Equilibrium climate sensitivity estimated by equilibrating climate models

Maria Rugenstein¹,²*, Jonah Bloch-Johnson³, Jonathan Gregory³,⁴, Timothy Andrews⁴, Thorsten Mauritsen⁵, Chao Li², Thomas L. Fröhlicher⁶,⁷, David Paynter⁸, Gokhan Danabasoglu⁹, Shuting Yang¹⁰, Jean-Louis Dufresne¹¹, Long Cao¹², Gavin A. Schmidt¹³, Ayako Abe-Ouchi¹⁴, Olivier Geoffroy¹⁵, and Reto Knutti¹

¹Institute for Atmospheric and Climate Science, ETH Zurich, CH-8092 Zurich, Switzerland
²Max-Planck-Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany
³NCAS, University of Reading, Reading
⁴UK Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB
⁵Stockholm University, SE-106 91 Stockholm, Sweden
⁶Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland
⁷Oeschger Centre for Climate Change Research, University of Bern, Switzerland
⁸Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA
⁹National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307
¹⁰Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen, Denmark
¹¹LMD/IPSLS Centre National de la Recherche Scientifique, Sorbonne Université, École Normale Supérieure, PSL Research University, École Polytechnique, Paris, France
¹²School of Earth Sciences, Zhejiang University, Hang Zhou, Zhejiang Province, 310027, China
¹³NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025
¹⁴Atmosphere and Ocean Research Institute, The University of Tokyo, Japan
¹⁵Centre National de Recherches Météorologiques, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Key Points:
• 27 simulations of 15 general circulation models are integrated to near equilibrium
• All models simulate a higher equilibrium warming than predicted by using extrapolation methods
• Tropics and mid-latitudes dominate the change of the feedback parameter on different timescales

*Max-Planck-Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany

Corresponding author: Maria Rugenstein, maria.rugenstein@mpimet.mpg.de

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019GL083898

©2019 American Geophysical Union. All rights reserved.
Abstract
The methods to quantify equilibrium climate sensitivity are still debated. We collect millennial-length simulations of coupled climate models and show that the global mean equilibrium warming is higher than those obtained using extrapolation methods from shorter simulations. Specifically, 27 simulations with 15 climate models forced with a range of CO₂ concentrations show a median 17% larger equilibrium warming than estimated from the first 150 years of the simulations. The spatial patterns of radiative feedbacks change continuously, in most regions reducing their tendency to stabilizing the climate. In the equatorial Pacific, however, feedbacks become more stabilizing with time. The global feedback evolution is initially dominated by the tropics, with eventual substantial contributions from the mid-latitudes. Time-dependent feedbacks underscore the need of a measure of climate sensitivity that accounts for the degree of equilibration, so that models, observations, and paleo proxies can be adequately compared and aggregated to estimate future warming.

1 Estimating equilibrium climate sensitivity
The equilibrium climate sensitivity (ECS) is defined as the global- and time-mean, surface air warming once radiative equilibrium is reached in response to doubling the atmospheric CO₂ concentration above pre-industrial levels. It is by far the most commonly and continuously applied concept to assess our understanding of the climate system as simulated in climate models and it is used to compare models, observations, and paleo-proxies (Knutti et al., 2017; Charney et al., 1979; Houghton et al., 1990; Stocker, 2013). Due to the large heat capacity of the oceans, the climate system takes millennia to equilibrate to a forcing, but performing such a long simulation with a climate model is often computationally not feasible. As a result, many modeling studies use extrapolation methods on short, typically 150-year long, simulations to project equilibrium conditions (Taylor et al., 2011; Andrews et al., 2012; Collins et al., 2013; Otto et al., 2013; Lewis & Curry, 2015; Andrews et al., 2015; Forster, 2016; Calel & Stainforth, 2017). These so-called effective climate sensitivities (Murphy, 1995; Gregory et al., 2004) are often reported as ECS values (Hargreaves & Annan, 2016; Tian, 2015; Brient & Schneider, 2016; Forster, 2016). Research provides evidence for decadal-to-centennial changes of feedbacks (e.g., Murphy (1995); Senior and Mitchell (2000); Gregory et al. (2004); Winton et al. (2010); Armour et al. (2013); Proistosescu and Huybers (2017); Paynter et al. (2018)) but the behavior on longer timescales has
Figure 1. Evolution of global and annual mean top of the atmosphere (TOA) imbalance and surface temperature anomalies (14 small panels). The first 150 years of step forcing simulations are depicted in light gray. For experiments which are not step forcing simulations only the period after stabilizing CO$_2$ concentrations is shown. The black line shows the linear regression of TOA imbalance and surface warming for the last 15% of warming. The panel on the lower right shows the ratio $\Delta T_{\text{best est}} / \Delta T_{\text{est 1-150}}$, see text for definitions. A dot at the lower end of the bar indicates with 90% confidence that $\Delta T_{\text{best est}}$ and $\Delta T_{\text{est 1-150}}$ obtained by resampling 10,000 times do not overlap. The gray hashed bar in the background is the median of all simulations (1.17). FAMOUS $\text{abrupt}4x$ ends outside of the depicted range at 1.53. Table 1 specifies the model versions and names, length of simulations, and numerical values for different climate sensitivity estimates.
not been compared among models. Here, we utilize LongRunMIP, a large set of millennia-
long coupled general circulations models (GCMs) to estimate the true equilibrium warming,
study the centennial-to-millennial behavior of the climate system under elevated radiative
forcing, and test extrapolation methods. LongRunMIP is a model intercomparison project
(MIP) of opportunity in that its initial contributions were preexisting simulations, without
a previously agreed upon protocol. The minimum contribution is a simulation of at least
1000 years with a constant CO$_2$ forcing level. The collection consists mostly of doubling or
quadrupling step forcing simulations (“abrupt2x”, “abrupt4x”, ...) as well as annual incre-
ments of 1% CO$_2$ increases reaching and sustaining doubled or quadrupled concentrations
(“1pct2x”, “1pct4x”). Table 1 lists the simulations and models used here, while M. Rugen-
stein et al. (2019) documents the entire modeling effort and each contribution in detail.

The equilibration of top of the atmosphere (TOA) radiative imbalance and surface
temperature anomaly of the simulations are depicted in Fig. 1. Throughout the manuscript,
we show anomalies as the difference to the mean of the unforced control simulation with
pre-industrial CO$_2$ concentrations. Light gray dots indicate annual means of the first 150
years of a step forcing simulation, requested by the Coupled Model Intercomparison Project
Phase 5 and 6 protocols (CMIP5 and CMIP6; Taylor et al. (2011); Eyring et al. (2016))
and widely used to infer ECS (Andrews et al., 2012; Geoffroy, Saint-Martin, Olivié, et al.,
2013). We refer to this timescale as “decadal to centennial”. Colors indicate the “centen-
nial to millennial” timescale we explore here. The diminishing distances to the reference
line at TOA = 0 indicate that most simulations archive near-equilibrium by the end of the
simulations. However, even if a simulation has an equilibrated TOA imbalance of near zero,
the surface temperature, surface heat fluxes, or ocean temperatures can still show a trend
(discussed in M. Rugenstein et al. (2019)).

