The radiative feedback continuum from Snowball Earth to an ice-free hothouse

Ian Eisenman[*1] **and Kyle C. Armour**[2]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA
[2]Department of Atmospheric Sciences and School of Oceanography, University of Washington, Seattle, WA, USA

Abstract
Paleoclimate records have been used to estimate the modern equilibrium climate sensitivity. However, this requires understanding how the feedbacks governing the climate response vary with the climate itself. Here we warm and cool a state-of-the-art climate model to simulate a continuum of climates ranging from a nearly ice-covered Snowball Earth to a nearly ice-free hothouse, and we compute the resulting changes in feedbacks. We find that the pre-industrial (PI) climate is near the stability optimum: warming leads to a less-stable (more-sensitive) climate, as does cooling of more than 2K. Under further cooling, we find that the total feedback becomes no longer stable, indicating the Snowball Earth bifurcation point. Physically interpreting the results using a radiative kernel analysis, we find that the decrease in stability for climates colder than the PI occurs mainly due to the albedo and lapse-rate feedbacks, and that the decrease in stability for climates warmer than the PI occurs mainly due to the cloud shortwave feedback. These results suggest a complex relationship between climate feedbacks and global temperature with a structure that is not well represented by including a term in the global energy budget that is quadratic in temperature, as has typically been assumed in previous studies relating feedbacks between different underlying climate states.

Corresponding author: Ian Eisenman, `eisenman@ucsd.edu`

1 Introduction

Recent community assessments (Sherwood et al., 2020; Forster et al., 2021) have substantially narrowed the estimated range of Earth's equilibrium climate sensitivity (ECS) for the first time in decades, leading to better constraints on future warming (Lee et al., 2021). This narrowing of the uncertainty in the ECS (which is defined as the equilibrium global-mean surface temperature response to CO_2 doubling from pre-industrial levels) was achieved in large part through the use of paleoclimate records from times when the climate was substantially different from today. In Sherwood et al. (2020), the ECS likelihoods derived from proxy reconstructions of temperatures and estimates of radiative forcing during the Last Glacial Maximum (LGM) and mid-Pliocene warm period (mPWP) provided the strongest line of evidence against high ECS values. In Forster et al. (2021), proxy reconstructions of LGM, mPWP, and Eocene temperatures also informed the strongest line of evidence against high ECS values: so-called "emergent constraints" wherein a relationship between temperature changes and ECS within an ensemble of Earth System Models (ESMs) is combined with observations or paleoproxy reconstructions of those temperature changes to derive a constraint on ECS.

A confounding factor in the use of paleoclimate records to inform the sensitivity of modern climate to greenhouse gas forcing is that the radiative feedbacks governing the climate response can vary with the underlying climate itself (e.g., Sherwood et al., 2020; Forster et al., 2021). That is, the use of paleoclimate records to constrain ECS requires understanding how modern radiative feedbacks (which govern ECS) relate to radiative feedbacks operating in climates much colder or much warmer than today.

Following previous work (e.g., Roe & Armour, 2011; Bloch-Johnson et al., 2015, 2021), Sherwood et al. (2020) represented the dependence of radiative feedbacks on the underlying climate by including a quadratic feedback term in the standard model of global energy balance used to relate reconstructions of temperature and climate forcing to modern-day ECS. This approach typically represents the net radiative feedback as becoming less negative (i.e., a more-sensitive climate) with global warming and more negative with global cooling (e.g., Sherwood et al., 2020). While higher-order terms that are cubic and beyond in surface temperature could be included, they are typically assumed to be small and omitted. This raises key questions regarding the range of temperatures over which this approximation applies, what causes it to fail outside this range, and relatedly how confident we can be in the structure of the radiative feedback dependence on global temperature over a wide range of climate states. The answers to these questions also have implications for emergent constraints, in which the mapping of feedbacks between past and future climate states is implicitly accounted for through the use of ESMs to simulate the paleoclimate states and ECS values on which the constraints rely.

Here we warm and cool a state-of-the-art ESM to simulate a continuum of climates ranging from a nearly ice-covered Snowball Earth to a nearly ice-free hothouse planet. We analyze how the radiative feedbacks depend on the underlying climate, and we physically interpret the results.

2 Climate model simulations

Using NCAR's Community Earth System Model Version 2 (CESM2, Danabasoglu et al., 2020) in its standard workhorse configuration, we ramp CO_2 concentrations over a range of 11.5 doublings. Specifically, we start from the end of a 500-year pre-industrial (PI) control simulation, which has a constant CO_2 concentration of 284.7 ppm, and we either increase or decrease the atmospheric CO_2 concentration at a rate of 1% per year (Fig. 1a). The Warming simulation, which extends the preexisting gradual CO_2 quadrupling simulation of Danabasoglu (2019a), is 279 years long and ends with a CO_2 concentration of 4,522 ppm, which is 16 times the PI value. The Cooling simulation is 514 years long and ends with a CO_2 concentration of 1.6 ppm, which is 1/175 times the PI value. See Supplementary Information (SI) Sec. S1 for details.

This leads to a 59K range in simulated global-mean surface temperature, with climates ranging from a nearly ice-covered Snowball Earth to a nearly ice-free hothouse planet. Averaged over the last decade of the PI control simulation, the global-mean surface temperature is 15°C, and the ice area is 11.4% of the global surface area. The latter includes sea ice, snow cover on land, and prescribed time-invariant glacial ice cover (see SI Sec. S1 for details), with sea ice covering 6.1% of the ocean (4.3% of the globe). In the Warming simulation, the annual-mean global-mean surface temperature increases by 18K to 33°C (Fig. 1b), and the annual-mean ice area decreases to 3.2% of the globe (Fig. 1c), with sea ice covering 0.0% of the ocean. In the Cooling simulation, the temperature decreases by 41K to -26°C (Fig. 1b), and the ice area increases to 68.7% of the globe (Fig. 1c), with sea ice covering 70.3% of the ocean.

The surface temperature in the deep tropics (averaged annually and over 10°S–10°N) is 28°C in the PI (Fig. 1e), and it reaches 42°C in the final decade of the Warming run (Fig. 1f) and 5°C in the final decade of the Cooling run (Fig. 1d). In the PI climate, the polar surface temperature (averaged annually and over both hemispheres poleward of 70°) is -24°C, and this region is largely covered with snow and ice (Fig. 1h). In the final decade of the Warming run, the polar temperature reaches 4°C, and the remaining ice cover is almost exclusively glacial ice, which is a specified surface type in CESM2 with an area that does not evolve during the simulations. In the final decade of the Cooling run, the polar temperature reaches -68°C, and the ice cover extends into the tropics.

3 Net radiative feedback and effective climate sensitivity

In order to evaluate the net radiative feedback over this continuum of climates, we adopt the standard model of global energy balance and climate feedbacks:

$$\Delta N = \Delta F_{GHG} + \Delta F_{net} = \Delta F_{GHG} + \lambda_{net}\,\Delta T, \tag{1}$$

with

$$\lambda_{net} \equiv \frac{\Delta F_{net}}{\Delta T} = \frac{\Delta N - \Delta F_{GHG}}{\Delta T}. \tag{2}$$

Here all quantities are averaged annually and globally: N is the top-of-atmosphere (TOA) net energy flux reported by the model (using top-of-model fields), F_{GHG} is an estimate of CO_2 radiative forcing relative to PI based on the line-by-line radiative transfer calculations of Byrne and Goldblatt (2014) (see SI Sec. S2 and Fig. S1), $F_{net} \equiv N - F_{GHG}$ is the net radiative response of the climate system, λ_{net} is the net radiative feedback parameter, and T is the surface temperature. The fluxes are defined to be positive in the downward direction, and the feedback parameter is negative for a stable climate. The modifier Δ is described below.

The radiative forcing F_{GHG} and resulting value of N are plotted in Figs. 2a,b, and the difference F_{net} is plotted in Fig. 2c. It can be readily seen that F_{net} does not depend linearly on T. Specifically, the slope of the F_{net} versus T curve (Fig. 2c) is most negative near the PI climate (black vertical dashed line), being less steep in warmer and colder simulated climates. In extremely cold climates, the slope becomes zero around $T = 0°C$ and then changes sign for climates with $T < 0°C$, implying that additional incremental levels of cooling lead to less energy coming into the climate system and hence more cooling.

We consider two approaches to define λ_{net} in Eq. (2), following Rugenstein and Armour (2021) (see SI Sec. S3 for details):

(i) The "effective feedback" $\lambda_{net}^{\mathrm{eff}}$, which describes the radiative feedback processes operating between a given climate state and the PI climate. In this case, we define Δ as

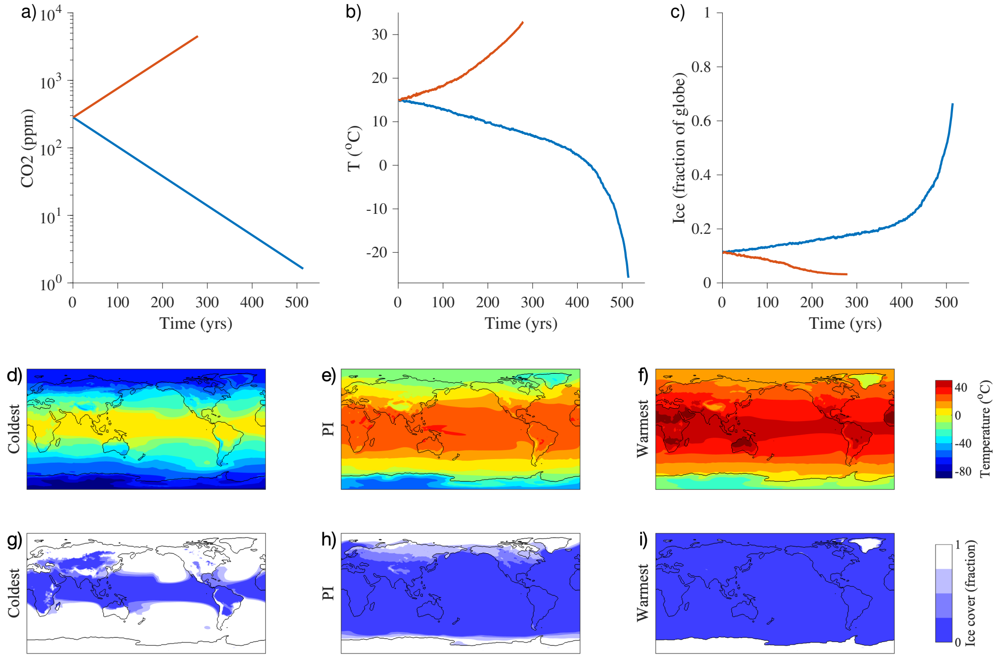

Figure 1. Forcing and climate response in CESM2 simulations. Time series of (a) specified atmospheric CO_2 volume mixing ratio, (b) annual-mean global-mean surface temperature, and (c) annual-mean global ice area (including sea ice, snow cover on land, and glacial ice), in the Warming simulation (red) and the Cooling simulation (blue). Also included are surface temperature maps averaged over the last decade of the (d) Cooling, (e) PI control, and (f) Warming simulations, as well as ice area maps averaged over the last decade of the (g) Cooling, (h) PI control, and (i) Warming simulations. Note that we use the relatively short averaging period of a single decade in these maps in order to better capture the full range given the rates of change near the end of the Warming and Cooling simulations.

the anomaly from the PI climate, and Eq. (2) is calculated from F_{net} after applying a polynomial smoothing. Note that this allows $\lambda_{net}^{\text{eff}}$ to vary smoothly even in the limit $\Delta T \to 0$, as described in SI Sec. S3 and shown in Fig. S2.

(ii) The "differential feedback" $\lambda_{net}^{\text{diff}}$, which describes the feedback processes operating within a given climate. Hence $\lambda_{net}^{\text{diff}}$ is the local tangent value of the slope in Fig. 2a. In this case, we define Δ as the anomaly associated with an incremental change in climate, and Eq. (2) is calculated using a regression of F_{net} versus T within a running window.

The effective feedback may be seen as most directly relevant to current discussions of ECS, since they often involve estimates of past climates compared with today, rather than estimates of past climate variability (e.g., Sherwood et al., 2020; Forster et al., 2021). On the other hand, the differential feedback reflects the radiative response to a temperature perturbation in a given underlying climate, and hence it may be somewhat easier to physically interpret.

