Climate sensitivity increases under higher CO_2 levels due to feedback temperature dependence

Article

Accepted Version

Climate sensitivity increases under higher CO_2 levels due to feedback temperature dependence

Jonah Bloch-Johnson[1], Maria Rugenstein[2,3], Martin B. Stolpe, Tim Rohrschneider[2], Yiyu Zheng[2], and Jonathan Gregory[1,4]

[1]NCAS, University of Reading, Reading, UK
[2]Max Planck Institute for Meteorology, Hamburg, Germany
[3]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA
[4]Met Office Hadley Centre, Exeter, UK

Key Points:

- Equilibrium warming per CO_2 doubling increases with CO_2 level for 13 of 14 climate models.
- Positive feedback temperature dependence explains most of the sensitivity increase.
- Nonlinear feedbacks increase the long-term risk of extreme warming under high CO_2 emissions.

Corresponding author: Jonah Bloch-Johnson, `j.bloch-johnson@reading.ac.uk`

Abstract

Equilibrium climate sensitivity - the equilibrium warming per CO_2 doubling - increases with CO_2 concentration for thirteen of fourteen coupled general circulation models for 0.5 to 8 times the preindustrial concentration. In particular, the abrupt4xCO_2 equilibrium warming is more than twice the 2xCO_2 warming. We identify three potential causes: nonlogarithmic forcing, feedback CO_2 dependence, and feedback temperature dependence. Feedback temperature dependence explains at least half of the sensitivity increase, while feedback CO_2 dependence explains a smaller share, and nonlogarithmic forcing decreases sensitivity in as many models as it increases it. Feedback temperature dependence is positive for ten out of fourteen models, primarily due to the longwave clear-sky feedback, while cloud feedbacks drive particularly large sensitivity increases. Feedback temperature dependence increases the risk of extreme or runaway warming, and is estimated to cause six models to warm at least an additional 3K under 8xCO_2.

Plain Language Summary

Increasing CO_2 reduces the rate at which energy leaves Earth, causing a net energy gain at its surface. The resulting warming increases the rate that energy leaves the planet. The planet stops warming once it regains balance. Studies usually assume that doubling atmospheric CO_2 always produces the same eventual global temperature rise (called the "equilibrium climate sensitivity"), whatever the starting CO_2 level. We show, on the contrary, that in nearly all the computer climate models we have examined, the extra warming for each doubling goes up as the CO_2 level increases. In most models, the warmer the climate becomes, the more it has to warm in order to balance a further CO_2 doubling, because warming becomes less effective at rebalancing the flow of energy. This effect increases projections of warming, especially for scenarios of greatest CO_2 increase.

1 Introduction

The *equilibrium climate sensitivity* (ΔT_{2x}) is the equilibrium global-mean surface warming per CO_2 doubling (Hansen et al., 1985; Stocker et al., 2013). ΔT_{2x} is often assumed to be constant (Stocker et al., 2013), allowing the equilibrium warming from different CO_2 increases to be characterized by a single metric, and for time series with various CO_2 changes to be used to estimate ΔT_{2x}. A constant ΔT_{2x} rests on two assumptions: each CO_2 doubling induces the same radiative forcing, and each unit of forcing induces the same equilibrium warming (i.e., that the net radiative feedback is constant). However, for low or high enough CO_2 concentrations, the net radiative feedback becomes positive, causing runaway glaciation (Hoffman et al., 1998) or a runaway greenhouse (Komabayasi, 1967; Ingersoll, 1969) respectively. Given these limits, will ΔT_{2x} remain constant across the range of CO_2 levels expected under future emissions scenarios?

Paleoclimatologists have investigated this question (Heydt et al., 2016; Farnsworth et al., 2019). Studies of the early Cenozoic find an increase in climate sensitivity with CO_2 concentration (Caballero & Huber, 2013; Anagnostou et al., 2016; Shaffer et al., 2016; Farnsworth et al., 2019; Zhu et al., 2019; Anagnostou et al., 2020), while studies of the Pleistocene disagree about whether sensitivity increases (three of four models in Crucifix, 2006; Yoshimori et al., 2009; Friedrich et al., 2016; Köhler et al., 2017; Snyder, 2019), stays the same (Martínez-Botí et al., 2015), or decreases (one of four models in Crucifix, 2006) with CO_2. However, different continental configurations may affect how sensitivity changes with CO_2 (Caballero & Huber, 2013; Wolf et al., 2018; Farnsworth et al., 2019).

While most studies of general circulation models under modern conditions have found that sensitivity increases with CO_2 (Hansen et al., 2005; Bitz et al., 2012; Block & Mauritsen, 2013; Caballero & Huber, 2013; Jonko et al., 2013; Meraner et al., 2013; Gregory et al., 2015; Rieger et al., 2017; Duan et al., 2019; Rohrschneider et al., 2019), some have found that it decreases (Stouffer & Manabe, 2003; Kutzbach et al., 2013) or remains roughly constant (Colman & McAvaney, 2009). However, these thirteen studies only evaluate ΔT_{2x} for models from five modelling centers. In most cases they use mixed-layer oceans, neglecting changes in ocean dynamics that can affect sensitivity (Kutzbach et al., 2013; Farnsworth et al., 2019).