Throughout the manuscript, we use “∆T$_{[\text{specification}]}$” for a true or estimated equilib-
rium warming, for a range of forcing levels not only CO$_2$ doubling (Table 1). We define the
best estimate of equilibrium warming, $\Delta T_{\text{best est}}$, as the temperature-axis intersect of the
regression of annual means of TOA imbalance and surface temperature anomaly over the
simulations’ final 15% of global mean warming (black lines in Fig. 1). The lower right panel
in Fig. 1 illustrates that all simulations eventually warm significantly more (measured by
$\Delta T_{\text{best est}}$) than predicted by the most commonly used method to estimate the equilibrium
temperature by extrapolating a least-square regression of the first 150 years of the same step
forcing simulation (Gregory et al., 2004; Flato et al., 2013), denoted here as “$\Delta T_{\text{est 1–150}}$”. }

©2019 American Geophysical Union. All rights reserved.
For simulations that have gradual forcings (e.g., 1pct2x), we use 150 year long step forcing simulations of the same model to calculate $\Delta T_{est\,1-150}$. The median increase of $\Delta T_{best\,est}$ over $\Delta T_{est\,1-150}$ is 17% for all simulations and 16% for the subset of CO$_2$ doubling and quadrupling simulations. While $\Delta T_{est\,1-150}$ implies a constant feedback parameter (the slope of the regression line), other extrapolation methods allow for a time-dependent feedback parameter, but still typically underestimate $\Delta T_{best\,est}$. Using years 20-150 in linear regression ($\Delta T_{est\,20-150}$; e.g., Andrews et al. (2015); Armour (2017)) results in a median equilibrium warming estimate which is 7% lower than $\Delta T_{best\,est}$, both for all simulations and the subset of CO$_2$ doubling and quadrupling. The two-layer model including ocean heat uptake efficacy ($\Delta T_{EBM-\epsilon}$; e.g., Winton et al. (2010); Geoffroy, Saint-Martin, Bellon, et al. (2013)) results in a multi model median equilibrium warming estimate which is 9% lower than $\Delta T_{best\,est}$, again both for all simulations and the subset of CO$_2$ doubling and quadrupling. Both methods are described and illustrated in the supplemental material.

$\Delta T_{best\,est}$ of any forcing level can be scaled down to doubling CO$_2$ levels to estimate equilibrium warming for CO$_2$ doubling. We do so by assuming that the temperature scales with the forcing level, which depends logarithmically on the CO$_2$ concentration (Myhre et al., 1998), and assuming no feedback temperature dependence (e.g. Mauritsen et al. (2018) and Rohrschneider et al. (2019), see discussion below). The estimate of equilibrium warming for CO$_2$ doubling range from 2.42 to 5.83 K (excluding FAMOUS abrupt4x at 8.55K; see Table 1 and Fig. 1). Note that the simulation abrupt4x of the model FAMOUS warms anomalously strongly. As this simulation represents a physically possible result, we do not exclude it from the analysis (see more details in SM Section 4). The results are qualitatively the same if $\Delta T_{best\,est}$ is defined by regressing over the last 20% instead of 15% of warming or instead time averaging the surface warming toward the end of every simulation without taking the information of the TOA imbalance into account. SM Section 1 discusses different options and choices to determine $\Delta T_{best\,est}$.

2 Global feedback evolution

Current extrapolation methods underestimate the equilibrium response because climate feedbacks change with the degree of equilibration (Murphy, 1995; Senior & Mitchell, 2000; Andrews et al., 2015; Knutti & Rugenstein, 2015; M. A. A. Rugenstein, Caldeira, & Knutti, 2016; Armour, 2017; Proistosescu & Huybers, 2017; Paynter et al., 2018). We define the global net TOA feedback as the local tangent in temperature-TOA space ($\delta$TOA/$\delta$T) com-
Figure 2.  

(a) Time evolution of global feedbacks in four characteristic models. Net TOA feedback (gray) is the sum of its components: the cloud effects in the shortwave (red) and longwave (blue), and clear sky feedbacks in the shortwave (salmon) and longwave (light blue). Circles at the right of each panel indicate the feedbacks arising from internal variability; shading and vertical lines shows the 2.5-97.5% confidence intervals. Panel titles give the model name and length of the simulation. Time periods of 1-20 years and 150-1000 years are shaded gray. (b) Feedback evolution in the step forcing simulations of CCSM3, CESM104, CNRMCM6, ECHAM5MPIOM, FAMOUS, GISSER, HadCM3L, HadGEM2, IPSL3M5A, MPIESM11, and MPIESM12, see Table 1 for naming convention. Lines show all simulations, dots represent median values and bars spans all but the two highest and two lowest simulations. SM Fig. 4 and 5 show the feedback evolution for all available simulations.
computed by a least square regression of all global and annual means of netTOA imbalance and surface temperature anomaly within a temperature bin, which is moved in steps of 0.1 K throughout the temperature space to obtain the continuous local slope of the point cloud (sketched out in SM Fig. 2a). We decompose the net TOA imbalance into its clear sky and cloud radiative effects (CRE; e.g., Wetherald and Manabe (1988); Soden and Held (2006); Ceppi and Gregory (2017)) in the shortwave and longwave (Fig. 2a). The feedbacks change continuously – not on obviously separable timescales – in some models more at the beginning of the simulations (e.g., CESM104), in some models after 150 years (e.g., GISSE2R) or, in some models, intermittently throughout the simulation (e.g., MPIESM11 or HadGEM2).

The shortwave CRE dominates the magnitude and the timing of the net feedback change, and can be counteracted by the longwave CRE. The reduction of the shortwave clear sky feedback associated with ice albedo, lapse rate, and water vapor is a function of temperature and occurs on centennial to millennial timescales. Longwave clear sky changes, when present, contribute to the increase of the sensitivity with equilibration time and temperature.

The net feedback parameter can be composed of a subtle balance of different components at any time and the forced signal is not obviously linked to the feedback arising from internal variability, defined by regressing all available annual and global means of TOA imbalance and surface temperature anomalies (relative to the mean) of the control simulations (circles in Fig. 2a; Roe (2009); Brown et al. (2014); Zhou et al. (2015); Colman and Hanson (2017)). Models which are more sensitive than other models – have feedbacks which are more positive – at the beginning of the simulation are generally also more sensitive towards the end. The model spread in the magnitude of feedbacks does not substantially reduce in time, while the feedback parameter change varies from negligible to an order of magnitude. We quantify the continuous changes across models by considering different time periods, namely years 1-20, 21-150, and 151-1000 (Fig. 2b), in each of which we regress all points. In addition to the increase of the feedback parameter between years 1-20 and 21-150, which has been documented for CMIP5 models (Geoffroy, Saint-Martin, Bellon, et al., 2013; Andrews et al., 2015; Proistosescu & Huybers, 2017; Ceppi & Gregory, 2017), there is a further increase from centennial to millennial timescales.

Previous research has shown that the change in feedbacks over time can come about through a dependence of feedback processes on the increasing temperature (Hansen et al., 1984; Jonko et al., 2013; Caballero & Huber, 2013; Meraner et al., 2013; Bloch-Johnson et al., 2015), due to evolving surface warming patterns and feedback processes ("pattern effect");
**Figure 3.** Multi-model mean normalized patterns of surface warming (local warming divided by global warming) between the average of (a) the control simulation and year 15-25, (b) year 15-25 and 140-160, (c) year 140-160 and 800-1000, and their differences (d and e) for the same models and simulations as in Fig. 2b. For models contributing several simulations, these are averaged. Stippling in panel d and e indicates that 9 out of 11 models agree in the sign of change.