The net feedback parameter calculated using each of these approaches is plotted in Figs. 2d,e. A striking result is that the PI climate is near the stability optimum. The differential feedback $\lambda_{net}^{\text{diff}}$, which indicates the stability of the climate system to perturbations, is most negative when the global-mean temperature is 2K cooler than the PI value (Fig. 2d). Starting from the PI, warming leads to less-stabilizing radiative feed-

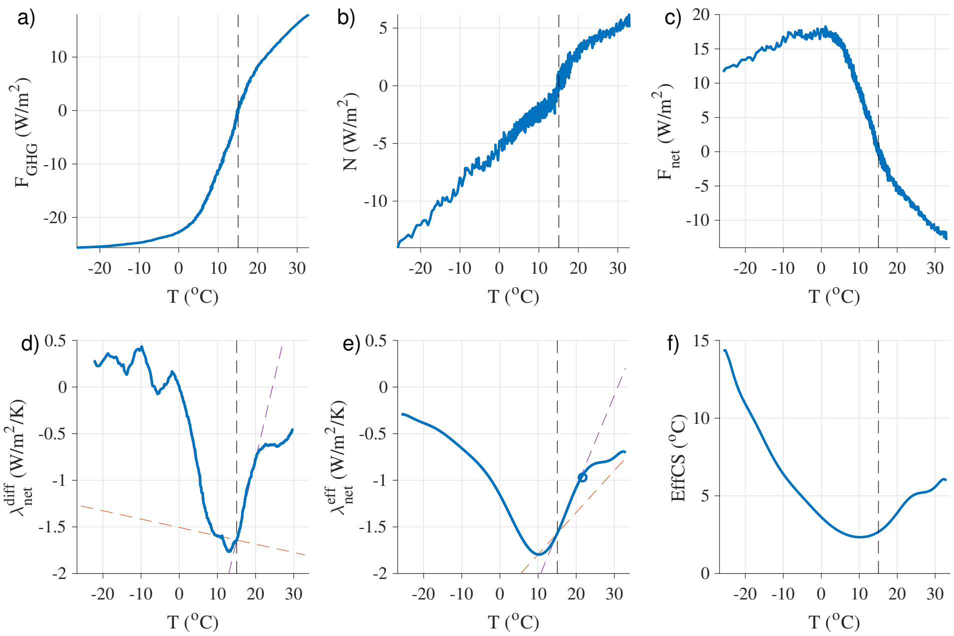

Figure 2. Dependence of the net feedback and effective climate sensitivity on the underlying climate. (a) CO_2 radiative forcing F_{GHG}. (b) TOA net energy flux N. (c) Net radiative response of the climate system, $F_{net} \equiv N - F_{GHG}$. (d) Net differential feedback parameter $\lambda_{net}^{\mathrm{diff}}$. (e) Net effective feedback parameter $\lambda_{net}^{\mathrm{eff}}$. The blue circle indicates the result from a previous analysis of an instantaneous CO_2 quadrupling simulation with the same climate model (Hahn et al., 2021). (f) The effective climate sensitivity EffCS. All quantities are plotted versus the global-mean surface temperature T. The dashed lines in panels (d) and (e) indicate a linear dependence of λ_{net} on ΔT that runs through the PI climate and either a climate 5K colder (red) or a climate 5K warmer (magenta). In all panels, the vertical dashed line indicates the PI climate.

backs and hence a more-sensitive climate, as does cooling of more than 2K. The effective feedback $\lambda_{net}^{\mathrm{eff}}$ shows similar behavior, being most negative when the global-mean temperature is 5K cooler than the PI value (Fig. 2e).

The TOA net energy flux when the climate has reached equilibrium is $N = 0$, as is approximately the case in the simulated PI climate (Fig. 2b). Hence from Eq. (2), the equilibrium warming response to a change in CO_2 is $\Delta T = -\Delta F_{GHG}/\lambda_{net}$. This is known as the ECS in the special case of a doubling of CO_2 from PI levels, as mentioned above. It is given by ECS $= -F_{2\times}/\lambda_{2\times}$, where $F_{2\times} = 4.2$ W/m^2 is the value of the radiative forcing ΔF_{GHG} when CO_2 is doubled from its PI value of 284.7 ppm, and $\lambda_{2\times}$ is the value of the feedback parameter $\lambda_{net}^{\mathrm{eff}}$ operating in this climate state. For other climate states, the effective climate sensitivity (EffCS) is similarly defined using the effective feedback parameter:

$$\mathrm{EffCS} \equiv \frac{-F_{2\times}}{\lambda_{net}^{\mathrm{eff}}}. \tag{3}$$

The EffCS is plotted in Fig. 2f. This shows that the sensitivity is lowest near the PI climate, with more-sensitive climates at warmer and much colder temperatures. The continuum of simulated climates spans a range of EffCS values from 2°C to 15°C. Note that the EffCS (Fig. 2f) scales as the inverse of $\lambda_{net}^{\mathrm{eff}}$ (Fig. 2e).

Under more-extreme cooling, the value of $\lambda_{net}^{\mathrm{diff}}$ in Fig. 2d becomes positive when the global-mean temperature drops below approximately 0°C, which is 15K colder than the PI climate. At this point there is a change in the sign of the slope of the F_{net} versus T curve in Fig. 2c: as the temperature drops below this point, incremental coolings are accompanied by incremental decreases in the level of heating by the net radiative response of the climate system. This corresponds to the Snowball Earth bifurcation point, beyond which the sea ice in the model irreversibly expands toward the equator. Note that this is the point at which the global-mean temperature and global ice area begin to abruptly change in the Cooling simulation (Fig. 1b,c). The implications of this change in the sign of $\lambda_{net}^{\mathrm{diff}}$ can be illustrated using a simple single-layer model of the climate system, which is described in SI Sec. S4. The positive value of $\lambda_{net}^{\mathrm{diff}}$ implies that the climate is transitioning across a range of temperatures for which the only equilibrium climate state is unstable (SI Fig. S3). Previous studies have demonstrated that bifurcations and bistability associated with the Snowball Earth climate occur in climate models of varying levels of complexity in certain ranges of CO_2 and solar luminosity (e.g., Marotzke & Botzet, 2007; Voigt & Marotzke, 2010; Roe & Baker, 2010; Voigt et al., 2011; Pierrehumbert et al., 2011). Note that $\lambda_{net}^{\mathrm{eff}}$ remains negative for all climates, in contrast with $\lambda_{net}^{\mathrm{diff}}$, which illustrates how the EffCS and $\lambda_{net}^{\mathrm{eff}}$ framework can give potentially misleading results about the stability of the underlying climate state because it is based on anomalies from the PI climate.

Under warming, the values of $\lambda_{net}^{\mathrm{eff}}$ and $\lambda_{net}^{\mathrm{diff}}$ increase monotonically. Notably, the climate remains stable ($\lambda_{net}^{\mathrm{diff}}$ is negative) even at extreme levels of global warming nearing 15K above the PI. Note that previous studies using idealized single-column radiative models have found that the net climate feedback becomes more negative with warming for climates warmer than approximately 25K above the PI (Seeley & Jeevanjee, 2021; Kluft et al., 2021).

Note that when the climate is forced to transiently evolve away from an equilibrated state, it is possible for the climate feedback parameter to become less negative due to the spatial pattern of surface temperature changes (Winton et al., 2010). In SI Sec. S4, we investigate the extent to which this may explain the results in Fig. 2d,e by using a standard two-layer model of the climate system (Held et al., 2010) that includes a term to represent the deep ocean heat uptake efficacy. The results show that although deep ocean heat uptake efficacy can cause $\lambda_{net}^{\mathrm{diff}}$ and $\lambda_{net}^{\mathrm{eff}}$ to become less negative under both warming and cooling as the climate gets farther from its equilibrated state, a moderate (CMIP5-mean) deep ocean heat uptake efficacy leads to far smaller changes in λ_{net} than we find in CESM2 (SI Fig. S4). Furthermore, even with a large ocean heat uptake efficacy, the two-layer model results in a “V”-shaped feedback dependence on temperature that is centered at the equilibrated climate (purple curve in SI Fig. S4), in contrast to the “U” shape centered at a temperature several degrees colder than the PI that we find in CESM2 (Fig. 2d,e). This suggests that the changes in λ_{net} shown in Fig. 2d,e are considerably outside of what would be expected from changing surface temperature patterns associated with deep ocean heat uptake, and that feedback nonlinearities with global temperature are the main cause of the dependence of the net feedback on the underlying climate in CESM2 over the simulated range considered here.

3.1 Linear representation of $\lambda_{net}(T)$

Many recent studies have suggested that colder climates are more stable than warmer climates, including climates considerably colder than the PI. Specifically, as summarized in Forster et al. (2021), paleoclimate records (von der Heydt et al., 2014; Anagnostou et al., 2016, 2020; Friedrich et al., 2016; Royer, 2016; Shaffer et al., 2016; Kohler et al., 2017; Snyder, 2019; Stap et al., 2019) and comprehensive climate models (Caballero & Huber, 2013; Jonko et al., 2013; Meraner et al., 2013; Good et al., 2015; Duan et al., 2019; Mauritsen et al., 2019; Stolpe et al., 2019; Zhu et al., 2019) suggest a general trend toward less-stabilizing radiative feedbacks (hence higher EffCS) with increasing global temperature over a range of climates spanning approximately 6K colder than today to approximately 10K warmer than today. However, the results presented here suggest that the PI climate is near a stability optimum, with warming and cooling beyond 2K both leading to less-stable climates (Fig. 2d). Similarly, warming and substantial cooling both lead to less-negative values of $\lambda_{net}^{\mathrm{eff}}$ and higher EffCS (Figs. 2e,f). While the climate at the temperature characteristic of the LGM (4-6K colder than the PI) is more stable than the simulated climates that are warmer than the PI, consistent with the studies mentioned above, we find that climates beyond about 6K colder than PI can be considerably less stable than climates warmer than the PI.

As noted in the Introduction, previous work has typically represented nonlinearities in the dependence of the net radiative response on the underlying climate by using a quadratic relationship with global temperature (e.g., Sherwood et al., 2020). In this case, Eq. (1) is replaced with

$$\Delta N = \Delta F_{GHG} + \lambda_0\,\Delta T + \frac{1}{2}\,\alpha\,\Delta T^2, \tag{4}$$

where λ_0 is the net feedback near the PI climate and α is a coefficient scaling the nonlinear radiative response. This implies a linear dependence on global temperature for both the effective feedback and the differential feedback:

$$\lambda_{net}^{\mathrm{eff}} = \lambda_0 + \frac{1}{2}\,\alpha\,\Delta T \quad \text{and} \quad \lambda_{net}^{\mathrm{diff}} = \lambda_0 + \alpha\,\Delta T.$$

Note that here we adopt the formalism used in Sherwood et al. (2020).

Sherwood et al. (2020) use the value $\alpha = 0.1$ W/m^2/K^2 (with an uncertainty of ±0.1 W/m^2/K^2) for the difference in the feedback at the LGM compared with the PI, and they implicitly assume no change in feedback between the PI and warmer climates. We include red dashed lines in Figs. 2d,e to represent a linear dependence of λ_{net} on T that goes through the PI climate ($T = 15°C$) and the climate with an LGM-like level of cooling ($T = 10°C$). The slopes of the curves correspond to values of $\alpha = -0.01$ W/m^2/K^2 for $\lambda_{net}^{\mathrm{diff}}$ and $\alpha = 0.05$ W/m^2/K^2 for $\lambda_{net}^{\mathrm{eff}}$. We also include for comparison magenta dashed lines that go through the PI climate and the climate at 5K of warming ($T = 20°C$), which have slopes that correspond to values of $\alpha = 0.17$ W/m^2/K^2 for $\lambda_{net}^{\mathrm{diff}}$ and $\alpha = 0.10$ W/m^2/K^2 for $\lambda_{net}^{\mathrm{eff}}$. Note that CESM2 has previously been shown to be among the ESMs with the largest values of α when assessed for temperature changes near the PI climate (Bloch-Johnson et al., 2021).

These results show that the value of α adopted by Sherwood et al. (2020) for the change in λ_{net} at the LGM is much larger than in the CESM2 results, because the "U" shape in Figs. 2d,e causes the feedback at 5K of cooling to be similar to the feedback at the PI. If we were to repeat the Sherwood et al. (2020) analysis using the value of α that we find here for the difference between the LGM and PI feedbacks, our lower value of α would imply a lower modern-day climate sensitivity than Sherwood et al. (2020) found, which amounts to a stronger constraint on the upper bound of the EffCS than they report. This is because the CESM2 results suggest that the LGM may be a more direct analogue to current warming than previously assumed, since the feedbacks are relatively similar. In other words, Sherwood et al. (2020) took λ_{net} to be more negative at the LGM than the modern value, whereas we find that the feedbacks are similar. So a given paleo estimate of the LGM value of λ_{net} implies a similar modern feedback value according to our results, whereas the analysis of Sherwood et al. (2020) would take it to imply a less negative modern feedback and hence a more sensitive modern climate.

For climates warmer than the PI, these results imply a larger value of α that is actually somewhat similar to the Sherwood et al. (2020) result. But here the value of α applies to warming rather than cooling. Overall, this suggests that feedback nonlinearities could be large for future warming, consistent with some other studies (e.g., Bloch-Johnson et al., 2021), while being relatively small for colder climates similar to the LGM.

These results suggest that the formulation of feedbacks as changing linearly with global temperature applies only over a narrow range of climates, and that because of the "U" shape of the relationship between the feedback and the underlying climate, comparing feedbacks between two climates depends sensitively on the temperatures of the climates.

The range of climates over which the quadratic term in the global energy budget (Eq. (4)) serves as a useful approximation can be seen by comparing the linear fits to the values of $\lambda_{net}^{\mathrm{diff}}$ and $\lambda_{net}^{\mathrm{eff}}$ (magenta lines in Figs. 2d,e). For the effective feedback $\lambda_{net}^{\mathrm{eff}}$, the quadratic term captures much of the variation in the feedback parameter for climates with T between about 3K colder and 8K warmer than the PI climate, thus serving as a decent approximation to feedback changes over a temperature range spanning the PI climate and CO_2 quadrupling but not spanning climates as cold as the LGM. Outside of this temperature range, the quadratic approximation fails spectacularly. For the differential feedback $\lambda_{net}^{\mathrm{diff}}$, the quadratic term provides a decent approximation over a similar temperature range. Including additional terms in the Taylor series expansion (i.e., order ΔT^3 and higher in Eq. (4)) would be expected to widen the range over which the expansion provides a useful approximation.

4 Individual radiative feedback parameters

In order to identify what physical processes are responsible for the decrease in stability under both cooling and warming from the PI (Fig. 2b), we begin by using a radiative kernel analysis to assess which individual feedback parameters are driving the changes. The radiative kernels were generated by Pendergrass et al. (2018) based on CAM5, which is the previous version of the atmospheric model in CESM2. Using these kernels, we compute the annual-mean global-mean change in the radiative response associated with changes in the surface temperature (F_P for Planck feedback), atmospheric lapse rate (F_L for lapse-rate feedback), humidity (F_w for water-vapor feedback), and surface albedo (F_α for albedo feedback). We compute the cloud radiative response (F_c) as the difference between the sum of the individual feedbacks and F_{net}; hence F_c also includes the residual (F_{res}) due to inaccuracies in the radiative kernel analysis (see SI Sec. S5 for details). Each of the resulting radiative responses is shown in SI Fig. S5. Note that the radiative kernel analysis effectively linearizes the simulated response to changing climate fields about a climate near the PI. Although it would be more accurate to use radiative kernels that vary with the climate (e.g., Jonko et al., 2012), the present analysis could be seen as a preliminary step toward building understanding of feedbacks across a wide continuum of climate changes by using a kernel that does not vary with climate, before considering how the radiative kernels change.

