Equilibrium climate sensitivity estimated by equilibrating climate models

Maria Rugenstein[1,2*], Jonah Bloch-Johnson[3], Jonathan Gregory[3,4], Timothy Andrews[4], Thorsten Mauritsen[5], Chao Li[2], Thomas L. Frölicher[6,7], David Paynter[8], Gokhan Danabasoglu[9], Shuting Yang[10], Jean-Louis Dufresne[11], Long Cao[12], Gavin A. Schmidt[13], Ayako Abe-Ouchi[14], Olivier Geoffroy[15], and Reto Knutti[1]

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

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 CO_2 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 CO_2 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

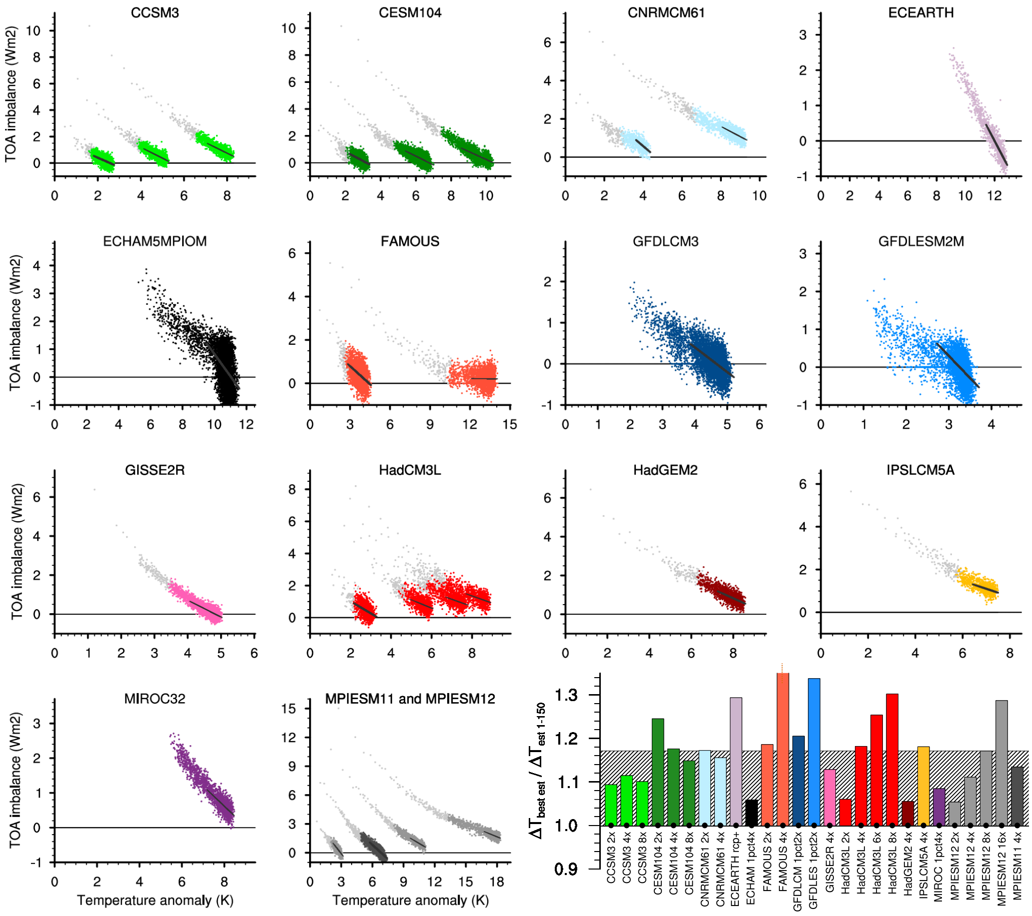


Figure 1. Evolution of global and annual mean top of the atmosphere (TOA) imbalance and surface temperature anomalies (14 small panels). The first 150 years of step forcing simulations are depicted in light gray. For experiments which are not step forcing simulations only the period after stabilizing 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 $\Delta T_{\mathrm{best\ est}}$ / $\Delta T_{\mathrm{est\ 1-150}}$, see text for definitions. A dot at the lower end of the bar indicates with 90% confidence that $\Delta T_{\mathrm{best\ est}}$ and $\Delta T_{\mathrm{est\ 1-150}}$ 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 *abrupt4x* 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 increments 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. Rugenstein 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 "centennial 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 "$\Delta T_{\textit{[specification]}}$" for a true or estimated equilibrium warming, for a range of forcing levels not only CO_2 doubling (Table 1). We define the best estimate of equilibrium warming, $\Delta T_{\mathrm{best\,est}}$, 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 $\Delta T_{\mathrm{best\,est}}$) 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 "$\Delta T_{\mathrm{est}\,1-150}$".

For simulations that have gradual forcings (e.g., *1pct2x*), we use 150 year long step forcing simulations of the same model to calculate $\Delta T_{\mathrm{est}\,1-150}$. The median increase of $\Delta T_{\mathrm{best\,est}}$ over $\Delta T_{\mathrm{est}\,1-150}$ is 17% for all simulations and 16% for the subset of CO_2 doubling and quadrupling simulations. While $\Delta T_{\mathrm{est}\,1-150}$ 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 $\Delta T_{\mathrm{best\,est}}$: Using years 20-150 in linear regression ($\Delta T_{\mathrm{est}\,20-150}$; e.g.,Andrews et al. (2015); Armour (2017)) results ina median equilibrium warming estimate which is 7% lower than $\Delta T_{\mathrm{best\,est}}$, both for all simulations and the subset of CO_2 doubling and quadrupling. The two-layer model including ocean heat uptake efficacy ($\Delta T_{\mathrm{EBM}-\epsilon}$; 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 then $\Delta T_{\mathrm{best\,est}}$, again both for all simulations and the subset of CO_2 doubling and quadrupling. Both methods are described and illustrated in the supplemental material.

$\Delta T_{\mathrm{best\,est}}$ 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 $\Delta T_{\mathrm{best\,est}}$ 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 $\Delta T_{\mathrm{best\,est}}$.

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 (δTOA/δT) com-

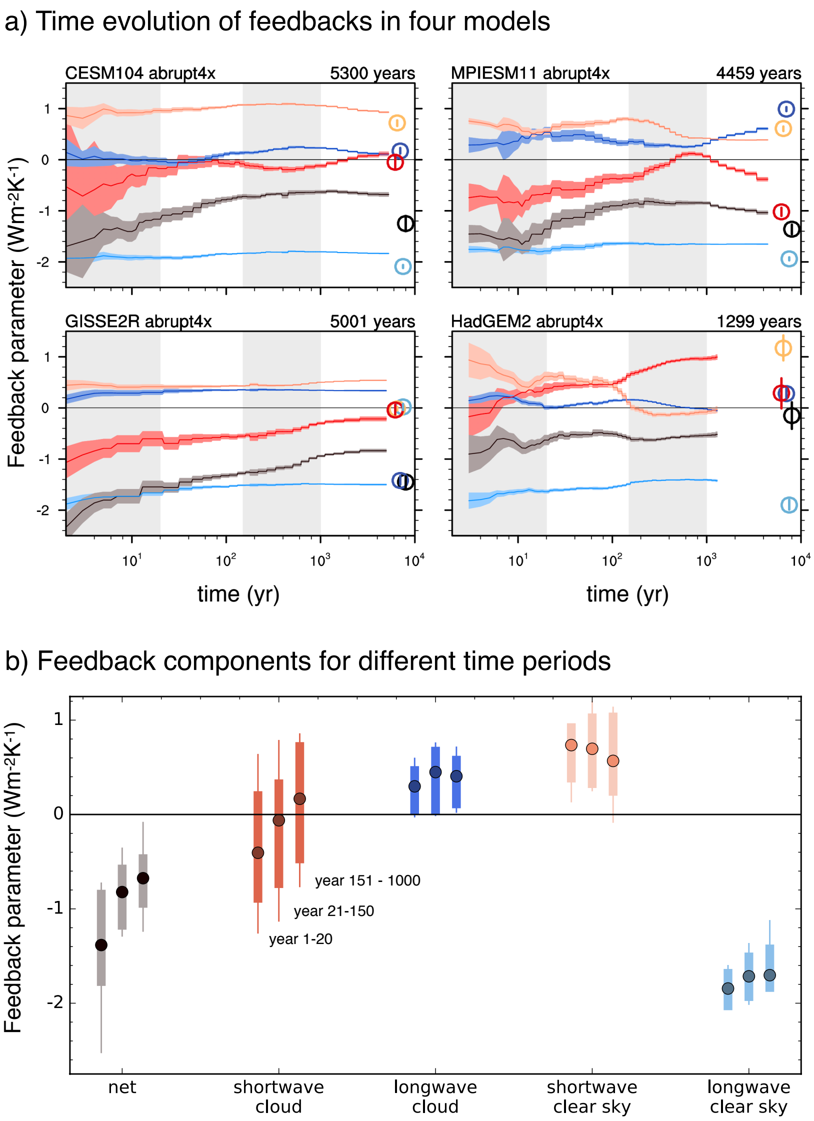


Figure 2. a) Time evolution of global feedbacks in four characteristic models. Net TOA feedback (gray) is the sum of its components: the cloud effects in the shortwave (red) and longwave (blue), and clear sky feedbacks in the shortwave (salmon) and longwave (light blue). Circles at the right of each 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, GISSE2R, HadCM3L, HadGEM2, IPSLCM5A, 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.

puted 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”;

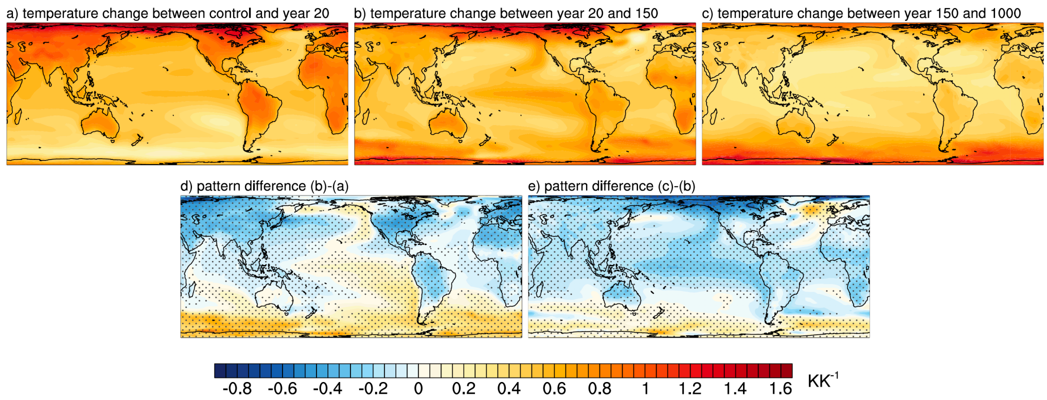


Figure 3. Multi-model mean normalized patterns of surface warming (local warming divided by global warming) between the average of (a) the control simulation and year 15-25, (b) year 15-25 and 140-160, (c) year 140-160 and 800-1000, and their differences (d and e) for the same models and simulations 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,

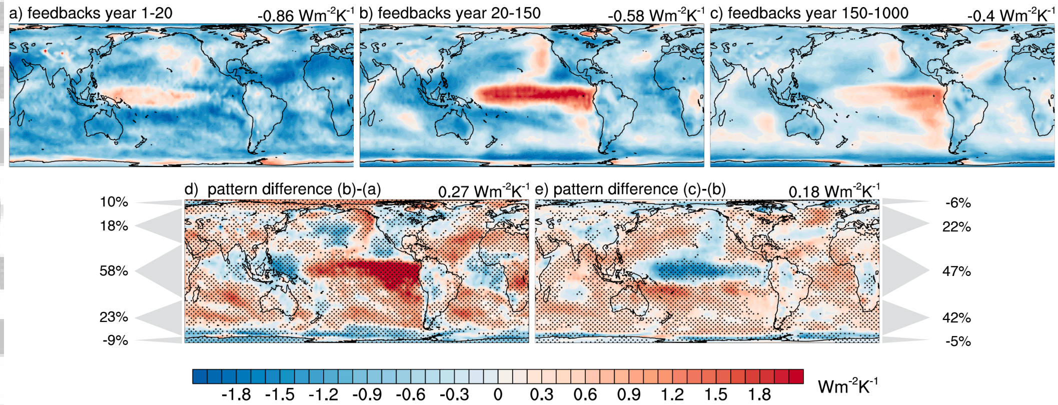


Figure 4. Time evolution of feedback patterns. Model-mean of local contribution to the change in global feedbacks (local TOA anomaly divided by global warming during the period indicated in the panel titles; see text for definitions) (a–c) and their differences (d, e). The global feedback value is shown 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.s 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; Meraner et al., 2013; Block & Mauritsen, 2013; Andrews et al., 2015; Ceppi & Gregory, 2019), especially in the Pacific (Andrews & Webb, 2018; Ceppi & Gregory, 2017).