Recently, two datasets have become available with coupled atmosphere-ocean general circulation model (AOGCM) simulations at multiple constant CO_2 levels initialized under preindustrial conditions (abrupt$n$xCO_2 simulations, where nxCO_2 refers to the increase relative to preindustrial CO_2 concentration): ten Coupled Model Intercomparison Project Phase 6 (CMIP6) models with abrupt0.5xCO_2 and abrupt2xCO_2 simulations run as part of NonLinMIP (Good et al., 2016) in addition to the standard abrupt4xCO_2 simulations (Eyring et al., 2016), and five models in the LongRunMIP archive (a collection of 1000+ year simulations of coupled AOGCMs; Rugenstein et al., 2019) with abrupt2xCO_2, abrupt4xCO_2, and abrupt8xCO_2 simulations. One model participated in both projects.

In this paper, we show that equilibrium climate sensitivity generally increases with CO_2 level (Section 2); that changes in radiative forcing are not large enough to explain this increase for most models (Section 3); and that the increase is instead caused by positive feedback temperature dependence, with some contribution from feedback CO_2 dependence (Section 4). We compare these three nonlinear terms and their causes (Section 5) and then summarize our findings (Section 6).

2 Equilibrium warming

Let T be the globally-averaged surface temperature, and $\Delta T \equiv T - T_{pi}$ be the warming relative to the preindustrial period. We define $\Delta T_{eq}(C)$ as the equilibrium warming caused by changing the CO_2 concentration from its preindustrial value ($pCO_{2,pi} \approx$ 280ppm) to a new value (pCO_2), where C is the number of CO_2 doublings relative to this preindustrial period,

$$C(pCO_2) \equiv \log_2\left(\frac{pCO_2}{pCO_{2,pi}}\right) \tag{1}$$

Under preindustrial conditions, $C_{pi} = 0$; in an abrupt2xCO_2 simulation, $C = 1$; and so forth. Table S1 is a glossary of all symbols used in this paper.

One condition for equilibrium is that the net top-of-atmosphere radiative flux N (downwards positive) is zero, on average. If we assume that N depends solely on C and T, then we can express a change in N in an abrupt$n$xCO_2 simulation as an initial change due to C and a subsequent change due to T:

$$\begin{aligned} N(C,T) - N(C_{pi},T_{pi}) &= (N(C,T_{pi}) - N(C_{pi},T_{pi})) + (N(C,T) - N(C,T_{pi})) && (2)\\ &= (N(C,T_{pi}) - N(C_{pi},T_{pi})) + \int_{T_{pi}}^{T_{pi}+\Delta T} \frac{\partial N(C,T)}{\partial T} dT && (3)\\ &= F(C_{pi},T_{pi},C) + \int_{T_{pi}}^{T_{pi}+\Delta T} \lambda(C,T) dT && (4) \end{aligned}$$

F is the *radiative forcing*, the change in N relative to a given initial condition (C_i, T_i) caused by C doublings of CO_2 while holding surface temperature fixed ($F(C_i, T_i, C) \equiv N(C_i + C, T_i) - N(C_i, T_i)$), and λ is the *radiative feedback*, the dependence of N on T ($\lambda(C,T) \equiv \partial N(C,T)/\partial T$), where the sign convention implies the feedback is typically negative. We can find $\Delta T_{eq}(C)$ by setting $N(C,T) = 0$:

$$F(C_{pi},T_{pi},C) = -\int_{T_{pi}}^{T_{pi}+\Delta T_{eq}(C)} \lambda(C,T) dT \tag{5}$$

where we assume $N(C_{pi}, T_{pi}) = 0$, since the preindustrial climate was roughly in equilibrium.

Under preindustrial concentrations, the spectral line shape of CO_2 absorption bands creates a logarithmic dependence of N on changes in pCO_2, so that the *forcing per CO_2 doubling* ($\tilde{F} \equiv \partial N/\partial C$) is often assumed to be constant (Myhre et al., 1998). Our definition of radiative forcing also includes adjustments of the atmosphere, land, and ocean to CO_2 changes that occur independently of subsequent changes in surface temperature (e.g., Sherwood et al., 2014; Kamae et al., 2015). This "effective radiative forcing" is also often assumed to be constant per CO_2 doubling (Forster et al., 2016), as is the radiative feedback (Hansen et al., 1985; Gregory et al., 2004). Substituting these constant terms into Eq. 5, we can solve for $\Delta T_{eq}(C)$:

$$\Delta T_{eq}(C) = -\frac{\tilde{F}}{\lambda} C \tag{6}$$

Assuming a constant $\tilde{F}$ and λ is equivalent to approximating $N(T, C)$ with the linear Taylor expansion of N around preindustrial values of C_{pi} and T_{pi} (i.e., $N(C, T) \approx \tilde{F}C + \lambda\Delta T$, where $C = \Delta C$ because $C_{pi} = 0$). The linear approximation of Eq. 6 is ubiquituous in climate science (e.g., Stocker et al., 2013; Knutti et al., 2017).