We define each individual feedback parameter as

$$\lambda_i \equiv \frac{\Delta F_i}{\Delta T}, \tag{5}$$

where the subscript i can indicate any individual feedback and Δ has the same two definitions as in Eq. (2).

The results (Fig. 3) indicate that the decrease in stability (i.e., $\lambda_{net}^{\mathrm{diff}}$ becoming less negative) for climates more than 2K colder than the PI is caused by the lapse-rate and albedo feedbacks, whereas the decrease in stability for climates warmer than the PI is caused mainly by the cloud feedback. The roles of these feedbacks occur robustly in both the differential feedback analysis (right column in Fig. 3) and the effective feedback analysis (left column in Fig. 3). Note that although there is some compensation between the lapse-rate feedback (λ_L) and the water-vapor feedback (λ_w), as expected, the changes in the combined feedback ($\lambda_L+\lambda_w$) are dominated by the lapse-rate feedback (see red dashed lines in third row of Fig. 3).

4.1 Physical interpretation of results

Here we interpret the results in Fig. 3. We focus on the differential feedback parameters, since they describe the physics of a given climate and hence may be more-readily understood than the effective feedback parameters.

The large range of simulated climate changes may be expected to be annually and zonally uniform to a first approximation. Hence we repeat the analysis in Fig. 3 taking the annual average and the zonal average of each kernel as well as each simulated climate field before multiplying the kernels by the climate fields (see SI Sec. S6 for details). We find that the result matches closely with the feedback parameters computed using the full 4-dimensional structure of the simulated climate and kernel fields (SI Fig. S6). This suggests that the zonal and seasonal patterns of temperature, surface albedo, humidity, and cloud changes do not play a substantial role in the variations in each feedback parameter shown in Fig. 3, allowing the specific factors driving the variations in each feedback parameter to be more-readily assessed by examining only the meridional and vertical structure of the fields.

The decrease in stability with cooling in cold climates is the main novel result of the present study, since previous work has discussed the decrease in stability with warming. Hence we begin by interpreting the lapse-rate and albedo feedbacks.

Lapse-rate feedback. The lapse-rate feedback describes the impact of changes in the vertical temperature structure. In the tropics today, deep convection occurs, and the temperature profile is close to being moist adiabatic. Warming causes the moist adiabatic lapse rate to decline. This is a negative local feedback, since it means that smaller changes in surface temperature are needed to bring about a given change in outgoing longwave radiation. On the other hand, in the present-day Arctic the planetary boundary layer is often capped by a temperature inversion and hence a very stable stratification, which suppresses vertical mixing and causes temperature changes at the surface not to be propagated aloft, which is a positive local feedback.

The inversion strength can be described by the difference in potential temperature between the 700-hPa level and the surface (cf. Wood & Bretherton, 2006), which is plotted in Fig. 4a. Across the range of simulated climates, ice-covered regions of the globe tend to have an inversion, as expected because the surface absorbs less solar radiation when it is covered with snow or ice, setting up the potential for a positive lapse-rate feedback in these regions. This leads to a less-negative global lapse-rate feedback as the climate cools and more of the globe resembles the present-day Arctic (Fig. 4a,b).

As the climate warms and sea ice is lost, the erosion of polar inversions leads to less-positive polar lapse-rate feedbacks (Fig. 4a,b). However, the lapse-rate feedback in the tropical region becomes less negative with warming for climates warmer than the PI. An analysis of a previous version of this model lead to fairly similar changes in the spatial pattern of the lapse-rate feedback parameter under varied levels of forced warming (Merlis et al., 2022). The mechanisms driving the changes in the tropical temperature profile that cause this are beyond the scope of the current study. The result is that the global lapse-rate feedback becomes somewhat less negative with warming, at least up to temperatures about 10 K warmer than the PI climate, although the cloud feedback dominates the changes in the net feedback parameter for climates warmer than the PI.

The temperature feedback radiative kernel has a spatial structure that varies vertically but which is fairly uniform horizontally, suggesting that lapse-rate feedback changes should approximately track changes in the vertical structure of globally-averaged atmo-

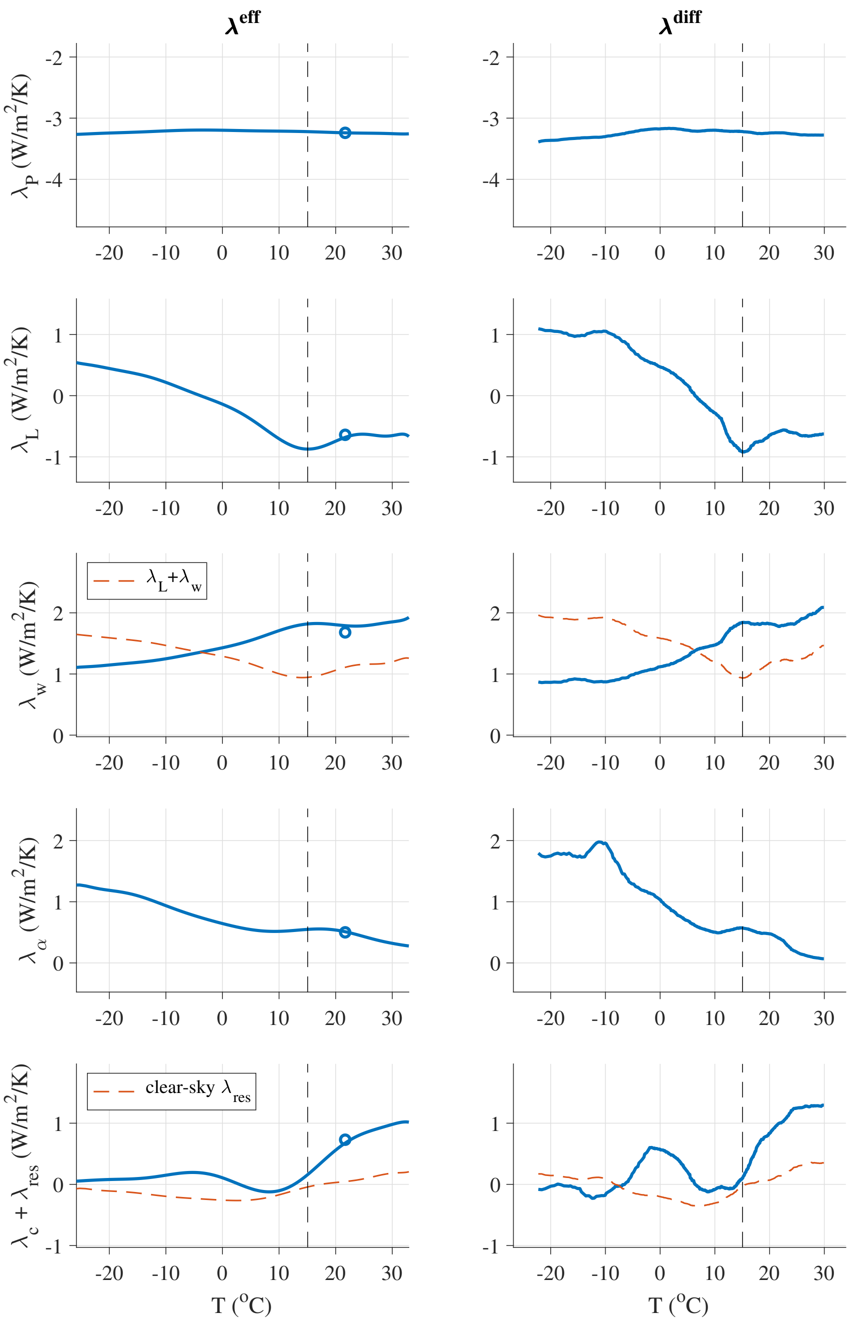

Figure 3. Individual feedback parameters computed using radiative kernels. Panels have different vertical ranges but the same vertical scale for comparison. The sum of the lapse-rate and water-vapor feedbacks is also indicated in the third row, and the clear-sky result for the residual is also indicated in the fifth row. The vertical dashed line in each panel indicates the PI climate. The blue circles indicate the results from a previous analysis of an instantaneous CO_2 quadrupling simulation with the same climate model (Hahn et al., 2021) for comparison.

spheric warming. Indeed, we get a similar result when we repeat the analysis using the global-mean temperature profile (red line in Fig. 4c), which removes the influence of horizontal variations in the efficiency of radiation to space but still retains vertical variations. This result implies that the changes in the lapse-rate feedback parameter under cooling global temperature are dictated primarily by the global-mean atmospheric temperature profile becoming more similar to the Arctic today, causing the global lapse-rate feedback to approach the positive value in the Arctic today.

Albedo feedback. The albedo feedback occurs because a warmer climate has less ice cover, and ice-free regions absorb more solar radiation rather than reflecting it back to space. We find that the albedo feedback increases approximately monotonically with cooling global temperature across the range of simulated climates. The albedo feedback radiative kernel has a spatial structure with values most negative in the low latitudes, where there is the most incident solar radiation. Nonetheless, we find that the migration of the ice edge into sunnier latitudes has a relatively limited influence on the variations in the albedo feedback parameter: we get a fairly similar result when we repeat the analysis using a spatially-uniform radiative kernel, which removes the influence of spatial variations in incident solar radiation as well as clouds and other factors (red line in Fig. 4e). In this case the albedo feedback parameter is approximated to be proportional to the sensitivity of the ice area to global temperature (i.e., the slope in Fig. 4d). This implies that the albedo feedback becomes more destabilizing primarily because the ice area expands more rapidly with cooling in colder climates.

This behavior continues in climates warmer than the PI, with the change in ice area per change in global temperature continuing to decrease as the climate warms (Fig. 4d), leading to a smaller albedo feedback in warmer climates (Fig. 4e). In the warmest simulated climates there is almost no remaining snow and sea ice (Fig. 1i), and the albedo feedback $\lambda_\alpha^{\text{diff}}$ approaches zero (Fig. 3).

Cloud feedback. Clouds cause shortwave cooling and longwave heating, and changes in clouds with climate lead to a feedback that can be either positive or negative. We find that the cloud feedback in CESM2 is approximately zero near the PI climate, but the feedback becomes increasingly destabilizing as the underlying climate warms. Previous work using CESM2 and earlier versions of this model similarly found that cloud feedbacks are more destabilizing in warmer climates (Caballero & Huber, 2013; Zhu et al., 2019; Zhu & Poulsen, 2020).

An important caveat associated with the changes in the cloud feedback shown in Fig. 3 is that this term includes the residual due to factors including inaccuracies in the radiative kernel analysis. One measure of this is the residual when the kernel analysis is repeated using clear-sky fields (see Sec. 5 below), which we find contributes about 25% of the diagnosed cloud feedback change between the PI and warmest simulated climates (red dashed line in bottom right panel of Fig. 3).

We also carry out an alternative test of the impact of clouds that does not rely on the radiative kernels. Instead, we redo the net feedback analysis in Fig. 2 using clear-sky fields reported by the model for the change in TOA net energy flux ΔN. The resulting values of $\lambda_{net}^{\text{eff}}$ and $\lambda_{net}^{\text{diff}}$ are plotted in SI Fig. S8. For both measures of the net feedback in SI Fig. S8, the feedback remains relatively constant in climates warmer than the PI when using clear-sky fields, whereas it becomes steadily less negative with warming when using all-sky fields. This suggests that cloud changes contribute substantially to the trend toward a less-negative net feedback for climates warmer than the PI, consistent with the kernel analysis results in Fig. 3.

This alternative approach also allows us to separate the influence of cloud shortwave effects from cloud longwave effects. We find that using clear-sky fields for only the longwave component of ΔN causes behavior resembling the all-sky results, whereas using clear-sky fields for only the shortwave component of ΔN causes behavior resembling the clear-sky results (SI Fig. S8). This suggests that the increase in the net feedback in warm climates is caused primarily by the cloud shortwave feedback, which is consistent with the results of previous studies (Caballero & Huber, 2013; Zhu et al., 2019; Zhu & Poulsen, 2020).

Planck feedback. The Planck feedback describes how warming the surface and atmospheric column above causes more outgoing longwave radiation to space due to the Stefan-Boltzmann law. This feedback remains relatively invariant across the range of simulated climates, although it becomes slightly more negative as the climate cools. Note that because we use a radiative kernel, we account only for changes in the Planck feedback due to the evolving pattern of surface temperature change, and we do not represent how the Planck feedback depends on global temperature. The Planck feedback radiative kernel is most negative in the warmest regions of the control climate (see SI Sec. S6). The meridional structure of the surface temperature evolution is shown in SI Fig. S7. Simulated surface temperature changes tends to be amplified in ice-covered regions (Fig. 4f), which is expected to occur primarily due to the albedo feedback and lapse-rate feedback. As the ice-covered regions expand equatorward, the amplification moves out of the polar region, which causes the Planck feedback to become slightly more negative (see SI Sec. S6 for details). Note that Fig. 4f indicates that polar amplification is not a ubiquitous feature of climate change within this wide range of climates.

Water-vapor feedback. The water-vapor feedback occurs because warmer air can hold more water vapor, which is a greenhouse gas. This feedback tends to be more positive in warmer climates, for reasons that can be explained using idealized one-dimensional radiative-convective equilibrium models (Meraner et al., 2013). Consistent with this, we find that the strength of the water-vapor feedback varies approximately monotonically with the underlying climate, becoming more positive with warming, although it becomes fairly constant in climates warmer than the PI.

