease as time passes. Therefore,  $N$  under 1pctCO<sub>2</sub> rises steeply with  $T$  at first (large  $dN/dT$ ), but at a continuously decreasing rate ( $dN/dT$  decreases, solid green line in figure 6). This slow decline in  $dN/dT$  and hence in ocean heat uptake efficiency, which contributes to the increase in TCRP under 1pctCO<sub>2</sub>, arises from the warming of the deep ocean on multi-decadal time scales.

## 7. Transient climate response in representative concentration pathways

Forster *et al.* [38] find in CMIP5 experiments that  $\rho$  for 2000–2050 is smaller (i.e. TCRP is greater) for RCP8.5 than for 1pctCO<sub>2</sub>, and smaller for RCP4.5 than for RCP8.5. The forcing in RCPs is dominated by greenhouse gases, especially CO<sub>2</sub>, throughout the twenty-first century, and increasingly so as time passes and anthropogenic aerosol forcing becomes relatively smaller. From our findings in §5, we might suppose that the  $\rho$  dependence on scenario could partly be due to  $\alpha$  decreasing with increasing CO<sub>2</sub>. However, the forcing used by Forster *et al.* to calculate  $\rho$  was estimated using the method of Forster & Taylor [39], which assumes a constant  $\alpha$ , according to  $F = N + \alpha T \Rightarrow \rho = F/T = N/T + \alpha$ . Hence, the  $\rho$  variation they find can only come from ocean heat uptake efficiency, although if  $F$  were re-evaluated in these scenarios using prescribed sea-surface conditions [12,30,40] variation might be revealed in  $\alpha$  as well.

The average rate of forcing increase from 2000 to 2050 is equivalent to 1.1% yr<sup>-1</sup> of CO<sub>2</sub> concentration in RCP8.5 and 0.7% yr<sup>-1</sup> in RCP4.5 (using data from [2]), both different from 1pctCO<sub>2</sub>. However, from equation (6.1), with  $T(0) = 0$  and  $N(0) = 0$ , we obtain

$$\kappa = \frac{N}{T} = \frac{\int_0^t dN(t')/dt' dt'}{\int_0^t dT(t')/dt' dt'} = \frac{\int_0^t N_s(t') dt'}{\int_0^t T_s(t') dt'},$$

which does not depend on  $S = dF/dt$ . Therefore,  $\kappa$  will be the same, at a given time, under a scenario of constant forcing increase at any constant rate  $S$ , implying that the differences in the average rate of forcing increase in 1pctCO<sub>2</sub> and RCPs do not explain the  $\rho$  variation noted by Forster *et al.*

Instead, we suggest that the explanation has two aspects. First, the value of  $\rho$  for 1pctCO<sub>2</sub> corresponds to the TCR, i.e. year 70, the time of  $2 \times \text{CO}_2$ , whereas by 2050 in the RCP experiments, which are preceded by historical simulations, the forcing has already been increasing for a couple of centuries. Hence, the  $\kappa$  and  $\rho$  for the RCPs are smaller than for the TCR because of the continuous decline of ocean heat uptake efficiency as time passes. Second, under RCP4.5, although the forcing continuously increases ( $dF/dt > 0$ ) during the twenty-first century, it tends to stabilize ( $d^2F/dt^2 < 0$ ). Referring to equation (2.1), this means that  $F(t') - F(t' - 1)$  is greater for smaller  $t'$ , giving greater weight to  $X_s$  for later times, when  $N_s/T_s$  is smaller. Therefore,  $dN/dT$  and hence  $\kappa$  and  $\rho$  decrease more rapidly in stabilization scenarios such as RCP4.5 than in scenarios such as RCP8.5 having constant or increasing  $dF/dt$ .