The linear approximation implies that the *equilibrium climate sensitivity* (ΔT_{2x}), the equilibrium warming per CO_2 doubling, is simply $-\tilde{F}/\lambda$, which, being a ratio of two constants, is itself a constant. It should therefore not matter how many CO_2 doublings are used to estimate it, since $\Delta T_{2x} = \Delta T_{eq}(C_1)/C_1 = \Delta T_{eq}(C_2)/C_2$. Fig. 1a shows instead that our estimates of $\Delta T_{eq}(C)/C$ increase with CO_2 concentration for thirteen of fourteen models. Colored bars show estimates made by extrapolating regressions of years 21 to 150 of N against ΔT to equilibrium ($N = 0$) for abrupt2^CxCO_2 simulations (Gregory et al., 2004, see also solid gray lines in Fig. S1). In these estimates, N and ΔT are anomalies: for LongRunMIP, we subtract the model's control simulation's mean value; for CMIP6, we subtract the linear fit of the control simulation after the branch point for the abrupt$n$xCO_2 simulations. We use only one ensemble member for each simulation.

Estimates of ΔT_{eq} typically increase with simulation length (Rugenstein et al., 2020; Dai et al., 2020; Dunne et al., 2020). While most CMIP6 simulations are only 150 years long, some are longer, and the LongRunMIP models are all at least 1000 years long. Black horizontal lines in Fig. 1a show estimates using years 101 to 750+ (see Table S2 for exact number of years). Here and in the following we use bootstrapping to estimate the 2.5^{th} to 97.5^{th} percentile range of uncertainty (gray and black vertical lines in Fig. 1; see Text S1). Black bars show multi-model mean values for the two experiments for which we have simulations of all models.

The sensitivity definition in Fig. 1a (i.e., $\Delta T_{2x}(C) \equiv \Delta T_{eq}(C)/C$) is often used to estimate ΔT_{2x} from abrupt4xCO_2 simulations, which our results suggest would lead to an average overestimate of at least 0.5K, even neglecting the outlier of FAMOUS. Equivalently, the nonlinearity of N leads to an average increase in equilibrium warming of at least 1K under 4xCO_2. Sherwood et al. (2020) suggested that using only the first 150 years to estimate ΔT_{eq} of an abrupt4xCO_2 simulation compensates for this overestimate. For our five models with 1000+ year abrupt2xCO_2 simulations, this compensation does not hold individually (CNRM-CM6-1's ΔT_{2x} would be 0.4K too small, FAMOUS's 1.8K too large), or on average (an 8% overestimate). If we define sensitivity instead as the equilibrium warming caused by successive CO_2 doublings ($\Delta T_{2x}(C) \equiv \Delta T_{eq}(C) - \Delta T_{eq}(C-1)$; Jonko et al., 2013), then changes in sensitivity are larger, with increases larger than 1K for seven models (Fig. S2). Alternatively, if we define sensitivity as the warming from doubling CO_2 relative to preindustrial conditions only ($\Delta T_{2x} \equiv \Delta T_{eq}(1)$; e.g., Knutti et al., 2017; Ceppi & Gregory, 2017), our results suggest that this metric may have a limited applicability.

The above shows that the equilibrium climate sensitivity is inconstant, and thus the linear approximation is inaccurate. To understand the increase in sensitivity, we take the quadratic Taylor expansion of N around (C_{pi}, T_{pi}):

$$N(C,T) \approx \frac{\partial N}{\partial C}\Big|_{\substack{C=C_{pi}\\T=T_{pi}}} C + \frac{\partial N}{\partial T}\Big|_{\substack{C=C_{pi}\\T=T_{pi}}} \Delta T + \frac{1}{2}\left(\frac{\partial^2 N}{\partial C^2}C^2 + \frac{\partial^2 N}{\partial T^2}(\Delta T)^2 + 2\frac{\partial^2 N}{\partial C \partial T} C\Delta T\right) \quad (7)$$

Substituting these new terms into Eq. 5, we have:

$$(\tilde{F}_{pi} + \frac{1}{2}\partial_C \tilde{F} C)C = -(\lambda_{pi} + \partial_C \lambda C + \frac{1}{2}\partial_T \lambda \Delta T_{eq})\Delta T_{eq} \quad (8)$$

where $\tilde{F}_{pi} \equiv \frac{\partial N}{\partial C}|_{C_{pi},T_{pi}}$ and $\lambda_{pi} \equiv \frac{\partial N}{\partial T}|_{C_{pi},T_{pi}}$ are the *preindustrial forcing per CO_2 doubling* and *preindustrial feedback* respectively, $\partial_C \tilde{F} \equiv \frac{\partial^2 N}{\partial C^2}$ is the CO_2 dependence of the forcing per doubling (which we call the *nonlinear forcing*), $\partial_C \lambda \equiv \frac{\partial^2 N}{\partial C \partial T}$ is the *feedback CO_2 dependence*, and $\partial_T \lambda \equiv \frac{\partial^2 N}{\partial T^2}$ is the *feedback temperature dependence*.