5 Caveats

The results in Fig. 2 rely on direct model output in addition to the estimated CO_2 radiative forcing (F_{GHG}), which is computed using the line-by-line radiative transfer calculations of Byrne and Goldblatt (2014). These instantaneous radiative forcing (IRF) calculations do not account for stratospheric temperature adjustment, although they give similar results for our purposes to the line-by-line radiative model results of Etminan et al. (2016) which do include stratospheric temperature adjustment (SI Fig. S1). Neither calculation allows for the rapid adjustments to the tropospheric temperature profile in response to CO_2 forcing that are needed to estimate the effective radiative forcing (ERF; Sherwood et al., 2015).

We assess the error associated with this approach by comparing with two separate estimates of the ERF associated with CO_2 quadrupling from the PI level in CESM2, noting that the error may be larger for climates farther from the PI. First, we use a preexisting CESM2 run (Danabasoglu, 2019c) that has the sea-surface temperature (SST) field fixed at PI values and CO_2 increased by $4\times$ in order to estimate the ERF based on the change in TOA net radiation fields. Second, we use the regression method of Gregory et al. (2004) to estimate the ERF based on the first 20 years of a preexisting CESM2 simulation in which CO_2 was instantaneously quadrupled from its PI value (Danabasoglu, 2019b). In the latter analysis, the ERF is obtained by extrapolating the relationship between global-mean TOA net energy flux and surface temperature to zero surface temperature anomaly. The results are 8.90 W/m^2 for the fixed-SST ERF estimate and 8.77 W/m^2 for the regression method ERF estimate, compared with 8.56 W/m^2 in the line-by-line radiative transfer code IRF estimate that we adopt in this analysis. The close agreement between the IRF estimate from the radiative transfer code and the ERF estimate from

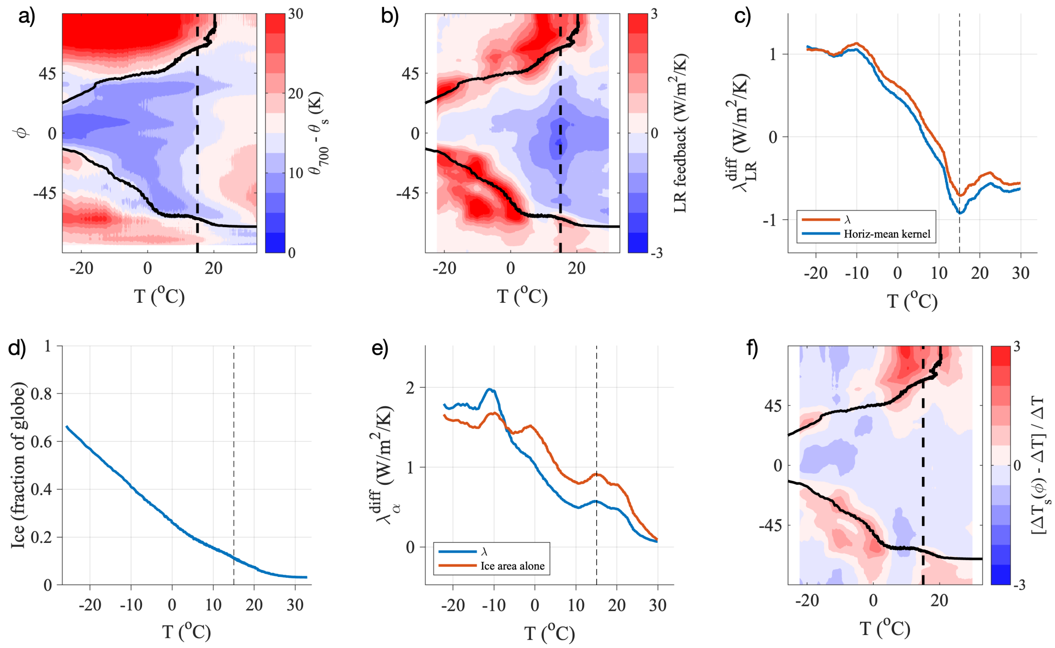

Figure 4. Physical interpretation of changes in individual feedback parameters. (a) Inversion strength, plotted as the difference in annual-mean zonal-mean potential temperature θ between the 700-hPa level and the surface. (b) Spatial structure of the annual-mean zonal-mean lapse-rate feedback parameter value (see SI Sec. S6). The spatial mean of this field gives the differential lapse-rate feedback parameter as estimated using annual-mean global-mean fields (SI Fig. S6). (c) Lapse-rate feedback parameter dependence on underlying climate. The red line is an approximation using only the global-mean atmospheric temperature profile, given by Eq. (S19) in SI Sec. S6, and the blue line is $\lambda_L^{\rm diff}$ repeated from Fig. 2. (d) Global ice area (as in Fig. 1c) plotted versus surface temperature. (e) Albedo feedback parameter dependence on underlying climate. The red line is an approximation using only the sensitivity of the total ice area to global-mean temperature (i.e., the slope of the curve in panel d), given by Eq. (S21) in SI Sec. S6, and the blue line is $\lambda_\alpha^{\rm diff}$ repeated from Fig. 2. (f) Pattern of amplified surface warming, shown as the change in the local departure of the zonal-mean temperature from the global-mean temperature, normalized by the change in the global-mean temperature (see SI Sec. S6). The black vertical dashed line in each panel indicates the PI climate. In panels a, b, and f, the black solid line indicates the 50% contour of the ice cover.

CESM2 may be coincidental given that CESM2, like most ESMs, shows substantial forcing adjustments from rapid changes in atmospheric temperature and cloud cover in response to CO_2 changes (e.g., Smith et al., 2020). However, this agreement gives confidence in the use of the IRF estimate (Fig. 2a) as an approximation to the ERF in CESM2 for our calculations.

Another consideration is whether radiative forcing should change with the underlying climate itself. Here we have adopted the standard definition of radiative forcing that assumes that CO_2 changes occur within a constant climate (i.e., fixed surface temperature), and hence that all radiatively-important atmospheric and surface field changes beyond rapid adjustments are part of the radiative feedback on surface temperature changes. However, another defensible choice for the differential feedback would be to define radiative forcing relative to the continuously evolving climate, in which case the CO_2 forcing would change depending on factors including changes in atmospheric water vapor, cloud cover, and the difference in temperature between the surface and the stratosphere (e.g., Jeevanjee et al., 2021; Romps et al., 2022). Calculating the radiative forcing under this alternative definition, which would require additional simulations, would modify the value of the differential feedback. Note that while this ambiguity in forcing definition is inherent to the differential feedback, the effective feedback only uses the standard radiative forcing definition adopted here because it is defined in terms of anomalies relative to the PI climate (Sherwood et al., 2015; Jeevanjee et al., 2021).

As noted above, the radiative kernel analysis does not allow the radiative response to perturbations in climate fields to evolve with the underlying climate because it effectively linearizes the simulated response about a climate near the PI. Furthermore, since the radiative kernels are set to zero above a fixed tropopause, radiative responses may not be calculated accurately in climates with a tropopause that is substantially higher than in the PI (e.g., Meraner et al., 2013).

To assess the accuracy of the kernel analysis, we re-ran the kernel analysis using clear-sky versions of the radiative kernels, which are included in the fields produced by Pendergrass et al. (2018). The residual between the sum of the clear-sky feedback parameters and the clear-sky TOA net energy flux reported by the model is indicated as a red dashed line in the bottom row of Fig. 3. This provides an estimate of the uncertainty in the analysis. Although not negligible, the values are relatively small. Note that cancelation between feedbacks may play a role in these relatively small residuals (cf. Koll & Cronin, 2018), especially for climates far from the PI.

Furthermore, for the lapse-rate and albedo feedbacks, which dominate net feedback changes in colder climates, we found that using a horizontally-averaged kernel produced similar results. That is, horizontal variations in the kernel between the warm tropics and cold poles have minimal influence on how feedbacks change across climate states; instead, feedback changes primarily track changes in the global ice extent and the globally-averaged vertical structure of the atmosphere. This insensitivity to capturing differences in the radiative kernels across the range of spatial variations in the control climate (from the tropics to the poles) suggests that changes in the radiative efficiency of the atmosphere across climate states may be of secondary importance, supporting the accuracy of this analysis which uses a kernel that does not vary with climate.

This analysis uses an approximately equilibrated PI climate, whereas the simulated climates that are increasingly warmer or colder than the PI are expected to be increasingly far from equilibrium. Hence it may be seen as a source of concern that the net climate feedback is found to be most negative near the PI and increasingly less negative in climates increasingly warmer or colder than the PI. However, a number of factors suggest that the level of equilibration is not substantially influencing the values of the climate feedback that we calculate. First, we identify simple and fairly basic physical processes that drive the increase in sensitivity with cooling (related to the lapse rate and albedo feedbacks), suggesting that this is likely to be a robust climate response, and the increase in sensitivity with warming has been previously identified as a robust feature of many climate models (e.g., Forster et al., 2021, their Fig. 7.11). Second, previous studies have found a loss of stability at the Snowball Earth bifurcation point, implying an increase in sensitivity as $\lambda_{net}^{\mathrm{diff}}$ approaches zero under extreme cooling. Third, the minimum climate feedback is in a climate that is approximately 2K colder than the PI, rather than being at the equilibrated PI climate. Fourth, this approach does not depend on the level of equilibration, at least when applied to a simplified representation of the climate system (SI Fig. S3). Fifth, we find that the impact of deep ocean heat uptake efficacy would not produce this shape (SI Fig. S4). And sixth, the clear-sky residual is relatively small (Fig. 3), showing that the alternative approach of using kernels, rather than the TOA balance used to generate the results in Figs. 2d,e, gives a similar result.

Moreover, we compared our results with a previous analysis (Hahn et al., 2021) of the CESM2 instantaneous CO_2 quadrupling simulation (Danabasoglu, 2019b). Hahn et al. (2021) used the same radiative kernels as the present study (Pendergrass et al., 2018), and their results include values for the feedback parameters around simulation year 100 (averaged over their simulation years 85-115), at which point the global-mean surface temperature is 6.6°C above the initial PI value. We indicate the feedback parameter values at this level of warming as blue circles in Fig. 3. The agreement with our analysis (blue lines in Fig. 3) adds some confidence to our interpretation that the relationships we find between feedback values and global temperatures do not depend strongly on the degree of equilibration. We similarly included a blue circle indicating their value for $\lambda_{net}^{\mathrm{eff}}$ in Fig. 2e, which agrees with our results (blue line in Fig. 2e).

Finally, the experimental design used here does not allow for slow feedbacks associated with factors including changes in ice sheets, the carbon cycle, and the deep ocean, which could modify the stability of the climate given sufficient time to adjust. These results should thus be interpreted as a measure of how the traditional fast feedbacks (i.e., Planck, water vapor, lapse rate, surface albedo, and clouds) depend on the underlying climate state, and they are relevant to studies that treat ice sheets and other slow feedbacks as external forcings (e.g., the LGM analysis of Sherwood et al., 2020). If ice sheets were allowed to change, it is expected that their distinct spatial structure of ERF would produce different relationships between climate feedbacks and global temperature changes than those under CO_2 forcing alone explored here (e.g., Zhu & Poulsen, 2021; Cooper et al., 2023). It is similarly expected that the results may differ if the model were allowed to approximately equilibrate to each level of CO_2, rather than using the 1% per year ramping adopted in the present study.

6 Summary and conclusions

As constraints on the modern-day ECS based on past warm and cold climates gain in prominence (e.g., Sherwood et al., 2020; Forster et al., 2021), it is becoming increasingly important to understand how and why climate feedbacks change over a wide range of climate states. In this study, we warmed and cooled a state-of-the-art climate model (NCAR CESM2) to simulate a continuum of climates ranging from a nearly ice-covered Snowball Earth to a nearly ice-free hothouse planet. We ramped CO_2 concentrations over a range of 11.5 doublings, which led to a 59K range in simulated annual-mean global-mean transient surface temperature changes.

Previous studies have represented the dependence of climate feedbacks on the underlying global temperature by approximating that the net feedback scales linearly, which is equivalent to including a quadratic term in the global energy budget (e.g., Sherwood et al., 2020). Our results suggest that this representation only approximately holds over a limited range of climates, spanning about 3K colder to 8K warmer than the PI climate. Importantly, LGM-like temperatures (4-6K colder than PI) fall outside of this range, suggesting that this representation is not accurate for assessing how LGM feedbacks relate to feedbacks in the modern-day or future climate, as has been done in previous analyses (e.g., Sherwood et al., 2020). The "U" shape of the relationship we find between the net feedback and global temperature implies a stronger constraint lowering the upper bound of the EffCS as inferred from LGM proxy reconstructions than reported by (Sherwood et al., 2020).

Since the relationship between the simulated net feedback and underlying climate is expected to depend on the choice of model, it would be useful to reproduce the present analysis using other ESMs. It is noteworthy that the 279-year and 514-year CO_2 ramping simulations generated for this analysis could be fairly straightforwardly repeated with a different ESM. This would be particularly valuable because paleoclimate constraints on the ECS all rely on mapping feedbacks between different climate states. Recent studies using CESM2 identified an apparent cold bias in the simulation of the LGM climate (Zhu et al., 2021) and warm bias in the simulation of the early Eocene (Zhu et al., 2020), and a new version of the model was developed with cloud feedbacks tuned to be less positive ("CESM2-PaleoCalibr", Zhu et al., 2022), which reduced the LGM bias and also resulted in a reduced modern-day ECS. Comparing the present analysis with a similar analysis that used CESM2-PaleoCalibr rather than CESM2 would further identify to what extent the tuning caused the dependence of the net feedback on the underlying climate to be shifted or restructured, which may shed further light on the way feedbacks in past climate states serve as analogs for feedbacks in the modern climate. That is, future work could determine whether identified biases in simulations of past warm climates using ESMs become reduced by changes in the value of the net feedback applying to all climates states (a vertical shift of the "U" shape in Figs. 2d,e) or by changes in the net feedback dependence on the underlying climate state (a change in the horizontal width of the "U" shape in Figs. 2d,e).