The decline in ocean heat uptake efficiency is one of the reasons why TCRP increases, as discussed in §§1 and 6. The zero-layer model suggests that projections of  $T$  can be made by scaling the TCR =  $F_{2\times}/\rho$ , according to  $T = F/F_{2\times} \times \text{TCR} = F/\rho = F/(\alpha + \kappa)$ . If  $\kappa$  declines, this formula will give an underestimate of  $T$ . Specifically, using  $\kappa_{2\times}$  for the time of  $2 \times \text{CO}_2$  under 1pctCO<sub>2</sub>, i.e. as

in the TCR, instead of the  $\kappa_{\text{RCP}}$  actually prevailing at the time in the RCP projection will lead to a fractional error in  $T$  of

$$\frac{F}{\alpha + \kappa_{2\times}} \frac{\alpha + \kappa_{\text{RCP}}}{F} - 1 = \frac{\kappa_{\text{RCP}} - \kappa_{2\times}}{\alpha + \kappa_{2\times}}.$$

In our set of 16 CMIP5 AOGCMs, the ensemble-mean  $\kappa_{\text{RCP}}$  for the time-mean of 2081–2100 is  $0.58 \text{ W m}^{-2} \text{ K}^{-1}$  under RCP8.5 and  $0.46 \text{ W m}^{-2} \text{ K}^{-1}$  under RCP4.5, while  $\kappa_{2\times} = 0.73 \text{ W m}^{-2} \text{ K}^{-1}$  and  $\alpha = 1.31 \text{ W m}^{-2} \text{ K}^{-1}$  (the latter assumed constant, table 1). Hence, scaling the TCR will underestimate  $T$  in 2081–2100 by 7% and 13% for RCP8.5 and RCP4.5, respectively, due to the decline in  $\kappa$ .

The estimated equilibrium warming  $F/\alpha$  exceeds the zero-layer estimate  $F/(\alpha + \kappa_{\text{RCP}})$  by 44% and 35% in 2081–2100. This shows the important influence of ocean heat uptake on the mean magnitude of  $T$ , which is the reason why the TCR gives a better indication of the expected  $T$  than the ECS =  $F_{2\times}/\alpha$ .

On the other hand, the model spread in  $T$  under 1pctCO<sub>2</sub> and RCPs is dominated by the spread in  $F$  and  $\alpha$ , with  $\kappa$  being comparatively uninfluential [11,38,41]. Consequently, the correlation across the models of  $T$  for 2016–2035 under RCP8.5 and RCP4.5 with the TCR (0.69 and 0.68, respectively) is not much higher than with the ECS (0.65 for both RCPs). For 2081–2100, the correlation of  $T$  with TCR (0.79 and 0.76) is *lower* than with the ECS (0.83 and 0.80). This is because  $\kappa$  declines at different rates in the various models, in such a way that using  $\kappa_{2\times}$  degrades the correlation. By contrast, the correlation of  $T$  with  $F_{2\times}/(\alpha + \kappa_{\text{RCP}})$ , i.e. using the actual  $\kappa_{\text{RCP}}$ , is higher than with the ECS (0.84 for both RCPs).

However, the correlation of  $T$  for 2081–2100 under RCP8.5 and RCP4.5 with  $T_{140}$  ( $T$  at the time of  $4 \times \text{CO}_2$  under 1pctCO<sub>2</sub>) is greater still (0.90 and 0.85). This suggests that the decline of  $\alpha$  or the non-logarithmic increase in CO<sub>2</sub> forcing (§5) might also affect the RCP projections, and therefore  $T_{140}$  might be a better model metric than the TCR for the magnitude of climate change projected for the end of the 21st century.

## 8. Discussion and conclusion

The 1pctCO<sub>2</sub> experiment (with the CO<sub>2</sub> concentration increasing at  $1\% \text{ yr}^{-1}$ ) has been used since the early 1990s in AOGCMs to make an idealized projection of time-dependent climate change. The standard metric for model comparison of the magnitude of climate change is the TCR, defined as the global mean surface air temperature change  $T$  under this scenario at the time of  $2 \times \text{CO}_2$  (after 70 years). In all CMIP5 AOGCMs, the global warming in 1pctCO<sub>2</sub> under the second doubling, from  $2 \times \text{CO}_2$  to  $4 \times \text{CO}_2$ , is greater than the TCR, by 40% in the model mean. We describe this as an increase in the TCRP, which we define as the increase in  $T$  per unit increase in forcing (in  $\text{K W}^{-1} \text{ m}^2$ , in any scenario).