The three nonlinear terms ($\partial_C \tilde{F}$, $\partial_C \lambda$, and $\partial_T \lambda$) can all cause the equilibrium climate sensitivity to change with CO_2 concentration. Solving for $\Delta T_{eq}(C)$, we have

$$\Delta T_{eq}(C) = \frac{-(\lambda_{pi} + \partial_C \lambda C) - \sqrt{(\partial_C \lambda^2 - \partial_T \lambda \partial_C \tilde{F})C^2 + 2(\lambda_{pi}\partial_C \lambda - \tilde{F}_{pi}\partial_T \lambda)C + \lambda_{pi}^2}}{\partial_T \lambda} \quad (9)$$

We ignore the other quadratic solution, which gives an unstable equilibrium for C. In the following sections, we consider the impact of these terms on ΔT_{eq}.

3 Radiative forcing

Direct forcing depends linearly on C for small C (Myhre et al., 1998, who estimate $F(C) = 3.71C$ Wm^{-2}; dashed black line, Fig. 1b). At higher CO_2 levels, new absorption bands make this dependence superlinear (Byrne & Goldblatt, 2014; Etminan et al., 2016). Using the left side of Eq. 8, we have

$$F(C_{pi}, T_{pi}, C) = \tilde{F}_{pi} C + \frac{1}{2}\partial_C \tilde{F} C^2 \quad (10)$$

Byrne and Goldblatt (2014) used line-by-line radiative calculations and a simple stratospheric adjustment model to estimate $\tilde{F}_{pi} = 3.69$ Wm^{-2} and $\partial_C \tilde{F} = 0.375$ Wm^{-2} for $0.7xCO_2$ to $36xCO_2$, implying an increase in forcing per doubling with CO_2 concentration (gray bars in Fig. 1b).

We estimate forcing per doubling for each simulation (colored bars, Fig. 1b) by regressing the first ten years of N vs. ΔT to $\Delta T = 0$ (dashed black lines in Fig. S1; Gregory et al., 2004). This estimate includes adjustments as well as direct effects. Forcing per doubling decreases with C about as often as it increases, so that nonlinear forcing cannot explain the general increase in sensitivity. For CO_2 levels for which we have simulations for all models ($2xCO_2$ and $4xCO_2$), the multi-model mean forcing per doubling slightly decreases with C, although this decrease is not statistically significant.

Sensitivity increases with CO_2 concentration by a greater factor than forcing per doubling for most models (Fig. 1c). While all simulations but one have superlinear warming (i.e., are right of the vertical dashed line), nine simulations have sublinear forcing (i.e., are below the horizontal dashed line). Thirteen out of seventeen simulations have a smaller forcing increase than a warming increase (i.e. fall below the 1-to-1 line), as do the multi-model means. Moreover, there is little correlation between the nonlinear warming and forcing factors ($R^2 = 0.05$), even ignoring models with anomalous sensitivity increases (FAMOUS and CESM2; $R^2 = 0.14$). Forcing does not play a large role in the sensitivity increase for most models, although it may for individual models (e.g., CESM1.0.4).

Using twenty years instead of ten to estimate F reduces uncertainty (Fig. S3a) but biases estimates of F low, because of an increase in the slope of N vs. ΔT over time (Fig. S3b), and has little effect on our findings in Fig. 1c (see Fig. S3c). Sensitivity also increases by a greater factor than would be implied by Byrne and Goldblatt (2014) (Fig. S3d). Our findings are also the same if we first estimate $\tilde{F}_{pi}$ and $\partial_C \tilde{F}$ for each model by fitting the quadratic function in Eq. 10 (Figs. S4a and S4b): $\partial_C \tilde{F}$ is positive for only half of the models, with multi-model mean values of $\tilde{F}_{pi} = 4.01$ $\mathrm{Wm^{-2}}$ and $\partial_C \tilde{F} = 0.017$ $\mathrm{Wm^{-2}}$.

4 Radiative feedback

If sensitivity is not proportional to forcing, then Eq. 5 implies the feedback is inconstant. Inconstant feedbacks are commonly associated with the "pattern effect," in which the slope of N vs. ΔT under constant forcing varies. This slope is the weighted average of the spatial pattern of feedbacks, where the weights are given by the spatial pattern of surface warming, which evolves primarily due to the warming delay in regions of deep ocean heat uptake (e.g., Senior & Mitchell, 2000; Armour et al., 2013; Andrews et al., 2015; Rose et al., 2014; Rugenstein et al., 2016; Zhou et al., 2017; Dong et al., 2019; Bloch-Johnson et al., 2020).