The simulations robustly shows 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 considered.

We find that the longwave clear sky feedback does moderately increase in many models 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 quantify 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 centennial 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 patters 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 $\Delta T_{est\,1-150}$ 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 sponsored 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 Programme (GA01101). TLF acknowledges support from the Swiss National Science Foundation 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 Excellence 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

Allen, M. R., & Frame, D. J. (2007). Call off the quest. *Science*, *318*(5850), 582-583. Retrieved from http://www.sciencemag.org/content/318/5850/582.short doi: 10.1126/science.1149988

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models. *Journal of Climate*, *28*(4), 1630-1648. Retrieved from http://dx.doi.org/10.1175/JCLI-D-14-00545.1

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*(9). Retrieved from http://dx.doi.org/10.1029/2012GL051607

Andrews, T., & Webb, M. J. (2018). The Dependence of Global Cloud and Lapse Rate Feedbacks on the Spatial Structure of Tropical Pacific Warming. *Journal of Climate*, *31*(2), 641-654. Retrieved from https://doi.org/10.1175/JCLI-D-17-0087.1

Armour, K. C. (2017). Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, *7*, 331 EP -. Retrieved from http://dx.doi.org/10.1038/nclimate3278

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-Varying Climate Sensitivity from Regional Feedbacks. *Journal of Climate*, *26*(13), 4518–4534. Retrieved from http://dx.doi.org/10.1175/JCLI-D-12-00544.1

Armour, K. C., Marshall, J., Scott, J. R., Donohoe, A., & Newsom, E. R. (2016). Southern ocean warming delayed by circumpolar upwelling and equatorward transport. *Nature Geosci*, *9*(7), 549–554. Retrieved from http://dx.doi.org/10.1038/ngeo2731

Barsugli, J. J., & Sardeshmukh, P. D. (2002). Global atmospheric sensitivity to tropical sst anomalies throughout the indo-pacific basin. *Journal of Climate*, *15*(23), 3427-3442. Retrieved from https://doi.org/10.1175/1520-0442(2002)015<3427:GASTTS>2.0.CO;2

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, *42*(12), 4973– 4980. Retrieved from http://dx.doi.org/10.1002/2015GL064240 (2015GL064240)

Bloch-Johnson, J., Rugenstein, M., & Abbot, D. S. (in revision). Spatial radiative feedbacks from interannual variability using multiple regression. *Journal of Climate*.

Block, K., & Mauritsen, T. (2013). Forcing and feedback in the MPI-ESM-LR coupled model under abruptly quadrupled CO_2. *Journal of Advances in Modeling Earth Systems*, *5*(4), 676–691. Retrieved from http://dx.doi.org/10.1002/jame.20041

Brient, F., & Schneider, T. (2016). Constraints on Climate Sensitivity from Space-Based Measurements of Low-Cloud Reflection. *Journal of Climate*, *29*(16), 5821-5835. Retrieved from https://doi.org/10.1175/JCLI-D-15-0897.1

Brown, P. T., Li, W., Li, L., & Ming, Y. (2014). Top-of-atmosphere radiative contribution to unforced decadal global temperature variability in climate models. *Geophysical Research Letters*, *41*(14), 5175–5183. Retrieved from http://dx.doi.org/10.1002/2014GL060625

Caballero, R., & Huber, M. (2013). State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(35), 14162–14167. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761583/

Calel, R., & Stainforth, D. A. (2017). On the Physics of Three Integrated Assessment Models. *Bulletin of the American Meteorological Society*, *98*(6), 1199-1216. Retrieved from https://doi.org/10.1175/BAMS-D-16-0034.1

Ceppi, P., & Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and earth's observed radiation budget. *Proceedings of the National Academy of Sciences*, *114*(50), 13126–13131. Retrieved from https://www.pnas.org/content/114/50/13126

Ceppi, P., & Gregory, J. M. (2019). A refined model for the Earth's global energy balance. *Climate Dynamics*. Retrieved from https://doi.org/10.1007/s00382-019-04825-x

Charney, J., Arakawa, A., Baker, D., Bolin, B., Dickinson, R. E., Goody, R., ... Wunsch, C. (1979). *Carbon Dioxide and Climate: A Scientific Assessment* (Tech. Rep.). Washington, DC.: National Academy of Science.

Collins, M., Knutti, R., Arblaster, J. M., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., ... Wehner, M. F. (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. In T. Stocker et al. (Eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Colman, R., & Hanson, L. (2017). On the relative strength of radiative feedbacks under climate variability and change. *Climate Dynamics*, *49*(5), 2115–2129. Retrieved from https://doi.org/10.1007/s00382-016-3441-8

Dong, Y., Proistosescu, C., Armour, K. C., & Battisti, D. S. (2019). Attributing Historical and Future Evolution of Radiative Feedbacks to Regional Warming Patterns using a Green's Function Approach: The Preeminence of the Western Pacific. *Journal of Climate*, *0*(0), null. Retrieved from https://doi.org/10.1175/JCLI-D-18-0843.1

Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, *43*(24), 12,614-12,623. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL071930

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved from https://www.geosci-model-dev.net/9/1937/2016/

Fedorov, A. V., Burls, N. J., Lawrence, K. T., & Peterson, L. C. (2015). Tightly linked zonal and meridional sea surface temperature gradients over the past five million years. *Nature Geoscience*, *8*, 975 EP -. Retrieved from https://doi.org/10.1038/ngeo2577

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., ... Rummukainen, M. (2013). Evaluation of Climate Models. In T. Stocker et al. (Eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Forster, P. M. (2016). Inference of climate sensitivity from analysis of earth's energy budget. *Annual Review of Earth and Planetary Sciences*, *44*(1), 85-106. Retrieved from http://dx.doi.org/10.1146/annurev-earth-060614-105156

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D., & Tytéca, S. (2013). Transient Climate Response in a Two-Layer Energy-Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs. *Journal of Climate*, *26*(6), 1859–1876. Retrieved from http://dx.doi.org/10.1175/JCLI-D-12-00196.1

Geoffroy, O., Saint-Martin, D., Olivié, D. J. L., Voldoire, A., Bellon, G., & Tytéca, S. (2013). Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments. *Journal of Climate*, *26*(6), 1841–1857. Retrieved from http://dx.doi.org/10.1175/JCLI-D-12-00195.1

Good, P., Andrews, T., Chadwick, R., Dufresne, J.-L., Gregory, J. M., Lowe, J. A., … Shiogama, H. (2016). nonlinMIP contribution to CMIP6: model intercomparison project for non-linear mechanisms: physical basis, experimental design and analysis principles (v1.0). *Geoscientific Model Development*, *9*(11), 4019–4028. Retrieved from https://www.geosci-model-dev.net/9/4019/2016/

Good, P., Lowe, J. A., Andrews, T., Wiltshire, A., Chadwick, R., Ridley, J. K., … Shiogama, H. (2015, 02). Nonlinear regional warming with increasing co2 concentrations. *Nature Clim. Change*, *5*(2), 138–142.

Gregory, J. M., & Andrews, T. (2016). Variation in climate sensitivity and feedback parameters during the historical period. *Geophysical Research Letters*, *43*(8), 3911–3920. Retrieved from http://dx.doi.org/10.1002/2016GL068406

Gregory, J. M., Andrews, T., & Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO_2. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *373*(2054). Retrieved from http://rsta.royalsocietypublishing.org/content/373/2054/20140417

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., … Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3). Retrieved from http://dx.doi.org/10.1029/2003GL018747

Grose, M. R., Gregory, J., Colman, R., & Andrews, T. (2018). What Climate Sensitivity Index Is Most Useful for Projections? *Geophysical Research Letters*, *45*(3), 1559–1566. Retrieved from http://dx.doi.org/10.1002/2017GL075742

Hansen, J., Lacis, A., Rind, D., Russel, G., Stone, P., Fung, I., … Lerner, J. (1984). Analysis of feedback mechanisms. In: Climate processes and climate sensitivity. In J. Hansen & T. Takahashi (Eds.), *Climate sensitivity: Analysis of feedback mechanisms* (Vol. 5, pp. 130–163). American Geophysical Union, Washington, DC: AGU Geophysical Monograph 29, Maurice Ewing.

Hargreaves, J. C., & Annan, J. D. (2016). Could the pliocene constrain the equilibrium climate sensitivity? *Climate of the Past*, *12*(8), 1591–1599. Retrieved from http://www.clim-past.net/12/1591/2016/

Haugstad, A. D., Armour, K. C., Battisti, D. S., & Rose, B. E. J. (2017). Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, *44*(14), 7455-7463. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074372`

Houghton, J., Jenkins, G. J., & Ephraums, J. J. (1990). Report prepared for Intergovernmental Panel on Climate Change by Working Group I. In *Climate Change: The IPCC Scientific Assessment* (p. 410). Cambridge University Press.

Jansen, M. F., Nadeau, L.-P., & Merlis, T. M. (2018). Transient versus Equilibrium Response of the Ocean's Overturning Circulation to Warming. *Journal of Climate*, *31*(13), 5147-5163. Retrieved from `https://doi.org/10.1175/JCLI-D-17-0797.1`

Jonko, A. K., Shell, K. M., Sanderson, B. M., & Danabasoglu, G. (2013). Climate Feedbacks in CCSM3 under Changing CO2 Forcing. Part II: Variation of Climate Feedbacks and Sensitivity with Forcing. *Journal of Climate*, *26*(9), 2784-2795. Retrieved from `http://dx.doi.org/10.1175/JCLI-D-12-00479.1`

Kang, S. M., & Xie, S.-P. (2014). Dependence of Climate Response on Meridional Structure of External Thermal Forcing. *Journal of Climate*, *27*(14), 5593–5600. Retrieved from `http://dx.doi.org/10.1175/JCLI-D-13-00622.1`

Knutson, T. R., & Manabe, S. (1995). Time-Mean Response over the Tropical Pacific to Increased $C0_2$ in a Coupled Ocean-Atmosphere Model. *Journal of Climate*, *8*(9), 2181-2199. Retrieved from `https://doi.org/10.1175/1520-0442(1995)008<2181:TMROTT>2.0.CO;2`

Knutti, R., & Rugenstein, M. A. A. (2015). Feedbacks, climate sensitivity and the limits of linear models. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *373*(2054). doi: 10.1098/rsta.2015.0146

Knutti, R., Rugenstein, M. A. A., & Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, *10*, 727 EP -. Retrieved from `http://dx.doi.org/10.1038/ngeo3017`

Kohyama, T., Hartmann, D. L., & Battisti, D. S. (2017). La Nina–like Mean-State Response to Global Warming and Potential Oceanic Roles. *Journal of Climate*, *30*(11), 4207-4225. Retrieved from `https://doi.org/10.1175/JCLI-D-16-0441.1`

Lewis, N., & Curry, J. A. (2015). The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dynamics*, *45*(3), 1009–1023. Retrieved from `http://dx.doi.org/10.1007/s00382-014-2342-y`

Li, C., Storch, J.-S., & Marotzke, J. (2013). Deep-ocean heat uptake and equilibrium climate response. *Climate Dynamics*, *40*(5-6), 1071-1086. Retrieved from `http://`

dx.doi.org/10.1007/s00382-012-1350-z

Liu, F., Lu, J., Garuba, O. A., Huang, Y., Leung, L. R., Harrop, B. E., & Luo, Y. (2018). Sensitivity of Surface Temperature to Oceanic Forcing via q-Flux Green's Function Experiments. Part II: Feedback Decomposition and Polar Amplification. *Journal of Climate*, *31*(17), 6745-6761. Retrieved from https://doi.org/10.1175/JCLI-D-18-0042.1