The zero-layer model of the Earth energy budget assumes that  $T$  scales with  $F$ , according to  $F = \rho T$ , where  $F$  is the effective radiative forcing and  $\rho = \alpha + \kappa$  is a constant, the sum of the climate feedback parameter  $\alpha$  and the ocean heat uptake efficiency  $\kappa$ . In this model, ocean heat uptake thus looks formally like a climate feedback; since  $\alpha$  and  $\kappa$  are both positive, the warming surface climate balances the forcing by losing heat at an increasing rate to space (through  $\alpha T$ ) and into the deep ocean (through  $\kappa T$ ). In the zero-layer model, the TCRP is a constant  $1/\rho$ . The inconstancy of TCRP implies that, although useful as an interpretative picture, the zero-layer model is not accurate. We find that the TCRP increases during the 1pctCO<sub>2</sub> experiment because both  $\alpha = (F - N)/T$  and  $\kappa = N/T$  decline, where  $N$  is the rate of heat uptake by the climate system, which is practically all stored in the ocean.

The abrupt4xCO<sub>2</sub> experiment was newly introduced in CMIP5. It has proved useful in clarifying the concepts of forcing and feedback [42], and allows the  $4 \times \text{CO}_2$  forcing to be evaluated by the regression method. Furthermore, the climate response to abrupt4xCO<sub>2</sub> can be used to estimate the response to any forcing scenario, by regarding it as a succession of step changes in forcing, and assuming that the responses to the steps combine linearly (in the sense of systems theory) [18].

Using the step model, we have shown that the decline of ocean heat uptake efficiency  $\kappa$  during 1pctCO<sub>2</sub> is a linear response. In the first couple of decades, ocean heat uptake efficiency is particularly large because the upper ocean is absorbing most of the heat, so that  $dT/dt$  is restrained by its heat capacity; this is the ‘cold start’ effect. Thereafter,  $\kappa$  continues to decline because the upper ocean loses heat into the deep ocean with decreasing efficiency as the deep ocean warms up. The thermal coupling between the upper and deep ocean, quantified by fitting the abrupt4xCO<sub>2</sub> response using a two-layer model, is strongly correlated with the zero-layer  $\kappa$ . This means that the model spread in  $\kappa$  is quantitatively accounted for by the spread in this thermal coupling. As an approximation to the two-layer model, the zero-layer model is most accurate at intermediate times, after the initial rapid warming of the upper layer, but before the deep layer has warmed substantially. Consequently, the TCR can be estimated within about 5% from  $T$  after about 20 years in the abrupt4xCO<sub>2</sub> scenario.

Explaining the spread in physical terms requires analysis of the processes of ocean vertical heat transport (advection, diffusion, etc.; e.g. [21,43]). Although the two-layer model is quite accurate and useful as an exploratory tool, it is of course a simplification; the behaviour of ocean heat uptake in three-dimensional ocean GCMs does not look like that of two discrete layers. Ocean heat uptake is often treated in simple climate models as vertical diffusion (e.g. [17,44,45]); this can reproduce AOGCM global mean heat uptake reasonably well on century time scales, but it too is not a good representation of the processes involved. Moreover, linearity is inaccurate in the ocean interior, especially in more complex scenarios [5].

In contrast to ocean heat uptake efficiency, the step model shows that the apparent decline of the climate feedback parameter  $\alpha$  during 1pctCO<sub>2</sub> is not a linear response. Using the two-layer model, we have shown that it could be explained by  $\alpha$  decreasing as CO<sub>2</sub> concentration increases, or by CO<sub>2</sub> forcing rising more rapidly with CO<sub>2</sub> concentration than the usually assumed logarithmic dependence. The physical reasons for these phenomena require analysis. Both of these possibilities have been demonstrated in some GCMs, but we cannot distinguish them with CMIP5 experiments. To meet this need, we recommend that abrupt2xCO<sub>2</sub> (instantaneously doubled CO<sub>2</sub>) experiments should be carried out as well as abrupt4xCO<sub>2</sub>, as has been proposed for the CFMIP subproject of CMIP6.