The framework in Section 2 does not account for spatially-varying feedbacks, which make $N(C,T)$ an ill-defined function, in that it can have multiple values: the same globally-averaged T with warmer temperatures in regions with strong negative feedbacks implies a lower N than if the surface temperature was spatially uniform. It is more accurate to define $N(C,\vec{T})$, where $\Delta\vec{T}$ is the spatial temperature pattern (Haugstad et al., 2017). This means that the equilibrium response cannot generally be estimated from the slope of N vs. ΔT, which may evolve differently at different forcing levels simply because the patterns of warming associated with each simulation are different. For example, it is possible for the slope of N vs. ΔT to change due to a pattern effect, but for the overall response to forcing to be linear, so that the equilibrium climate sensitivity is constant (Rohrschneider et al., 2019).

To create a tractable framework, we assume that every globally-averaged surface temperature T is associated with a unique equilibrium pattern, $\vec{T}_{eq}(T)$, which is the pattern when T is in equilibrium (stable or unstable) for some C. We then substitute N with $N_{eq}(C,T) \equiv N(C,\vec{T}_{eq}(T))$ in our above definitions of λ and F. This substitution does not affect our forcing definition, as forcing is typically defined with respect to an equilibrated state, but ensures that any change in the feedback implies a change in the proportionality of $F(C)$ to $\Delta T_{eq}(C)$, and vice versa, as expected from Eq. 5. It also implies that the only way in which the pattern effect affects the equilibrium climate sensitivity is through changes in the equilibrium pattern of warming.

From Eq. 8, we have:

$$\lambda(C,T) = \lambda_{pi} + \partial_C \lambda C + \partial_T \lambda \Delta T \tag{11}$$

where $\lambda_{pi} \equiv \partial N/\partial T|_{pi}$ is the preindustrial feedback, $\partial_C \lambda \equiv \partial \lambda/\partial C = \partial^2 N/\partial C \partial T$ represents the feedback CO_2 dependence, and $\partial_T \lambda \equiv \partial \lambda/\partial T = \partial^2 N/\partial T^2$ represents the feedback temperature dependence (Roe & Armour, 2011; Bloch-Johnson et al., 2015).

Feedback CO_2 dependence quantifies the effect of additional atmospheric CO_2 on radiative feedbacks, such as damping the Planck feedback by making more frequencies optically thick (Seeley & Jeevanjee, 2020). It can also include effects due to forcing adjustments. The pattern effect prevents us from comparing the slope of N vs. ΔT across forcing levels to estimate $\partial_C \lambda$. Instead, we use additional experiments for five coupled AOGCMs, CESM1.2.2, CESM2*, CNRM-CM6-1*, HadGEM2, and HadGEM3-GC31-LL* (starred models are from our main analysis; see Table S3 and Text S2), to estimate

$\partial_C\lambda$. Since $\partial\lambda/\partial C \equiv \partial^2 N/\partial C\partial T = \partial\tilde{F}/\partial T$, feedback CO_2 dependence is also the dependence of the forcing per doubling on the reference temperature. We use pairs of experiments initialized at a colder temperature (T_{cold}) and a warmer temperature (T_{warm}) and the same initial CO_2 concentration C_i to estimate forcing from the same amount of CO_2 doubling C:

$$\partial_C\lambda = \partial_T\tilde{F} \approx \frac{1}{\Delta T}\frac{\Delta F(C_i, T_i, C)}{C} = \frac{F_{warm} - F_{cold}}{(T_{warm} - T_{cold})C} \tag{12}$$

where $F_{warm} \equiv F(C_i, T_{warm}, C)$ and $F_{cold} \equiv F(C_i, T_{cold}, C)$.

F_{cold} and F_{warm} can be estimated using pairs of abrupt simulations (i.e., an abrupt4xCO_2 simulation to estimate F_{cold}, and a simulation where CO_2 is abruptly lowered from 4xCO_2 to preindustrial values to estimate $-F_{warm}$) or from two pairs of fixed-SST experiments (Hansen et al., 2005) at two different temperatures and CO_2 concentrations. $\partial_C\lambda$ has a multi-model mean value of $\partial_C\lambda_{mean} = 0.0256$ $Wm^{-2}K^{-1}$ and a range of 0.0057 to 0.049 $Wm^{-2}K^{-1}$, suggesting that feedback CO_2 dependence is generally positive, increasing sensitivity with CO_2 concentration.