Senior and Mitchell (2000); Winton et al. (2010); Armour et al. (2013); M. A. A. Rugenstein, Gregory, et al. (2016); Gregory and Andrews (2016); Haugstad et al. (2017); Paynter et al. (2018), or both at the same time (Rohrschneider et al., 2019). There is no published method which clearly differentiates between time/pattern and temperature/state dependence and simulations with several forcing levels are needed to disentangle them. The relationship between forcing and CO$_2$ concentrations is a matter of debate (Etminan et al., 2016) and further complicates the analysis, as time, temperature, and forcing level dependence might compensate to some degree (Gregory et al., 2015). As not all models contributed several forcing levels, we focus in the following on robust pattern changes in surface temperatures and feedbacks, which occur in most or all simulations irrespective of their overall temperature anomaly or forcing level.

**3 Pattern evolution of surface warming and feedbacks**

The evolution of surface warming patterns during the decadal, centennial, and millennial periods displays a fast establishment of a land-sea warming contrast, Arctic amplification, and the delayed warming over the Southern Ocean that have been studied on annual to centennial timescales (Fig. 3; Senior and Mitchell (2000); Li et al. (2013); Collins et al. (2013); Armour et al. (2016)). Arctic amplification does not change substantially,
Figure 4. Time evolution of feedback patterns. Model-mean of local contribution to the change in global feedbacks (local TOA anomaly divided by global warming during the period indicated in the panel titles; see text for definitions) (a–c) and their differences (d, e). The global feedback value is shown in the panel title. Regionally aggregated contributions to the global values are indicated with percent numbers and gray triangles (22°S-22°N, 22°S/N-66°S/N, 66°S/N-90°S/N, representing 40%, 27%, and 4% of the global surface area respectively). Model and simulations selection, weighting, and stippling is the same as in Fig. 3. SM Fig. 6–12 shows all TOA components.

whereas Antarctic amplification strengthens by approximately 50% on centennial to millennial timescales (Salzmann, 2017; M. Rugenstein et al., 2019). The warming in the northern North Atlantic reflects the strengthening of the Atlantic meridional overturning circulation, after the initial decline (Stouffer & Manabe, 2003; Li et al., 2013; M. A. A. Rugenstein, Sedláček, & Knutti, 2016; Rind et al., 2018; Jansen et al., 2018).

In the Pacific, at all times, the temperatures in absolute terms are higher in the West compared to the East Pacific. The eastern equatorial Pacific warms more than the warm pool in most simulations, a phenomenon reminiscent of the positive phase of the El-Niño-Southern-Oscillation (ENSO) (“ENSO-like warming” (Song & Zhang, 2014; Andrews et al., 2015; Luo et al., 2017; Tierney et al., 2019)). This tendency can last several millennia, but significantly reduces or stops in most simulations after a few hundred years. Similar to the Equatorial east Pacific, the south east Pacific warms more than the warm pool (Zhou et al., 2016; Andrews & Webb, 2018). However, models display a large variance in the timescales of warming in these two regions, i.e. the warm pool can initially warm faster or slower than the south east Pacific. Across the Pacific, the change in surface warming pattern is reminiscent of the Interdecadal Pacific Oscillation (IPO; Fig. 3d). In many models, the reduction
of the Walker circulation coincides with the decadal to centennial ENSO/IPO-like warming pattern, but it does not obviously coincide with surface warming pattern changes on the millennial timescale, indicating that subtropical ocean gyre advection and upwelling play a more prominent role on longer timescales (Knutson & Manabe, 1995; Song & Zhang, 2014; Fedorov et al., 2015; Andrews & Webb, 2018; Luo et al., 2017; Zhou et al., 2017; Kohyama et al., 2017). The mechanisms and spread of model responses in the Pacific are still under investigation.

Feedbacks defined as the local tangent in temperature-TOA space as used in Fig. 2a contain a signal from both the internal variability and the forced response. In order to isolate the forced response, we take the difference of the means at the beginning and end of the time periods discussed above. We call this definition of feedbacks the finite difference approach, as it represents a change across a time period (SM Fig. 2b). Fig. 4 shows the local contribution to the global net TOA changes (defined as the local change in TOA imbalance divided by the global temperature change.) for the same time periods and models as used in Fig. 3. In the initial years, the atmosphere restores radiative balance through increased radiation to space almost everywhere, except in the western-central Pacific (Fig. 4a), whereas on decadal to centennial timescales, the structure of the feedbacks mirrors the surface temperature evolution and develops a pattern reminiscent of ENSO/IPO (Fig. 4b). The cloud response dominates the pattern change, although for CMIP5 models, changes on decadal and centennial timescales have been attributed to changing lapse rate feedbacks as well (SM Fig. 6-8 and Andrews et al. (2015); Andrews and Webb (2018); Ceppi and Gregory (2017)). For the millennial timescales, our models show that feedbacks become less negative almost everywhere, switching from slightly negative to positive in parts of the Southern Ocean and North Atlantic region, and become less destabilizing in the Tropical Pacific (Fig. 4c). The feedback pattern change from decadal to centennial timescales (Fig. 4d) is reversed in many regions on centennial to millennial timescales (Fig. 4e), particularly in the entire Pacific basin, the Atlantic, and parts of Asia and North America. This “pattern flip” is dominated by longwave CRE (SM Fig. 8) and mirrors, in the Pacific, the reduction in ENSO/IPO-like surface warming patterns discussed for the surface temperature evolution.

Note that the local temperature is not part of the calculation of the local contribution in feedback changes. Due to the far-field effects of local feedbacks (e.g., Rose et al. (2014); Kang and Xie (2014); M. A. A. Rugenstein, Caldeira, and Knutti (2016); Zhou et al. (2016, 2017); Ceppi and Gregory (2017); Liu et al. (2018); Dong et al. (2019)), the relation between
the local feedback contribution (Fig. 4) and the local temperatures (Fig. 3) is not straightforward. There is strong correspondence between changes of TOA fluxes and temperature patterns in the Pacific on decadal to millennial timescales: Stronger (weaker) local warming coincides with a more positive (negative) local feedback contribution. However, there is no clear correspondence directly after the application of the forcing, or over land and the Southern Ocean through time. SM Fig. 13 and 14 show overlays of Fig. 3 and 4 for a better comparison. A local correspondence does not necessarily indicate a strong local feedback (i.e. local TOA divided by local surface temperature change), as both the local TOA and the surface in one region could be forced by another region. A closer investigation of local and far-field influence of feedbacks is under investigation (Bloch-Johnson et al., in revision).

Although the spatial patterns of changing temperature and radiative feedbacks vary among models, the large scale features discussed here occur robustly across most models and forcing levels, and also occur in the 1pct2x and 1pct4x simulations, which are not included in the figures.

4 Regions accounting for changing global feedbacks

We quantify the contribution of the tropics, extra-tropics, and polar regions to the global feedback change (Fig. 4d,e) by adding up all feedback contributions of the respective areas indicated by the gray triangles and expressing them as percentages of the total. We note that the total is the global feedback parameter, i.e., the slope of the point clouds in Fig. 1 which is indicated on the top right of each panel. These percentages reflect the role played by TOA fluxes in each region, which is not the same as the role played by surface warming in each region, as noted above. Whereas the tropics account for the bulk of the change (58% on decadal to centennial and 47% on centennial to millennial timescales), the mid-latitudes become more important with time (Northern and Southern Hemisphere combined for 41% on decadal to centennial and for 66% on centennial to millennial timescales). The high latitudes, dominated by the shortwave clear sky feedback (SM Fig. 12), play only a minor role in influencing the global response at all timescales. The regional accounting of global feedback changes permits us to test competing explanations regarding the spatial feedback pattern by placing them in a common temporal framework. Primary regions controlling the global feedback evolution have been suggested to be the Southern Hemisphere mid to high latitudes (Senior & Mitchell, 2000), the Northern Hemisphere subpolar regions (Rose & Rayborn, 2016; Trossman et al., 2016), and the Tropics (Jonko et al., 2013; Mer-
The simulations robustly show a delayed warming in the Southern Hemisphere relative
to the Northern Hemisphere throughout the millennia-long integrations, which correlates
with the time evolution of net TOA and shortwave CRE (not shown). This behavior lends
support to the hypothesis of Senior and Mitchell (2000) who propose that feedbacks change
through time due to the slow warming rates of the Southern Ocean relative to the upper
atmospheric levels. This reduced lapse rate increases atmospheric static stability (and thus,
the shortwave cloud response) in the transient part of the simulation, but decreasingly less
so towards equilibrium.