Luo, Y., Lu, J., Liu, F., & Garuba, O. (2017). The Role of Ocean Dynamical Thermostat in Delaying the El Niño–Like Response over the Equatorial Pacific to Climate Warming. *Journal of Climate*, *30*(8), 2811-2827. Retrieved from https://doi.org/10.1175/JCLI-D-16-0454.1

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... Roeckner, E. (2018). Developments in the mpi-m earth system model version 1.2 (mpi-esm 1.2) and its response to increasing co2. *Journal of Advances in Modeling Earth Systems*, *0*(ja). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001400

Meraner, K., Mauritsen, T., & Voigt, A. (2013). Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, *40*(22), 5944–5948. Retrieved from http://dx.doi.org/10.1002/2013GL058118

Millar, R. J., Otto, A., Forster, P. M., Lowe, J. A., Ingram, W. J., & Allen, M. R. (2015). Model structure in observational constraints on transient climate response. *Climatic Change*, *131*(2), 199–211. Retrieved from https://doi.org/10.1007/s10584-015-1384-4

Murphy, J. M. (1995). Transient Response of the Hadley Centre Coupled Ocean-Atmosphere Model to Increasing Carbon Dioxide. Part 1: Control Climate and Flux Adjustment. *Journal of Climate*, *8*(1), 36–56. Retrieved from http://dx.doi.org/10.1175/1520-0442(1995)008<0036:TROTHC>2.0.CO;2

Myhre, G., Highwood, E. J., Shine, K. P., & Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophysical Research Letters*, *25*(14), 2715-2718. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98GL01908

Otto, A., Otto, F. E. L., Boucher, O., Church, J., Hegerl, G., Forster, P. M., ... Allen, M. R. (2013). Energy budget constraints on climate response. *Nature Geosci*, *6*(6), 415–416. Retrieved from http://dx.doi.org/10.1038/ngeo1836

Paynter, D., Frölicher, T. L., Horowitz, L. W., & Silvers, L. G. (2018). Equilibrium Climate Sensitivity Obtained From Multimillennial Runs of Two GFDL Climate Models. *Journal of Geophysical Research: Atmospheres*, *123*(4), 1921-1941. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JD027885

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*(7). Retrieved from http://advances.sciencemag.org/content/3/7/e1602821

Rind, D., Schmidt, G. A., Jonas, J., Miller, R., Nazarenko, L., Kelley, M., & Romanski, J. (2018). Multicentury Instability of the Atlantic Meridional Circulation in Rapid Warming Simulations With GISS ModelE2. *Journal of Geophysical Research: Atmospheres*, *123*(12), 6331-6355. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JD027149

Roe, G. (2009). Feedbacks, timescales, and seeing red. *Annual Review of Earth and Planetary Sciences*, *37*(1), 93-115. Retrieved from http://dx.doi.org/10.1146/annurev.earth.061008.134734

Rohrschneider, T., Stevens, B., & Mauritsen, T. (2019, 02). On simple representations of the climate response to external radiative forcing. *Climate Dynamics*. doi: 10.1007/s00382-019-04686-4

Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N., & Koll, D. D. B. (2014). The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, *41*(3), 1071–1078. Retrieved from http://dx.doi.org/10.1002/2013GL058955

Rose, B. E. J., & Rayborn, L. (2016). The Effects of Ocean Heat Uptake on Transient Climate Sensitivity. *Current Climate Change Reports*, 1–12. Retrieved from http://dx.doi.org/10.1007/s40641-016-0048-4

Rose, B. E. J., & Rencurrel, M. C. (2016). The Vertical Structure of Tropospheric Water Vapor: Comparing Radiative and Ocean-Driven Climate Changes. *Journal of Climate*, *29*(11), 4251-4268. Retrieved from http://dx.doi.org/10.1175/JCLI-D-15-0482.1

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., ... Yang, S. (2019). LongRunMIP – motivation and design for a large collection of millennial-length GCM simulations. *Bulletin of the American Meteorological Society*, *0*(0), null. Retrieved from https://doi.org/10.1175/BAMS-D-19-0068.1

Rugenstein, M. A. A., Caldeira, K., & Knutti, R. (2016). Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophysical Research Letters*, *43*(18), 9877–9885. Retrieved from http://dx.doi.org/10.1002/2016GL070907

Rugenstein, M. A. A., Gregory, J. M., Schaller, N., Sedláček, J., & Knutti, R. (2016). Multiannual Ocean–Atmosphere Adjustments to Radiative Forcing. *Journal of Climate*, *29*(15), 5643-5659. Retrieved from http://dx.doi.org/10.1175/JCLI-D-16-0312.1

Rugenstein, M. A. A., Sedláček, J., & Knutti, R. (2016). Nonlinearities in patterns of long-term ocean warming. *Geophysical Research Letters*, *43*(7), 3380–3388. Retrieved from http://dx.doi.org/10.1002/2016GL068041

Salzmann, M. (2017). The polar amplification asymmetry: role of Antarctic surface height. *Earth System Dynamics*, *8*(2), 323–336. Retrieved from https://www.earth-syst-dynam.net/8/323/2017/

Senior, C. A., & Mitchell, J. F. B. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, *27*(17), 2685–2688. Retrieved from http://dx.doi.org/10.1029/2000GL011373

Shiogama, H., Stone, D., Emori, S., Takahashi, K., Mori, S., Maeda, A., … Allen, M. R. (2016). Predicting future uncertainty constraints on global warming projections. *Scientific Reports*, *6*, 18903 EP -.

Soden, B. J., & Held, I. M. (2006). An Assessment of Climate Feedbacks in Coupled Ocean-Atmosphere Models. *Journal of Climate*, *19*(14), 3354-3360. Retrieved from http://dx.doi.org/10.1175/JCLI3799.1

Song, X., & Zhang, G. J. (2014). Role of Climate Feedback in El Nino-like SST Response to Global Warming. *Journal of Climate*. Retrieved from http://dx.doi.org/10.1175/JCLI-D-14-00072.1

Stocker, T. F. e. a. (2013). *Climate Change 2013: The Physical Science Basis.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Stouffer, R., & Manabe, S. (2003). Equilibrium response of thermohaline circulation to large changes in atmospheric CO_2 concentration. *Climate Dynamics*, *20*(7-8), 759-773. Retrieved from http://dx.doi.org/10.1007/s00382-002-0302-4

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2011). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. Retrieved from http://dx.doi.org/10.1175/BAMS-D-11-00094.1

Tian, B. (2015). Spread of model climate sensitivity linked to double-intertropical convergence zone bias. *Geophysical Research Letters*, *42*(10), 4133–4141. Retrieved from http://dx.doi.org/10.1002/2015GL064119

Tierney, J. E., Haywood, A. M., Feng, R., Bhattacharya, T., & Otto-Bliesner, B. L. (2019). Pliocene Warmth Consistent With Greenhouse Gas Forcing. *Geophysical Research Letters*, *46*(15), 9136-9144. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083802

Trossman, D. S., Palter, J. B., Merlis, T. M., Huang, Y., & Xia, Y. (2016). Large-scale ocean circulation-cloud interactions reduce the pace of transient climate change. *Geophysical Research Letters*, *43*(8), 3935–3943. Retrieved from http://dx.doi.org/10.1002/2016GL067931

Wetherald, R. T., & Manabe, S. (1988). Cloud feedback processes in a general circulation model. *Journal of the Atmospheric Sciences*, *45*(8), 1397–1416. Retrieved from http://dx.doi.org/10.1175/1520-0469(1988)045<1397:CFPIAG>2.0.CO;2

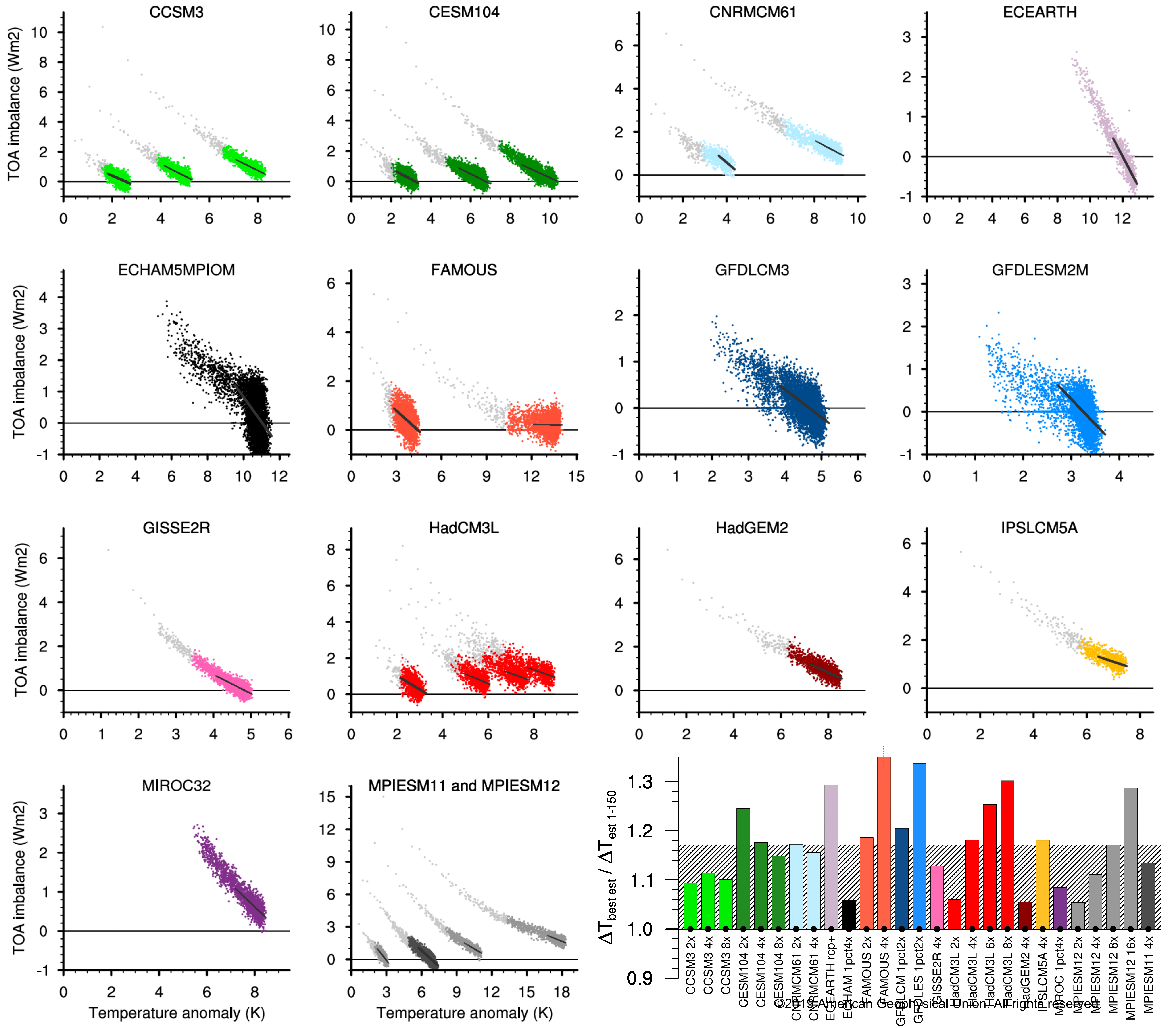
Winton, M., Takahashi, K., & Held, I. M. (2010). Importance of Ocean Heat Uptake Efficacy to Transient Climate Change. *Journal of Climate*, *23*(9), 2333-2344. Retrieved from http://dx.doi.org/10.1175/2009JCLI3139.1

Zhou, C., Zelinka, M. D., Dessler, A. E., & Klein, S. A. (2015). The relationship between inter-annual and long-term cloud feedbacks. *Geophysical Research Letters*. Retrieved from http://dx.doi.org/10.1002/2015GL066698 (2015GL066698)

Zhou, C., Zelinka, M. D., & Klein, S. A. (2016). Impact of decadal cloud variations on the Earth's energy budget. *Nature Geosci*, *9*(12), 871–874. Retrieved from http://dx.doi.org/10.1038/ngeo2828