To estimate each model's feedback temperature dependence, we perform a least squares fit of Eq. 8 using estimates of $\tilde{F}_{pi}$ and $\partial_C\tilde{F}$ from the previous section, as well as model-specific estimates of $\partial_C\lambda$ when available, or otherwise $\partial_C\lambda_{mean}$. We perform this fit using pairs of C and ΔT_{eq} for each simulation, including the pair $C = 0$ and $\Delta T_{eq} = 0$ for the control simulation, giving estimates of λ_{pi} and $\partial_T\lambda$ (colored dots, Fig. 2). We find that ten of the fourteen models have positive feedback temperature dependence, with a multi-model mean value of $\partial_T\lambda_{mean} = 0.029$ $Wm^{-2}K^{-2}$ and a range of -0.14 to 0.109 $Wm^{-2}K^{-2}$.

With positive feedback temperature dependence, warming increases the feedback, leading to further warming, and so on. Under sufficient forcing, runaway warming occurs (Zaliapin & Ghil, 2010; Bloch-Johnson et al., 2015), specifically when Eq. 9 has no real solution ($\partial_T\lambda > (\lambda_{pi}+\partial_C\lambda C)^2/(\partial_C\tilde{F}C^2+2\tilde{F}_{pi}C)$), as shown by the light gray region for 8xCO_2 and dark gray region for 4xCO_2 (assuming that radiative forcing follows Byrne and Goldblatt (2014) and $\partial_C\lambda = \partial_C\lambda_{mean}$). FAMOUS falls in the latter region, and its abrupt4xCO_2 simulation does appear to lose its negative feedback (Fig. S1); four models lie in the 8xCO_2 runaway region. Climates in the gray regions do not actually warm infinitely, but simply warm sufficiently that the quadratic approximation breaks. Higher-order terms determine the temperature at which stability is regained, or if stability is lost in the first place. Models close to these runaway regions experience a sensitivity increase at the associated forcing level: the six models with black outlines experience an estimated increase of equilibrium warming under 8xCO_2 of at least 3K, given each model's forcing and $\partial_C\lambda$ estimates.

High estimated sensitivity ($\Delta T_{4x}/2 > 4.5K$) has been found in twenty CMIP6 models (Table S4). Of the six models with $\Delta T_{4x}/2 > 4.5K$ that appear in our study (i.e., models right of the dotted line in Fig. 2), four have $\Delta T_{2x} < 4.5K$ (i.e., are left of the dashed line). These models reconcile the moderate ΔT_{2x} implied by observations, paleoclimate, and processed-based analysis (Sherwood et al., 2020) with the sensitivity increases seen in paleoclimate studies of the warm Cenozoic (Caballero & Huber, 2013; Pierrehumbert, 2013; Anagnostou et al., 2016; Shaffer et al., 2016; Farnsworth et al., 2019).

To test the assumptions behind Fig. 2, we recalculate it with default values of $\partial_C\lambda =$ 0 and 0.05 $Wm^{-2}K^{-1}$ (Fig. S5a and S5b, respectively). This shifts the estimates of $\partial_T\lambda$ in the opposite direction as $\partial_C\lambda$, but also shifts the thresholds in the same manner, so that qualitatively the results are unchanged. Estimating forcing using years 1-20 instead of 1-10 has little effect (Fig. S5c), nor does using the direct estimate of $F(C)$ instead of $(\tilde{F}_{pi}+\frac{1}{2}\partial_C\tilde{F}C)C$ on the left side of Eq. 8 (Fig. S5d). Fig. S5e shows how $\partial_T\lambda$ evolves as more years are used to estimate the equilibrium warming. While more years do not greatly affect the results relative to each other, using years 101-1000 instead of 21-150 increases the magnitude of $\partial_T \lambda$ (excepting FAMOUS, which appears to be in a state of runaway). Since feedback temperature dependence should continue to affect the slope of N vs. ΔT beyond year 150 (Rugenstein et al., 2020), our estimates of CMIP6 models' $|\partial_T \lambda|$ and sensitivity changes may both be biased low.

5 Causes of sensitivity increases

Fig. 3a compares the contribution of the three nonlinear terms to each model's change in equilibrium climate sensitivity, $\Delta\Delta T_{2x} \equiv \Delta T_{4x}/2 - \Delta T_{2x}$. Using Eq. 9 to express equilibrium warming as a function of the quadratic approximation coefficients, $\Delta T_{eq}(C; \tilde{F}_{pi}, \lambda_{pi}, \partial_C \tilde{F}, \partial_C \lambda, \partial_T \lambda)$, we define:

$$\Delta\Delta T_{2x,\partial_C \tilde{F}} \equiv \Delta T_{eq}(2; \tilde{F}_{pi}, \lambda_{pi}, \partial_C \tilde{F}, 0, 0)/2 - \Delta T_{eq}(1; \tilde{F}_{pi}, \lambda_{pi}, \partial_C \tilde{F}, 0, 0) \quad (13)$$

$$\Delta\Delta T_{2x,\partial_C \lambda} \equiv \Delta T_{eq}(2; \tilde{F}_{pi}, \lambda_{pi}, 0, \partial_C \lambda, 0)/2 - \Delta T_{eq}(1; \tilde{F}_{pi}, \lambda_{pi}, 0, \partial_C \lambda, 0) \quad (14)$$

$$\Delta\Delta T_{2x,\partial_T \lambda} \equiv \Delta T_{eq}(2; \tilde{F}_{pi}, \lambda_{pi}, 0, 0, \partial_T \lambda)/2 - \Delta T_{eq}(1; \tilde{F}_{pi}, \lambda_{pi}, 0, 0, \partial_T \lambda) \quad (15)$$