The results presented here are a first step toward mapping feedback changes over a wide range of climates. They place past and future climate changes in a broader context, with implications for our understanding of what physical mechanisms cause the sensitivity of each radiative feedback to the underlying climate state.

Data availability

Model output from the Warming and Cooling simulations is available at https://eisenman-group.github.io. The kernels used in this analysis were downloaded from https://github.com/apendergrass/cam5-kernels. Source data for the line plots in Figs. 1–4 are provided with this paper.

Code availability

Code to compute the differential and effective net feedback parameters (Fig. 2d,e) from the simulation output, which can similarly be used with the kernels to compute the individual feedback parameters (Fig. 3), is available at https://eisenman-group.github.io.

Acknowledgments

Without implying their endorsement, we thank Vince Cooper, Lily Hahn, Jack Bauchop, Ivan Mitevski, Jonah Bloch-Johnson, Brian Rose, Nadir Jeevanjee, Nick Lutsko, Matt Long, David Neelin, Angie Pendergrass, and Brendan Byrne for helpful discussions at various points during the course of this work. This work was supported by US National Science Foundation grants OCE-2048590, AGS-1752796, and OCE-2002276, and National Oceanic and Atmospheric Administration MAPP Program award NA20OAR4310391.

References

Anagnostou, E., John, E. H., Babila, T. L., Sexton, P. F., Ridgwell, A., Lunt, D. J., ... Foster, G. L. (2020). Proxy evidence for state-dependence of climate sensitivity in the Eocene greenhouse. *Nature Comm.*, *11*(1), 4436. doi: 10.1038/s41467-020-17887-x

Anagnostou, E., John, E. H., Edgar, K. M., Foster, G. L., Ridgwell, A., Inglis, G. N., ... Pearson, P. N. (2016). Changing atmospheric CO2 concentration was the primary driver of early Cenozoic climate. *Nature*, *533*(7603), 380–+. doi: 10.1038/nature17423

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophys. Res. Lett.*, *42*(12), 4973–4980. doi: 10.1002/2015GL064240

Bloch-Johnson, J., Rugenstein, M., Stolpe, M. B., Rohrschneider, T., Zheng, Y. Y., & Gregory, J. M. (2021). Climate sensitivity increases under higher CO2 levels due to feedback temperature dependence. *Geophys. Res. Lett.*, *48*(4), e2020GL089074. doi: 10.1029/2020GL089074

Byrne, B., & Goldblatt, C. (2014). Radiative forcing at high concentrations of well-mixed greenhouse gases. *Geophys. Res. Lett.*, *41*(1), 152–160. doi: 10.1002/2013GL058456

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*, *110*(35), 14162–14167. Retrieved from `http://www.pnas.org/cgi/doi/10.1073/pnas.1303365110` doi: 10.1073/pnas.1303365110

Cooper, V. T., Armour, K. C., Hakim, G., Tierney, J., Osman, M. B., Proistosescu, C., . . . Chmielowiec, P. (2023). Glacial maximum pattern effects reduce climate sensitivity estimates. *Submitted preprint*. doi: 10.31223/X5VD56

Danabasoglu, G. (2019a). *NCAR CESM2 model output prepared for CMIP6 CMIP 1pctco2.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.7497` doi: 10.22033/ESGF/CMIP6.7497

Danabasoglu, G. (2019b). *NCAR CESM2 model output prepared for CMIP6 CMIP abrupt-4xCO2.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.7519` doi: 10.22033/ESGF/CMIP6.7519

Danabasoglu, G. (2019c). *NCAR CESM2 model output prepared for CMIP6 RFMIP piClim-4xCO2.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.7706` doi: 10.22033/ESGF/CMIP6.7706

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., . . . Strand, W. G. (2020). The Community Earth System Model version 2 (CESM2). *J. Adv. In Modeling Earth Systems*, *12*(2), e2019MS001916. doi: 10.1029/2019MS001916

Duan, L., Cao, L., & Caldeira, K. (2019). Estimating contributions of sea ice and land snow to climate feedback. *Journal of Geophysical Research: Atmospheres*, *124*(1), 199–208. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029093` doi: 10.1029/2018JD029093

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. *Geophys. Res. Lett.*, *43*(24), 12614–12623. doi: 10.1002/2016GL071930

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., . . . Zhang, H. (2021). The earth's energy budget, climate feedbacks, and climate sensitivity [Book Section]. In V. Masson-Delmotte et al. (Eds.), *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (p. 923–1054). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/9781009157896.009

Friedrich, T., Timmermann, A., Tigchelaar, M., Timm, O. E., & Ganopolski, A. (2016). Nonlinear climate sensitivity and its implications for future greenhouse warming. *Science Advances*, *2*(11), e1501923. doi: 10.1126/sciadv.1501923

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., & 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. *J. Climate*, *26*(6), 1859–1876. doi:

10.1175/JCLI-D-12-00196.1

Good, P., Lowe, J. A., Andrews, T., Wiltshire, A., Chadwick, R., Ridley, J. K., … Shiogama, H. (2015). Nonlinear regional warming with increasing CO2 concentrations. *Nature Climate Change*, *5*(2), 138–142. doi: 10.1038/nclimate2498

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. *Geophys. Res. Lett.*, *31*(3), L03205. doi: 10.1029/2003GL018747

Hahn, L. C., Armour, K. C., Zelinka, M. D., Bitz, C. M., & Donohoe, A. (2021). Contributions to Polar Amplification in CMIP5 and CMIP6 models. *Frontiers In Earth Sci.*, *9*. doi: 10.3389/feart.2021.710036

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F. R., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Climate*, *23*(9), 2418–2427. doi: 10.1175/2009JCLI3466.1

Jeevanjee, N., Seeley, J. T., Paynter, D., & Fueglistaler, S. (2021). An analytical model for spatially varying clear-sky CO2 forcing. *J. Climate*, *34*(23), 9463–9480. doi: 10.1175/JCLI-D-19-0756.1

Jonko, A. K., Shell, K. M., Sanderson, B. M., & Danabasoglu, G. (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. *J. Climate*, *25*(15), 5260–5272. doi: 10.1175/JCLI-D-11-00524.1

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. doi: 10.1175/JCLI-D-12-00479.1

Kluft, L., Dacie, S., Brath, M., Buehler, S. A., & Stevens, B. (2021). Temperature-dependence of the clear-sky feedback in radiative-convective equilibrium. *Geophys. Res. Lett.*, *48*(22), e2021GL094649. doi: 10.1029/2021GL094649

Kohler, P., Stap, L. B., von der Heydt, A. S., de Boer, B., van de Wal, R. S. W., & Bloch-Johnson, J. (2017). A state-dependent quantification of climate sensitivity based on paleodata of the last 2.1 million years. *Paleoceanography*, *32*(11), 1102–1114. doi: 10.1002/2017PA003190

Koll, D. D. B., & Cronin, T. W. (2018). Earth's outgoing longwave radiation linear due to H2O greenhouse effect. *Proc. National Acad. Sciences United States Am.*, *115*(41), 10293–10298. doi: 10.1073/pnas.1809868115

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., … Zhou, T. (2021). Future global climate: Scenario-based projections and near-term information [Book Section]. In V. Masson-Delmotte et al. (Eds.), *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (p. 553–672). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/9781009157896.006

Marotzke, J., & Botzet, M. (2007). Present-day and ice-covered equilibrium states in a comprehensive climate model. *Geophys. Res. Lett.*, *34*(16), L16704. doi: 10.1029/2006GL028880

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., … Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing co2. *J. Adv. In Modeling Earth Systems*, *11*(4), 998–1038. doi: 10.1029/2018MS001400

Meraner, K., Mauritsen, T., & Voigt, A. (2013). Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.*, *40*(22), 5944–5948. doi: 10.1002/2013GL058118

Merlis, T. M., Feldl, N., & Caballero, R. (2022). Changes in Poleward Atmospheric Energy Transport over a Wide Range of Climates: Energetic and Diffusive

Perspectives and A Priori theories. *J. Climate*, *35*(20), 2933–2948. doi: 10.1175/JCLI-D-21-0682.1

Pendergrass, A. G., Conley, A., & Vitt, F. M. (2018). Surface and top-of-atmosphere radiative feedback kernels for cesm-cam5. *Earth System Sci. Data*, *10*(1), 317–324. doi: 10.5194/essd-10-317-2018

Pierrehumbert, R. T., Abbot, D. S., Voigt, A., & Koll, D. (2011). Climate of the neoproterozoic. *Ann. Rev. Earth Planetary Sciences, Vol 39*, *39*, 417–460. doi: 10.1146/annurev-earth-040809-152447

Roe, G. H., & Armour, K. C. (2011). How sensitive is climate sensitivity? *Geophys. Res. Lett.*, *38*, L14708. doi: 10.1029/2011GL047913

Roe, G. H., & Baker, M. B. (2010). Notes on a catastrophe: A feedback analysis of Snowball Earth. *J. Climate*, *23*(17), 4694–4703. doi: 10.1175/2010JCLI3545.1

Romps, D. M., Seeley, J. T., & Edman, J. P. (2022). Why the forcing from carbon dioxide scales as the logarithm of its concentration. *J. Climate*, *35*(13), 4027–4047. doi: 10.1175/JCLI-D-21-0275.1

Royer, D. L. (2016). Climate Sensitivity in the Geologic Past. *Annual Review of Earth and Planetary Sciences*, *44*(1), 277–293. Retrieved from `http://www.annualreviews.org/doi/10.1146/annurev-earth-100815-024150` doi: 10.1146/annurev-earth-100815-024150

Rugenstein, M. A. A., & Armour, K. C. (2021). Three flavors of radiative feedbacks and their implications for estimating equilibrium climate sensitivity. *Geophys. Res. Lett.*, *48*(15), e2021GL092983. doi: 10.1029/2021GL092983

Seeley, J. T., & Jeevanjee, N. (2021). H2o windows and co2 radiator fins: A clear-sky explanation for the peak in equilibrium climate sensitivity. *Geophys. Res. Lett.*, *48*(4), e2020GL089609. doi: 10.1029/2020GL089609

Shaffer, G., Huber, M., Rondanelli, R., & Pepke Pedersen, J. O. (2016). Deep time evidence for climate sensitivity increase with warming. *Geophysical Research Letters*, *43*(12), 6538–6545. doi: 10.1002/2016GL069243

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., & Stevens, B. (2015). Adjustments in the forcing-feedback framework for understanding climate change. *Bulletin Am. Meteorol. Soc.*, *96*(2), 217–228. doi: 10.1175/BAMS-D-13-00167.1

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., … Zelinka, M. D. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence. *Rev. Geophysics*, *58*(4), e2019RG000678. doi: 10.1029/2019RG000678

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., … Forster, P. M. (2020). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chem. Phys.*, *20*(16), 9591–9618. doi: 10.5194/acp-20-9591-2020

Snyder, C. W. (2019). Revised estimates of paleoclimate sensitivity over the past 800,000 years. *Climatic Change*, *156*(1), 121–138. Retrieved from `https://doi.org/10.1007/s10584-019-02536-0` doi: 10.1007/s10584-019-02536-0

Stap, L. B., Kohler, P., & Lohmann, G. (2019). Including the efficacy of land ice changes in deriving climate sensitivity from paleodata. *Earth System Dynam.*, *10*(2), 333–345. doi: 10.5194/esd-10-333-2019

Stolpe, M. B., Medhaug, I., Beyerle, U., & Knutti, R. (2019, oct). Weak dependence of future global mean warming on the background climate state. *Climate Dynamics*, *53*(7), 5079–5099. Retrieved from `https://doi.org/10.1007/s00382-019-04849-3` doi: 10.1007/s00382-019-04849-3

Voigt, A., Abbot, D. S., Pierrehumbert, R. T., & Marotzke, J. (2011). Initiation of a Marinoan Snowball Earth in a state-of-the-art atmosphere-ocean general circulation model. *Climate Past*, *7*(1), 249–263. doi: 10.5194/cp-7-249-2011

Voigt, A., & Marotzke, J. (2010). The transition from the present-day climate to a modern Snowball earth. *Climate Dynam.*, *35*(5), 887–905. doi: 10.1007/s00382-009-0633-5

von der Heydt, A. S., Kohler, P., van de Wal, R. S. W., & Dijkstra, H. A. (2014). On the state dependency of fast feedback processes in (paleo) climate sensitivity. *Geophys. Res. Lett.*, *41*(18), 6484–6492. doi: 10.1002/2014GL061121

Winton, M. (2011). Do climate models underestimate the sensitivity of Northern Hemisphere sea ice cover? *J. Climate*, *24*(15), 3924–3934. doi: 10.1175/2011JCLI4146.1

Winton, M., Takahashi, K., & Held, I. M. (2010). Importance of ocean heat uptake efficacy to transient climate change. *J. Climate*, *23*(9), 2333–2344. doi: 10.1175/2009JCLI3139.1

Wood, R., & Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *J. Climate*, *19*(24), 6425–6432. doi: 10.1175/JCLI3988.1

Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Gettelman, A., Bacmeister, J. T., Neale, R. B., ... Kay, J. E. (2022). LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2. *J. Adv. In Modeling Earth Systems*, *14*(4), e2021MS002776. doi: 10.1029/2021MS002776

Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Poulsen, C. J., Tierney, J. E., Lofverstrom, M., & DiNezio, P. (2021). Assessment of equilibrium climate sensitivity of the Community Earth System Model version 2 through simulation of the Last Glacial Maximum. *Geophys. Res. Lett.*, *48*(3), e2020GL091220. doi: 10.1029/2020GL091220

Zhu, J., & Poulsen, C. J. (2020). On the increase of climate sensitivity and cloud feedback with warming in the Community Atmosphere Models. *Geophys. Res. Lett.*, *47*(18), e2020GL089143. doi: 10.1029/2020GL089143

Zhu, J., & Poulsen, C. J. (2021). Last Glacial Maximum (LGM) climate forcing and ocean dynamical feedback and their implications for estimating climate sensitivity. *Climate Past*, *17*(1), 253–267. doi: 10.5194/cp-17-253-2021

Zhu, J., Poulsen, C. J., & Otto-Bliesner, B. L. (2020). High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nature Climate Change*, *10*(5), 378–379. doi: 10.1038/s41558-020-0764-6

Zhu, J., Poulsen, C. J., & Tierney, J. E. (2019). Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks. *Sci. Adv.*, *5*(9), eaax1874. doi: 10.1126/sciadv.aax1874

Supplementary Information for "The radiative feedback continuum from Snowball Earth to an ice-free hothouse"

Ian Eisenman and Kyle Armour

S1 Simulation details

We use NCAR CESM2 in its standard workhorse configuration. The atmospheric component is CAM6, and the ocean component is POP2. The atmosphere and ocean both have nominal horizontal resolutions of 1°, and there are 32 vertical levels in the atmosphere and 60 vertical levels in the ocean.