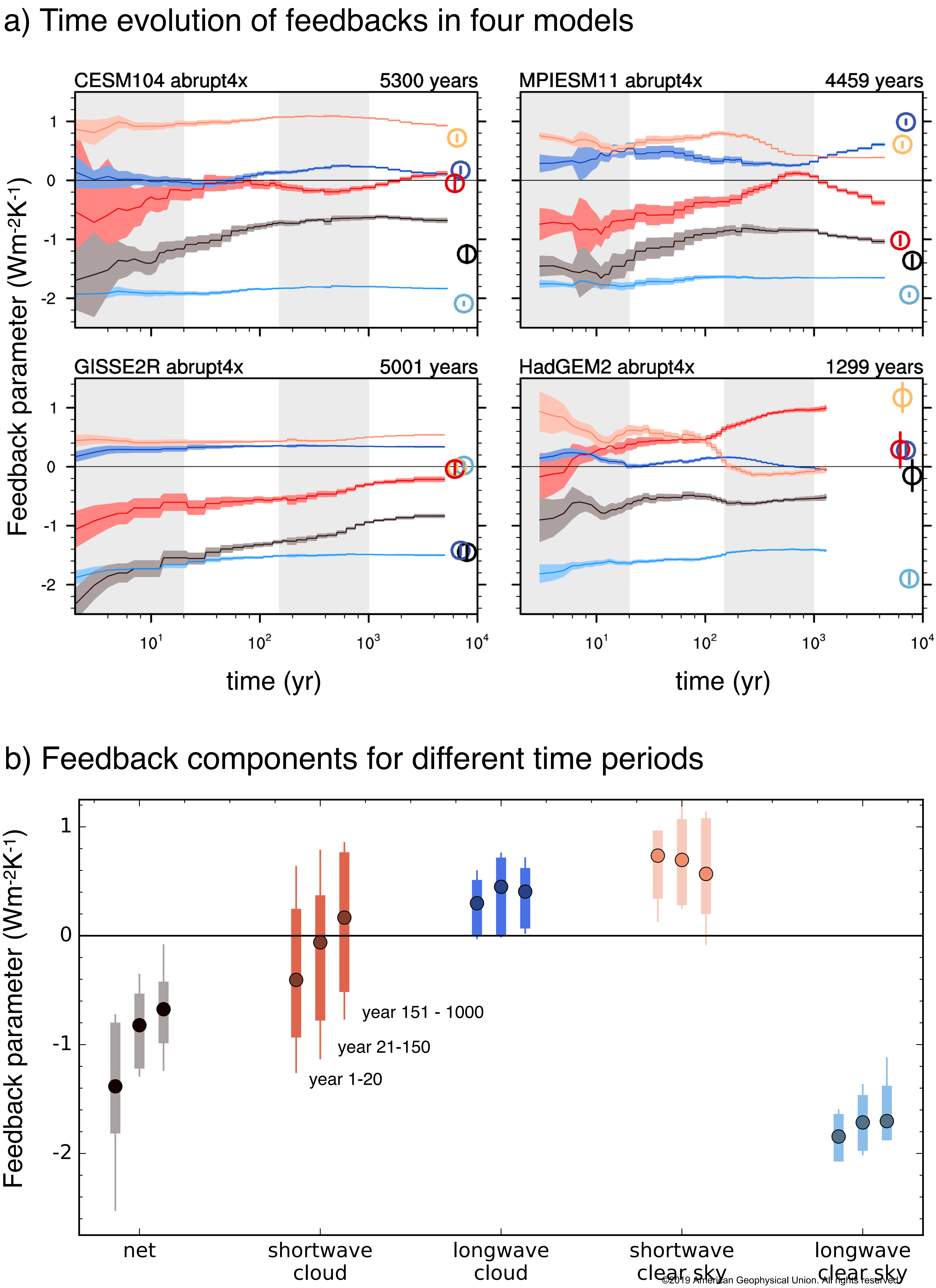
Zhou, C., Zelinka, M. D., & Klein, S. A. (2017). Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a green's function approach. *Journal of Advances in Modeling Earth Systems*, *9*(5), 2174–2189. Retrieved from http://dx.doi.org/10.1002/2017MS001096

Figure 1.



CCSM3
CESM104
CNRMCM61
ECEARTH
ECHAM5MPIOM
FAMOUS
GFDLCM3
GFDLESM2M
GISSE2R
HadCM3L
HadGEM2
IPSLCM5A
MIROC32
MPIESM11 and MPIESM12
TOA imbalance (Wm2)
Temperature anomaly (K)
$\Delta T_{\text{best est}} / \Delta T_{\text{est 1-150}}$
CCSM3 2x
CCSM3 4x
CCSM3 8x
CESM104 2x
CESM104 4x
CESM104 8x
CNRMCM61 2x
CNRMCM61 4x
ECEARTH rcp+
ECHAM 1pct4x
FAMOUS 2x
FAMOUS 4x
GFDLCM 1pct2x
GFDLES 1pct2x
GISSE2R 4x
HadCM3L 2x
HadCM3L 4x
HadCM3L 6x
HadCM3L 8x
HadGEM2 4x
IPSLCM5A 4x
MIROC 1pct4x
MPIESM12 2x
MPIESM12 4x
MPIESM12 8x
MPIESM12 16x
MPIESM11 4x

Figure 2.



a) Time evolution of feedbacks in four models
CESM104 abrupt4x
5300 years
MPIESM11 abrupt4x
4459 years
GISSE2R abrupt4x
5001 years
HadGEM2 abrupt4x
1299 years
Feedback parameter ($Wm^{-2}K^{-1}$)
time (yr)
b) Feedback components for different time periods
year 151 - 1000
year 21-150
year 1-20
net
shortwave cloud
longwave cloud
shortwave clear sky
longwave clear sky

Figure 3.

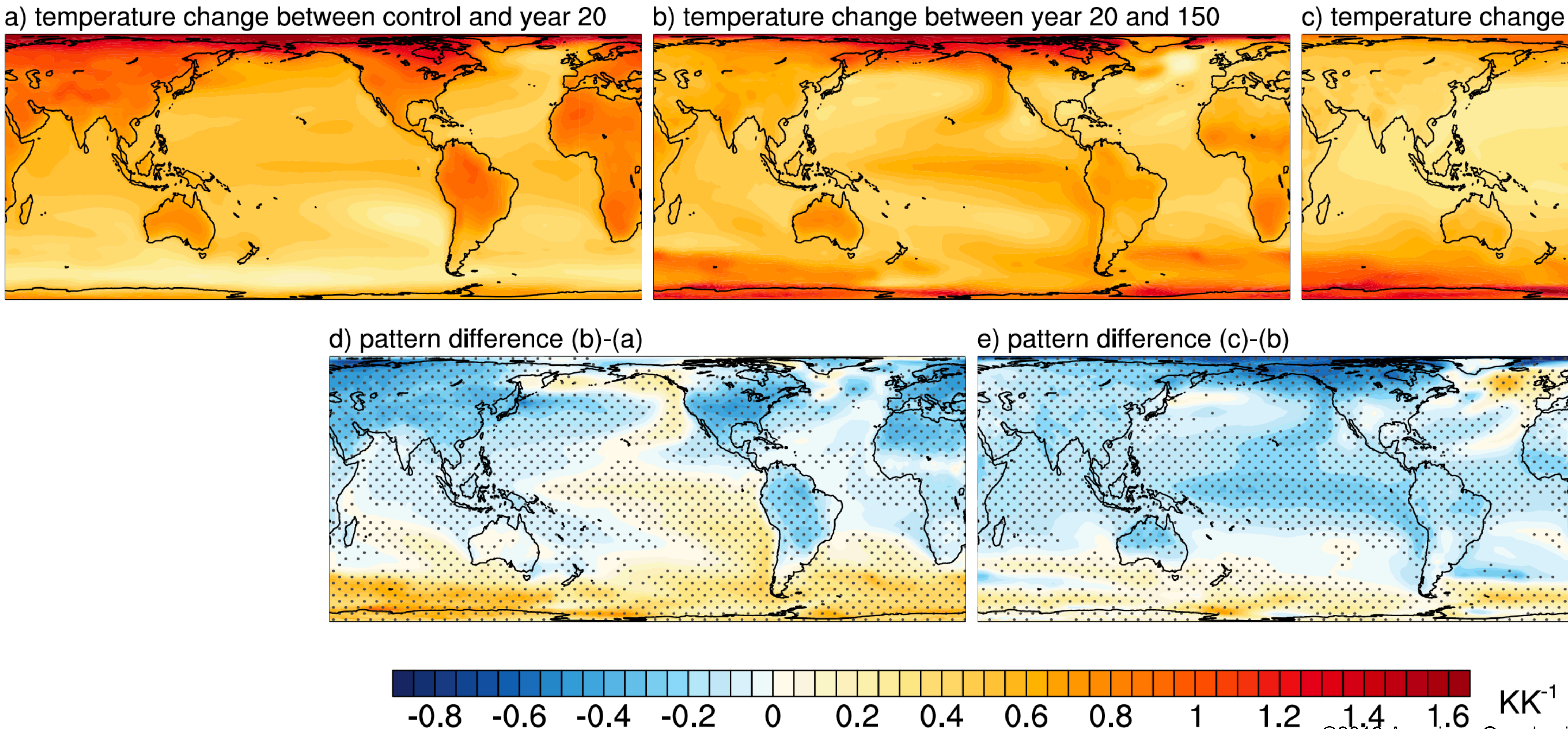

a) temperature change between control and year 20
b) temperature change between year 20 and 150
c) temperature change between year 150 and 1000
d) pattern difference (b)-(a)
e) pattern difference (c)-(b)
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
KK-1

Figure 4.

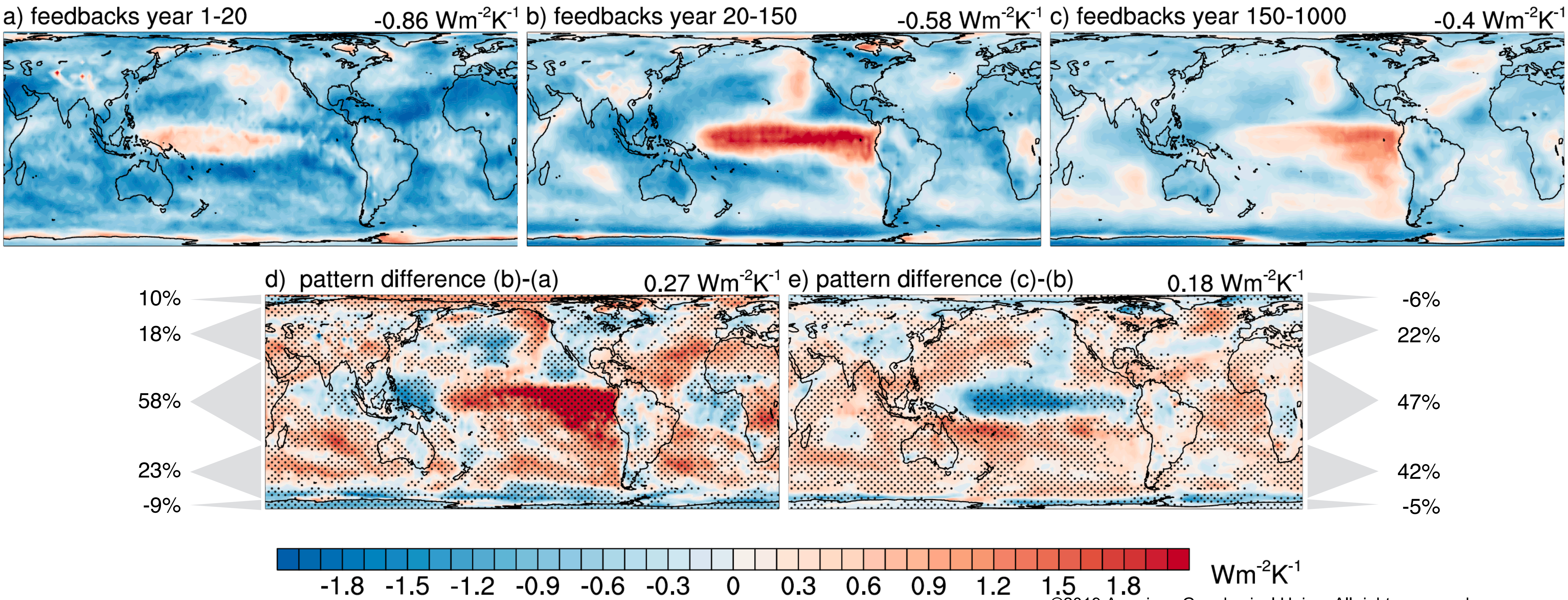

a) feedbacks year 1-20
-0.86 Wm-2K-1
b) feedbacks year 20-150
-0.58 Wm-2K-1
c) feedbacks year 150-1000
-0.4 Wm-2K-1
d) pattern difference (b)-(a)
0.27 Wm-2K-1
e) pattern difference (c)-(b)
0.18 Wm-2K-1
10%
18%
58%
23%
-9%
-6%
22%
47%
42%
-5%
-1.8
-1.5
-1.2
-0.9
-0.6
-0.3
0
0.3
0.6
0.9
1.2
1.5
1.8
Wm-2K-1

