Manuscript accepted in *Geophysical Research Letters*

Drivers of continued surface warming after cessation of carbon emissions

Richard G. Williams[1], Vassil Roussenov[1], Thomas L. Frölicher[2,3] and Philip Goodwin[4]

[1] Department of Earth, Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, L69 3GP, UK.

[2] Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland.

[3] Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.

[4] School of Ocean and Earth Sciences, University of Southampton, Southampton, UK.

Corresponding author: Richard Williams (ric@liverpool.ac.uk)

Key Points:

- Continued warming explained using theory from a heat balance and a carbon inventory budget
- Peak global warming delayed after the cessation of carbon emissions
- Timing of peak warming controlled by the weakening of ocean heat uptake

Abstract

The climate response after cessation of carbon emissions is examined here, exploiting a single equation connecting surface warming to cumulative carbon emissions. The multi-centennial response to an idealized pulse of carbon is considered by diagnosing a 1000 year integration of an Earth system model (GFDL ESM2M) and an ensemble of efficient Earth system model simulations. After emissions cease, surface temperature evolves according to (i) how much of the emitted carbon remains in the atmosphere and (ii) how much of the additional radiative forcing warms the surface rather than the ocean interior. The peak in surface temperature is delayed in time after carbon emissions cease through the decline in ocean heat uptake, which in turn increases the proportion of radiative forcing warming the surface. Eventually, after many centuries, surface temperature declines as the radiative forcing decreases through the excess atmospheric CO_2 being taken up by the ocean and land.

1 Introduction

There are a wide range of climate projections for our future climate, usually based upon the surface temperature change for a cumulative carbon emission [*Allen et al.,* 2009; *Matthews et al.,* 2009; *Zickfeld et al.,* 2009; *Gillet et al.,* 2013; *Collins et al.,* 2013]. This connection provides the basis of forming budgets of the maximum carbon emission compatible with surface warming targets [*Meinshausen et al.,* 2009; *Matthews et al.,* 2012; *Collins et al.,* 2013]. However, it is currently unclear how the climate system responds when carbon emissions cease, either with cumulative carbon emission remaining at a maximum value or decreasing in time with carbon capture. The surface temperature change is either viewed as reaching a maximum close to when carbon emissions cease [*Ricke and Caldeira*, 2014], or remains nearly constant [*Matthews and Caldeira*, 2008; *Gillet et al.,* 2011] or surface temperature continues to increase [*Plattner et al.,* 2008; *Frölicher et al.,* 2014; *Ehlert and Zickfeld*, 2017] over several centuries. In addition, there may be climate effects occurring after the surface temperature reaches a maximum, involving continued ocean warming and associated sea level rise [*Plattner et al.,* 2008; *Frölicher and Joos*, 2010; *Gillet et al.,* 2011]. The drivers of the climate adjustment after emissions cease are unclear, although there are heuristic arguments of a partial compensation between the decline in ocean heat uptake and the reduced radiative forcing from ocean and terrestrial uptake of atmospheric CO_2 [*Solomon et al*., 2009].

In this study, we apply a theoretical relationship between global-mean surface warming and cumulative carbon emissions to understand the surface warming response on a multi-centennial timescale [*Goodwin et al*., 2015; *Williams et al.,* 2016]. The surface warming response is considered on a timescale of many centuries after carbon emissions cease. During this period, carbon is exchanged between the atmosphere, ocean and terrestrial biosphere; our analysis ignores sediment and weathering interactions that become significant on longer multi-millennial timescales [*Archer et al.,* 2009].

Our theory connecting the change in global-mean surface temperature and the cumulative carbon emission is next set out (Section 2). The theory is then applied to understand the surface warming response to a 1000 year long Earth system model experiment, examining the response to carbon emissions restricted to within 100 years [*Frölicher and Paynter*, 2015] (Section 3). The surface warming response is interpreted in terms of changes in atmospheric CO_2 linked to changes in the global carbon inventories and changes in the relationship between surface warming and radiative forcing. The surface warming response from the single Earth system model is compared with the response from a large ensemble of projections from an efficient Earth system model [*Goodwin,* 2016; *Goodwin et al.,* 2017] (Section 4). Finally, the key outcomes of the study are summarized (Section 5).

2. Theoretical context

If changes in radiative forcing are only driven by changes in atmospheric CO_2, the global-mean change in surface air temperature, $\Delta T(t)$ (in K), may be connected via a single equation to cumulative carbon emissions [*Goodwin et al.*, 2015], by assuming an empirical surface heat budget and combined with changes in global carbon inventories:

$$\Delta T(t) = \frac{a}{\lambda}\left(1 - \frac{\varepsilon(t)N(t)}{R(t)}\right)\left(\frac{I_{em}(t) - \Delta I_{ter}(t) + I_{Usat}(t)}{I_B}\right). \qquad (1)$$

Here, the surface temperature change, $\Delta T(t)$, may be interpreted in terms of the product of two sets of non-dimensional time-dependent terms (each set enclosed in the brackets). The first bracket represents the fraction of radiative forcing used to warm the surface, involving the radiative forcing into the climate system, $R(t)$ (in Wm^{-2}), the planetary heat uptake, $N(t)$ (in Wm^{-2}), effectively from ocean heat uptake, and a non-dimensional weighting of the heat uptake, $\varepsilon(t)$ [*Winton et al.,* 2010]. The second bracket represents a carbon budget related to the change in atmospheric CO_2 [*Goodwin et al.,* 2015], involving the cumulative carbon emissions, $I_{em}(t)$, the change in the terrestrial carbon inventory, $\Delta I_{ter}(t)$, and the carbon

under-saturation of the global ocean, $I_{Usat}(t)$, and modulated by the buffered atmosphere and the ocean carbon inventory, I_B (all in PgC) [*Goodwin et al.,* 2007]. The change since the preindustrial is denoted by Δ. The magnitude of the surface temperature change is then obtained by the product of these two sets of terms multiplied by the radiative forcing coefficient for atmospheric CO_2, a (in Wm^{-2}) [*Forster et al.,* 2013] and divided by the climate feedback parameter, λ (in $(Wm^{-2})K^{-1}$) [*Knutti and Hegerl*, 2008; *Gregory et al.,* 2004].

If cumulative carbon emissions reach a maximum value, $I_{em}(t_{cease})$ at a time t_{cease}, and then remain unchanged, the surface temperature change may continue to increase or decrease, based upon the evolution of the weighted ocean heat uptake, $\varepsilon(t)N(t)$, the radiative forcing since the preindustrial, $R(t)$, the ocean undersaturation of carbon, $I_{Usat}(t)$ and the change in the terrestrial carbon sink, $\Delta I_{ter}(t)$:

$$\Delta T(t) = \frac{a}{\lambda}\left(1 - \frac{\varepsilon(t)N(t)}{R(t)}\right)\left(\frac{I_{em}(t_{cease}) - \Delta I_{ter}(t) + I_{Usat}(t)}{I_B}\right). \qquad (2)$$

Surface temperature either (i) increases in time through a decline in the ratio of the weighted ocean heat uptake and radiative forcing, $\varepsilon(t)N(t)/R(t)$, or (ii) declines in time through the atmospheric CO_2 decreasing from an increase in the terrestrial carbon sink, $\Delta I_{ter}(t)$, or a decrease in the ocean undersaturation of carbon, $I_{Usat}(t)$. Thus, two sets of terms representing the thermal and carbon response of the system act in partial opposition, as argued heuristically by *Solomon et al.* [2009].

Eventually the ocean approaches a long-term equilibrium with the atmosphere for both heat and carbon at a later time $t_{equilib}$. This long-term equilibrium state is obtained when both the ocean heat uptake, $N(t)$, and the global carbon undersaturation, $I_{Usat}(t)$, asymptote to zero in (2). For this equilibrium state (prior to any feedback from ocean sediments and weathering), the change in surface temperature is then given by the net cumulative carbon emissions to the atmosphere and ocean, $I_{em}(t_{cease}) - \Delta I_{ter}(t_{equilib})$, multiplied by the long-term equilibrium climate response to carbon emissions, $a/(\lambda I_B)$ [*Williams et al.,* 2012],

$$\Delta T(t \to t_{equilib}) \approx \frac{a}{\lambda I_B}\left(I_{em}(t_{cease}) - \Delta I_{ter}(t_{equilib})\right). \qquad (3)$$

3. Model analyses

Our theory is now exploited to understand the surface warming response of a 1000 year integration of fully coupled carbon cycle-climate Earth system model.