To evaluate the forcings in CMIP5 historical and RCP experiments, Forster *et al.* [38] used the method of Forster & Taylor [39], which depends on the assumption that  $\alpha$  is constant and the same for all forcing agents, and uses  $F = N + \alpha T$  with  $\alpha$  for CO<sub>2</sub> to estimate forcings due to other agents given  $N$  and  $T$  from the AOGCM. If  $\alpha$  declines with increasing CO<sub>2</sub>, this method will increasingly underestimate the AOGCM forcing. In general, if  $\alpha$  cannot be assumed to be constant, it can be computed only if  $F$  has been separately diagnosed. This is the purpose of the proposal by the Radiative Forcing Model Intercomparison Project (RFMIP) subproject of CMIP6 to perform experiments using AGCMs with prescribed sea-surface conditions and various forcings. We expect that the results will yield important new information about the dependence of  $\alpha$  on time, climate state and the nature of the forcing agent.

In the step-model estimate of 1pctCO<sub>2</sub>, the TCRP increases, but not as rapidly as it does in the AOGCMs, because the step model reproduces only the linear part of the effect, relating to ocean heat uptake, not the nonlinear part, relating to CO<sub>2</sub> forcing or climate feedback. Consequently, under 1pctCO<sub>2</sub>  $T$  warms initially more quickly in the step model than in the AOGCMs, but later on more slowly (figure 4). It is notable, however, that the step model converges on the AOGCM results at  $4 \times \text{CO}_2$ , after 140 years. This is because the step model is constructed from the AOGCM abrupt4xCO<sub>2</sub> experiment, and hence uses the AOGCM  $\alpha$  and  $F$  for  $4 \times \text{CO}_2$ , whatever the values that prevail in the AOGCM earlier in the 1pctCO<sub>2</sub> experiment. Since the step model correctly reproduces the heat uptake  $N$  as a function of  $T$  throughout the experiment, the energy budget  $F = N + \alpha T$  is satisfied at  $4 \times \text{CO}_2$  in the step model as in the AOGCM.

Long & Collins [17] calculate the ECS  $F_{2\times}/\alpha$  as a function of time in abrupt4xCO<sub>2</sub> and 1pctCO<sub>2</sub> (their fig. 2 shows MIROC-ESM as an example). (They correctly call this quantity the ‘effective climate sensitivity’, the distinction from ECS being that  $\alpha$  is estimated from a non-equilibrium state, but the definition is otherwise the same; see [46].) They show that ECS increases during

1pctCO<sub>2</sub>. This is because of the same nonlinear effects that explain the inaccuracy of the step model and the apparent decline of  $\alpha$ . ECS also increases slightly under  $4 \times \text{CO}_2$ . This is due to the tendency of  $\alpha$  to decline under fixed CO<sub>2</sub> (e.g. [25]), because of evolving patterns of sea-surface temperature or other effects leading to a nonlinear dependence of global mean radiative fluxes on global mean surface temperature change (e.g. [28,29,47]). This inconstancy of the climate feedback parameter under fixed CO<sub>2</sub> must also contribute to the increase of TCRP under 1pctCO<sub>2</sub>, but it is much less important than the other effects we have described.

The pronounced inconstancy of TCRP limits the usefulness of the TCR as a model metric, though it remains relevant as an indicator of the approximate magnitude of  $T$  in response to future anthropogenic forcing. Assuming that AOGCMs are qualitatively realistic, we expect that the TCRP of the real world also increases during time-dependent climate change forced by increasing CO<sub>2</sub>. Therefore, the TCRP inferred from observed climate change could be smaller than that applicable to future climate change [48]; this could be part of the reason for observational estimates of TCR being lower than model-based ones [2]. Furthermore, scaling the TCR by  $F/F_2$  will underestimate future  $T$  in response to higher CO<sub>2</sub> [4], though the error would be partly alleviated by using the correct dependence of  $F$  on CO<sub>2</sub>, if it is not accurately logarithmic as usually assumed. The underestimate will be more serious for stabilization scenarios, in which  $\kappa$  declines more quickly. The decline of  $\kappa$  at different rates in different models means that  $T$  in the late twenty-first century under RCP4.5 and RCP8.5 correlates less strongly with the TCR than with the ECS (which does not depend on  $\kappa$ ), and less strongly with the ECS than with  $T_{140}$  ( $T$  at the time of  $4 \times \text{CO}_2$ , at year 140, under 1pctCO<sub>2</sub>). Hence,  $T_{140}$  might be a better metric for model intercomparison than the TCR or the ECS for the expected magnitude of  $T$  in the late twenty-first century.