Table 1: Estimates of equilibrium

Table 1: Estimates of equilibrium warming and their uncertainties for each simulation

Model long name **Short name**	Simulation	Length	Control length	$\Delta T_{best\ est}$ Best estimate of equilibrium warming
CCSM3 **CCSM3**	abrupt2x	3000	1530	2.57 (2.55 - 2.58)
	abrupt4x	2120		5.45 (5.42 - 5.48)
	abrupt8x	1450		8.85 (8.80 - 8.89)
CESM 1.0.4 **CESM104**	abrupt2x	2500	1320	3.2 (3.18 - 3.23)
	abrupt4x	5300		6.75 (6.73 - 6.76)
	abrupt8x	4000		10.36 (10.34 - 10.39)
CNRM-CM6-1 **CNRMCM6**	abrupt2x	750	2000	4.83 (4.71 - 4.98)
	abrupt4x	1850		11.14 (11.03 - 11.26)
EC-Earth-PISM **ECEARTH**	*abrupt4x	150	508	
	RCP8.5+	1270		11.98 (11.97 - 12.00)
ECHAM5/MPIOM **ECHAM5**	*abrupt4x	1000	100	11.04 (11.02 - 11.07)
	1pct4x	6080		11.66 (11.44 - 11.92)
FAMOUS **FAMOUS**	abrupt2x	3000	3000	4.40 (4.4 - 4.46)
	abrupt4x	3000		17.09 (16.04 - 19.19)
GFDL-CM3 **GFDLCM3**	*abrupt4x	150	5200	
	1pct2x	5000		4.67 (4.66 - 4.68)
GFDL-ESM2M **GFDLESM2M**	*abrupt4x	150	1340	
	1pct2x	4500		3.25 (3.24 - 3.25)
GISS-E2-R **GISSE2R**	abrupt4x	5000	5225	4.83 (4.87 - 4.88)
HadCM3L **HadCM3L**	abrupt2x	1000	1000	3.34 (3.27 - 3.42)
	abrupt4x	1000		6.89 (6.76 - 7.06)
	abrupt6x	1000		9.29 (9.06 - 9.59)
	abrupt8x	1000		10.82 (10.57 - 11.12)
HadGEM2-ES **HadGEM2**	abrupt4x	1328	239	9.54 (9.44 - 9.65)
IPSL-CM5A-LR **IPSLCM5A**	abrupt4x	1000	1000	9.53 (9.32 - 9.77)
MIROC 3.2 **MIROC32**	*abrupt4x	150	681	-
	1pct4x	2000		8.97 (8.93 - 9.01)
MPI-ESM 1.2 **MPIESM12**	abrupt2x	1000	1237	2.94 (2.92 - 2.96)
	abrupt4x	1000		6.71 (6.67-6.7)
	abrupt8x	1000		11.98 (11.91 - 12.06)
	abrupt16x	1000		21.88 (21.64 - 22.15)
MPI-ESM 1.1 **MPIESM11**	abrupt4x	4459	2000	6.84 (6.84 - 6.85)

warming and their uncertainties for each simulation

$\Delta T_{est\ 1-150}$ Estimated equilibrium warming from year 1-150	$\Delta T_{EBM-\epsilon\ 1-150}$ Energy balance model with ocean heat uptake efficacy	$\Delta T_{est\ 20-150}$ Estimated equilibrium warming from year 20-150	Estimated equilibrium warming for doubling CO_2
2.35 (2.26 - 2.47)	2.75	2.49 (2.34 - 2.79)	2.57
4.89 (4.76 - 5.01)	5.32	5.23 (5.03 - 5.48)	2.73
8.04 (7.90 - 8.18)	8.52	8.44 (8.26 - 8.68)	2.95
2.57 (2.50 - 2.67)	2.57	2.62 (2.50 - 2.82)	3.20
5.74 (5.61 - 5.89)	6.08	6.30 (6.07 - 6.61)	3.38
9.02 (8.89 - 9.18)	9.53	9.64 (9.45 - 9.88)	3.46
4.12 (3.92 - 4.33)	4.43	4.6 (4.29 - 5.03)	4.83
9.64 (9.44 - 9.84)	10.02	9.64 (9.32 - 10.05)	5.70
6.61 (6.46 - 6.72)	7.03	6.81 (6.65 - 6.99)	-
-	-	-	4.27
10.43 (10.1 - 10.89)	9.90	9.88 (9.44 - 10.73)	5.83
-	-	-	5.52
3.71 (3.56 - 3.93)	4.00	4.10 (3.76 - 4.75)	4.40
11.20 (11.00 - 11.42)	12.19	11.87 (11.51 - 12.33)	8.55
7.75 (7.42 - 8.02)	8.40	8.27 (7.96 - 8.67)	-
-	-	-	4.67
4.86 (4.76 - 5.00)	5.07	5.20 (4.96 - 5.51)	-
-	-	-	3.25
4.28 (4.12 - 4.42)	4.49	4.62 (4.51 - 4.74)	2.44
3.15 (3.00 - 3.36)	3.34	3.48 (3.21 - 3.98)	3.34
5.83 (5.71 - 5.99)	6.30	6.50 (6.15 - 7.10)	3.45
7.41 (7.21 - 7.61)	8.10	8.14 (7.60 - 9.29)	3.60
8.31 (8.13 - 8.47)	9.25	9.49 (8.97 - 10.41)	3.61
9.04 (8.59 - 9.37)	10.82	11.04 (10.23 - 12.22)	4.77
8.07 (7.93 - 8.24)	8.40	8.56 (8.25 - 8.93)	4.76
8.27 (8.18 - 8.39)	8.78	8.46 (8.21 - 8.75)	-
-	-	-	4.49
2.79 (2.71 - 2.87)	2.98	2.88 (2.76 - 3.04)	2.94
6.04 (5.92 - 6.15)	6.54	6.46 (6.29 - 6.68)	3.35
10.23 (10.06 - 10.35)	10.65	10.50 (10.35 - 10.65)	4.00
17.00 (16.48 - 17.41)	18.81	18.54 (18.19 - 18.89)	5.47
6.03 (5.96 - 6.11)	6.33	6.38 (6.20 - 6.58)	3.42

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

List of Figures

1 Climate sensitivity definitions: $\Delta T_{\mathrm{best\,est}}$, $\Delta T_{\mathrm{est\,1-150}}$, 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 $\Delta T_{\mathrm{best\,est}}$ as the **best estimate of the equilibrium warming**. For the equilibrated *abrupt2x* simulations $\Delta T_{\mathrm{best\,est}}$ 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). $\Delta T_{\mathrm{best\,est}}$ 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 is it 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 $\Delta T_{\mathrm{best\,est}}$, e.g., taking fewer years into account would increase $\Delta T_{\mathrm{best\,est}}$. We chose this definition of $\Delta T_{\mathrm{best\,est}}$ to apply the same criterium for all simulations, independent how close to equilibrium they are. One could determine $\Delta T_{\mathrm{best\,est}}$ 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. Non of our results here depends qualitatively on the exact definition of $\Delta T_{\mathrm{best\,est}}$. Most other choices would increase the difference between $\Delta T_{\mathrm{best\,est}}$ 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 $\Delta T_{[specification]}$, where *specification* indicates the method of estimation.

- $\mathbf{\Delta T}_{\mathrm{est\,1-150}}$ is an estimate of equilibrium warming based on the **energy balance model** $\mathrm{N} = \mathrm{F} - \lambda\,\mathrm{T}$, where N is the TOA imbalance (in $\mathrm{Wm^{-2}}$), F is the forcing (in $\mathrm{Wm^{-2}}$, intersect of the regression line with the vertical axis at T = 0), λ the feedback parameter (in $\mathrm{Wm^{-2}K^{-1}}$, 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 $\Delta T_{\mathrm{est\,1-150}}$, 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 $\Delta T_{\mathrm{est\,1-150}}$ is often referred to as "effective climate sensitivity" (Gregory et al., 2004).

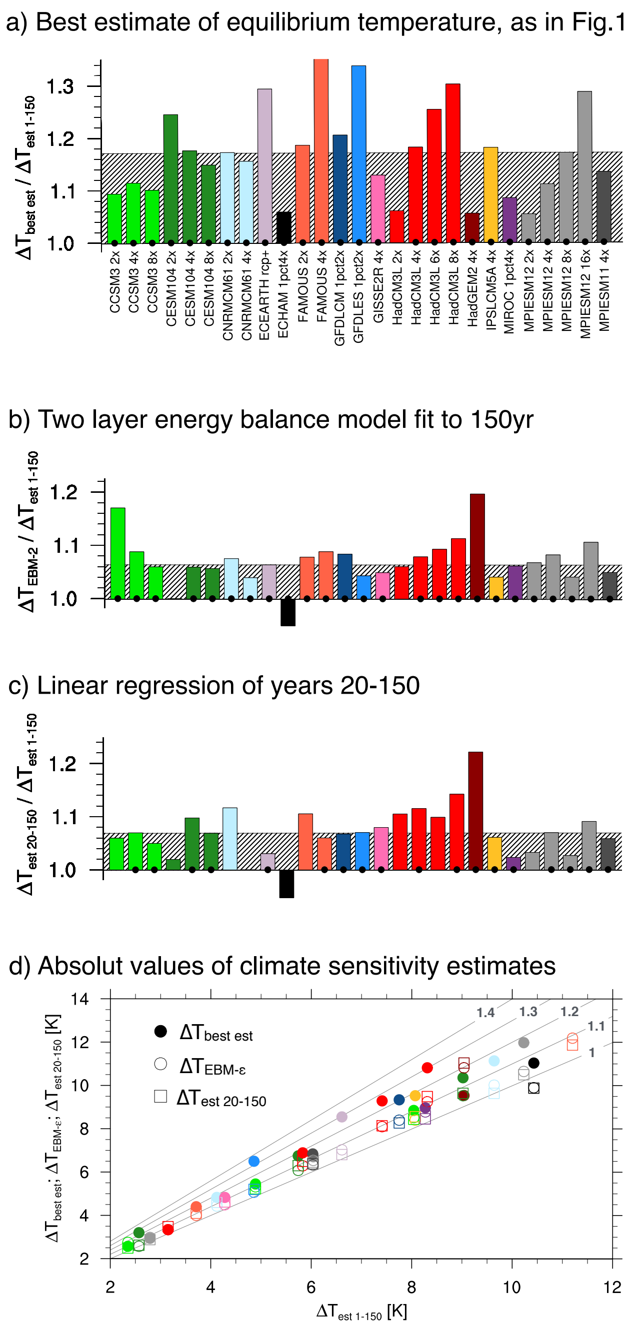
- $\mathbf{\Delta T}_{\mathrm{EBM-\epsilon}}$, another estimate of equilibrium warming, is calculated with an **energy balance model including ocean heat uptake efficacy** (Winton et al., 2010; Geoffroy et al., 2013; Armour, 2017). This model takes into account the effect of the ocean heat uptake on 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 (Winton et al., 2010; Rugenstein et al., 2016a; Rose and Rencurrel, 2016; Haugstad et al., 2017). We follow the procedure detailed out in Geoffroy et al. 2013 to obtain $\Delta\mathrm{T}_{\mathrm{EBM}-\epsilon}$, using annual means of year 1-150 for each forcing level separately.