The extra-tropical cloud response in the model-mean is non-negligible in the Southern
Ocean and North Atlantic on decadal to centennial timescales, as proposed by Rose and
Rencurrel (2016) and Trossman et al. (2016). However, it comes to dominate the global
response only on centennial to millennial timescales and when both hemispheres are consid-
ered.

We find that the longwave clear sky feedback does moderately increase in many mod-
els as the temperature or the forcing level increases, mainly in the tropics and Northern
Hemisphere mid-latitudes (Fig. 2a, SM Fig. 4, SM Fig. 5). This is in accordance with the
proposed argument that the tropics govern the global feedback evolution because the water
vapor feedback increases with warming (Jonko et al., 2013; Meraner et al., 2013; Block &
Mauritsen, 2013; Andrews et al., 2015), possibly following the rising tropical tropopause
(Meraner et al., 2013; Mauritsen et al., 2018).

Recent work has focused on the relative influence of the Pacific, specifically the relative
influence of temperatures of the warm pool versus compared to other regions. Feedbacks in
regions of atmospheric deep convections have a far-field and global effect, while feedbacks
in regions of atmospheric subsidence have only a local or regional influence (Barsugli &
Sardeshmukh, 2002; Zhou et al., 2017; Andrews & Webb, 2018; Ceppi & Gregory, 2019;
Dong et al., 2019). With the available fields in the LongRunMIP archive, we cannot quan-
tify the relative importance of water vapor and lapse rate feedbacks. However, the short and
longwave cloud response (SM Fig. 6–8) in the models qualitatively agree with the proposed
change of tropospheric stability patterns on decadal to centennial timescales (Andrews &
Webb, 2018; Ceppi & Gregory, 2017), especially in the Pacific region. In contrast, on centen-
nial to millennial timescales, the tropical Pacific response becomes less important compared to the mid-latitudes and the net tropical CRE does not change anymore (SM Fig. 6).

5 Implications

We demonstrate that the evolution of the global feedback response is dominated by the mid-latitudes on centennial to millennial and the tropics on decadal to centennial timescales. The global net feedback change is a result of a subtle balance of different regions and different TOA components at all times; even more so in single simulations than in the model mean shown here. This motivates process-based feedback studies in individual models as well as multi-model ensembles to draw robust conclusions and increase physical understanding of processes. To relate the timescales and model behavior to the observational record and paleo proxies a better understanding of a) the atmospheric versus oceanic drivers of surface temperature patterns in both, the coupled climate models and the real world and b) the local and far field interactions of tropospheric stability, clouds, and surface temperatures need to be achieved. Note that climate models have typical and persistent biases in regions we identify as important, mainly the Equatorial Pacific, Southern Ocean and ocean upwelling regions. The pattern effect of the real world might act on timescales which are different than the ones of the climate models.

Our results show that radiative feedbacks, usually called “fast”, act continuously less stabilizing on the climate system as the models approach equilibrium. As a result, the equilibrium warming is higher than estimated with common extrapolation methods from short simulations for all models and simulations in the LongRunMIP archive. ECS has been historically used as a model characterization (Charney et al., 1979), but some studies propose that it is not the most adequate measure for estimating changes expected over the next decades and until the end of the century (e.g., Otto et al. (2013); Shiogama et al. (2016); Knutti et al. (2017)). Alternative climate sensitivity measures are the effective climate sensitivity computed on different timescales, the transient climate response to gradually increasing CO$_2$ (TCR), or the transient climate response to cumulative carbon emissions (e.g., Allen and Frame (2007); Millar et al. (2015); Gregory et al. (2015); Grose et al. (2018)). Beyond not being an accurate indicator of the equilibrium response, these alternative climate sensitivity measures capture the models in different degrees of equilibration. We show that it is an open question how different measures of sensitivity relate to each other. A recent study shows that $\Delta T_{\text{est}}^{1-150}$ correlates better than TCR with end-of-21st-century warming.
across model (Grose et al. (2018), see also Gregory et al. (2015)). Thus, we underscore
the need of comparing models, observations, and paleo proxies on well-defined measures of
climate sensitivity, which ensure they are in the same state of equilibration.

Acknowledgments

Fields shown in this paper can be accessed on https://data.iac.ethz.ch/longrunmip/
GRL/. See www.longrunmip.org and M. Rugenstein et al. (2019) for more details on each
simulation and available variables, not shown here.

We thank Urs Beyerle, Erich Fischer, Jeremy Rugenstein, Levi Silvers, and Martin Stolpe
for technical help and comments on the manuscript.

MR is funded by the Alexander von Humboldt Foundation. NCAR is a major facility spon-
sored by the US National Science Foundation under Cooperative Agreement No. 1852977.
TA was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Pro-
gramme (GA01101). TLF acknowledges support from the Swiss National Science Founda-
tion under grant PP00P2_170687, from the EU-H2020 project CCiCC, and from the Swiss
National Supercomputing Centre (CSCS). CL was supported through the Clusters of Ex-
cellence CliSAP (EXC177) and CLICCS (EXC2037), University Hamburg, funded through
the German Research Foundation (DFG). SY was partly supported by European Research
Council under the European Community’s Seventh Framework Programme (FP7/2007-
2013)/ERC grant agreement 610055 as part of the ice2ice project.
References


©2019 American Geophysical Union. All rights reserved.
a) Time evolution of feedbacks in four models

- CESM104 abrupt4x, 5300 years
- MPIESM11 abrupt4x, 4459 years
- GISSE2R abrupt4x, 5001 years
- HadGEM2 abrupt4x, 1299 years

b) Feedback components for different time periods

- Year 1-20
- Year 21-150
- Year 151-1000

Feedback parameter (Wm\(^{-2}\)K\(^{-1}\))
Table 1: Estimates of equilibrium warming and their uncertainties for each simulation