**Data accessibility.** The CMIP5 data analysed in this work are available through <http://pcmdi9.llnl.gov/esgf-web-fe>.

**Authors' contributions.** J.M.G. designed the study, carried out the data analysis and drafted the manuscript. T.A. and P.G. made substantial intellectual contributions to the concepts described in the manuscript and reviewed it critically. All the authors approved of the version to be published.

**Competing interests.** The authors declare that they have no competing interests.

**Funding.** The research leading to these results has received funding from the European Research Council (under the European Community's Seventh Framework Programme FP7/2007-2013, ERC grant agreement no. 247220, project 'Seachange') and the NCAS-Climate Programme for J.M.G.'s contribution, and from the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101) for all the authors.

**Acknowledgements.** We are grateful to Piers Forster, Bjorn Stevens, Nic Lewis and other colleagues at Ringberg 2015 for useful discussions, to Eric Wolff and colleagues for organizing the Royal Society discussion meeting on feedbacks on climate in the Earth system, and to Reto Knutti and an anonymous referee for their encouraging and helpful review comments. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available the output from their models (listed in table 1). For CMIP, the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

## References

1. Cubasch U, Meehl GA, Boer GJ, Stouffer RJ, Dix M, Noda A, Senior CA, Raper SCB, Yap KS. 2001 Projections of future climate change. In *Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (eds JT Houghton, Y Ding, DJ Griggs, M Noguer, P van der Linden, X Dai, K Maskell, CI Johnson), pp. 525–582. Cambridge, UK: Cambridge University Press.
2. Collins M *et al.* 2013 Long-term climate change: projections, commitments and irreversibility. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds TF Stocker *et al.*), pp. 1029–1136. Cambridge, UK: Cambridge University Press.