- $\Delta\mathbf{T}_{\mathrm{est}\,20-150}$ is again a least-square linear regression of annual means of TOA imbalance and surface temperature anomaly, exactly like $\Delta\mathbf{T}_{\mathrm{est}\,1-150}$, but only taking into account years 20-150 (Andrews et al., 2015; Armour, 2017). 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. $\Delta\mathrm{T}_{\mathrm{eff}\,20-150}$ might be regarded as an approximation to the more physically motivated $\Delta\mathrm{T}_{\mathrm{EBM}-\epsilon}$, 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 $\Delta\mathrm{T}_{\mathrm{eff}\,1-150}$. $\Delta\mathrm{T}_{\mathrm{EBM}-\epsilon}$ and $\Delta\mathrm{T}_{\mathrm{eff}\,20-150}$ 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 $\Delta\mathrm{T}_{\mathrm{best\,est}}$ 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

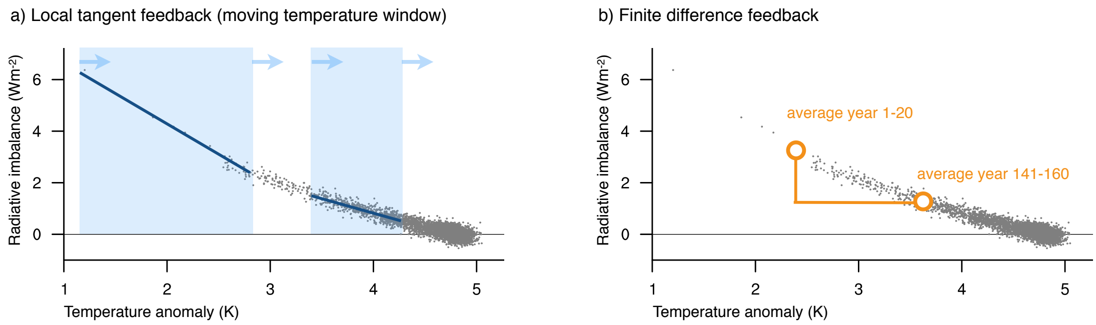
a) As in main text Fig. 1, b) for $\Delta T_{EBM-\epsilon}$, and c) for $\Delta T_{eff\,20-150}$. The hashed bars denotes the median of all simulations (1.17 (a), 1.06 (b), 1.07 (c)). Black dots in panel a-c indicate that the 90% confidence interval of the two measures do not overlap d) absolute values 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 LongRunMIPensemble. 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, GFDLESM2M, 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).



SM Fig. 2: Schematic of feedback definitions

Local tangent feedback

Main text Fig. 1 indicates that the feedback parameter λ 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 0.1 K 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:

$$(\overline{\mathrm{TOA}}_{\mathrm{local}\,p2} - \overline{\mathrm{TOA}}_{\mathrm{local}\,p1})/(\overline{\mathrm{T}}_{\mathrm{global}\,p2} - \overline{\mathrm{T}}_{\mathrm{global}\,p1}), \tag{1}$$

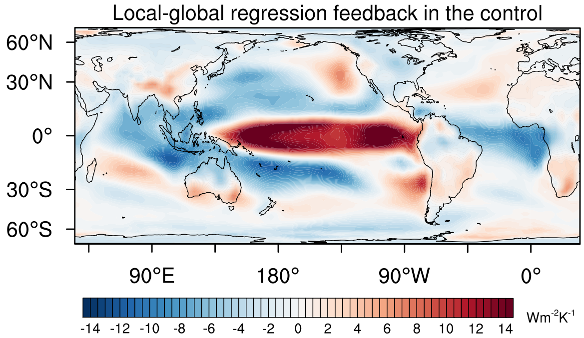
where $p1$ and $p2$ refer to a time average at the beginning and end of a period, and TOA and T to the top of the atmosphere and surface temperature annual anomalies. This approach is similar to the local tangent used in main text Fig. 2b and SM Fig. 5, but avoids taking into account a different number 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

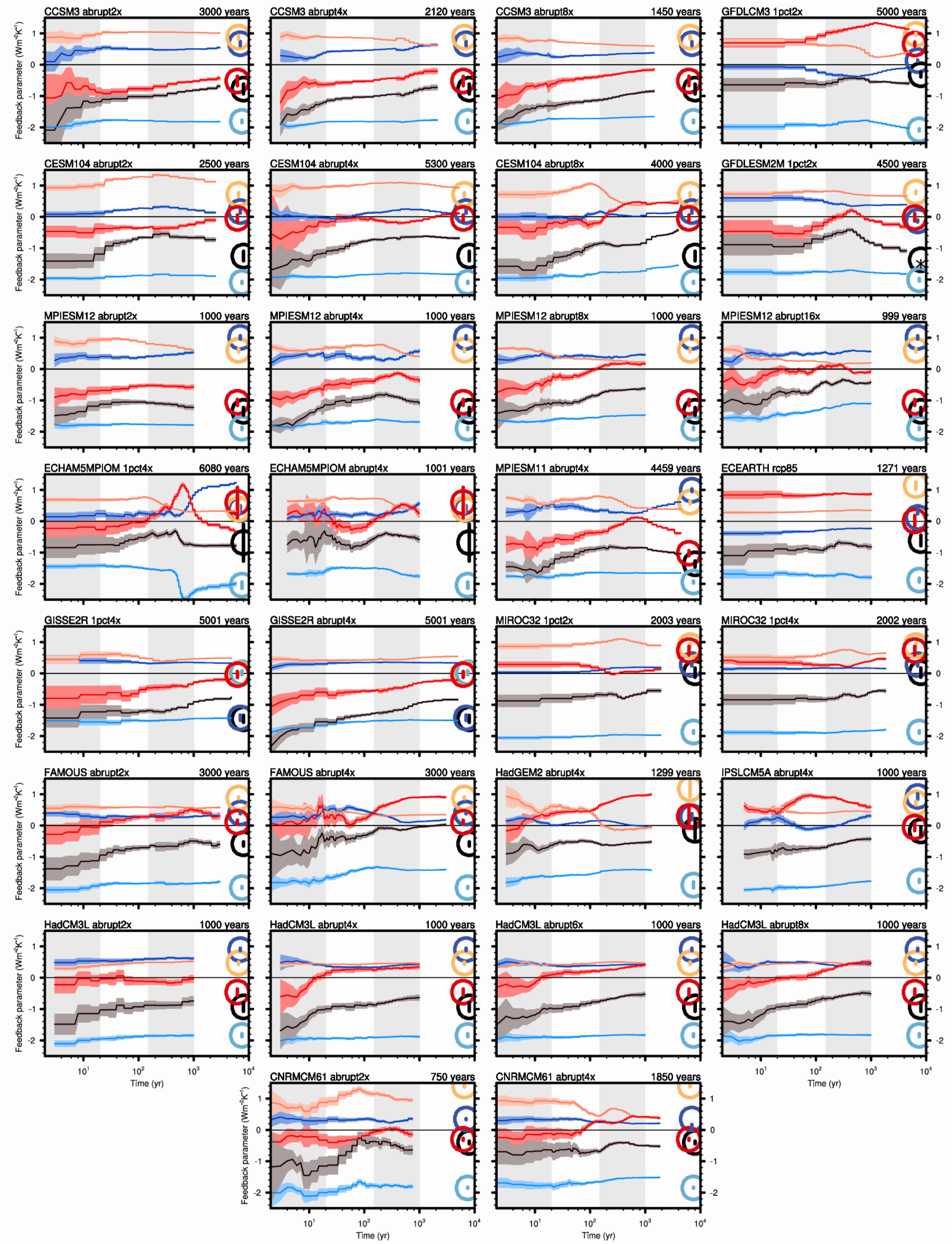
Although the spatial patterns of changing temperature and radiative feedbacks vary a lot among models, they are qualitatively consistent:

- 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 $\overline{p1}$ was the average of years 30-50 and $\overline{p2}$ 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.

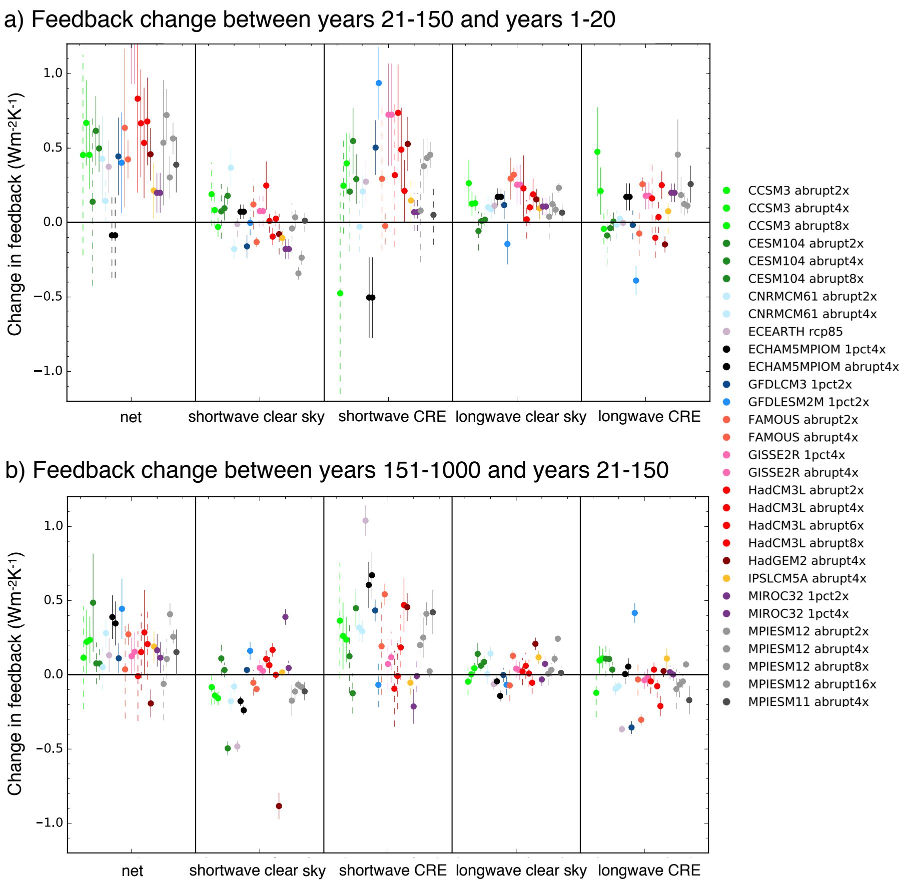
2.2 Time evolution of the net feedback parameter and its components in all simulations

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.