Feedback temperature dependence is the dominant term for the three models with the largest sensitivity increases, accounts for 69% of the average increase, and contributes the largest term to the median increase (where FAMOUS is excluded from the averages, as the quadratic model suggests it experiences runaway warming under 4xCO_2). Feedback CO_2 dependence contributes a small, positive increase in sensitivity, while nonlinear forcing decreases sensitivity about as much and as often as it increases it.

To better understand these sensitivity increases, we estimate the flux components of the preindustrial feedback and feedback temperature dependence (Fig. 3b-d; see Fig. S6 for all components and uncertainties) by substituting individual top-of-atmosphere fluxes for N in the above derivations (see Text S3). We consider longwave vs. shortwave and noncloud vs. cloud components. For longwave fluxes, noncloud vs. cloud components are estimated using clear-sky fluxes and cloud radiative effect. For shortwave fluxes, to avoid cloud masking (Soden et al., 2004) we instead use approximate partial radiative perturbation (APRP; Taylor et al., 2007) for models with sufficient data available, including most CMIP6 models. For all other models we use clear-sky fluxes and cloud radiative effect as with the longwave.

The longwave noncloud feedback typically has positive temperature dependence (colored circles, Fig. 3b) due to an increasing water vapor feedback (Crucifix, 2006; Colman & McAvaney, 2009; Meraner et al., 2013). While some studies found that this increase is balanced by a strengthening negative lapse rate feedback (Boer et al., 2005; Colman & McAvaney, 2009; Yoshimori et al., 2009; Caballero & Huber, 2013), in recent studies the water vapor feedback dominates (Block & Mauritsen, 2013; Jonko et al., 2013; Meraner et al., 2013; Rieger et al., 2017), and Meraner et al. (2013) found a positive $\partial_T \lambda_{LW noncloud}$ for most CMIP5 models. Our findings contradict recent papers that find a constant longwave clear-sky feedback (Koll & Cronin, 2018; Zhang et al., 2020), though we agree that the value of the preindustrial feedback is likely close to -2 $Wm^{-2}K^{-1}$.

The shortwave noncloud feedback (colored circles, Fig. 3c) is the sum of a surface term (Fig. S6e) and an atmosphere term (Fig. S6f). The former represents a positive ice albedo feedback, which typically saturates, giving a negative temperature dependence (Colman & McAvaney, 2009; Block & Mauritsen, 2013; Jonko et al., 2013; Meraner et al., 2013; Rieger et al., 2017; Duan et al., 2019). The noncloud atmosphere term represents a positive water vapor feedback, which typically has a positive temperature dependence. Their sum has a positive preindustrial feedback with negligible temperature dependence (Fig. 3c). The SW noncloud outliers are models for which clear-sky fluxes were used instead of APRP (circles with black dots, Fig. 3c). Comparison of clear-sky vs. APRP estimates of the SW noncloud component suggests that cloud masking biases generally increases the uncertainty of the SW noncould component (Fig. S6c vs. S6g).

While the cloud feedback has multi-model mean values close to zero, it has more intermodel spread than the other two components (Fig. 3d) and has positive temperature dependence for most models. For CESM2, this occurs because its negative mixed-phase cloud feedback saturates (Tan et al., 2016; Frey & Kay, 2018; Bjordal et al., 2020). The spread in cloud feedback explains the range of nonlinearity in Fig. 3a. The average longwave noncloud feedback on its own (gray circle in Fig. 3b) would experience too little warming for its temperature dependence to matter (i.e., $\Delta\Delta T_{2x} = \Delta T_{4x}/2 - \Delta T_{2x} \approx 0.17$K assuming forcing from Byrne and Goldblatt (2014) and average $\partial_C \lambda$). Adding the shortwave noncloud feedback does not change the temperature dependence, but makes the preindustrial feedback more positive (gray triangle in Fig. 3b), causing more warming, increasing the nonlinearity (i.e., $\Delta\Delta T_{2x} \approx 0.33$K). Adding the average cloud feedback causes little change (gray square in Fig. 3b). For individual models, cloud feedbacks can move the climate into nonlinear regions, either by increasing the preindustrial feedback (CanESM5), or by increasing the feedback temperature dependence (CESM2 and FAMOUS). On the other hand, GISS-E2-2-G's cloud feedback temperature dependence is anomalously negative, and therefore it is the only model for which sensitivity decreases with CO_2 concentration.