3. Flato G *et al.* 2013 Evaluation of climate models. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds TF Stocker *et al.*), pp. 741–866. Cambridge, UK: Cambridge University Press.
4. Gregory JM, Forster PM. 2008 Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res.* **113**, D23105. (doi:10.1029/2008JD010405)
5. Bouttes N, Gregory JM, Lowe JA. 2013 The reversibility of sea level rise. *J. Clim.* **26**, 2502–2513. (doi:10.1175/JCLI-D-12-00285.1)
6. Rhein M *et al.* 2013 Observations: ocean. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds TF Stocker *et al.*), pp. 255–316. Cambridge, UK: Cambridge University Press.
7. Church JA *et al.* 2013 Sea level change. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds TF Stocker *et al.*), pp. 1137–1216. Cambridge, UK: Cambridge University Press.
8. Palmer MD, McNeall DJ. 2013 Internal variability of Earth's energy budget simulated by CMIP5 climate models. *Environ. Res. Lett.* **9**, 034016. (doi:10.1088/1748-9326/9/3/034016)
9. Gregory JM, Mitchell JFB. 1997 The climate response to CO<sub>2</sub> of the Hadley Centre coupled AOGCM with and without flux adjustment. *Geophys. Res. Lett.* **24**, 1943–1946. (doi:10.1029/97GL01930)
10. Raper SCB, Gregory JM, Stouffer RJ. 2002 The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *J. Clim.* **15**, 124–130. (doi:10.1175/1520-0442(2002)015<0124:TROCSA>2.0.CO;2)
11. Kuhlbrodt T, Gregory JM. 2012 Ocean heat uptake and its consequences for the magnitude of sea level rise and climate change. *Geophys. Res. Lett.* **39**, L18608. (doi:10.1029/2012GL052952)
12. Held IM, Winton M, Takahashi K, Delworth T, Zeng F, Vallis GK. 2010 Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Clim.* **23**, 2418–2427. (doi:10.1175/2009JCLI3466.1)
13. Merlis TM, Held IM, Stenchikov GL, Zeng F, Horowitz LW. 2014 Constraining transient climate sensitivity using coupled climate model simulations of volcanic eruptions. *J. Clim.* **27**, 7781–7795. (doi:10.1175/JCLI-D-14-00214.1)
14. Wigley TML. 1987 Relative contributions of different trace gases to the greenhouse effect. *Clim. Monitor* **16**, 14–29.
15. Shine KP, Derwent RG, Wuebbles DJ, Morcrette JJ. 1990 Radiative forcing of climate. In *Climate change. The IPCC scientific assessment* (eds JT Houghton, GJ Jenkins, JJ Ephraums), ch. 5, pp. 45–68. Cambridge, UK: Cambridge University Press.
16. Myhre G, Highwood EJ, Shine KP, Stordal F. 1998 New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.* **25**, 2715–2718. (doi:10.1029/98GL01908)
17. Long DJ, Collins M. 2013 Quantifying global climate feedbacks, responses and forcing under abrupt and gradual CO<sub>2</sub> forcing. *Clim. Dyn.* **41**, 2471–2479. (doi:10.1007/s00382-013-1677-0)
18. Good P, Gregory JM, Lowe JA. 2011 A step-response simple climate model to reconstruct and interpret AOGCM projections. *Geophys. Res. Lett.* **38**, L01703. (doi:10.1029/2010GL045208)
19. Good P, Gregory JM, Lowe JA, Andrews T. 2013 Abrupt CO<sub>2</sub> experiments as tools for predicting and understanding CMIP5 representative concentration pathway projections. *Clim. Dyn.* **40**, 1041–1053. (doi:10.1007/s00382-012-1410-4)
20. Good P *et al.* 2015 Nonlinear regional warming with increasing CO<sub>2</sub> concentration. *Nat. Clim. Change* **5**, 138–142. (doi:10.1038/nclimate2498)
21. Gregory JM. 2000 Vertical heat transports in the ocean and their effect on time-dependent climate change. *Clim. Dyn.* **16**, 501–515. (doi:10.1007/s003820000059)
22. Geoffroy O, Saint-Martin D, Oliv   DJL, Voldoire A, Bellon G, Tyt  ca S. 2013 Transient climate response in a two-layer energy-balance model. I. Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *J. Clim.* **26**, 1841–1857. (doi:10.1175/JCLI-D-12-00195.1)
23. Dickinson RE. 1981 Convergence rate and stability of ocean–atmosphere coupling schemes with a zero-dimensional climate model. *J. Atmos. Sci.* **38**, 2112–2120. (doi:10.1175/1520-0469(1981)038<2112:CRASOO>2.0.CO;2)