3.1 Model formulation and surface warming simulation

We use output from a 1000 year simulation of the global coupled carbon-climate Earth System Model developed at the Geophysical Fluid Dynamics Laboratory (GFDL ESM2M) [*Dunne et al.*, 2012; *Dunne et al.,* 2013]. The physical core of the model is an updated version of the CM2.1 [*Delworth et al.*, 2006]. The atmospheric model (AM2) has a horizontal resolution of approximately 2° and 24 vertical levels. The ocean model (MOM4p1) consists of 50 vertical levels and has a horizontal resolution of 1° or less. The ocean biogeochemical component (TOPAZv2) includes 30 tracers to represent cycles of carbon, oxygen, and the major macronutrients and iron. The land model (LM3.0) has five different vegetation pools and two soil carbon pools, and changes in these pools are simulated through phenology, natural mortality and fire.

Following *Frölicher and Paynter* [2015], an idealised global warming simulation is examined where the model is forced by prescribed atmospheric CO_2 increasing at an annual rate of 1% from the preindustrial 286 ppm to 745 ppm until global-mean surface air temperature increases by 2°C since the preindustrial (occurring in year 99 of the simulation). After that, emissions of carbon are set to zero. All other non-CO_2 greenhouse gases are kept at preindustrial levels.

3.2 Diagnostic approach

Based on our theory (1) [*Goodwin et al.,* 2015; *Williams et al.,* 2016], we diagnose the surface warming response in terms of changes in global carbon inventories and an empirical heat budget.

3.2.1 Carbon inventory diagnostics

Our theory relies upon a global carbon inventory holding such that a cumulative carbon emission, $I_{em}(t)$, drives a change in the carbon inventories,

$$I_{em}(t) = \Delta I_{atm}(t) + \Delta I_{ter}(t) + \Delta I_{ocean}(t), \quad (4)$$

where $\Delta I_{atm}(t)$, $\Delta I_{ter}(t)$ and $\Delta I_{ocean}(t)$ are the changes in the atmospheric, terrestrial and ocean carbon inventories since the preindustrial (all defined in PgC). Cumulative carbon emissions compatible with the prescribed atmospheric CO_2 concentration are diagnosed from the sum of changes in the prescribed atmospheric CO_2 inventory and the simulated time-integrated atmosphere-ocean and atmosphere-land carbon fluxes [*Jones et al.,* 2013; *Frölicher and Paynter*, 2015].

The ocean carbon inventory is held as dissolved inorganic carbon, $\Delta I_{ocean}(t) = V\Delta DIC(t)$, here V is the global volume (m^3) and $DIC(t)$ is the global volume-weighted dissolved inorganic carbon concentration. A saturated component, $\Delta C_{sat}(t)$, of the $DIC(t)$ is diagnosed (e.g. *Lauderdale et al.* [2013]) based upon its potential temperature, salinity and alkalinity, and instantaneous atmospheric $CO_2(t)$. The carbon undersaturation of the global ocean, I_{Usat}, is then defined by $I_{Usat}(t) = V(\Delta C_{sat}(t) - \Delta DIC(t))$ [*Goodwin et al.,* 2015]. Our theory for the carbon inventory changes uses the buffered atmosphere and ocean carbon inventory, $I_B = I_{atmos} + VC_{sat} / B$ [*Goodwin et al.,* 2007], where I_{atmos} is the atmospheric carbon inventory, V is the ocean volume, C_{sat} is the saturated dissolved inorganic carbon (mol m^{-3}) and B is the buffer or Revelle factor [*Williams and Follows*, 2011]; for ΔI_{ocean}, I_{Usat} and I_B to be plotted in gC, there is a further multiplication of 12 g mol^{-1}.

3.2.2 Radiative forcing and empirical heat budget diagnostics

The radiative forcing, $R(t)$ in Wm^{-2}, is diagnosed from the increase in atmospheric CO_2 since the preindustrial [*Myhre et al.,* 1998],

$$R(t) = a\Delta \ln CO_2(t), \quad (5)$$

where a=4.85 Wm^{-2} [*Forster et al.,* 2013], and $\Delta \ln CO_2(t) = \ln(CO_2(t)) - \ln(CO_2(t_o))$ with the preindustrial $CO_2(t_o)$ take as 286 ppm, and a positive R represents the heat input into the climate system since the preindustrial.

The climate feedback parameter, λ in $Wm^{-2}K^{-1}$, is diagnosed from a long-term empirical heat budget,

$$R(t) = \lambda \Delta T(t), \quad (6)$$

with a value of λ=1.01 $Wm^{-2}K^{-1}$ using a least-squares regression using annual-mean data from years 800 to 1000 of the 1000-yr simulation [*Gregory and Forster,* 2008]. On shorter timescales, the radiative forcing drives a planetary heat uptake, $N(t)$ in Wm^{-2}, [*Gregory et al.,* 2004], which is dominated by the ocean heat uptake, such that

$$R(t) = \lambda \Delta T(t) - \varepsilon(t) N(t), \quad (7)$$

where $\varepsilon(t)$ is a non-dimensional weighting of ocean heat uptake, referred to as the efficacy [*Winton et al.,* 2010, 2013; *Paynter and Frölicher*, 2015]. The product $\varepsilon(t)N(t)$ is diagnosed from the mismatch of $R(t)$ and $\lambda\Delta T(t)$. Alternatively, the radiative forcing may be assumed to drive a surface warming and planetary heat uptake with a time-varying climate feedback parameter, $\alpha(t)$, such that

$$R(t) = \alpha(t)\Delta T(t) - N(t), \qquad (8)$$

where $N(t)$ is taken from the tendency in the global ocean heat content; see Fig. S1 for diagnostics of both these equivalent empirical heat budgets together with the time variation in $\varepsilon(t)$ and $\alpha(t)$.

3.3 Analyses of a 1000 year Earth system model integration

The global warming response is interpreted in terms of the carbon system response, the radiative heat response and their combined effect.

3.3.1 Carbon system response

The carbon emissions drive an increase in atmospheric CO_2 from preindustrial 286 to 745 ppm (black line in Fig. 1a), and an increase in the terrestrial (green line in Fig. 1b) and ocean carbon inventories (blue line in Fig. 1b). After emissions cease, the terrestrial and ocean systems initially both take up atmospheric CO_2, but eventually the terrestrial uptake weakens and the ocean provides the only net carbon sink after 200 years (Fig. 1b,c). The terrestrial carbon uptake is controlled by the land biosphere involving a competition between primary production, respiration and disturbances from fire. By the end of the simulation, there is a small overall loss of carbon from the land biosphere due to the carbon loss from fire emissions dominating over the carbon gain due to elevated net ecosystem production.

Over the 1000 year integration, there is an overall decline in the air-borne fraction, $\Delta I_{atm}/I_{em}$ (black line in Fig. 1d), an increase in the ocean-borne fraction, $\Delta I_{ocean}/I_{em}$ (blue line in Fig. 1d) and an initial increase and then a later decline in the land-borne fraction, $\Delta I_{ter}/I_{em}$ (green line in Fig. 1d) [*Frölicher and Paynter*, 2015].

The change in atmospheric CO_2 may be directly connected to the net cumulative carbon emission to the combined atmosphere and ocean, $I_{em}(t) - \Delta I_{ter}(t)$, plus a term measuring the carbon undersaturation of the global ocean, $I_{Usat}(t)$ [*Goodwin et al.,* 2015], such that

$$\Delta \ln CO_2(t) = \frac{1}{I_B}\left(I_{em}(t) - \Delta I_{ter}(t) + I_{Usat}(t)\right), \qquad (9)$$

where $\Delta I_{ter}(t)$ represents the change in the terrestrial carbon inventory and I_B is the buffered atmosphere and ocean carbon inventory, measuring the available carbon in the combined atmosphere and ocean taking into account carbonate buffering, and its pre-industrial value is 3487 PgC [*Goodwin et al.,* 2007; *Williams et al.,* 2017]. Our theoretical prediction for atmospheric CO_2 from (9) agrees well with the actual model response (Fig. 1a) over the first 400 years. There is a subsequent misfit due to the buffered carbon inventory, $I_B = I_{atmos} + VC_{sat}/B$, increasing from its pre-industrial value of 3487 PgC to 4054 PgC over the last 200 years of the record; this increase in I_B is from the effect of the rise in I_{atmos} and C_{sat} exceeding the effect of the rise in B due to the large cumulative carbon emission of nearly 2000 PgC [*Goodwin et al.*, 2007].