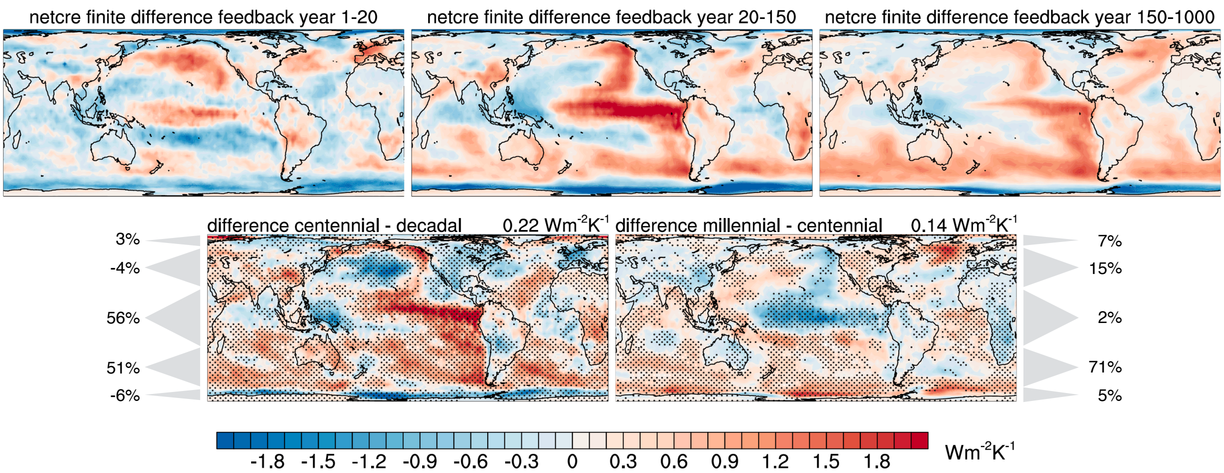


SM Fig. 4: As Fig. 2a, but showing all simulations' feedback evolution through time

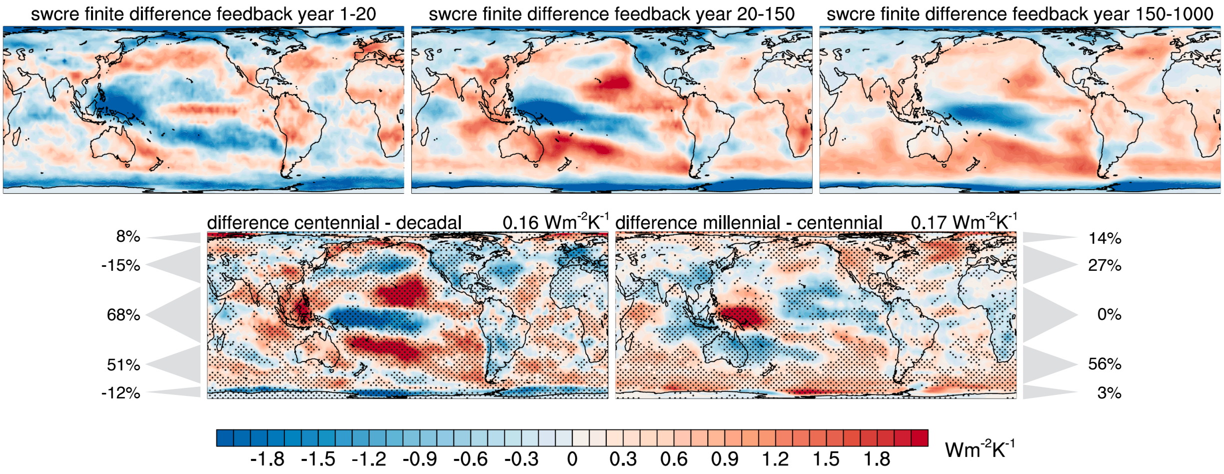
Gray bars indicate the time periods used in the main text. For the "ramp" simulations, year 1 is the first year after the ramp. The feedbacks are more equilibrated at that point compared to the step forcing simulations. Circles on the right side of each panel indicate the feedbacks based on the control simulation. Lines within the circles and shading show the 2.5-97.5% confidence interval.



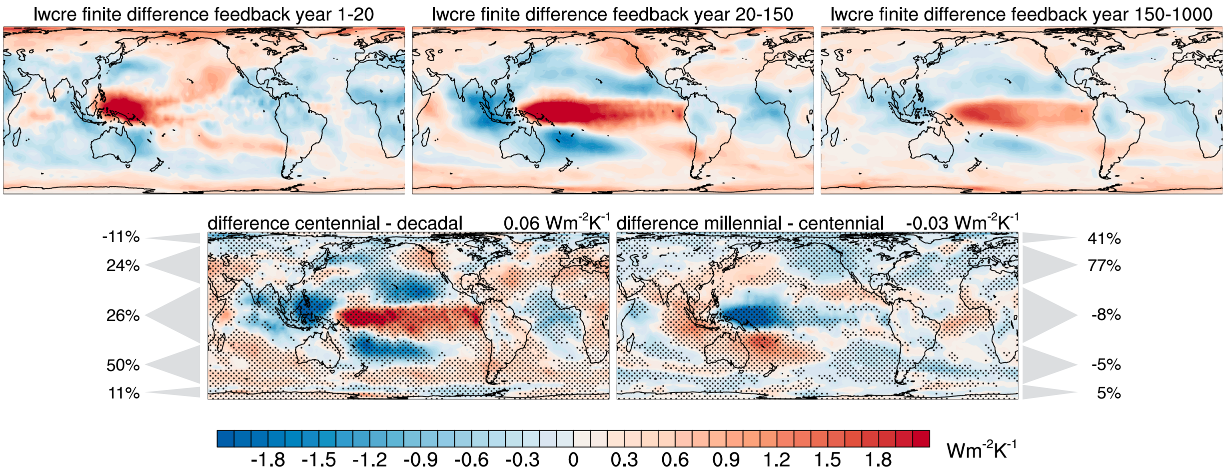
SM Fig. 5: Similar to Fig. 2b, but showing all simulations' feedback changes and their uncertainties The dots are the *difference* of the feedbacks between the two time scales indicated in the header. The lines are 5-95% uncertainty intervals, solid lines mean the difference is significant, dashed lines mean it is not significant.



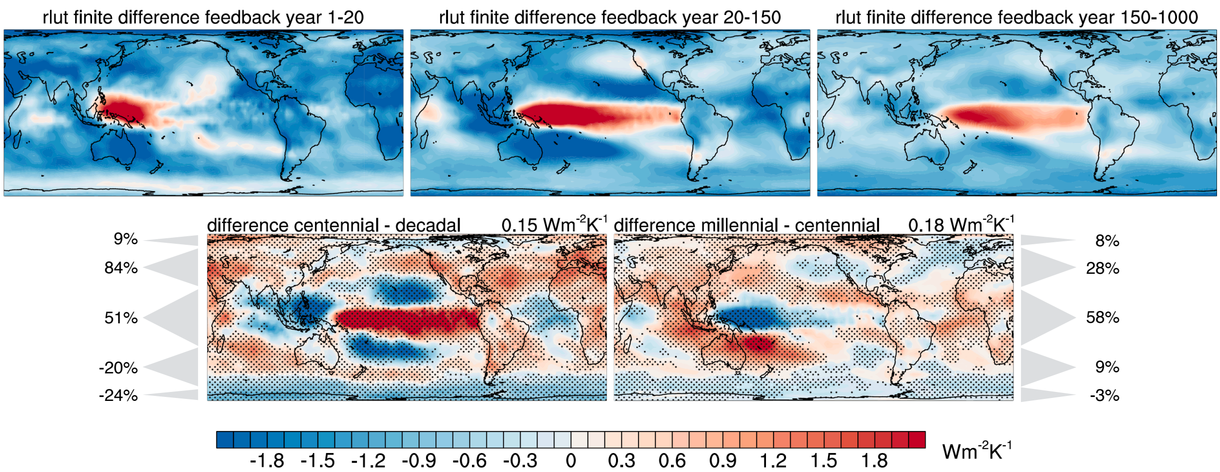
SM Fig. 6: As in Fig. 4, but showing the net cloud radiative effect only.



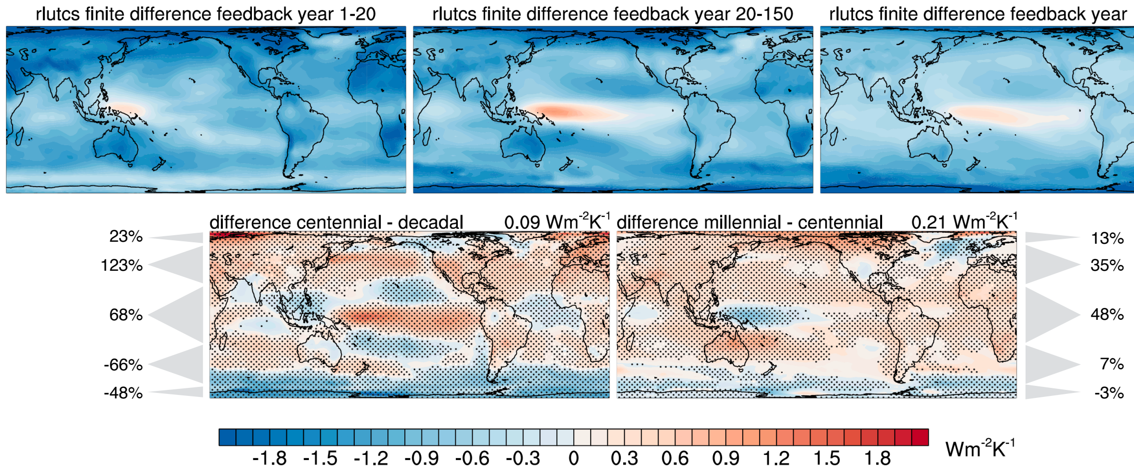
SM Fig. 7: As in Fig. 4, but showing the shortwave cloud radiative effect only.



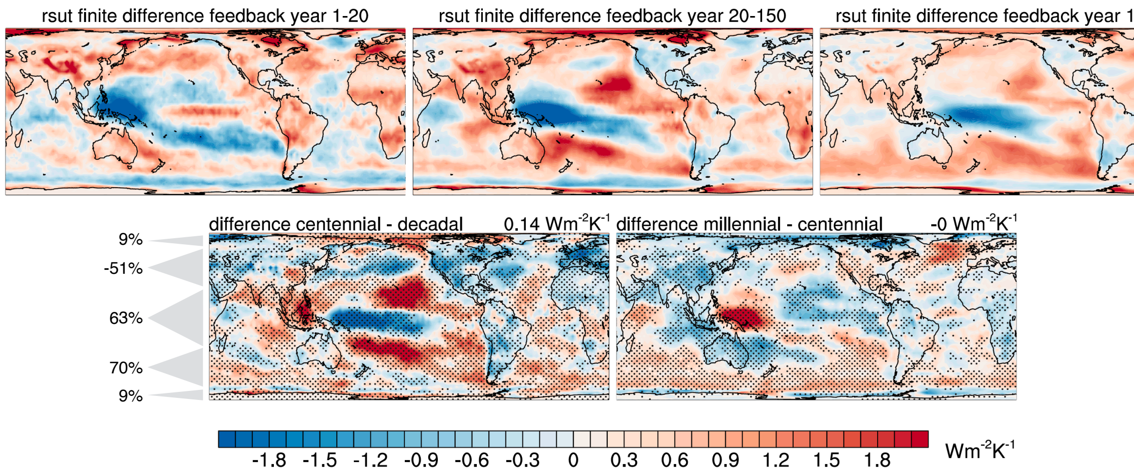
SM Fig. 8: As in Fig. 4, but showing the longwave cloud radiative effect only.



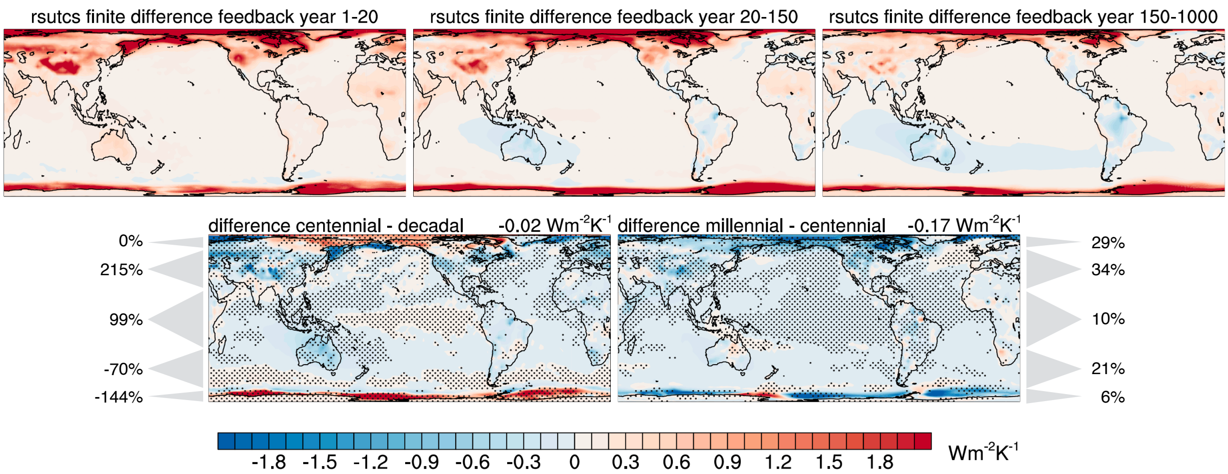
SM Fig. 9: As in Fig. 4, but showing the net longwave component only.