<table>
<thead>
<tr>
<th>Model long name</th>
<th>Simulation</th>
<th>Length</th>
<th>Control length</th>
<th>$\Delta T_{\text{best est}}$ Best estimate of equilibrium warming</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSM3</td>
<td>abrupt2x</td>
<td>3000</td>
<td>1530</td>
<td>2.57 (2.55 - 2.58)</td>
</tr>
<tr>
<td></td>
<td>abrupt4x</td>
<td>2120</td>
<td></td>
<td>5.45 (5.42 - 5.48)</td>
</tr>
<tr>
<td>CCSM3</td>
<td>abrupt8x</td>
<td>1450</td>
<td></td>
<td>8.85 (8.80 - 8.89)</td>
</tr>
<tr>
<td>CESM 1.0.4</td>
<td>abrupt2x</td>
<td>2500</td>
<td>1320</td>
<td>3.2 (3.18 - 3.23)</td>
</tr>
<tr>
<td>CESM104</td>
<td>abrupt4x</td>
<td>5300</td>
<td></td>
<td>6.75 (6.73 - 6.76)</td>
</tr>
<tr>
<td></td>
<td>abrupt8x</td>
<td>4000</td>
<td></td>
<td>10.36 (10.34 - 10.39)</td>
</tr>
<tr>
<td>CNRM-CM6-1</td>
<td>abrupt2x</td>
<td>750</td>
<td>2000</td>
<td>4.83 (4.71 - 4.98)</td>
</tr>
<tr>
<td>CNRMCm6</td>
<td>abrupt4x</td>
<td>1850</td>
<td></td>
<td>11.14 (11.03 - 11.26)</td>
</tr>
<tr>
<td>EC-Earth-PISM</td>
<td>*abrupt4x</td>
<td>150</td>
<td>508</td>
<td>11.98 (11.97 - 12.00)</td>
</tr>
<tr>
<td>ECEARTH</td>
<td>RCP8.5+</td>
<td>1270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECHAMS/MPIOM</td>
<td>*abrupt4x</td>
<td>1000</td>
<td>100</td>
<td>11.04 (11.02 - 11.07)</td>
</tr>
<tr>
<td>ECHAM5</td>
<td>1pct4x</td>
<td>6080</td>
<td></td>
<td>11.66 (11.44 - 11.92)</td>
</tr>
<tr>
<td>FAMOUS</td>
<td>abrupt2x</td>
<td>3000</td>
<td>3000</td>
<td>4.40 (4.4 - 4.46)</td>
</tr>
<tr>
<td>FAMOUS</td>
<td>abrupt4x</td>
<td>3000</td>
<td></td>
<td>17.09 (16.04 - 19.19)</td>
</tr>
<tr>
<td>GFDL-CM3</td>
<td>*abrupt4x</td>
<td>150</td>
<td>5200</td>
<td>4.67 (4.66 - 4.68)</td>
</tr>
<tr>
<td>GFDLCM3</td>
<td>1pct2x</td>
<td>5000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td>*abrupt4x</td>
<td>150</td>
<td>1340</td>
<td>3.25 (3.24 - 3.25)</td>
</tr>
<tr>
<td>GFDLESMM2M</td>
<td>1pct2x</td>
<td>4500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GISS-E2-R</td>
<td>abrupt4x</td>
<td>5000</td>
<td>5225</td>
<td>4.83 (4.87 - 4.88)</td>
</tr>
<tr>
<td>GISSE2R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadCM3L</td>
<td>abrupt2x</td>
<td>1000</td>
<td>1000</td>
<td>3.34 (3.27 - 3.42)</td>
</tr>
<tr>
<td>HadCM3L</td>
<td>abrupt4x</td>
<td>1000</td>
<td></td>
<td>6.89 (6.76 - 7.06)</td>
</tr>
<tr>
<td></td>
<td>abrupt6x</td>
<td>1000</td>
<td></td>
<td>9.29 (9.06 - 9.59)</td>
</tr>
<tr>
<td></td>
<td>abrupt8x</td>
<td>1000</td>
<td></td>
<td>10.82 (10.57 - 11.12)</td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>abrupt4x</td>
<td>1328</td>
<td>239</td>
<td>9.54 (9.44 - 9.65)</td>
</tr>
<tr>
<td>HadGEM2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>abrupt4x</td>
<td>1000</td>
<td>1000</td>
<td>9.53 (9.32 - 9.77)</td>
</tr>
<tr>
<td>IPSLCM5A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIROC 3.2</td>
<td>*abrupt4x</td>
<td>150</td>
<td>681</td>
<td>8.97 (8.93 - 9.01)</td>
</tr>
<tr>
<td>MIROC32</td>
<td>1pct4x</td>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPI-ESM 1.2</td>
<td>abrupt2x</td>
<td>1000</td>
<td>1237</td>
<td>2.94 (2.92 - 2.96)</td>
</tr>
<tr>
<td>MPIESM12</td>
<td>abrupt4x</td>
<td>1000</td>
<td></td>
<td>6.71 (6.67 - 6.7)</td>
</tr>
<tr>
<td></td>
<td>abrupt8x</td>
<td>1000</td>
<td></td>
<td>11.98 (11.91 - 12.06)</td>
</tr>
<tr>
<td></td>
<td>abrupt16x</td>
<td>1000</td>
<td></td>
<td>21.88 (21.64 - 22.15)</td>
</tr>
<tr>
<td>MPI-ESM 1.1</td>
<td>abrupt4x</td>
<td>4459</td>
<td>2000</td>
<td>6.84 (6.84 - 6.85)</td>
</tr>
<tr>
<td>MPIESM11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta T_{\text{test 1-150}}$</td>
<td>$\Delta T_{\text{EBM-}\epsilon \ 1-150}$</td>
<td>$\Delta T_{\text{test 20-150}}$</td>
<td>Estimated equilibrium warming for doubling CO$_2$</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>Estimated equilibrium warming from year 1-150</td>
<td>Energy balance model with ocean heat uptake efficacy</td>
<td>Estimated equilibrium warming from year 20-150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.35 (2.26 - 2.47)</td>
<td>2.75</td>
<td>2.49 (2.34 - 2.79)</td>
<td>2.57</td>
<td></td>
</tr>
<tr>
<td>4.89 (4.76 - 5.01)</td>
<td>5.32</td>
<td>5.23 (5.03 - 5.48)</td>
<td>2.73</td>
<td></td>
</tr>
<tr>
<td>8.04 (7.90 - 8.18)</td>
<td>8.52</td>
<td>8.44 (8.26 - 8.68)</td>
<td>2.95</td>
<td></td>
</tr>
<tr>
<td>2.57 (2.50 - 2.67)</td>
<td>2.57</td>
<td>2.62 (2.50 - 2.82)</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td>5.74 (5.61 - 5.89)</td>
<td>6.08</td>
<td>6.30 (6.07 - 6.61)</td>
<td>3.38</td>
<td></td>
</tr>
<tr>
<td>9.02 (8.89 - 9.18)</td>
<td>9.53</td>
<td>9.64 (9.45 - 9.88)</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>4.12 (3.92 - 4.33)</td>
<td>4.43</td>
<td>4.6 (4.29 - 5.03)</td>
<td>4.83</td>
<td></td>
</tr>
<tr>
<td>9.64 (9.44 - 9.84)</td>
<td>10.02</td>
<td>9.64 (9.32 - 10.05)</td>
<td>5.70</td>
<td></td>
</tr>
<tr>
<td>6.61 (6.46 - 6.72)</td>
<td>7.03</td>
<td>6.81 (6.65 - 6.99)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.27</td>
<td></td>
</tr>
<tr>
<td>10.43 (10.1 - 10.89)</td>
<td>9.90</td>
<td>9.88 (9.44 - 10.73)</td>
<td>5.83</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.52</td>
<td></td>
</tr>
<tr>
<td>3.71 (3.56 - 3.93)</td>
<td>4.00</td>
<td>4.10 (3.76 - 4.75)</td>
<td>4.40</td>
<td></td>
</tr>
<tr>
<td>11.20 (11.00 - 11.42)</td>
<td>12.19</td>
<td>11.87 (11.51 - 12.33)</td>
<td>8.55</td>
<td></td>
</tr>
<tr>
<td>7.75 (7.42 - 8.02)</td>
<td>8.40</td>
<td>8.27 (7.96 - 8.67)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.67</td>
<td></td>
</tr>
<tr>
<td>4.86 (4.76 - 5.00)</td>
<td>5.07</td>
<td>5.20 (4.96 - 5.51)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.25</td>
<td></td>
</tr>
<tr>
<td>4.28 (4.12 - 4.42)</td>
<td>4.49</td>
<td>4.62 (4.51 - 4.74)</td>
<td>2.44</td>
<td></td>
</tr>
<tr>
<td>3.15 (3.00 - 3.36)</td>
<td>3.34</td>
<td>3.48 (3.21 - 3.98)</td>
<td>3.34</td>
<td></td>
</tr>
<tr>
<td>5.83 (5.71 - 5.99)</td>
<td>6.30</td>
<td>6.50 (6.15 - 7.10)</td>
<td>3.45</td>
<td></td>
</tr>
<tr>
<td>7.41 (7.21 - 7.61)</td>
<td>8.10</td>
<td>8.14 (7.60 - 9.29)</td>
<td>3.60</td>
<td></td>
</tr>
<tr>
<td>8.31 (8.13 - 8.47)</td>
<td>9.25</td>
<td>9.49 (8.97 - 10.41)</td>
<td>3.61</td>
<td></td>
</tr>
<tr>
<td>9.04 (8.59 - 9.37)</td>
<td>10.82</td>
<td>11.04 (10.23 - 12.22)</td>
<td>4.77</td>
<td></td>
</tr>
<tr>
<td>8.07 (7.93 - 8.24)</td>
<td>8.40</td>
<td>8.56 (8.25 - 8.93)</td>
<td>4.76</td>
<td></td>
</tr>
<tr>
<td>8.27 (8.18 - 8.39)</td>
<td>8.78</td>
<td>8.46 (8.21 - 8.75)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.49</td>
<td></td>
</tr>
<tr>
<td>2.79 (2.71 - 2.87)</td>
<td>2.98</td>
<td>2.88 (2.76 - 3.04)</td>
<td>2.94</td>
<td></td>
</tr>
<tr>
<td>6.04 (5.92 - 6.15)</td>
<td>6.54</td>
<td>6.46 (6.29 - 6.68)</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td>10.23 (10.06 - 10.35)</td>
<td>10.65</td>
<td>10.50 (10.35 - 10.65)</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>17.00 (16.48 - 17.41)</td>
<td>18.81</td>
<td>18.54 (18.19 - 18.89)</td>
<td>5.47</td>
<td></td>
</tr>
<tr>
<td>6.03 (5.96 - 6.11)</td>
<td>6.33</td>
<td>6.38 (6.20 - 6.58)</td>
<td>3.42</td>
<td></td>
</tr>
</tbody>
</table>
Supporting information for