The atmospheric CO_2 response in (9) is primarily determined by the increase in cumulative carbon emissions, $I_{em}(t)$, and the change in the carbon undersaturation of the global ocean, $I_{Usat}(t)$ (Fig. 1b, red & grey dashed lines). Over the integration, the dissolved inorganic carbon, $\Delta DIC(t)$, increases as the ocean takes up the additional anthropogenic carbon added to the climate system (Fig. 1c, black line). At the same time, the saturated carbon in the ocean, $\Delta C_{sat}(t)$ (Fig. 1c, blue line), and atmospheric CO_2 increase together during the period of ongoing carbon emissions and then decrease after carbon emissions cease. The difference between the saturated carbon and dissolved inorganic carbon in the ocean, $\Delta C_{sat}(t) - \Delta DIC(t)$, increases rapidly during emissions and then strongly declines as the ocean takes up more carbon

(Fig. 1c, dashed grey line). Accordingly, the global ocean carbon undersaturation, $I_{Usat}(t)=V(\Delta C_{sat}(t)-\Delta DIC(t))$, declines after emissions cease (Fig. 1b, dashed grey line). The normalized fraction, $I_{Usat}(t)/I_{em}(t)$, also strongly declines over the integration (Fig. 1d, dashed grey line), which is due to the increasing effectiveness of the ocean and terrestrial system in taking up carbon on a multi-centennial timescale; similar responses for $I_{Usat}(t)/I_{em}(t)$ also occur in climate projections up to year 2100 from 11 different Earth system models [*Williams et al.,* 2017].

3.3.2 Surface warming response

In the model experiment, carbon emissions are chosen to cease when surface temperature increases by 2K relative to the preindustrial (this transition occurs at 98 years, $t=t_{cease}$) [*Frölicher and Paynter*, 2015]. After this period, there is still a further increase in surface temperature, rising typically by 0.6K over the subsequent 600 years (Fig. 2a).

To understand this delayed warming, we consider the empirical radiative heat budget (7). The increase in radiative forcing, $R(t)$, follows the rise in atmospheric CO_2, peaking at 4.5 Wm^{-2} at year 98 and then declining to 2.5 Wm^{-2} by year 1000 (Fig. 2b, black line). This input of heat drives both a surface climate response, $\lambda\Delta T(t)$ (blue line in Fig. 2b), and a weighted planetary heat uptake, $\varepsilon(t)N(t)$, dominated by the ocean heat uptake (red line in Fig. 2b). Initially a larger fraction of the radiative forcing drives ocean heat uptake and a warming of the ocean interior, while only the smaller remaining fraction drives surface warming. Eventually the ocean heat uptake declines and a larger fraction of the radiative forcing then drives a stronger surface warming. This decrease in the ocean heat uptake, $N(t)$ (Fig. 2c, purple line) is evident in the decrease in the rate of temperature rise for the bulk ocean (Figs. 2a, dashed line).

Our theory in equation (1) provides a way to understand the surface warming response to carbon emissions [*Goodwin et al.,* 2015]; this equation is based on combining the relationship between radiative forcing and the logarithm of atmospheric CO_2 (5), the empirical heat budget (7) and the relationship between the logarithm of atmospheric CO_2 and the carbon inventory changes (9). The theoretical prediction of surface temperature from (1, blue line in Fig. 2a) closely matches the actual modeled surface temperature for the first 400 years (black line Fig. 2a), then the theoretical prediction exceeds the model response (due to the overestimate of atmospheric CO_2 over the latter part of the record; see Fig 1a).

The theoretical prediction in (1) is made up of the product of two non-dimensional sets of time-dependent terms: the fraction of the radiative forcing used for surface warming, $(1-\varepsilon(t)N(t)/R(t))$, and the change in the logarithm of atmospheric CO_2, $(I_{em}(t)-\Delta I_{ter}(t)+I_{Usat}(t))/I_B$. The contribution for atmospheric CO_2 peaks when the cumulative emissions reach a maximum and then slightly declines (Fig. 2c, blue line). In contrast, the fraction of radiative forcing used for surface warming initially declines while cumulative carbon emissions increase, but then progressively increases towards 1 (Fig. 2c, red line). Thus, the delay in surface warming after emission peak is due to the decline in ocean heat uptake; in accord with the model diagnostics by *Frölicher et al.* [2014].

3.3.3 Transient climate response to carbon emissions

Climate projections are often understood in terms of the transient climate response to cumulative CO_2 emissions (TCRE), which is often found to be nearly constant for individual Earth system models [*Allen et al.,* 2009; *Matthews et al.,* 2009; *Zickfeld et al.,* 2009; *Gillet et al.,* 2013]. The TCRE, defined here by the ratio of the surface air temperature rise, $\Delta T(t)$, and the cumulative carbon emission, $I_{em}(t)$, is equivalent to the product of the surface warming dependence on radiative forcing, $\Delta T/R$, and the radiative forcing dependence of cumulative carbon emissions, R/I_{em} [*Williams et al.,* 2016]:

$$\mathrm{TCRE}(t) \equiv \frac{\Delta T(t)}{I_{em}(t)} = \left(\frac{\Delta T(t)}{R(t)}\right)\left(\frac{R(t)}{I_{em}(t)}\right), \qquad (10)$$

which may be written from (1) [*Goodwin et al.,* 2015] as

$$\mathrm{TCRE}(t) \equiv \frac{\Delta T(t)}{I_{em}(t)} = \left\{\frac{1}{\lambda}\left(1 - \frac{\varepsilon(t)N(t)}{R(t)}\right)\right\}\left\{\frac{a}{I_B}\left(\frac{I_{em}(t) - \Delta I_{ter}(t) + I_{Usat}(t)}{I_{em}(t)}\right)\right\}. \qquad (11)$$

The TCRE varies over the 1000 year integration: decreasing from an initial value of typically 1.2 K$(10^3\mathrm{PgC})^{-1}$ to close to 1K$(1000\mathrm{PgC})^{-1}$ when emissions cease and then increases towards 1.3K$(10^3\mathrm{PgC})^{-1}$ (Fig. 2d, black line), which is close to the equilibrium value expected from $a/(\lambda I_B)$ = 1.37 K $(10^3\mathrm{PgC})^{-1}$ (Fig. 2d, blue dashed line) [*Williams et al.,* 2012]. There is significant interannual and decadal variability throughout the record.

The dependence of radiative forcing on carbon emissions, R/I_{em}, continually declines from a peak value of 4 $\mathrm{Wm}^{-2}(10^3\mathrm{PgC})^{-1}$ to a value of 1.4 $\mathrm{Wm}^{-2}(10^3\mathrm{PgC})^{-1}$ by year 1000 (Fig. 3b). This evolution follows the response of the normalised carbon undersaturation of the global ocean, $I_{Usat}(t)/I_{em}(t)$ (Fig. 1d, grey dashed line). The carbon system is not yet at an equilibrium after year 1000 (as evident from the ratio of $I_{Usat}(t)/I_{em}(t)$ being non zero). The dependence of surface warming on radiative forcing, $\Delta T / R$, is the same as the non-dimensional term $(1-\varepsilon(t)N(t)/R(t))$ modulated by λ^{-1} (Fig. 3c). The surface warming dependence on radiative forcing increases in time as the weighted ocean heat uptake, $\varepsilon(t)N(t)$, declines in time. The thermal system is close to an equilibrium after year 1000, as evident from $\varepsilon(t)N(t)$ being close to zero.

Hence, there are changes in the TCRE with time, but these changes are relatively small compared with the changes in the separate contributions linked to changes in the surface warming dependence on radiative forcing, $\Delta T/R$, and the radiative forcing dependence on cumulative carbon emissions, R/I_{em}. In comparison, there are broadly similar changes in time for climate projections to year 2100 in an ensemble of 11 Earth system models [*Williams et al.*, 2017], although there are significant inter-model differences in the magnitudes of the TCRE and its dependences for the individual models.

4 Uncertainty analysis of an efficient Earth system model

To provide a wider context to the Earth system model projection, we now compare the GFDL model response to a large ensemble of projections from an efficient Earth system model, designed to be consistent with observations.