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

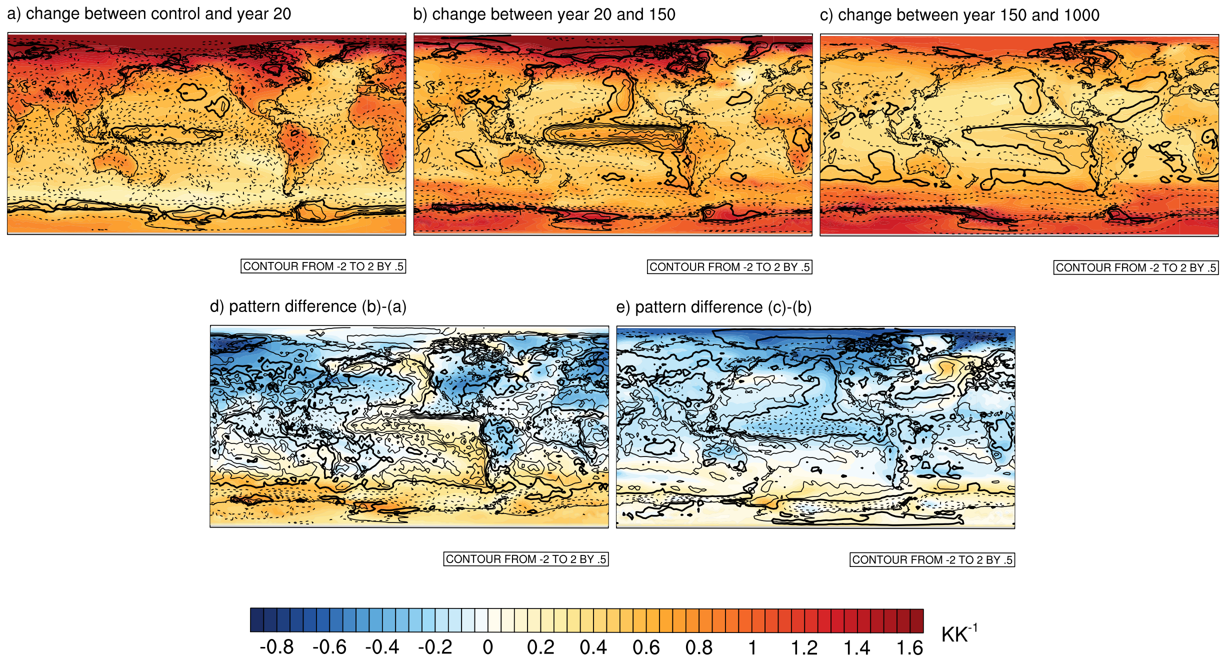


SM Fig. 11: As in Fig. 4, but showing the net shortwave component only.



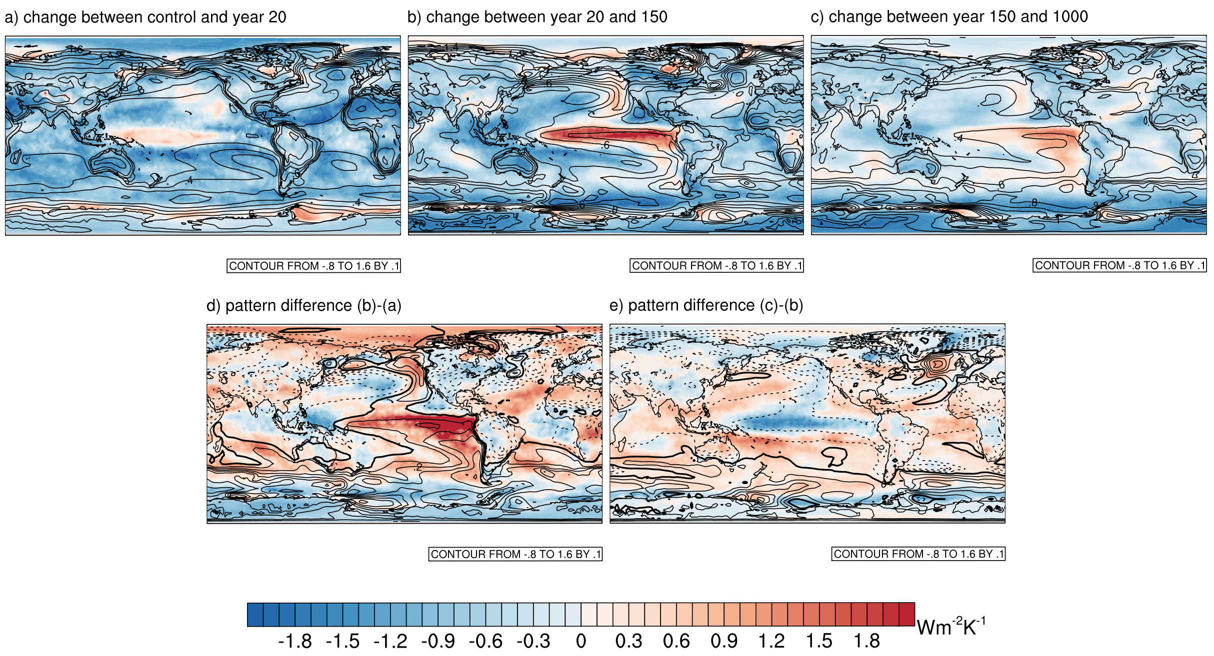
SM Fig. 12: As in Fig. 4, but showing the shortwave clear sky component only.

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; $Wm^{-2}K^{-1}$) 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; $Wm^{-2}K^{-1}$) 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

Andrews, T., J. M. Gregory, and M. J. Webb, 2015: The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models. *Journal of Climate*, **28 (4)**, 1630–1648, URL `http://dx.doi.org/10.1175/JCLI-D-14-00545.1`.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, **39 (9)**, URL `http://dx.doi.org/10.1029/2012GL051607`.

Armour, K. C., 2017: Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, **7**, 331 EP –, URL `http://dx.doi.org/10.1038/nclimate3278`.

Bloch-Johnson, J., R. T. Pierrehumbert, and D. S. Abbot, 2015: Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, **42 (12)**, 4973– 4980, doi:10.1002/2015GL064240, URL `http://dx.doi.org/10.1002/2015GL064240`, 2015GL064240.

Brown, P. T., W. Li, L. Li, and Y. Ming, 2014: Top-of-atmosphere radiative contribution to unforced decadal global temperature variability in climate models. *Geophysical Research Letters*, **41 (14)**, 5175–5183, URL `http://dx.doi.org/10.1002/2014GL060625`.

Caballero, R. and M. Huber, 2013: State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proceedings of the National Academy of Sciences of the United States of America*, **110 (35)**, 14 162–14 167, URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761583/`.

Ceppi, P. and J. M. Gregory, 2017: Relationship of tropospheric stability to climate sensitivity and earth's observed radiation budget. *Proceedings of the National Academy of Sciences*, **114 (50)**, 13 126–13 131, doi: 10.1073/pnas.1714308114, URL `https://www.pnas.org/content/114/50/13126`.

Charney, J., et al., 1979: Carbon Dioxide and Climate: A Scientific Assessment. Tech. rep., National Academy of Science, Washington, DC.

Geoffroy, O., D. Saint-Martin, G. Bellon, A. Voldoire, D. Olivié, and S. Tytéca, 2013: Transient Climate Response in a Two-Layer Energy-Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs. *Journal of Climate*, **26 (6)**, 1859–1876, URL http://dx.doi.org/10.1175/JCLI-D-12-00196.1.

Gregory, J. M., T. Andrews, and P. Good, 2015: The inconstancy of the transient climate response parameter under increasing CO_2. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **373 (2054)**, URL http://rsta.royalsocietypublishing.org/content/373/2054/20140417.

Gregory, J. M., et al., 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, **31 (3)**, URL http://dx.doi.org/10.1029/2003GL018747.

Hansen, J., et al., 2005: Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, **110 (D18)**, URL http://dx.doi.org/10.1029/2005JD005776.

Haugstad, A. D., K. C. Armour, D. S. Battisti, and B. E. J. Rose, 2017: Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, **44 (14)**, 7455–7463, doi:10.1002/2017GL074372, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074372.

Jonko, A. K., K. M. Shell, B. M. Sanderson, and G. Danabasoglu, 2012: Climate Feedbacks in CCSM3 under Changing CO2 Forcing. Part I: Adapting the Linear Radiative Kernel Technique to Feedback Calculations for a Broad Range of Forcings. *Journal of Climate*, **25 (15)**, 5260–5272, URL http://dx.doi.org/10.1175/JCLI-D-11-00524.1.

Knutti, R. and M. A. A. Rugenstein, 2015: Feedbacks, climate sensitivity and the limits of linear models. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **373 (2054)**, doi:10.1098/rsta.2015.0146.

Meraner, K., T. Mauritsen, and A. Voigt, 2013: Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, **40 (22)**, 5944–5948, URL http://dx.doi.org/10.1002/2013GL058118.

Roe, G., 2009: Feedbacks, timescales, and seeing red. *Annual Review of Earth and Planetary Sciences*, **37 (1)**, 93–115, URL http://dx.doi.org/10.1146/annurev.earth.061008.134734.

Rohrschneider, T., B. Stevens, and T. Mauritsen, 2019: On simple representations of the climate response to external radiative forcing. *Climate Dynamics*, doi:10.1007/s00382-019-04686-4.

Rose, B. E. J. and M. C. Rencurrel, 2016: The Vertical Structure of Tropospheric Water Vapor: Comparing Radiative and Ocean-Driven Climate Changes. *Journal of Climate*, **29 (11)**, 4251–4268, URL http://dx.doi.org/10.1175/JCLI-D-15-0482.1.

Rugenstein, M., et al., 2019: LongRunMIP – motivation and design for a large collection of millennial-length GCM simulations. *Bulletin of the American Meteorological Society*, **0 (0)**, null, URL https://doi.org/10.1175/BAMS-D-19-0068.1.

Rugenstein, M. A. A., K. Caldeira, and R. Knutti, 2016a: Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophysical Research Letters*, **43 (18)**, 9877–9885, URL http://dx.doi.org/10.1002/2016GL070907.

Rugenstein, M. A. A., J. M. Gregory, N. Schaller, J. Sedláček, and R. Knutti, 2016b: Multiannual Ocean–Atmosphere Adjustments to Radiative Forcing. *Journal of Climate*, **29 (15)**, 5643–5659, URL http://dx.doi.org/10.1175/JCLI-D-16-0312.1.

Shell, K. M., J. T. Kiehl, and C. A. Shields, 2008: Using the Radiative Kernel Technique to Calculate Climate Feedbacks in NCAR's Community Atmospheric Model. *Journal of Climate*, **21 (10)**, 2269–2282, doi: 10.1175/2007JCLI2044.1, URL `https://doi.org/10.1175/2007JCLI2044.1`, `https://doi.org/10.1175/2007JCLI2044.1`.

Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *Journal of Climate*, **21 (14)**, 3504–3520, doi:10.1175/2007JCLI2110.1, URL `http://dx.doi.org/10.1175/2007JCLI2110.1`.

Wetherald, R. T. and S. Manabe, 1988: Cloud feedback processes in a general circulation model. *Journal of the Atmospheric Sciences*, **45 (8)**, 1397–1416, doi:10.1175/1520-0469(1988)045<1397:CFPIAG>2.0.CO;2, URL `http://dx.doi.org/10.1175/1520-0469(1988)045<1397:CFPIAG>2.0.CO;2`.

Winton, M., K. Takahashi, and I. M. Held, 2010: Importance of Ocean Heat Uptake Efficacy to Transient Climate Change. *Journal of Climate*, **23 (9)**, 2333–2344, doi:10.1175/2009JCLI3139.1, URL `http://dx.doi.org/10.1175/2009JCLI3139.1`.

Zhou, C., M. D. Zelinka, A. E. Dessler, and S. A. Klein, 2015: The relationship between inter-annual and long-term cloud feedbacks. *Geophysical Research Letters*, URL `http://dx.doi.org/10.1002/2015GL066698`, 2015GL066698.