*Equilibrium climate sensitivity estimated by equilibrating climate models*

Maria Rugenstein, Jonah Bloch-Johnson, Jonathan Gregory, Timothy Andrews, Thorsten Mauritsen, Chao Li, Thomas L. Frölicher, David Paynter, Gokhan Danabasoglu, Shuting Yang, Jean-Louis Dufresne, Long Cao, Gavin Schmidt, Ayako Abe-Ouchi, Olivier Geoffroy, Reto Knutti

Contents

1 Climate sensitivity definitions: $\Delta T_{\text{best est}}, \Delta T_{\text{est 1-150}},$ ECS, and other measures 2

2 Feedbacks 5
   2.1 Definitions 5
   2.2 Time evolution of the net feedback parameter and its components in all simulations 7
   2.3 Surface temperature patterns and local contribution to the feedback parameter 12

3 Simulation of FAMOUS with quadrupled CO$_2$ concentration 13

List of Figures

1 Ratios of different estimates of climate sensitivity 4
2 Schematic of feedback definitions 5
3 Local contribution to the global feedback in the control simulation 7
4 As Fig. 2a, but showing all simulations’ feedback evolution through time 8
5 Similar to Fig. 2b, but showing all simulations’ feedback changes and their uncertainties 9
6 As in Fig. 4, but showing the net cloud radiative effect only 9
7 As in Fig. 4, but showing the shortwave cloud radiative effect only 10
8 As in Fig. 4, but showing the longwave cloud radiative effect only 10
9 As in Fig. 4, but showing the net longwave component only 10
10 As in Fig. 4, but showing the longwave clear sky component only 11
11 As in Fig. 4, but showing the net shortwave component only 11
12 As in Fig. 4, but showing the shortwave clear sky component only 11
13 Main text Fig. 4 over Fig. 3 12
14 Main text Fig. 3 over Fig. 4 12
1 Climate sensitivity definitions: $\Delta T_{\text{best est}}$, $\Delta T_{\text{est 1-150}}$, ECS, and other measures

The term Equilibrium Climate Sensitivity (ECS) refers to the equilibrium global and annual mean surface temperature response to a doubling of CO$_2$ above preindustrial levels after a step forcing, incorporating only the “fast”, “Charney feedbacks” (Planck, surface albedo, water vapor, lapse rate, and clouds feedbacks). This definition excludes the use of extrapolations, energy balance models, statistical models, or scaling between forcing levels. Under this notion, ECS can not be time dependent, while other methods that approximate ECS from shorter time scales may be. Originally, ECS was defined for experiments with “slab/mixed-layer” oceans, in which equilibration takes only about 30 years (Charney et al., 1979). For clarity, the same definition should apply to GCM experiments, although they have to be far longer.

We define $\Delta T_{\text{best est}}$ as the best estimate of the equilibrium warming. For the equilibrated abrupt2x simulations $\Delta T_{\text{best est}}$ is ECS. The minimum requirement to contribute a simulation to LongRunMIP is that the simulations must have at least a thousand years with a constant forcing level and an unforced control simulation. Depending on the forcing level and model physics, 1000 years are not enough to fully equilibrate the CO$_2$ forcing (main text Fig.1). $\Delta T_{\text{best est}}$ is determined by regressing global annual means of TOA imbalance and surface temperature anomaly over the final 15% of warming. We average the final 50 years of each model integration, determine 85% of that value, and regress all year of the model integration which follow the first occurrence of that value. Alternatives would be to first smooth or fit the temperature curve until it is monotonic before determining the number of years for the regression. The number of years/points in this warming range depends on the internal variability, forcing level, and degree of equilibration of the simulations. For simulations with a strong curvature towards the end of the integration time (FAMOUS abrupt4x, CESM104 abrupt8x, CNRMCM61 abrupt4x, and IPSL abrupt4x), these choices make a difference in the estimates of equilibrium warming. Our choices result in a low estimate of $\Delta T_{\text{best est}}$, e.g., taking fewer years into account would increase $\Delta T_{\text{best est}}$.

We chose this definition of $\Delta T_{\text{best est}}$ to apply the same criterium for all simulations, independent how close to equilibrium they are. One could determine $\Delta T_{\text{best est}}$ also from the model output temperature alone (without the use of TOA or extrapolation methods). Whether or not a simulation is equilibrated in one field could be defined with a threshold criteria, i.e. the temperature increase or the TOA imbalance over a certain period should be (close to) zero. None of our results here depends qualitatively on the exact definition of $\Delta T_{\text{best est}}$.

Most other choices would increase the difference between $\Delta T_{\text{best est}}$ and measures of effective climate sensitivity determined from fewer years. Rugenstein et al. (2019) discusses model biases, energy leakages, and drifts in the LongRunMIP simulations and how to contribute to the archive.