4.1 Formulation of the efficient Earth system model

The Warming Acidification and Sea level Projector (WASP) model [*Goodwin*, 2016] is used to consider uncertainties. A large number of initial ensemble configurations (10^8) are generated using a Monte Carlo approach, with each configuration containing a unique combination of 18 parameter values for quantities including a, λ and I_B (Supplementary Information). Each initial configuration undergoes historical forcing from year 1765 to 2016, and is then assessed against observations for surface warming, ocean heat content change and carbon inventory changes [*Goodwin*, 2016; *Goodwin et al*., 2017]. A total of 10,470 simulations are found to be observationally consistent. These observationally consistent ensemble-members are then re-initialised to a preindustrial state and forced with a 1% annual rise in atmospheric CO_2 via carbon emissions into the atmosphere. As each simulation reaches a global surface temperature anomaly of +2.0K the carbon emissions cease, and atmospheric CO_2 is allowed to vary according to the subsequent carbon exchanges between the atmosphere, ocean and terrestrial systems. Each separate simulation is integrated for 1000 years.

4.2 Response of the efficient Earth system model

The observation-consistent ensemble of model simulations reveals wide variation in how the TCRE and its separate dependences for surface warming on radiative forcing and radiative forcing on carbon emissions (Fig. 3, grey shading). The GFDL model response after emissions ceases generally lies within the responses of the efficient model ensemble (Fig. 3, black line).

Theory suggests that the TCRE, $\Delta T/I_{em}$, asymptotes towards the model value of $a/(\lambda I_B)$ in (2), while $\Delta T/R$ approaches the model value of λ^{-1} and R/I_{em} approaches the model value of a/I_B. In the large WASP ensemble, each simulation has a unique combination of a, I_B, and λ values [*Goodwin* 2016; *Goodwin et al*, 2017]. The simulated TCRE for the GFDL model in the 1000-year integration remains slightly below the value of $a/(\lambda I_B)$ of 1.37 K $(1000\text{PgC})^{-1}$ (Fig. 2d).

The GFDL simulation for the TCRE, $\Delta T/I_{em}$, lies broadly centrally within the WASP ensemble (Fig. 3a) (at the 54th percentile of the WASP ensemble from year 900 to 1000). This agreement is due to the GFDL model having a relatively low value of a/I_B offsetting a high value of λ^{-1} compared with the WASP ensemble (Fig. 3b,c); the GFDL model has a=4.85Wm^{-2} and I_B=3487PgC, placing the GFDL model value of a/I_B on the 13th percentile of the WASP ensemble, while the GFDL model has a value of λ of 1.01 Wm^{-2}K^{-1}, placing the GFDL model value of λ^{-1} on the 87th percentile of the WASP ensemble.

In the earlier stages of the integration, from years 50 to 300, the GFDL simulation of the TCRE, $\Delta T/I_{em}$, is lower than both the values of $a/(\lambda I_B)$ and the response of the WASP ensemble (Fig. 3a). This response is due to the relatively high value of ocean heat uptake, $N(t)$, in the GFDL model, which then leads to a relatively low values for both the TCRE, $\Delta T/I_{em}$ and $\Delta T/R$ during this period (Fig. 3a,c).

The different transient evolution of the TCRE in the GFDL model, relative to the WASP ensemble (Fig. 3a), is thus caused by differences between the transient heat uptake, $N(t)$, responses of the two models, since the GFDL model carbon reservoirs behavior is consistent within the WASP model ensemble (Fig. 3b). This comparison highlights how the theory provides a process-based way to identify the factors determining how the TCRE evolves over time.

5. Conclusions

The multi-centennial climate response after cessation of carbon emissions may be understood using theory connecting surface warming to cumulative carbon emissions [*Goodwin et al.,* 2015; *Williams et al.,* 2016], involving an empirical heat budget and global changes in carbon inventories. Our theory is compared with climate model experiments using an integration [*Frölicher and Paynter*, 2015] of an Earth system model (GFDL ESM2M) and a large ensemble (10^4) of efficient Earth system models simulations [*Goodwin*, 2016]. The surface temperature responses involves two competing contributions:
1. The surface temperature continues to increase even after a peak in atmospheric CO_2 through a decline in the fraction of heat taken up by the ocean interior, $\varepsilon(t)N(t)/R(t)$, which increases the fraction of the radiative forcing used to drive surface warming;
2. The surface temperature only decreases when there is a sufficient decrease in atmospheric CO_2, achieved either through an increase in the terrestrial carbon inventory, $\Delta I_{ter}(t)$, or a decline in the carbon undersaturation of the global ocean, $I_{Usat}(t)$.

The surface warming reaches a peak value that occurs later than the cessation of emissions due to the decline in ocean heat uptake, which initially dominates over the opposing effect of enhanced carbon uptake from the combined ocean and terrestrial system. On a millennial timescale, prior to any sediment and weathering feedback, the final surface warming is determined by the cumulative carbon emission to the combined atmosphere and ocean multiplied by a term depending upon 3 climate parameters, $a/(\lambda I_B)$ [*Williams et al.,* 2012].

This study reveals how the theory may be used to understand the response of Earth system models. Within our theory, these inter-model differences are represented by differences in ocean heat and carbon uptake, terrestrial carbon cycling, the climate sensitivity and the efficacy, representing the non-dimensional weighting of heat uptake. Time dependence in the feedbacks [*Knutti and Rugenstein*, 2017] are likely to lead to the climate sensitivity or efficacy evolving in time, such as in how climate sensitivity may vary through regional feedbacks [*Armour et al.,* 2013] and in how the efficacy may alter with changes in ocean circulation [*Winton et al.,* 2013; *Rose and Rayborn*, 2016].

The delayed warming after carbon emissions cease investigated here provides an additional challenge as to how policymakers should approach restricting warming to pre-defined targets, such as the Paris Agreement 1.5 and 2.0 K targets above preindustrial [*Meinshausen et al.,* 2009; *Matthews et al.,* 2012; *Ehlert and Zickfeld,* 2017]. Policy makers are usually focusing on the risks of climate hazards over the next century, but there are longer-term climate threats [*Friedlingstein et al*., 2011; *Frölicher et al.,* 2014] from continued surface warming after carbon emissions cease. The time-dependent drivers of these climate threats after emissions cease are then primarily related to the long-term climate impacts of how the ocean takes up heat and carbon.

Acknowledgments

RGW, VR and PG acknowledge support from the UK Natural Environmental Research Council, NE/N009789/1, PG also acknowledges support from UK NERC grant NE/P01495X/1. TLF acknowledges the financial support provided the Swiss National Science Foundation grant PP00P2_170687. We thank two anonymous referees for their constructive feedback. This study provides analyses of climate model experiments and the basic model data sets are available via the British Ocean Data Centre (enquiries@bodc.ac.uk) and the methodology to produce the derived products are described here.

References

Allen, M. R., D. J. Frame, C. Huntingford, C. D. Jones, J. A. Lowe, M. Meinshausen and N. Meishausen (2009), Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, 458, 1163–1166.

Archer, D., M. Eby, V. Brovkin, A. Ridgwell, L. Cao, U. Mikolajewicz, K. Caldeira, K. Matsumoto, G. Munhoven, A. Montenegro and K. Tokos (2009), Atmospheric lifetime of fossil fuel carbon dioxide. *Annu. Rev. Earth Planet. Sci.,* 37, 117-134, doi: 10.1146/annurev.earth.031208.100206.

Armour, K. C., C. M. Bitz and G. H. Roe (2013), Time-varying climate sensitivity from regional feedbacks. *J. Climate*, 26, 4518-4534, doi: 10.1175/JCLI-D-12-00544.1.

Collins M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski *et al.* (2013). *Chapter 12 - Long-term climate change: Projections, commitments and irreversibility.* In: *Climate Change 2013: The Physical Science Basis. IPCC Working Group I Contribution to AR5.* T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, UK and New York, USA.

Delworth, T. L. *et al.* (2006), GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics, *J. Clim.*, *19*(5), 643–674, doi:10.1175/JCLI3629.1.

Dunne, J. P. *et al.* (2012), GFDL's ESM2 Global Coupled Climate-Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics, *J. Clim.*, *25*, 6646–6665, doi:10.1175/JCLI-D-11-00560.1.

Dunne, J.P., J.G. John, E. Shevliakova, R.J. Stouffer, J.P. Krasting, S.L. Malyshev, P.C. Milly, L.T. Sentman, A.J. Adcroft, W. Cooke, K.A. Dunne, S.M. Griffies, R.W. Hallberg, M.J. Harrison, H. Levy, A.T. Wittenberg, P.J. Phillips, and N. Zadeh (2013), GFDL's ESM2 Global coupled climate-carbon Earth system model. Part II: Carbon system formulation and baseline simulation

characteristics. *J. Clim.,* **26**, 2247–2267, doi:10.1175/JCLI-D-12-00150.1.