24. Gregory JM, Ingram WJ, Palmer MA, Jones GS, Stott PA, Thorpe RB, Lowe JA, Johns TC, Williams KD. 2004 A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.* **31**, L03205. (doi:10.1029/2003gl018747)
25. Andrews T, Gregory JM, Webb MJ, Taylor KE. 2012 Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.* **39**, L09712. (doi:10.1029/2012GL051607)
26. Winton M, Takahashi K, Held IM. 2010 Importance of ocean heat uptake efficacy to transient climate change. *J. Clim.* **23**, 2333–2344. (doi:10.1175/2009JCLI3139.1)
27. Geoffroy O, Saint-Martin D, Bellon G, Voldoire A, Oliv   DJL, Tyt  ca S. 2013 Transient climate response in a two-layer energy-balance model. II. Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *J. Clim.* **26**, 1859–1876. (doi:10.1175/JCLI-D-12-00196.1)
28. Armour KC, Bitz CM, Roe GH. 2013 Time-varying climate sensitivity from regional feedbacks. *J. Clim.* **26**, 4518–4534. (doi:10.1175/JCLI-D-12-00544.1)
29. Andrews T, Gregory JM, Webb MJ. 2015 The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Clim.* **28**, 1630–1648. (doi:10.1175/JCLI-D-14-00545.1)
30. Hansen J *et al.* 2005 Efficacy of climate forcings. *J. Geophys. Res.* **110**, D18104. (doi:10.1029/2005JD005776)
31. Colman R, McAvaney B. 2009 Climate feedbacks under a very broad range of forcing. *Geophys. Res. Lett.* **36**, L01702. (doi:10.1029/2008GL036268)
32. Jonko AK, Shell KM, Sanderson BM, Danabasoglu G. 2012 Climate feedbacks in CCSM3 under changing CO<sub>2</sub> forcing. II. Variation of climate feedbacks and sensitivity with forcing. *J. Clim.* **26**, 2784–2795. (doi:10.1175/JCLI-D-12-00479.1)
33. Caballero R, Huber M. 2013 State-dependent climate sensitivity in past warm climate and its implications for future climate projections. *Proc. Natl Acad. Sci. USA* **110**, 14 162–14 167. (doi:pnas.1303365110)
34. Byrne B, Goldblatt C. 2014 Radiative forcing at high concentrations of well-mixed greenhouse gases. *Geophys. Res. Lett.* **41**, 152–160. (doi:10.1002/2013GL058456)
35. Giorgetta MA *et al.* 2013 Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project Phase 5. *J. Adv. Model Earth Syst.* **5**, 572–597. (doi:10.1002/jame.20038)
36. Hasselmann K, Sausen R, Maier-Reimer E, Voss R. 1993 On the cold start problem in transient simulations with coupled atmosphere–ocean models. *Clim. Dyn.* **9**, 53–61. (doi:10.1007/BF00210008)
37. Keen AB, Murphy JM. 1997 Influence of natural variability and the cold start problem on the simulated transient response to increasing CO<sub>2</sub>. *Clim. Dyn.* **13**, 847–864. (doi:10.1007/s003820050201)
38. Forster PM, Andrews T, Good P, Gregory JM, Jackson LS, Zelinka M. 2013 Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.* **118**, 1–12. (doi:10.1002/jgrd.50174)
39. Forster PMDF, Taylor KE. 2006 Climate forcings and climate sensitivities diagnosed from coupled climate model integrations. *J. Clim.* **19**, 6181–6194. (doi:10.1175/JCLI3974.1)
40. Andrews T. 2014 Using an AGCM to diagnose historical effective radiative forcing and mechanisms of recent decadal climate change. *J. Clim.* **27**, 1193–1209. (doi:10.1175/JCLI-D-13-00336.1)
41. Tomassini L, Geoffroy O, Dufresne JL, Idelkadi A, Cagnazzo C, Block K, Mauritsen T, Giorgetta M, Quaas J. 2013 The respective roles of surface temperature driven feedbacks and tropospheric adjustment to CO<sub>2</sub> in CMIP5 transient climate simulations. *Clim. Dyn.* **41**, 3103–3126. (doi:10.1007/s00382-013-1682-3)
42. Sherwood S, Bony S, Boucher O, Bretherton C, Forster P, Gregory J, Stevens B. 2015 Adjustments in the forcing-feedback framework for understanding climate change. *Bull. Am. Meteorol. Soc.* **96**, 217–228. (doi:10.1175/BAMS-D-13-00167.1)
43. Exarchou E, Kuhlbrodt T, Gregory JM, Smith RS. 2015 Ocean heat uptake processes: a model intercomparison. *J. Clim.* **28**, 887–908. (doi:10.1175/jcli-d-14-00235.1)
44. Sokolov AP, Forest CE, Stone PH. 2003 Comparing oceanic heat uptake in AOGCM transient climate change experiments. *J. Clim.* **16**, 1573–1582. (doi:10.1175/1520-0442-16.10.1573)

45. Huntingford C, Cox PM. 2000 An analogue model to derive additional climate change scenarios from existing GCM simulations. *Clim. Dyn.* **16**, 575–586. (doi:10.1007/s003820000067)
46. Planton S. 2013 Appendix III: glossary. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds TF Stocker *et al.*), p. 1451. Cambridge, UK: Cambridge University Press. (doi:10.1017/CBO9781107415324.031)
47. Meraner K, Mauritsen T, Voigt A. 2013 Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.* **40**, 5944–5948. (doi:10.1002/2013GL058118)
48. Bloch-Johnson J, Pierrehumbert RT, Abbot D. 2015 Feedback temperature dependence determines the risk of high warming. *Geophys. Res. Lett.* **42**, 4973–4980. (doi:10.1002/2015GL064240)