We denote other estimates of ECS with $\Delta T_{\text{specification}}$, where specification indicates the method of estimation.

- $\Delta T_{\text{est 1-150}}$ is an estimate of equilibrium warming based on the energy balance model $N = F - \lambda T$, where $N$ is the TOA imbalance (in Wm$^{-2}$), $F$ is the forcing (in Wm$^{-2}$, intersect of the regression line with the vertical axis at $T = 0$), $\lambda$ the feedback parameter (in Wm$^{-2}$K$^{-1}$, the slope of regression line) and $T$ the temperature anomaly (in K). This model assumes that the global net feedback parameter is constant, which means that for any given change in surface temperature the restoration strength of the TOA imbalance is the same (Gregory et al., 2004; Roe, 2009; Andrews et al., 2012; Knutti and Rugenstein, 2015). To estimate $\Delta T_{\text{est 1-150}}$, we extrapolate the least-square linear regression of annual means of $N$ and $T$ for year 1-150 to $N=0$, i.e. we do not use information of the forcing. Note that for doubling CO$_2$ concentrations $\Delta T_{\text{est 1-150}}$ is often referred to as “effective climate sensitivity” (Gregory et al., 2004).

- $\Delta T_{\text{EBM-}}$ another estimate of equilibrium warming, is calculated with an energy balance model including ocean heat uptake efficacy (Winton et al., 2010; Geoffroy et al., 2013; Armour, 2017). This model takes into account the effect of the ocean heat uptake on the strength of the radiative feedbacks.
The model distinguishes between the constant feedback parameter valid for the equilibrium state and a transient feedback parameter, depending on the ocean heat uptake. The efficacy factor can also be interpreted as the effect of changing patterns of surface heat flux or surface temperature \cite{Winton2010, Rugenstein2016a, Rose2016, Haugstad2017}. We follow the procedure detailed out in Geoffroy et al. 2013 to obtain $\Delta T_{\text{EBM}} - \epsilon$, using annual means of year 1-150 for each forcing level separately.

- $\Delta T_{\text{est} \ 20-150}$ is again a least-square linear regression of annual means of TOA imbalance and surface temperature anomaly, exactly like $\Delta T_{\text{est} \ 1-150}$, but only taking into account years 20-150 \cite{Andrews2015, Armour2017}. The assumption is that atmospheric feedbacks only change as long as the ocean mixed layer is out of equilibrium (10-20 years) and are constant afterwards. $\Delta T_{\text{eff} \ 20-150}$ might be regarded as an approximation to the more physically motivated $\Delta T_{\text{EBM}} - \epsilon$, which does not prescribe the time scale of the mixed layer but finds it through a fitting method.

SM Fig.1 relates the different measures of climate sensitivity to $\Delta T_{\text{eff} \ 1-150}$. $\Delta T_{\text{EBM}} - \epsilon$ and $\Delta T_{\text{eff} \ 20-150}$ raise the climate sensitivity estimate, but less than half of the true value. There is a tendency that for lower forcing magnitudes, the estimates are closer to the true $\Delta T_{\text{best est}}$ value than for higher forcing magnitudes, probably because the simulations are already more equilibrated after 150 years.

Table 1 in the main text lists all numerical values of SM Fig.1 and uncertainty estimates of the regressions.
SM Fig. 1: Ratios of different estimates of climate sensitivity

a) Best estimate of equilibrium temperature, as in Fig.1

b) Two layer energy balance model fit to 150yr

c) Linear regression of years 20-150

d) Absolute values of climate sensitivity estimates

a) As in main text Fig.1, b) for $\Delta T_{EBM-\epsilon}$, and c) for $\Delta T_{eff\,20-150}$. The hashed bars denote the median of all simulations (1.17 (a), 1.06 (b), 1.07 (c)). Black dots in panel a-c indicate that the 90% confidence interval of the two measures do not overlap d) absolute values of panels a-c. $\Delta T_{best\,est}$ FAMOUS abrupt4x lies outside the depicted range in panel a (1.53) and d (11.2 K, 17.09 K).
2 Feedbacks

2.1 Definitions

Aside from the kernel (Shell et al., 2008; Soden et al., 2008) and partial radiative perturbation (Wetherald and Manabe, 1988; Meraner et al., 2013) methods, the most common way to estimate feedbacks is to linearly regress the TOA imbalance against the surface temperature anomaly (Gregory et al., 2004; Andrews et al., 2012). To facilitate comparison to existing literature (Andrews et al., 2015), we adopt the time periods of years 1-20 and 21-150 following a step forcing, and add a third time period, years 150-1000 years, which is covered by all simulations in the LongRunMIP ensemble. Rugenstein et al. (2019) shows some quantities for the simulations with integration time scales above 1000 years.

Fig. 2b, 3, and 4 in the main text use these time periods for the step forcing simulations only. SM Fig. 5 shows all simulations, including the ramp simulations. For the ramp simulations (ECEARTH, GFDLCM3, GFDLES2M, MIROC32) the first two time periods are taken from the model’s 150 year long abrupt4x simulation, while the third time period starts after the end of the ramp until year 1000 (see SM Table 1).

![Schematic of feedback definitions](image)

SM Fig. 2: Schematic of feedback definitions

**Local tangent feedback**

Main text Fig. 1 indicates that the feedback parameter $\lambda$ does not change on clear time scales, meaning the time periods chosen above are ad hoc rather than representing points of inflection. Thus, to estimate the degree of continuous change of the curvature in the temperature-TOA space we linearly regress global and annual means of the net TOA imbalance for a limited temperature anomaly window (blue shading in SM Fig. 2a), which is moved in steps of 0.1K throughout the temperature space to obtain the continuous local slope of the point cloud (Gregory et al., 2004; Knutti and Rugenstein, 2015; Rugenstein et al., 2016b). The slope is then plotted as a function of time based on the time at which that temperature anomaly window first occurs (main text Fig. 2a and SM Fig. 4).

We adjust the width of the temperature window over which the TOA imbalance is regressed based on the simulation (e.g., abrupt2x have smaller windows than abrupt16x) and temperature anomaly (early in the equilibration there are fewer points to regress, thus, the temperature window has to be wider than towards the end of the simulation, SM Fig. 2). The last regression starts at least 1K before the equilibrium temperature is reached so that the warming signal still dominates over the internal variability. Nevertheless, the closer to equilibration the larger the amount of internal variability regressed. We choose a conservative temperature window width; meaning, the variability in main text Fig. 2a and SM Fig. 4 does not capture a large amount of internal variability. The 2.5-97.5% uncertainty (shading) reflects the uncertainty of the regression of each bin in the temperature-TOA space. A reduced temperature window width would increase the overall change of the estimated feedback through time, especially beginning with larger magnitudes. The models’ feedbacks arising from internal variability – obtained by regressing the annual and global means of TOA onto temperature
anomalies in the control simulation – are indicated with circles at the right side of each panel. In the global mean, there is no obvious connection between the feedback arising from internal variability and the forced response (?)

**Finite difference feedback and internal variability**

We are using an additional feedback definition, which represents a change across a certain time period:

\[
\frac{(\text{TOA}_{\text{local}p2} - \text{TOA}_{\text{local}p1})}{(T_{\text{global}p2} - T_{\text{global}p1})},
\]

where \(p1\) and \(p2\) refer to a time average at the beginning and end of a period, and TOA and T to the top of the atmosphere and surface temperature annual anomalies. This approach is similar to the local tangent used in main text Fig. 2b and SM Fig. [5] but avoids taking into account a different number of years for the three time periods, which implies different degrees of internal variability can realize.