Ehlert, D. and K. Zickfeld (2017), What determines the warming commitment after cessation of CO2 emissions? *Environ. Res. Lett.*, 12, 015002, doi:10.1088/1748-9326/aa564a.

Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka (2013), Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.*, 118, 1139–1150, doi:10.1002/jgrd.50174.

Friedlinstein, P., S. Solomon, G.-K. Plattner, R. Knutti, P Ciais and M. R. Raupach (2011), Long-term climate implications of twenty-first century options for carbon dioxide emission mitigation. *Nature Climate Change*, 1, 457-461, doi: 10.1038/nclimate1302.

Frölicher, T. L. and F. Joos (2010), Reversible and irreversible impacts of greenhouse gas emissions in multi-century projections with the NCAR global coupled carbon cycle-climate model. *Climate Dynamics*, 35, 1439-1459.

Frölicher, T. L., and D. J. Paynter (2015), Extending the relationship between global warming and cumulative carbon emissions to multi-millennial timescales, *Environ. Res. Lett.*, *10*(7), 75002, doi:10.1088/1748-9326/10/7/075002.

Frölicher, T. L., Winton M. and Sarmiento J. L. (2014), Continued global warming after CO_2 emissions stoppage. *Nat. Clim. Change,* **4**, 40–4.

Gillet, N. P., V. K. Arora, K. Zickfeld, S. J. Marshall and W. J. Merryfield (2011), Ongoing climate change following a complete cessation of carbon dioxide emissions. *Nature Geo*., **4**, 83-87.

Gillet, N. P., V. K. Arora, D. Matthews, and M. R. Allen (2013), Constraining the ratio of global warming to cumulative CO2 emissions using CMIP5 simulations. *J. Climate*, 26, 6844–6858.

Goodwin P., (2016), How historic simulation–observation discrepancy affects future warming projections in a very large model ensemble, *Clim. Dyn*., 47, 2219–2233, doi: 10.1007/s00382-015-2960-z.

Goodwin, P., I. D. Haigh, E. J. Rohling and A. Slangen (2017), A new approach to projecting 21st century sea-level changes and extremes, *Earth's Future*, 5, 240–253, doi:10.1002/2016EF000508.

Goodwin, P., R. G. Williams, M. J. Follows, and S. Dutkiewicz (2007), Ocean-atmosphere partitioning of anthropogenic carbon dioxide on centennial timescales. *Global. Biogeochem. Cycl.*, 21, GB1014.

Goodwin, P., R. G. Williams, and A. Ridgwell (2015), Sensitivity of climate to cumulative carbon emissions due to compensation of ocean heat and carbon uptake. *Nature Geosci.*, 8, 29–34.

Gregory, J. M., W.J. Ingram, M.A. Palmer, G.S. Jones, P.A. Stott, R.B. Thorpe, J.A. Lowe, T.C. Johns and K.D. Williams (2004), A new method for diagnosing radiative forcing and climate sensitivity. *Geophys Res Lett.*, 31, L03 205, doi:10.1029/2003GL018747.

Gregory, J. M., and P. M. Forster (2008), Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res.*, 113, D23, doi:10.1029/2008JD010405.

Jones, C., E. Robertson, V. Arora, P. Friedlingstein, E. Shevliakova, L. Bopp, V. Brovkin, T. Hajima, E. Kato, M. Kawamiya, S. Liddicoat, K. Lindsay, C.H. Reick, C. Roelandt, J. Segschneider, and J. Tjiputra (2013), Twenty-first-century compatible CO_2 emissions and airborne fraction simulated by CMIP5 Earth System models under four Representative Concentration Pathways. *J. Climate*, 26, 4398–4413.

Knutti, R., and G. C. Hegerl (2008), The equilibrium sensitivity of the Earths temperature to radiation changes. *Nature Geosci.*, 1, 735–743.

Knutti, R. and M. A. A. Rugenstein (2015), Feedbacks, climate sensitivity and the limits of linear models. *Phil Trans R Soc Lond A*., 373, doi:10.1098/rsta.2015.0146.

Lauderdale, J. M., A. C. Naveira Garabato, K. I. C. Oliver, M. J. Follows and R. G. Williams (2013), Wind-driven changes in the Southern Ocean residual circulation, ocean carbon reservoirs and atmospheric CO_2. *Climate Dynamics*, 41, 2145-2164, doi: 10.1007/s00382-012-1650-3

Matthews, H. D., and K. Caldeira (2008), Stabilizing climate requires near-zero emissions. *Geophys. Res. Lett.*, 35, L04705, doi: 10.1029/2007GL032388.

Matthews, H. R., N. P. Gillett, P. A. Stott, and K. Zickfeld (2009), The proportionality of global warming to cumulative carbon emissions. *Nature*, 459, 829–833.

Matthews, H. R., S. Solomon, and R. Pierrehumbert (2012), Cumulative carbon as a policy framework for achieving climate stabilization. *Phil. Trans. R. Soc.*, A370, 4365–4379.

Meinshausen, M., N. Meinshausen, W. Hare, S. C. B. Raper, K. Frieler, R. Knutti, D. J. Frame and M. R. Allen (2009), Greenhouse-gas emission targets for limiting global warming to 2^{o}C. *Nature*, 458, 1158–1162.

Myhre, G., E. J. Highwood, K. P. Shine, and F. Stordal (1998), New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.*, 25, 2715–2718.

Paynter, D., and T. L. Frölicher (2015), Sensitivity of radiative forcing, ocean heat uptake, and climate feedback to changes in anthropogenic greenhouse gases and aerosols. *J. Geophysical Res.: Atmos.*, 120, 9837–9854.

Plattner, G., R. Knutti, F. Joos, T.F. Stocker, W. von Bloh, V. Brovkin, D. Cameron, E. Driesschaert, S. Dutkiewicz, M. Eby, N.R. Edwards, T. Fichefet, J.C. Hargreaves, C.D. Jones, M.F. Loutre, H.D. Matthews, A. Mouchet, S.A. Müller, S. Nawrath, A. Price, A. Sokolov, K.M. Strassmann, and A.J. Weaver (2008), Long-term climate commitments projected with climate-carbon cycle models. *J. Climate*, **21**, 2721–2751, doi:10.1175/2007JCLI1905.1

Ricke, K.L. and K. Caldeira (2014), Maximum warming occurs about one decade after a carbon dioxide emission. *Environ. Res. Lett.,* 9, 124002, doi:10.1088/1748-9326/9/12/124002

Rose, B. E. J. and L. Rayborn (2016). The effects of ocean heat uptake on transient climate sensitivity, *Current Climate Change Reports*, 2, 190-201, doi: 10.1007/s40641-016-0048-4.

Solomon, S., G.-K. Plattner, R. Knutti, and P. Friedlingstein (2009), Irreversible climate change due to carbon dioxide emissions. *Proc. Natl. Acad. Sci. USA*, 106, 1704–1709.

Williams, R. G., P. Goodwin, A. Ridgwell, and P. L. Woodworth (2012), How warming and steric sea level rise relate to cumulative carbon emissions. *Geophys. Res. Lett.*, 39, L19 715, doi: 10.1029/2012GL052771.

Williams, R. G., P. Goodwin, V. M. Roussenov, and L. Bopp (2016), A framework to understand the Transient Climate Response to Emissions. *Environ. Res. Lett.*, 11, doi:10.1088/1748-9326/11/1/015003.

Williams, R. G., V.M. Roussenov, P. Goodwin, L. Resplandy and L. Bopp (2017), Sensitivity of global warming to carbon emissions: effects of heat and carbon uptake in a suite of Earth system models. *J. Climate*, in press, doi:10.1088/1748-9326/11/1/015003.

Williams, R.G. and M.J. Follows (2011), *Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms*. Cambridge Univ. Press, 416 pp. ISBN: 9780521843690

Winton M., S.M. Griffies, B. Samuels, J. L. Sarmiento and T. L. Frölicher (2013), Connecting changing ocean circulation with changing climate, *J. Climate,* 26, 2268–78.

Winton, M., K. Takahashi, and I. M. Held (2010), Importance of ocean heat uptake efficacy to transient climate change. *J. Climate*, 23, 2333–2344.