The Warming and Cooling simulations are both branched from the end of year 500 of a previously run pre-industrial (PI) control simulation (Danabasoglu et al., 2020) with the forcing fixed at 1850 levels. The atmospheric CO_2 concentration is increased or decreased at a rate of 1% per year from the start of each simulation. For the first 150 years of the Warming run, we use the pre-existing CESM2 "1pctCO2" simulation that is part of the CMIP6 archive (Danabasoglu, 2019a), which we extend to simulate further warming by branching to a cloned case. The Cooling run is identical to the Warming run except that the CO_2 change has the opposite sign.

Warming run details. This run initially failed during year 151 with the error "bounding bracket for pH solution not found" from `co2calc.F90`. Adjusting the POP time step from the default value `dt_count=48` to `dt_count=60` during years 151-152 caused this error to no longer occur. After year 279, there was an error in `lnd_import_export.F90` that the coupler was receiving an output of NaN from the land model. We were not able to resolve this error by reducing the CAM time step and ended the run after year 279.

Cooling run details. In year 279, this run failed with the error "bounding bracket for pH solution not found" from `co2calc.F90`, which was not resolved by increasing `dt_count`. So we commented out the line in the model code that called this error, which may lead to unreliable simulated pH. In year 333, when the CO_2 concentration reached approximately 10 ppm, the land component of the model failed with the error "CO2 is outside of an expected range" in `lnd_import_export.F90`, and we commented out the line in the model code that called this error. At the end of the 514-year run, there was an error with the iron flux being out of range in `marbl_diagnostics_mod.F90`, which we were not able to resolve by simply commenting out the line in the model code that called this error.

Quantities analyzed. For CO_2, we use the atmospheric field `co2vmr`, which is the CO_2 volume mixing ratio. For surface temperature, we use the atmospheric field `TS`, which is the radiative surface temperature. For the measure of inversion strength in Fig. 4a, we compute the potential temperature from the atmospheric temperature `T` at vertical level 23, which is at approximately 700 hPa on the model hybrid vertical coordinate. For ice cover, we take the maximum of the fields `FSNO` and `PCT_GLACIER/100`, multiply this value by `landfrac`, and then add `ICEFRAC`. Here `FSNO` is the fraction of ground covered by snow reported by the land model, `PCT_GLACIER` is the percent of ground covered by glaciers which is included in the surface dataset input used by the land model, `landfrac` is the fraction of the grid box covered by land reported by the land model, and `ICEFRAC` is the fraction of the grid box covered by sea ice reported by the atmospheric model. We compute the net energy flux N as `FSNT` − `FLNT`, with `FSNT` and `FLNT` the top-of-model net longwave and solar fluxes reported by the atmospheric model.

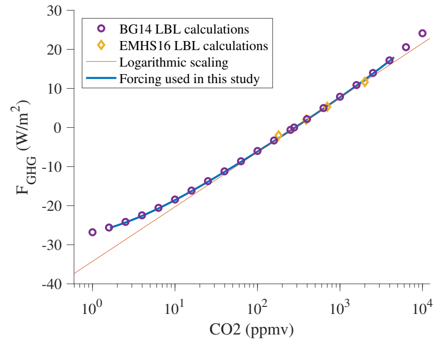

Figure S1. CO_2 radiative forcing. Circles indicate values from the Byrne and Goldblatt (2014) supplemental data file "text03.txt", diamonds indicate values from the Etminan et al. (2016) supplemental data table S1, the thin orange line indicates a logarithmic scaling of $F_{GHG} = F_{2\times} \log_2 (C/C_0)$ with $F_{2\times} = 4.2$ W/m^2 and C the varying CO_2 concentration which is scaled by the PI value $C_0 = 284.7$ ppm, and the thick blue line indicates the CO_2 radiative forcing used in this study (see Sec. S2 above). Here the data from Etminan et al. (2016) includes their 4 runs with their default concentrations of CH_4 and N_20 and varied CO_2 concentrations, and both the data from Etminan et al. (2016) and the data from Byrne and Goldblatt (2014) are shifted vertically such that the forcing is zero at 284.7 ppm.

S2 CO_2 forcing

Byrne and Goldblatt (2014) used a line-by-line radiative transfer code to calculate forcing from CO_2 (and other greenhouse gases). The publication includes a supplemental data text file ("text03.txt") that has radiative forcing associated with CO_2 concentrations varying from 1 ppm to 100,000 ppm. Although this is a considerably wider range of CO_2 concentrations than mentioned in their actual paper, the supplemental data values are valid output from their radiative model (Brendan Byrne, personal communication, January 2021).

The CO_2 in our simulations ranges from 1.6 ppm to 3422 ppm. We calculate the associated radiative forcing F_{GHG} using a cubic interpolation of the relationship between the radiative forcing associated with the global-mean annual-mean profile ("GAM" in "text03.txt") and the logarithm of the CO_2 concentration ("CO2" in "text03.txt"), which is shown in Fig. S1.

S3 Calculation of effective and differential feedback parameters

For the effective feedback parameters, we smooth each radiative response time series (F_{net} or F_i) using a least-squares fit to a 12th-order polynomial in $(T - T_0)$ that is constrained to go through (T_0, F_0), where T_0 and F_0 are the surface temperature and radiative response (F_{net} or F_i) averaged over years 480-499 of the PI simulation. This allows the ratio in Eq. (2) to be smooth even in the limit $T \to T_0$. This smoothing of F_{net}, and the resulting values of $\lambda_{net}^{\text{eff}}$ and EffCS, are plotted in Fig. S2 next to the raw unsmoothed annual-mean simulation output.

For the differential feedback parameters, we regress the radiative response (F_{net} or F_i) on the surface temperature T. We use a total-least-squares (TLS) regression, rather than a standard ordinary-least-squares (OLS) regression, because the radiative response (F_{net} or F_i) and temperature (T) both play the role of dependent variables. A TLS regression accounts for errors in both variables, whereas an OLS regression accounts for

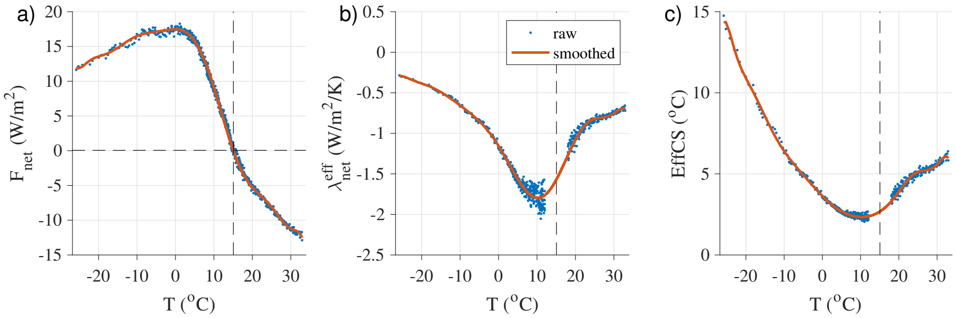

Figure S2. Smoothing of model output for $\lambda_{net}^{\text{eff}}$ calculation. (a) Net radiative response F_{net}. (b) Net effective feedback $\lambda_{net}^{\text{eff}}$. (c) Effective climate sensitivity EffCS. In each panel, the blue dots indicate the results with no smoothing of the raw annual-mean model output, and the red lines indicate the results after smoothing F_{net} with a 12th-order polynomial that is constrained to go through the PI reference climate values (black dashed lines in panel a). Because the numerator and denominator of Eq. 2 both asymptote to 0 at the PI climate, leading to large values of the ratio, the raw output is not plotted in panels b and c for T within 3K of the PI value.

errors in one variable and treats the other as an independent variable. The TLS regression depends on the choice of units, and we normalize each variable by the standard deviation of the residuals of the time trend, following Winton (2011). We compute the TLS regression in a running window of variable duration that spans temperatures in the range $\pm 3K$.

S4 Idealized models

The interpretation of these results may be aided by considering idealized models that roughly mimic the CESM2 simulations. Here we consider first a single-layer model, and then a two-layer model, both of which are represented by simple ordinary differential equations.

Single-layer model. We begin with a single-layer model approximation to the terms in Eq. (1). We set $\Delta N = c\,\frac{dT}{dt}$, where t is time and $c = 15$ W yr/K/m^2 is the effective heat capacity describing the relationship between T and energy absorbed in the climate system with a value based on fitting the CESM2 results. We use a fourth-order polynomial approximation of the relationship between ΔF_{net} and T in Fig. 2c: $\Delta F_{net}(T) = p_1\,\Delta T + p_2\,\Delta T^2 + p_3\,\Delta T^3 + p_4\,\Delta T^4$, where we define ΔT as the departure from the PI value of T as in the calculation of $\lambda_{net}^{\text{eff}}$; and similarly ΔF_{net} is the departure from the value in the PI climate, which we approximate to be zero such that $F_{net} = \Delta F_{net}$. The coefficients are $p_1 = -1.7$ W/m^2/K, $p_2 = 0.029$ W/m^2/K^2, $p_3 = 0.0042$ W/m^2/K^3, and $p_4 = 6.2\times10^{-5}$ W/m^2/K^4. This idealized representation of Eq. (1) takes the form of a nonlinear ordinary differential equation:

$$\frac{dT}{dt} = f(\Delta F_{GHG}, T) \equiv \frac{1}{c}\,\left[\Delta F_{GHG} + \Delta F_{net}(T)\right]. \tag{S1}$$

The associated feedback parameters $\lambda_{net}^{\text{diff}}$ and $\lambda_{net}^{\text{eff}}$ can be readily derived analytically in terms of the fit parameters in ΔF_{net}:

$$\lambda_{net}^{\text{diff}} \equiv \frac{dF_{net}}{dT} = p_1 + 2\,p_2\,\Delta T + 3\,p_3\,\Delta T^2 + 4\,p_4\,\Delta T^3, \tag{S2}$$

$$\lambda_{net}^{\text{eff}} \equiv \frac{\Delta F_{net}}{\Delta T} = p_1 + p_2\,\Delta T + p_3\,\Delta T^2 + p_4\,\Delta T^3. \tag{S3}$$

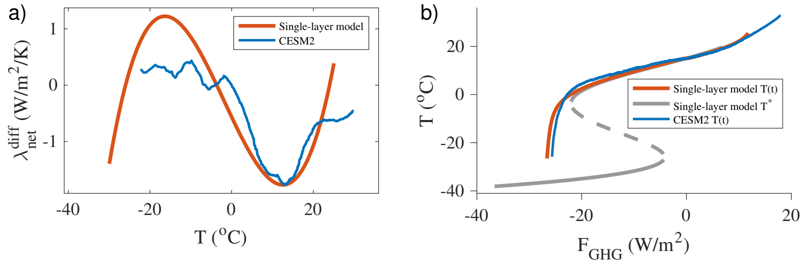

Figure S3. Single-layer idealized model result. (a) The polynomial representation of the feedback parameter $\lambda_{net}^{\rm diff}$ (Eq. (S2)) used in the idealized model (red), with the CESM2 results (as in Fig. 2d) included for comparison (blue). (b) The dependence of the temperature on the forcing. The time-evolving temperature simulated by the idealized model is shown in red. The steady-state solutions are shown in gray, with solid lines for stable solutions and a dashed line for the unstable solution. The CESM2 results are included for comparison (blue).

This system has steady-state solutions T^* that solve $0 = f(\Delta F_{GHG}, T^*)$, and the stability of these fixed points is dictated by $\frac{df}{dT} = \frac{1}{c}\frac{dF_{net}}{dT} = \frac{1}{c}\lambda_{net}^{\rm diff}$ evaluated at $T = T^*$. The idealized polynomial representation of $\lambda_{net}^{\rm diff}$ vs T in Eq. (S2) is shown in Fig. S3a.

Beginning from the fixed point with $\Delta F_{GHG} = 0$ (representing the PI), we increase and decrease the forcing as $\Delta F_{GHG} = \pm a\, t$, with $a = \pm 0.055$ W/m^2/yr based on fitting the $\pm 1\%$ per year ramping of CO_2 in CESM2. The resulting time-evolving temperature is plotted versus the forcing in Fig. S3b (red line). Steady-state solutions are indicated in gray, with solid lines for stable solutions and a dashed line for the unstable solution. The CESM2 simulation results are included for comparison (blue line). Here the time-evolving temperature is computed from Eq. (S1) using numerical time stepping, and the steady-state solutions are computed using a polynomial root finder. It should be emphasized that this idealized model is presented as a tool to help explain the CESM2 results, rather than to add any quantitative information to the analysis; for example, the temperature associated with the cold stable state in Fig. S3b results from extrapolation outside the range of the CESM2 simulations and hence is sensitive to the details of the polynomial fit.