**Main text Fig.** 4 uses Eq. [1]. For the first time period, \(\overline{p2}\) is the average of years 10-30, while \(\overline{p1}\) is the forcing value, obtained by extrapolating the regression of the first 6 years to the vertical axis (Hansen et al., 2005). Defining \(\overline{p1}\) as the average of the first three, five, or ten years results in a similar pattern. For the second time period, \(\overline{p2}\) is the average of years 140-160, while \(\overline{p1}\) is the average of year 10-30. For the third time period, \(\overline{p2}\) is the average of years 800-1000 (to include models which only cover 1000 years), while \(\overline{p1}\) is again the average of years 130-160. The pattern looks similar for slight changes in the averaging periods. Global numbers in main text Fig. 2b and Fig. 4 do not agree due the differences in the feedback definition, the exact number of years taken into account, the use of mean versus median, and (most importantly) the weighting of the models.

**Main text Fig.** 3 uses, for consistency, the same time periods, but the overall pattern evolution looks similar for other periods. The patterns show the warming distribution (unit K/K) which occurred since the last time period. Rugenstein et al. (2019) shows the absolute anomalies (unit K) of the surface temperature evolution (unit K).

We use the finite difference approach to filter out feedbacks associated with internal variability. In the control simulation, the restoring feedbacks are dominated by an ENSO/IPO dominated feedback pattern, with a small contribution of the high latitudes and the North Atlantic region (SM Fig. [3]). Using the local tangent approach over a number of years in different stages of the simulation will capture different relative amounts of forced response and internal variability, but only the feedbacks associated with the forced response are relevant for estimating the ECS. Intriguingly, the spatial contribution (but not the magnitude) of the centennial time scale feedbacks (main text Fig. 4b) resembles the feedback pattern in unperturbed simulations, across the Pacific and Indian ocean basins, but not in high latitudes (SM Fig. [3] Brown et al. (2014); Zhou et al. (2015); Ceppi and Gregory (2017)).

The characteristic feedback evolution in Fig. 4a-c and the “pattern flip” in Fig. 4d,e occurs also if we (1) regress over the same amount of temperature anomaly (i.e. the first, second, and third 1/3 of warming achieved in the first 1000 years of each simulation) or (2) use the linear regression of all years in each period (years 1-20, 21-150, 151-1000), i.e. using the same feedback definition as used in Fig. 2b, as opposed to the finite difference approach. The “pattern flip” does not occur, when (1) comparing regressions of the same number of years, e.g., subsequent 200 years, as a similar amount of internal variability is regressed and dominates the overall response or (2) when an increasing number of years are regressed, e.g., years 1-20, 1-150, 1-500, 1-1000 (as this feedback measure does not concern a change between two time periods within the simulation, but always relative to the same control state). Thus, a remaining open question is why the forced response in the Pacific on century (as opposed to millennial) time scales is so similar to the unforced feedback pattern.
SM Fig. 3: Local contribution to the global feedback in the control simulation
Model-mean of annual-mean local net TOA anomaly regressed against annual and global mean temperature anomaly. Note the change in scale compared to the other feedback maps.

Robust behavior across models
Although the spatial patterns of changing temperature and radiative feedbacks vary a lot among models, they are qualitatively consistent:

- in simulations with different forcing level for the same model, e.g., MPIESM12 Abrupt2x, 4x, 8x, and 16x;
- in a sub-sample of simulations across models, e.g., in only the Abrupt2x or only the Abrupt4x simulations;
- in a sub-sample of models, e.g. picking randomly ten different simulations;
- when weighting or not weighting the models according to the number of simulations they contributed;
- for different averaging periods or time periods, e.g., if for Fig. 3b and 4b $\overline{pT}$ was the average of years 30-50 and $\overline{pT}$ was the average of years 160-180.

However, regionally – on an ocean basin or continental scale or below – all these methodological choices result in different local feedback strengths for each model and the model average. This emphasizes the necessity of a suite of models and a careful use and comparison of feedback definitions.

2.2 Time evolution of the net feedback parameter and its components in all simulations
SM Fig. 4 and 5 show all simulations from main text Fig. 2a and b, while SM Fig. 6-12 show all components of the net TOA balance shown in main text Fig. 4.
SM Fig. 4: As Fig. 2a, but showing all simulations’ feedback evolution through time

Gray bars indicate the time periods used in the main text. For the “ramp” simulations, year 1 is the first year after the ramp. The feedbacks are more equilibrated at that point compared to the step forcing simulations. Circles on the right side of each panel indicate the feedbacks based on the control simulation. Lines within the circles and shading show the 2.5-97.5% confidence interval.
SM Fig. 5: Similar to Fig. 2b, but showing all simulations’ feedback changes and their uncertainties. The dots are the difference of the feedbacks between the two time scales indicated in the header. The lines are 5-95% uncertainty intervals, solid lines mean the difference is significant, dashed lines mean it is not significant.

SM Fig. 6: As in Fig. 4, but showing the net cloud radiative effect only.
SM Fig. 7: As in Fig. 4, but showing the shortwave cloud radiative effect only.

SM Fig. 8: As in Fig. 4, but showing the longwave cloud radiative effect only.

SM Fig. 9: As in Fig. 4, but showing the net longwave component only.
SM Fig. 10: As in Fig. 4, but showing the longwave clear sky component only.

SM Fig. 11: As in Fig. 4, but showing the net shortwave component only.

SM Fig. 12: As in Fig. 4, but showing the shortwave clear sky component only.
2.3 Surface temperature patterns and local contribution to the feedback parameter

SM Fig. 13: Main text Fig. 4 over Fig. 3
Local contribution to feedback parameter (main text Fig. 4; $\text{Wm}^{-2}\text{K}^{-1}$) in black contours over surface temperature pattern (main text Fig. 3; K/K) in colored contours.

SM Fig. 14: Main text Fig. 3 over Fig. 4
Surface temperature pattern (main text Fig. 3; K/K) in black contours over local contribution to feedback parameter (main text Fig. 4; $\text{Wm}^{-2}\text{K}^{-1}$) in colored contours.
3 Simulation of FAMOUS with quadrupled CO$_2$ concentration

The simulation *abrupt4x* from the model FAMOUS warms extraordinarily (Fig. 1 and Table 1 in the main text). After 3000 years of simulation it does not equilibrate. However, we have not found any error in the model integration which could have caused its behavior. The warming response is mainly due to the short wave cloud radiative effect (SM Fig.4), which is positive throughout the simulation, and the short wave clear sky feedback, which increases more strongly than in other models within the first 200 years. The ocean stratifies strongly and takes up heat much slower than other simulations (not shown). The simulation shows a similar change in patterns to the other models and to the *abrupt2x* simulation of FAMOUS. When the fields are normalized with the global temperature, they are qualitatively and quantitatively similar to other models, see e.g. Fig. 6 in Rugenstein et al. (2019). Thus, we conclude that the results are physically possible, of course with the caveats that apply to any of GCMs in the LongRunMIP or CMIP5/6 archives. Even if the evolution of the climate system looks unlikely and out of the usual range, we hope the simulation will be useful to study state dependence (Meraner et al., 2013; Jonko et al., 2012; Caballero and Huber, 2013; Bloch-Johnson et al., 2015; Gregory et al., 2015; Rohrschneider et al., 2019), as opposed to the time dependence (“pattern effect”) we focus on in this manuscript.

References