Zickfeld, K., M. Eby, H. D. Matthews, and A. J. Weaver (2009), Setting cumulative emissions targets to reduce the risk of dangerous climate change. *PNAS*, 106, 16 129–16 134.

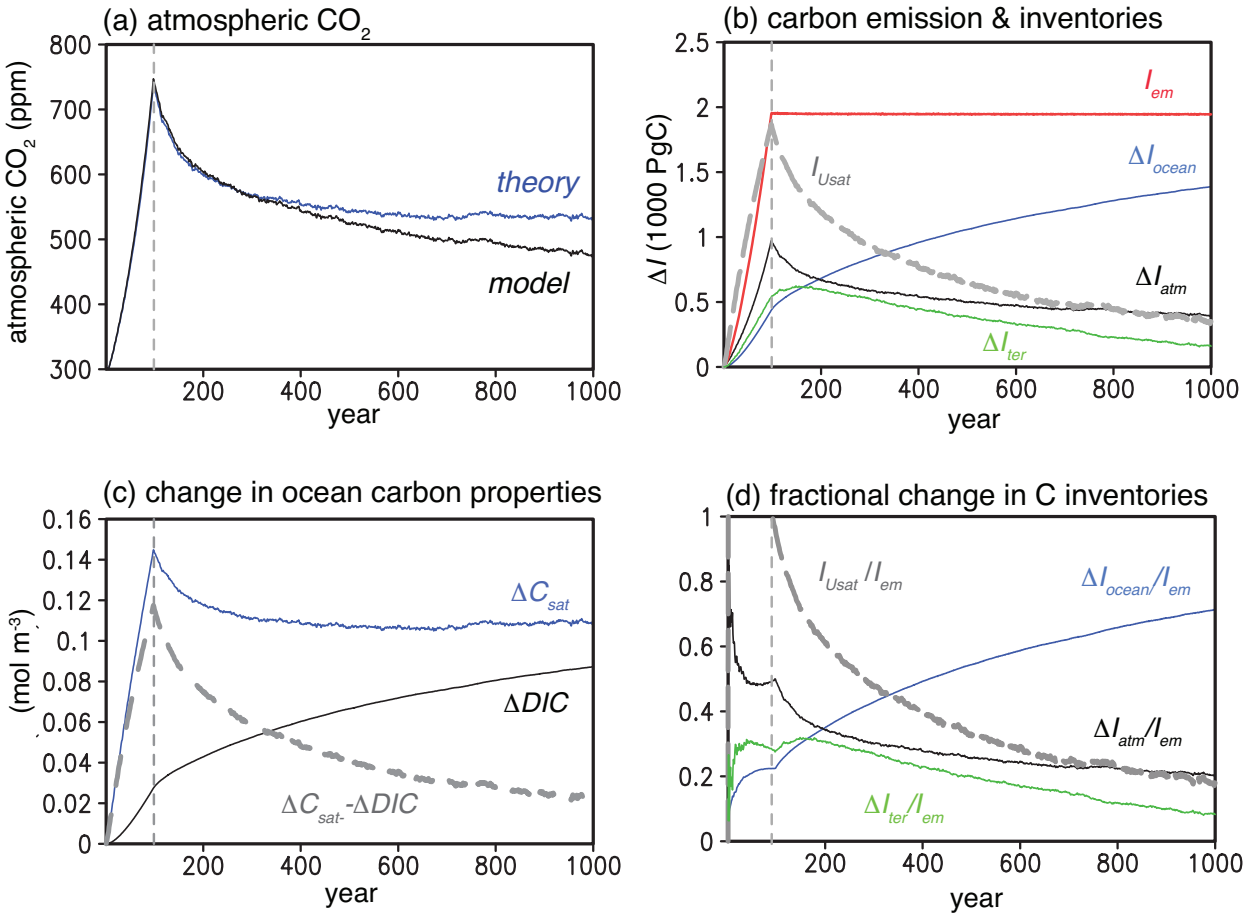


Figure 1. The 1000 year experiment with the GFDL-ESM2M Earth system model: (a) change in mixing ratio for atmospheric carbon dioxide (ppm, black), peaking in year 98 and then decreasing progressively, and the prediction from our theory (ppm, blue); (b) imposed cumulative carbon emission, I_{em} (red, 1000 PgC) with a linear increase up a 2K warming is reached at year 98 and then no further increase (vertical dashed line denotes the cessation of emissions) and the change in global carbon inventories (in 1000 PgC) for the atmosphere, $\Delta I_{atm}(t)$ (black), the ocean $\Delta I_{ocean}(t)$ (blue), and the terrestrial system $\Delta I_{ter}(t)$ (green) together with the carbon undersaturation of the global ocean, $I_{Usat}(t)$ (grey dashed); (c) the change in the ocean dissolved inorganic carbon, $\Delta DIC(t)$ (black, mol m^{-3}), and the change in the saturated dissolved inorganic carbon, $\Delta C_{sat}(t)$ (blue, mol m^{-3}) and their difference, $\Delta C_{sat}(t)$- $\Delta DIC(t)$ (dashed grey), where the global ocean carbon undersaturation is given by $I_{Usat}(t)=V(\Delta C_{sat}(t)- \Delta DIC(t))$ with V the global ocean volume; and (d) change in global carbon inventories divided by the cumulative carbon emissions for the atmosphere, $\Delta I_{atm}(t)/I_{em}(t)$ (black), the ocean $\Delta I_{ocean}(t)/I_{em}(t)$ (blue), and the terrestrial system $\Delta I_{ter}(t)/I_{em}(t)$ (green), together with the normalised carbon undersaturation of the global ocean, $I_{Usat}(t)/I_{em}(t)$ (dashed grey).