This helps to illustrate how the analysis used in this study does not depend on how equilibrated the climate system is with the evolving value of ΔF_{GHG}. Furthermore, the steady-state solutions of the ordinary differential equation can be readily found, indicating an unstable state at temperatures colder than the Snowball Earth bifurcation point, with a stable Snowball Earth state existing at even colder temperatures (beyond the range of climates simulated with CESM2). The time evolution of this simple system (Fig. S3b) helps illustrate how the positive values of $\lambda_{net}^{\rm diff}$ indicate times when the climate is transiently evolving across temperatures for which the only steady-state solution is unstable, rather than for example indicating an exponentially growing departure from an unstable climate state.

Two-layer model. We use the two-layer model of Held et al. (2010), which takes the form

$$c_s\,\frac{dT}{dt} = \Delta F_{GHG} + \lambda_0\,\Delta T + \epsilon\,\gamma\,(\Delta T_d - \Delta T) \quad \text{(S4)}$$

$$c_d\,\frac{dT_d}{dt} = \gamma\,(\Delta T - \Delta T_d). \quad \text{(S5)}$$

Here c_s is the heat capacity of the ocean surface layers that respond rapidly to the atmosphere, and we approximate this layer to be characterized by the surface temperature T, with ΔT the departure from the PI value of T as above; c_d and T_d are the heat

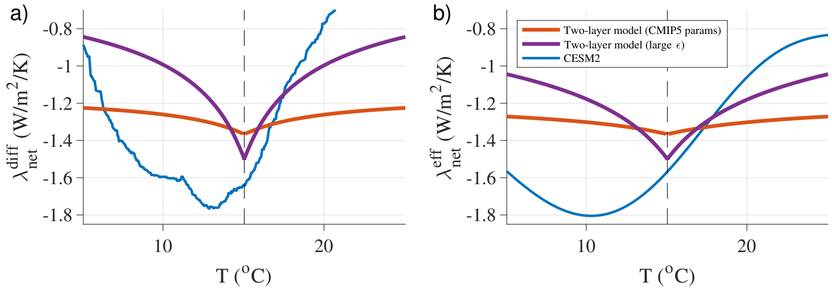

Figure S4. Two-layer idealized model results, showing the (a) differential and (b) effective net climate feedback. We include two parameter sets: moderate parameters estimated from the CMIP5 ensemble mean (red) and parameters adjusted to have a large deep ocean heat uptake efficacy (magenta). The CESM2 simulation results (as in Fig. 2d,e) are also included in blue.

capacity and temperature associated with the deeper ocean, with ΔT_d being the departure from the PI value of T_d; γ is a coefficient governing the heat-exchange between the two layers; λ_0 is a constant reference value of $\lambda^{\mathrm{eff}}_{net}$; and ϵ is a factor to account for the deep ocean heat uptake efficacy associated with changes in the relationship between F_{net} and T due to the spatial pattern of surface temperature changes as the climate evolves toward equilibrium. Note that the extent to which $\epsilon > 1$ represents the level of efficacy.

In this case the TOA net energy flux is equal to the change in heat content of the surface and deep layers, $\Delta N = c_s \frac{dT}{dt} + c_d \frac{dT_d}{dt} = c_s \frac{dT}{dt} + \gamma(\Delta T - \Delta T_d)$, and the deep ocean heat uptake efficacy influences the climate response as $\Delta F_{net} = \lambda_0 \, \Delta T - (\epsilon - 1)\, \gamma \, (\Delta T - \Delta T_d)$. Note that here $F_{net} = \Delta F_{net}$ as for the single-layer model. With these relationships, Eqs. (S4)-(S5) are equivalent to Eq. (1). The resulting feedback parameters are

$$\lambda^{\mathrm{diff}}_{net} \equiv \frac{dF_{net}}{dT} = \lambda_0 - (\epsilon - 1)\, \gamma \left(1 - \frac{dT_d}{dT}\right) \tag{S6}$$

$$\lambda^{\mathrm{eff}}_{net} \equiv \frac{\Delta F_{net}}{\Delta T} = \lambda_0 - (\epsilon - 1)\, \gamma \left(1 - \frac{\Delta T_d}{\Delta T}\right). \tag{S7}$$

We use parameter values $\epsilon = 1.28$, $\lambda_0 = -1.18$ W/m^2/K, $c_s = 8.2$ W yr/K/m^2, $c_d = 109$ W yr/K/m^2, and $\gamma = 0.67$ W/m^2/K, which are estimated from CMIP5 ensemble-mean simulation results (Geoffroy et al., 2013). As above, we begin from an equilibrated state and then increase and decrease the forcing as $\Delta F_{GHG} = \pm a\, t$ with $a = \pm 0.055$ W/m^2/yr in order to mimic the $\pm 1\%$ per year ramping of CO_2 in CESM2. The time-evolving temperature is computed using numerical time stepping, and the term $\frac{dT_d}{dT}$ in Eq. (S6) is computed as the ratio of time derivatives, which we express in terms of the forcing and temperatures using Eq. (S4)-(S5). The resulting dependence of $\lambda^{\mathrm{diff}}_{net}$ and $\lambda^{\mathrm{eff}}_{net}$ on T is shown in Fig. S4 (red line). We also consider the impact of a large deep ocean heat uptake efficacy by using $\epsilon = 2.5$ and adjusting the reference value of the feedback to $\lambda_0 = -0.5$ W/m^2/K, which is plotted in magenta. The CESM2 simulation result is included for comparison (blue line).

S5 Radiative kernel analysis

The radiative kernel fields were computed by Pendergrass et al. (2018) with the Parallel Offline Radiative Transfer model updated for compatibility with NCAR CAM5. The associated dataset includes monthly-mean radiative kernels associated with (1) surface temperature, (2) atmospheric temperature, (3) water vapor, and (4) surface albedo. The kernels, which vary as a function of space and time of year, represent the quantity $K_i \equiv$

$\partial R/\partial v_i$, where R is the TOA net radiative response and v_i is the relevant component of the simulated climate. All kernels are set to zero above the tropopause, which is approximated as a linear function of the cosine of the latitude. The dataset also includes kernels computed using clear-sky radiative fields. We use the mean annual cycle averaged over years 480-499 of the PI simulation as the reference climate.

We define the annual-mean, zonal-mean, meridional-mean, and vertical-integration operations as

$$\begin{aligned} \langle\,\cdot\,\rangle_t \equiv \frac{1}{(1\text{ yr})}\int_0^{1\text{ yr}} \cdot\, dt, \;\; \langle\,\cdot\,\rangle_\theta \equiv \frac{1}{(360^\circ)}\int_0^{360^\circ} \cdot\, d\theta, \\ \langle\,\cdot\,\rangle_\phi \equiv \frac{1}{2}\int_{-90^\circ}^{90^\circ} \cdot\, w(\phi)\, d\phi, \;\; \{\,\cdot\,\}_p \equiv \int_{p_t}^{p_s} \cdot\, dp. \end{aligned} \tag{S8}$$

Here t is time, θ is longitude, ϕ is latitude, p is vertical pressure level, p_s is the surface pressure, p_t is the approximate tropopause pressure, and averages are performed on CESM2 model levels unless otherwise noted. Note that following Pendergrass et al. (2018), we do not use the simulated varying p-field in the model, instead using a specified pressure field as a function of space and time of year based on a control simulation and the CAM hybrid grid. The gaussian weight $w(\phi) \approx \frac{\pi}{180^\circ} \cos\phi$ gives the area-weighting for each latitude; note that it departs slightly from a simple $\cos\phi$ scaling due to the details of the model grid. In what follows, a series of subscripts will indicate that series of averaging operations.

Note that although the CESM2 runs in the present study use CAM6, whereas the kernels are computed based on CAM5, both model versions have the same horizontal resolution. However, CAM6 has 32 levels and CAM5 has 30 levels, with the difference in vertical levels being confined exclusively to the stratosphere (the vertical levels are identical below the 88 hPa level). Since the kernel analysis is confined to the troposphere, the additional vertical resolution in the stratosphere in CAM6 does not require any interpolation of model fields.

The annual-mean global-mean radiative responses associated with each feedback (F_i), which are shown in Fig. S5, are computed by multiplying the monthly-mean simulation output during a given year with the radiative kernel and then averaging over time and space.

For the Planck feedback, this takes the form

$$\Delta F_P = \langle\, K_P(t,\theta,\phi)\, \Delta T_{s,2D}(t,\theta,\phi)\,\rangle_{t,\theta,\phi}\,, \tag{S9}$$

where K_P is the kernel and $T_{s,2D}$ is the surface temperature field.

For the lapse-rate feedback, it is

$$\Delta F_L = \Big\langle\, \{\, K_L(t,\theta,\phi,p)\, \Delta T_a'(t,\theta,\phi,p)\,\}_p \Big\rangle_{t,\theta,\phi}\,, \tag{S10}$$

where K_L is the kernel and $\Delta T_a'$ the departure of the 3D temperature change from the surface temperature change.

We use the “logarithmic” water-vapor kernel K_w in the Pendergrass et al. (2018) dataset, for which the radiative response takes the form

$$\Delta F_w = \left\langle \left\{ K_w(t,\theta,\phi,p)\, \Delta Q(t,\theta,\phi,p) \,/\, \left[\frac{\Delta Q}{\Delta T_a}\right]_h \right\}_p \right\rangle_{t,\theta,\phi}\,, \tag{S11}$$

where the term $\left[\frac{\Delta Q}{\Delta T_a}\right]_h$ describes the change in specific humidity under constant relative humidity and is a function of the 3D temperature field T_a, and the kernel K_w gives

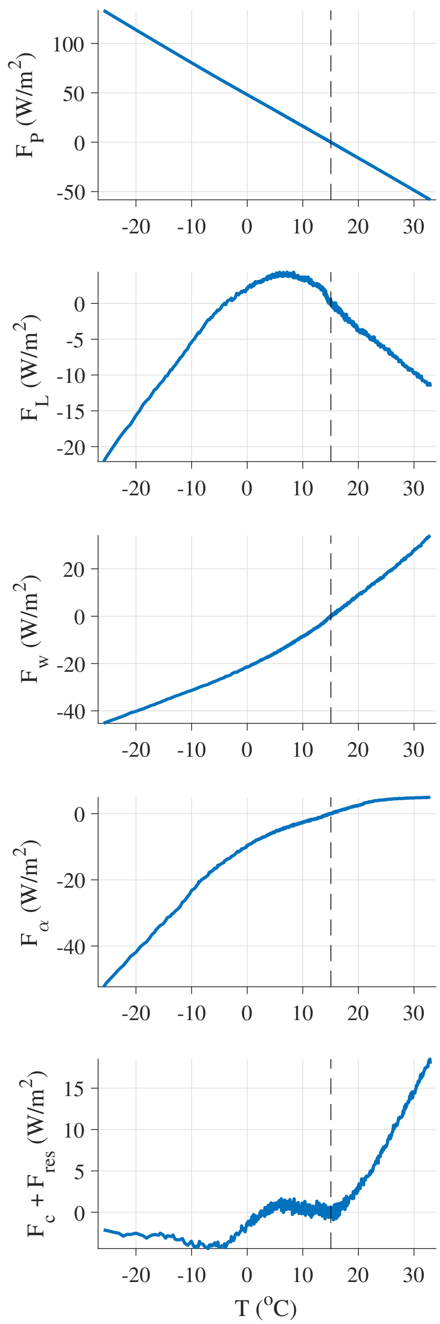

Figure S5. Radiative response associated with each feedback, computed using the radiative kernels.

the change in TOA radiation per change in atmospheric temperature that would occur if the specific humidity Q increased with the relative humidity h remaining as it is in the reference climate. Note that this representation does account for changes in relative humidity: there is essentially a normalization factor associated with constant relative humidity that is multiplied into K_w and divided out of the simulated field, such that it cancels in Eq. (S11).

For the albedo feedback, the radiative response is

$$\Delta F_\alpha = \langle\, K_\alpha(t,\theta,\phi)\, \Delta\alpha(t,\theta,\phi)\,\rangle_{t,\theta,\phi}\,, \tag{S12}$$

where K_α is the kernel and $\Delta\alpha$ is the change in the surface albedo, which is computed as $\alpha = S_{up}/S_{down}$ with S_{up} and S_{down} the upward and downward shortwave radiation at the surface.

S6 Approximating with annual-mean zonal-mean analysis

In Section 4.1 of the main text, we interpret the results with the aid of a simplified analysis that uses annual-mean zonal-mean radiative kernels and simulated fields. We define the annual and zonal average of the kernels as

$$\begin{aligned} \tilde{K}_P(\phi) &\equiv \langle K_P(t,\theta,\phi)\rangle_{t,\theta}\,, \;\; \tilde{K}_L(\phi,p) \equiv \langle K_L(t,\theta,\phi,p)\rangle_{t,\theta}\,, \\ \tilde{K}_w(\phi,p) &\equiv \langle K_w(t,\theta,\phi,p)\rangle_{t,\theta}\,, \;\; \tilde{K}_\alpha(\phi) \equiv \langle K_\alpha(t,\theta,\phi)\rangle_{t,\theta}\,, \end{aligned} \tag{S13}$$

and we similarly define the annual and zonal average of the relevant simulated climate fields as

$$\tilde{T}_s(\phi) \equiv \langle T_{s,2D}(t,\theta,\phi)\rangle_{t,\theta}\,, \;\; \tilde{T}_a(\phi,p) \equiv \langle T_a(t,\theta,\phi,p)\rangle_{t,\theta}\,, \;\; \tilde{\alpha}(\phi) \equiv \frac{\langle S_{up}(t,\theta,\phi)\rangle_{\phi,p}}{\langle S_{down}(t,\theta,\phi)\rangle_{\phi,p}}. \tag{S14}$$

Note that for the annual-mean zonal-mean albedo field $\tilde{\alpha}$, this uses the ratio of the means rather than the mean of the ratio, which is important for the approximate match between Fig. 3 and Fig. S6.

For the water-vapor feedback, we further approximate that the relative humidity remains as in the reference climate. Under this approximation, the terms in Eq. (S11) involving humidity simplify to $\Delta Q \,/\, \left[\frac{\Delta Q}{\Delta T_a}\right]_h = \Delta T_a$.

The resulting approximate radiative responses, which we indicate as $\tilde{F}_i$, are

$$\begin{aligned} \Delta\tilde{F}_P &= \left\langle \tilde{K}_P(\phi)\,\Delta\tilde{T}_s(\phi)\right\rangle_\phi, \;\; \Delta\tilde{F}_L = \left\langle \left\{ \tilde{K}_L(\phi,p)\,\Delta\tilde{T}'_a(\phi,p)\right\}_p \right\rangle_\phi, \\ \Delta\tilde{F}_w &= \left\langle \left\{ \tilde{K}_w(\phi,p)\,\Delta\tilde{T}_a(\phi,p)\right\}_p \right\rangle_\phi, \;\; \Delta\tilde{F}_\alpha(\phi) = \left\langle \tilde{K}_\alpha(\phi)\,\Delta\tilde{\alpha}(\phi)\right\rangle_\phi, \end{aligned} \tag{S15}$$

where $\Delta\tilde{T}'_a(\phi,p) \equiv \Delta\tilde{T}_a(\phi,p) - \Delta\tilde{T}_s(\phi)$ is the departure of the atmospheric temperature change from the surface temperature change.

The feedback parameters $\tilde{\lambda}_i$ are computed from these approximate radiative responses as above in Eq. (5). The residual term, $\tilde{\lambda}_c+\tilde{\lambda}_{res}$, is computed as above using ΔN and ΔF_{GHG}, with the radiative responses F_i replaced with the annual-mean zonal-mean analysis values $\tilde{F}_i$. This leads to feedback parameter values $\tilde{\lambda}_i$ that match fairly closely with the feedback parameters λ_i that were computed using the full 4D structure of the simulated climate and kernel fields (Fig. S6).

Planck feedback. The outgoing radiation can be written according to the Stefan-Boltzmann law as $\epsilon\,\sigma\,\tilde{T}_s^4$, where ϵ is the emissivity associated with the atmosphere making the surface less efficient at emitting radiation to space. The kernel K_P describes the

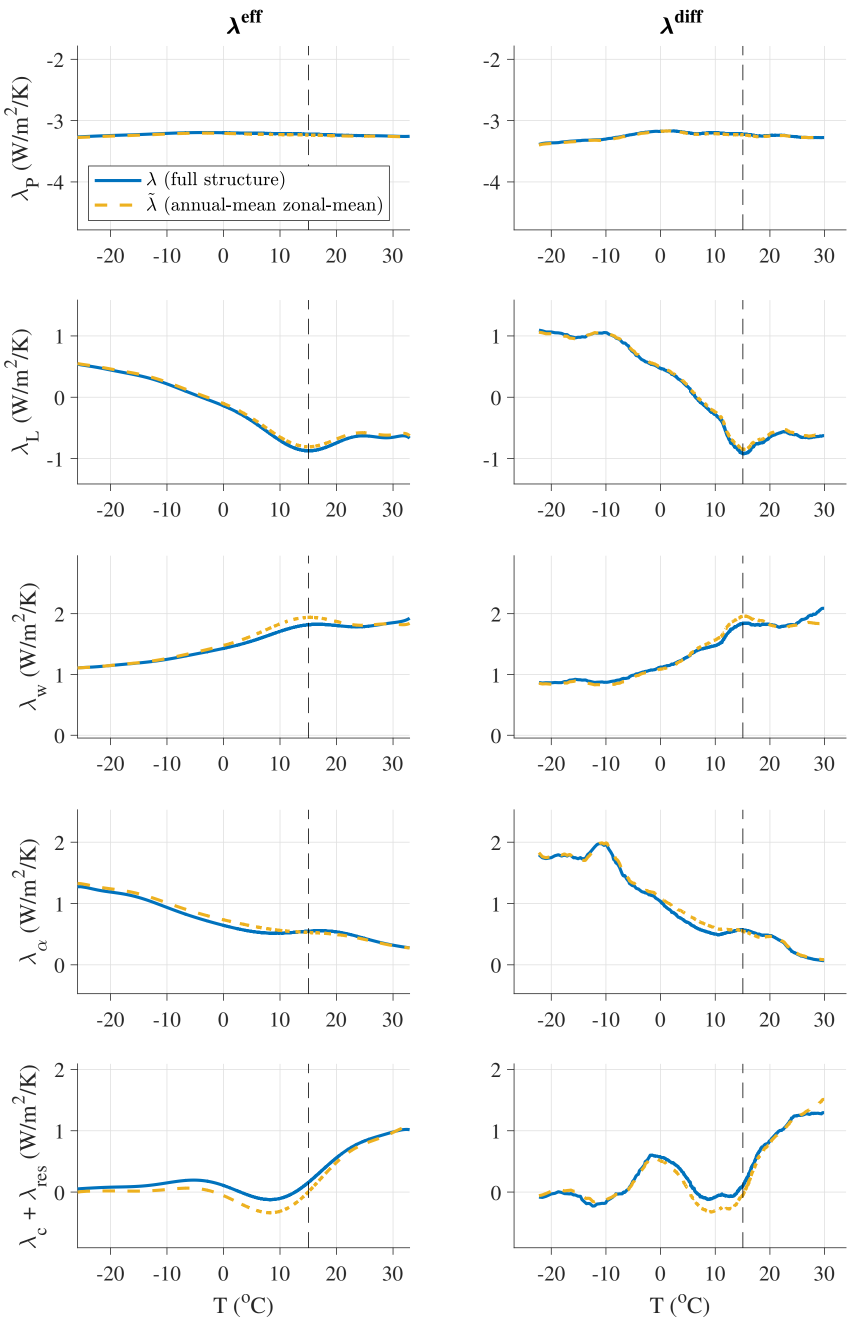

Figure S6. As in Fig. 3, but also including the results computed using annual-mean zonal-mean kernels and simulated climate fields (orange dashed lines).

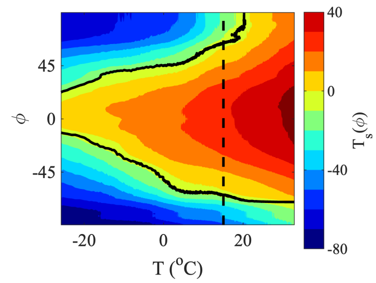

Figure S7. Meridional structure of simulated surface temperature changes. The black solid line indicates the 50% contour of the ice cover, and the black vertical dashed line indicates the PI climate.

change in incoming radiation per change in surface temperature, and hence the annual-mean zonal-mean kernel can be written as

$$\tilde{K}_P(\phi) = -4\,\epsilon(\phi)\,\sigma\,\left[\tilde{T}_{s,PI}(\phi)\right]^3, \tag{S16}$$

where $\tilde{T}_{s,PI}$ indicates the annual and zonal average of the surface temperature field $\tilde{T}_s$ in the reference climate. The emissivity ϵ varies in space due to factors including cloudiness, but the kernel can be fairly well approximated (not shown) using a uniform value of $\epsilon = 0.61$, which is based on matching the global-mean values of $\tilde{K}_P(\phi)$ and $\left[\tilde{T}_{s,PI}\right]^3$. Hence the kernel $\tilde{K}_P$ has a more-negative value in locations with a warmer surface temperature in the reference climate.

The differential Planck feedback parameter can be written as

$$\tilde{\lambda}_P^{\text{diff}} = \frac{\Delta\tilde{F}_P}{\Delta T} \approx \left\langle \tilde{K}_P(\phi)\,\frac{\Delta\tilde{T}_s(\phi)}{\Delta T} \right\rangle_\phi = \left\langle \tilde{K}_P(\phi) \right\rangle_\phi + \left\langle \tilde{K}_P(\phi)\,\frac{\Delta\tilde{T}_s(\phi) - \Delta T}{\Delta T} \right\rangle_\phi, \tag{S17}$$

which shows that the feedback is equal to the global-mean value of the kernel, plus a correction associated with locations where the temperature change departs from the global-mean temperature change. The evolution of the annual-mean zonal-mean surface temperature $\tilde{T}_s(\phi)$ is plotted in Fig. S7. The temperature departure term in Eq. (S17) $\frac{\Delta\tilde{T}_s(\phi)-\Delta T}{\Delta T}$, which is computed from $\tilde{T}_s(\phi)$ using the same TLS regression procedure as in the computation of the differential feedback parameters described in Sec. S3 above, is plotted in Fig. 4f. Note that the approximately equal sign in Eq. (S17) indicates that the TLS regression operation is being approximated as linear (OLS regression is linear whereas TLS regression is not).

Lapse-rate feedback. The lapse-rate feedback parameter can be written as

$$\tilde{\lambda}_L^{\text{diff}} = \frac{\Delta\tilde{F}_L}{\Delta T} \approx \left\langle \left\{ \tilde{K}_L(\phi,p)\,\frac{\Delta\tilde{T}_a'(\phi,p)}{\Delta T} \right\}_p \right\rangle_\phi. \tag{S18}$$

The quantity inside the meridional averaging operation is plotted in Fig. 4b. As in Eq. (S17), the ratio is computed using TLS regression, and the approximately equal sign indicates that this operation is being approximated as linear.

We repeat the analysis neglecting horizontal variations in the temperature profile (and kernel), in which case the parameter is approximated as

$$\tilde{\lambda}_L^{\text{diff}} \approx \left\{ \left\langle \tilde{K}_L(\phi,p) \right\rangle_\phi \left\langle \frac{\Delta\tilde{T}_a'(\phi,p)}{\Delta T} \right\rangle_\phi \right\}_p. \tag{S19}$$

Since the meridional average is performed before the vertical integration for the calculation in Eq. (S19), we carry out this meridional average on pressure levels rather than on model levels. The result is plotted in Fig. 4c (red line), which shows that changes in the globally-averaged temperature profile dominate the variations in the lapse-rate feedback parameter.

Albedo feedback. The radiative response associated with the albedo feedback can be written as

$$\tilde{F}_\alpha(\phi) = \underbrace{\left\langle \tilde{K}_\alpha(\phi) \right\rangle_\phi \langle \tilde{\alpha}(\phi) \rangle_\phi}_{\text{constant } \tilde{K}_\alpha} + \underbrace{\left\langle \left(\tilde{K}_\alpha(\phi) - \left\langle \tilde{K}_\alpha(\phi) \right\rangle_\phi \right) \tilde{\alpha}(\phi) \right\rangle_\phi}_{\text{effect of } \tilde{K}_\alpha \text{ variations}}. \tag{S20}$$

The first term on the right-hand side describes the influence of changes in the global-mean albedo alone, and the second term describes the effect of higher levels of incident solar radiation in low latitudes (as well as other factors that cause spatial variations in the kernel). The first term is scaled by the global-mean value of the kernel, $\left\langle \tilde{K}_\alpha(\phi) \right\rangle_\phi = -140$ W/m^2, whose magnitude is about 40% of the global-mean insolation, 340 W/m^2. If we neglect spatial variations in the kernel, the albedo feedback parameter can be approximated using the first term in Eq. (S20) alone as

$$\tilde{\lambda}_\alpha^{\text{diff}} \approx \left\langle \tilde{K}_\alpha(\phi) \right\rangle_\phi \frac{\Delta \langle \tilde{\alpha}(\phi) \rangle_\phi}{\Delta T}. \tag{S21}$$

This can be further simplified by using the relationship $\Delta \langle \tilde{\alpha}(\phi) \rangle_\phi \approx \delta_\alpha \, \Delta A_{ice}$. Here A_{ice} is the global ice area that is plotted in Figs. 1c and 4d, which includes sea ice, snow cover on land, and glacial ice, and is measured as a fraction of the globe; and $\delta_\alpha = 0.72$ is the surface albedo jump, which is determined here by regression between ice area A_{ice} and global-mean albedo $\langle \tilde{\alpha}(\phi) \rangle_\phi$. Inserting this into Eq. (S21) leads to

$$\tilde{\lambda}_\alpha^{\text{diff}} \approx \left\langle \tilde{K}_\alpha(\phi) \right\rangle_\phi \delta_\alpha \frac{\Delta A_{ice}}{\Delta T}. \tag{S22}$$

In this representation, the albedo feedback parameter is approximated as the sensitivity of the total ice area to global mean temperature, which is the slope $\frac{\Delta A_{ice}}{\Delta T}$ in Fig. 4d, scaled by a constant value. Fig. 4e shows that this approximation captures much of the variation in the albedo feedback parameter.

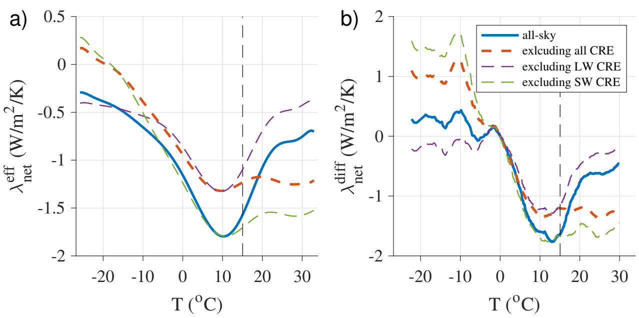

Figure S8. As in Fig. 2d,e, but including results computed using clear-sky fields. The net feedback parameter shown in Fig. 2d,e (shown here as blue curves) is calculated using Eq. (2) in the main text, which can be written as $\lambda_{net} \equiv \Delta F_{net} / \Delta T = \Delta \left(\texttt{FSNT} - \texttt{FLNT} - F_{GHG}\right) / \Delta T$, where `FSNT` and `FLNT` are the top-of-model longwave (LW) and shortwave (SW) fluxes reported by the atmospheric model. We exclude cloud radiative effects (CRE) by replacing `FSNT` and `FLNT` with clear-sky fields reported by the model (`FSNTC` and `FLNTC`), which is indicated by the red dashed lines. Next we exclude only LW or SW CRE by replacing only `FLNT` (magenta) or only `FSNT` (green) with clear-sky fields.