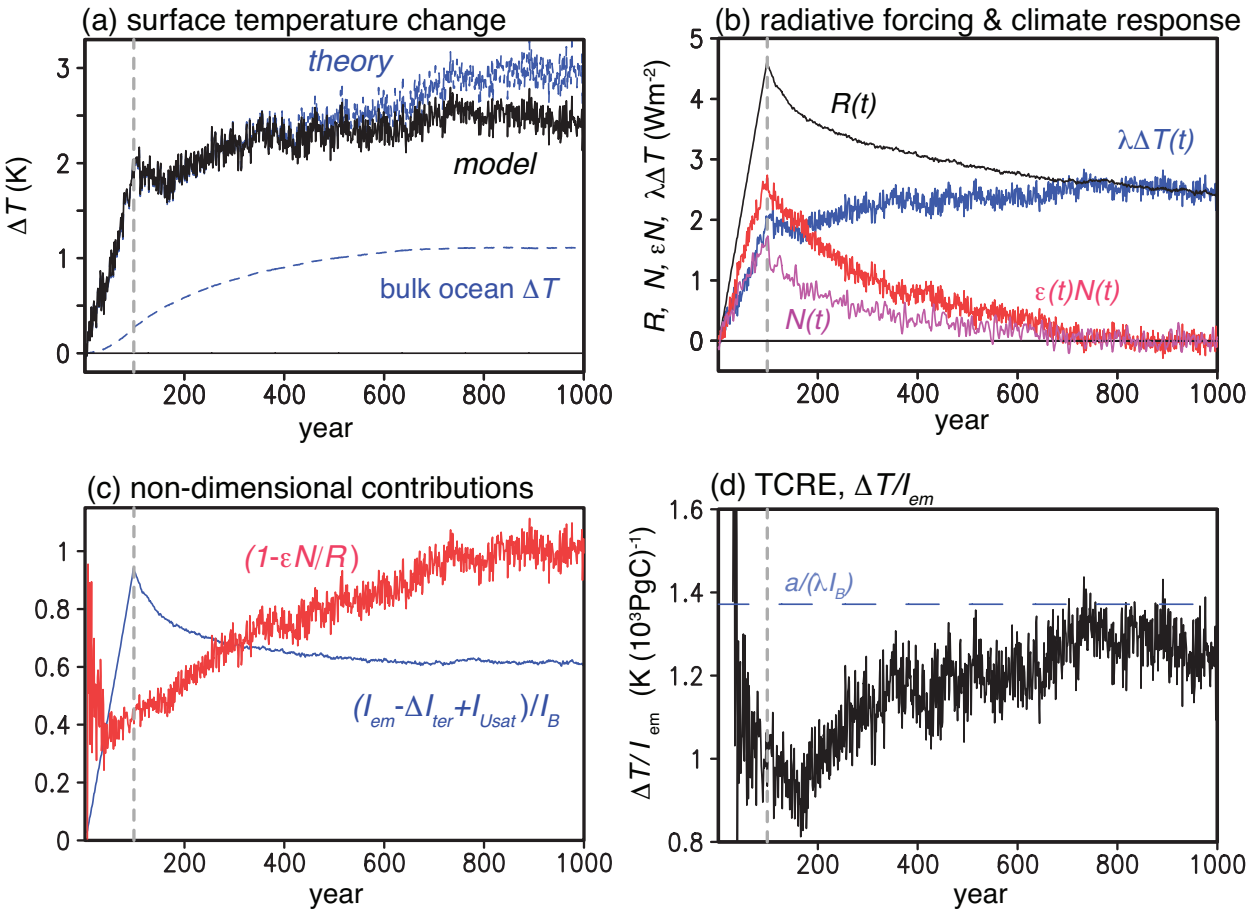


Figure 2. Contributions to the surface warming in the 1000 year experiment with the GFDL-ESM2M Earth system model: (a) the change in global-mean surface air temperature from the model, $\Delta T(t)$ (black, K) and the predicted temperature change from our theory (1) (blue, K), $(a/(\lambda I_B))((1-\varepsilon(t)N(t)/R(t)) (I_{em}(t)-\Delta I_{ter}(t)+I_{Usat}(t))$, together with the change in the bulk ocean temperature (dashed blue, K); (b) change in radiative forcing from atmospheric CO_2, $R(t)$ (black), surface climate response, $\lambda\Delta T(t)$ (blue), global ocean heat uptake, $N(t)$ (purple) and weighted ocean heat uptake, $\varepsilon(t)N(t)$ (red), (all in Wm^{-2}); (c), the non-dimensional contributions to the predicted surface temperature change, $\Delta T(t)$, depending on changes in the carbon inventories, $(I_{em}(t)-\Delta I_{ter}(t)+I_{Usat}(t))/I_B$ (blue) and the global heat uptake, $(1-\varepsilon(t)N(t)/R(t))$ (red); and (d) the Transient climate response to emissions (TCRE), $\Delta T/I_{em}$ (K $(10^3 PgC)^{-1}$) versus year together with the equilibrium value expected from $a/(\lambda I_B)$= 1.37 K $(1000\ PgC)^{-1}$ with a=4.85 Wm^{-2}, I_B=3487 PgC and λ=1.014 $Wm^{-2}K^{-1}$.

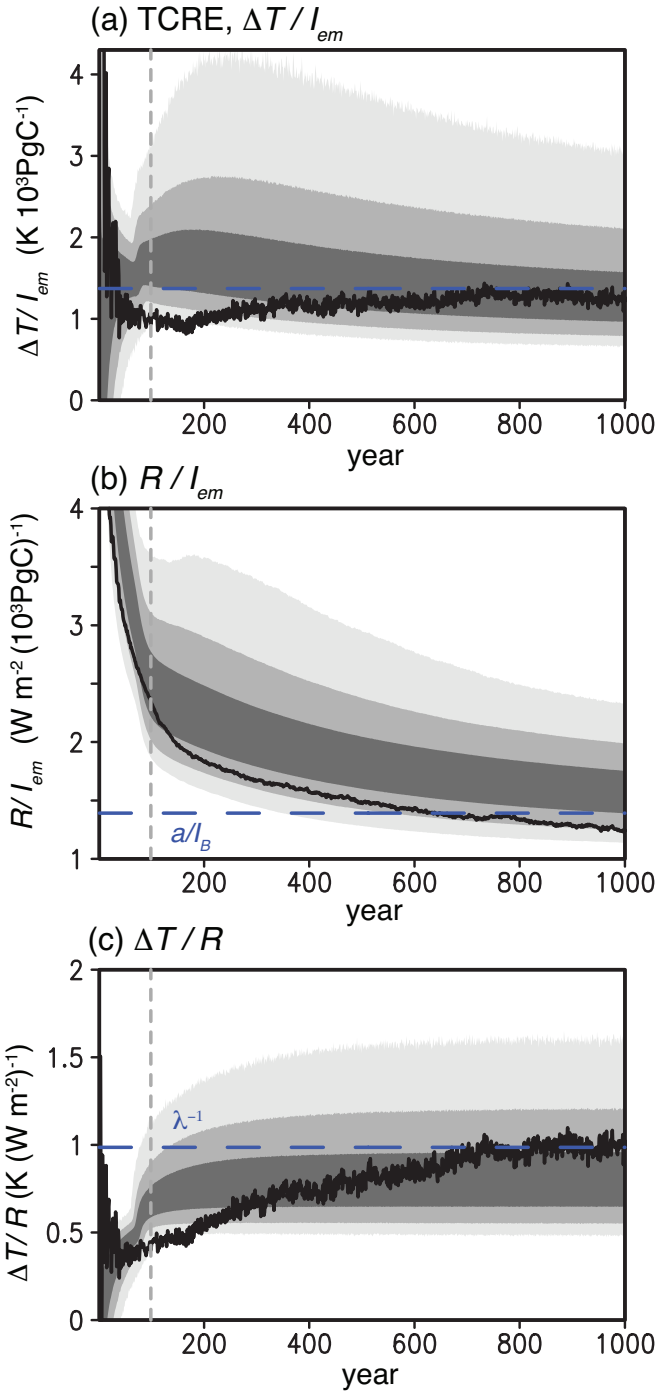


Figure 3. The 1000 year experiment with the GFDL-ESM2M Earth system model (black lines) and the efficient Earth system model (grey shading): (a) the Transient climate response to emissions (TCRE), $\Delta T/I_{em}$ (K (10^3PgC)$^{-1}$) versus year; (b) dependence of radiative forcing on cumulative carbon emissions, R/I_{em} (Wm^{-2}(10^3PgC)$^{-1}$); and (c) dependence of surface temperature on radiative forcing, $\Delta T/R$ (K (Wm^{-2})$^{-1}$). The projections of a large ensemble of 10,740 efficient Earth system model are denoted by grey shading (darkest, medium and light indicate the mean +/- 1, 2 and 3 standard deviations respectively), vertical dashed line denotes the cessation of emissions and blue dashed lines denote the expected long-term equilibrium values for $a/(\lambda I_B)$, a/I_B and λ^{-1} (as in Fig. 2d).

Geophysical Research Letters

Supporting Information for

Drivers of continued surface warming after cessation of carbon emissions

Richard G. Williams[1], Vassil Roussenov[1], Thomas L. Frölicher[2,3] and Philip Goodwin[4]

[1] Department of Earth, Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, L69 3GP, UK.

[2] Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland.

[3] Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.

[4] School of Ocean and Earth Sciences, University of Southampton, Southampton, UK.

Contents of this file

Text S1.

The empirical heat budget may be written in two equivalent ways:

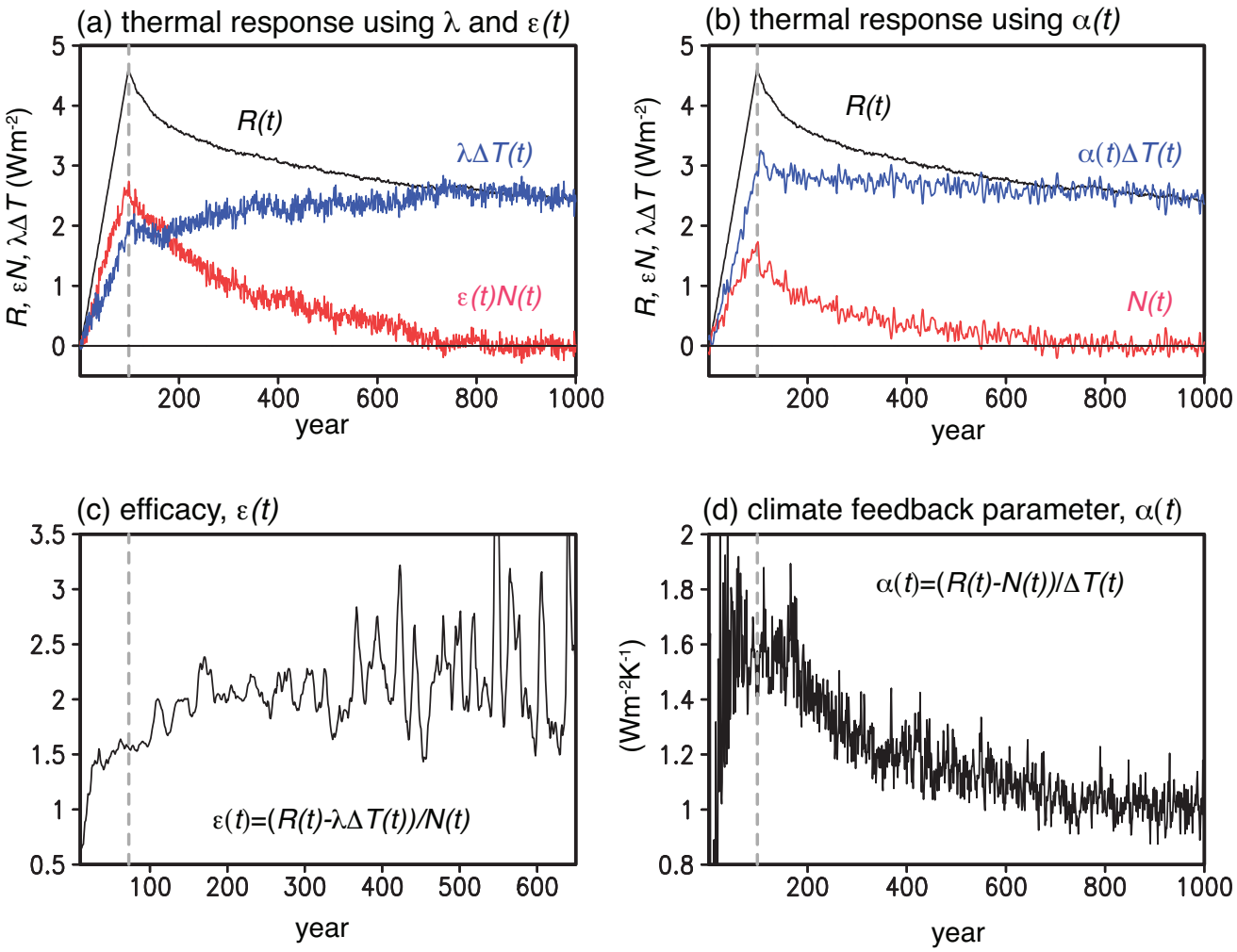
$$R(t) = \lambda \Delta T(t) + \varepsilon(t) N(t) = \alpha(t) \Delta T(t) + N(t)$$

where λ is the equilibrium climate feedback parameter ($Wm^{-2}K^{-1}$) and is chosen to be fixed in time, $\alpha(t)$ is a climate feedback parameter ($Wm^{-2}K^{-1}$) allowed to vary in time, $\varepsilon(t)$ is a non-dimensional weighting of planetary heat uptake allowed to vary in time, $R(t)$ is the radiative forcing since the preindustrial (Wm^{-2}), $\Delta T(t)$ is the global-mean surface air temperature since the preindustrial (K), and N is the planetary heat uptake (Wm^{-2}) that is effectively measured from ocean heat uptake.

The radiative forcing from atmospheric CO_2, $R(t)$, drives a surface climate response and planetary heat uptake, which may be represented by $\lambda \Delta T(t)$ plus the weighted ocean heat uptake, $\varepsilon(t)N(t)$ (Fig. S1a,c), or by $\alpha(t)\Delta T(t)$ plus the ocean heat uptake, $N(t)$ (Fig. S1b,d). If the latter form is adopted, then the relationship between surface temperature change and carbon emissions given in equation (1) may be rewritten as

$$\Delta T(t) = \frac{a}{\alpha(t) I_B} \left(1 - \frac{N(t)}{R(t)} \right) \left(I_{em}(t) - \Delta I_{ter}(t) + I_{Usat}(t) \right)$$

Time dependence in the feedbacks [*Knutti and Rugenstein*, 2017] are likely to lead to the climate feedback parameter, $\alpha(t)$, and efficacy, $\varepsilon(t)$, evolving in time; such as in how the climate feedback parameter varies through regional feedbacks [*Armour et al.,* 2013] or in how the efficacy alters with changes in ocean circulation [*Winton et al.,* 2013; *Rose and Rayborn*, 2016].



Supplementary Figure 1. Empirical heat budget over time: (a) the radiative forcing from atmospheric CO_2, *R(t)* (black), balances the surface climate response, $\lambda\Delta T(t)$ (blue), plus the weighted ocean heat uptake, $\varepsilon(t)N(t)$ (red), and (b) the radiative forcing from atmospheric CO_2, *R(t)* (black), balances the surface climate response, $\alpha(t)\Delta T(t)$ (blue), plus the global ocean heat uptake, *N(t)* (red); all in Wm^{-2}. In addition, in (c) the time-dependent non-dimensional weighting of the heat uptake, the efficacy, $\varepsilon(t)$, over the first 600 years when the product $\varepsilon(t)N(t)$ is non zero, and in (d) the time-variation of the time-dependent climate feedback parameter, $\alpha(t)$ ($Wm^{-2}K^{-1}$) over the entire record.

Text S2.

The WASP Earth system model ensemble

The Warming Acidification and Sea level Projector (WASP) is used to generate an initial ensemble of 10^8 simulations in a Monte Carlo approach following *Goodwin* [2016] and *Goodwin et al*. [2017]. A total of 18 model parameters are varied between simulations [*Goodwin*, 2016; *Goodwin et al*, 2017]: a; λ; I_B; 5 ocean tracer uptake timescales [*Goodwin*, 2017]; 2 equilibrium warming ratios (the ratio of warming between global air temperatures and sea surface temperatures, and the ratio of warming between sea surface temperatures and sub-surface ocean temperatures); the fraction of anthropogenic heat content increase that enters the ocean; 2 uncertainty scaling parameters related to uncertainty in radiative forcing from agents other than CO_2 (one for other greenhouse gasses and one for aerosols); and 2 non-dimensional efficacies relating the relative warming of processes to that incurred by CO_2 radiative forcing (ε representing the relative warming effect of ocean heat uptake and ε_{aero} representing the relative warming effect of radiative forcing from aerosols).

Each of the 10^8 simulations are forced with historic CO_2 and radiative forcing from other agents, from 1765 up to 2016, and assessed against reconstructions of historic changes to the real climate system (for surface warming, ocean heat content change and ocean and terrestrial carbon uptake) derived from observations. Following the methodology of *Goodwin* [2016] and *Goodwin et al* [2017], observational-consistent ranges for each constraint (e.g. surface warming) are based on the 90% confidence ranges for the quantity in the real climate system. A WASP simulation is accepted as observationally consistent if the simulated observables (based on surface warming, ocean heat content changes and ocean and terrestrial carbon uptake) fit either within the ranges of all observational constraints, or fit within all but one constraint and are less than 50% outside the range of the final constraint. This assessment allows the tails of the distributions for surface warming and ocean heat content to be included within the observation-consistent WASP ensemble.

After the observational tests, some 10,740 simulations remain, or 0.01% of the initial Monte Carlo simulations. These observationally-consistent ensemble members are then separately re-initialised and forced with the 1% annual rise in CO_2 experiments to compare to the GFDL model.

References

Armour, K. C., C. M. Bitz and G. H. Roe (2013), Time-varying climate sensitivity from regional feedbacks. *J. Climate*, 26, 4518-4534, doi: 10.1175/JCLI-D-12-00544.1.

Goodwin P. (2016), How historic simulation–observation discrepancy affects future warming projections in a very large model ensemble, *Clim. Dyn*., 47, 2219–2233, doi: 10.1007/s00382-015-2960-z.

Goodwin, P., I. D. Haigh, E. J. Rohling and A. Slangen (2017), A new approach to projecting 21st century sea-level changes and extremes, *Earth's Future*, 5, 240–253, doi:10.1002/2016EF000508.

Knutti, R. and M. A. A. Rugenstein (2015), Feedbacks, climate sensitivity and the limits of linear models. *Phil Trans R Soc Lond A*., 373, doi:10.1098/rsta.2015.0146.

Rose, B. E. J. and L. Rayborn (2016). The effects of ocean heat uptake on transient climate sensitivity, *Current Climate Change Reports*, 2, 190-201, doi: 10.1007/s40641-016-0048-4.

Winton M., S.M. Griffies, B. Samuels, J. L. Sarmiento and T. L. Frölicher (2013), Connecting changing ocean circulation with changing climate, *J. Climate,* 26, 2268–7