We briefly discuss the flux components of the other two nonlinear terms (Fig. S7). The LW clear-sky term of the nonlinear forcing is negative for eleven of fourteen models (Fig. S7a). Since the direct LW clear-sky forcing depends superlinearly on CO_2 doubling (Byrne & Goldblatt, 2014), this negative term is due either to oversimplifications in the model's radiative scheme, or to adjustments. The other components vary in sign, with the largest source of intermodel spread coming from the cloud components. Since APRP accounts for cloud masking, the SW cloud spread must also be due to forcing adjustments. Adjustments thus play a first-order role in determining nonlinear forcing. The LW clear-sky component of feedback CO_2 dependence is positive for all five models (Fig. S7b), likely due to a blocked Planck feedback. SW cloud contributes the largest source of intermodel spread, so that forcing adjustments also play a first-order role in this nonlinearity.

6 Conclusions

Equilibrium climate sensitivity increases with CO_2 concentration for thirteen of fourteen models, contradicting the linear approximation of global energy balance, which assumes a constant forcing per CO_2 doubling and a constant radiative feedback. On average, climate models experience at least a degree of additional equilibrium warming under 4xCO_2 due to this sensitivity increase. Using a quadratic approximation allows us to capture the sensitivity increase using three second-order terms: nonlinear forcing, feedback CO_2 dependence, and feedback temperature dependence.

Feedback temperature dependence explains 69% of the sensitivity increase, and explains more of the median increase than any other term. Most importantly, it explains the particularly large increase seen in a handful of models, as positive feedback temperature dependence can cause runaway increases in sensitivity. Four models are predicted to experience runaway warming under CO_2 concentrations eight times larger than the preindustrial, and six models are projected to experience at least three additional degrees of equilibrium warming under this concentration. Feedback temperature dependence plays a key role in determining the risk of extreme warming in the coming centuries.

Ten of fourteen models have positive feedback temperature dependence, primarily due to the longwave clear-sky feedback. Models with large sensitivity increases have cloud feedbacks with either anomalously positive temperature dependence or anomalously positive preindustrial values. Feedback CO_2 dependence plays a smaller role, but results from five models suggests that it is likely positive, increasing sensitivity, primarily due

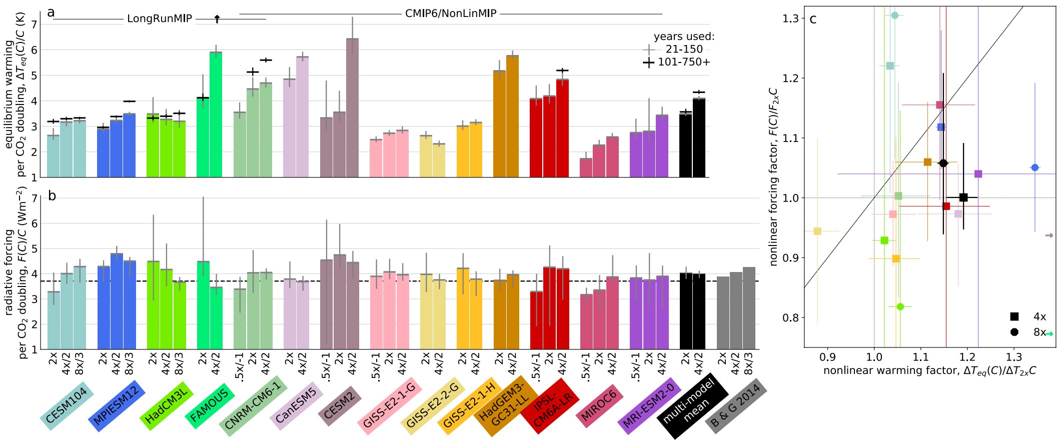

Figure 1. a. Equilibrium warming per CO_2 doubling ($\Delta T_{eq}(C)/C$) for abrupt-2^Cx simulations estimated using years 21 to 150 (colored bars and gray horizontal lines) and years 101 to n (where n is at least 750 years and given in Table S2; black horizontal lines). Vertical lines in panels a and b and all lines in panel c give the 2.5^{th} to 97.5^{th} percentile range of uncertainty (see Text S1). FAMOUS abrupt4xCO_2 is an outlier, with $\Delta T_{4x}/2 = 7.6$K when 1000 years are used.
b. Radiative forcing per CO_2 doubling ($F(C)/C$) for abrupt-2^Cx simulations estimated using years 1 to 10 (colored bars and gray horizontal lines. The dashed black line shows the Myhre et al. (1998) assumption of linear $F(C)$, while the gray bars give the analytic formula from Byrne and Goldblatt (2014).
c. Colored squares (octagons) show the factor by which equilibrium warming and forcing for an abrupt4xCO_2 (abrupt8xCO_2) simulation exceeds the linear extrapolation of its model's abrupt2xCO_2 values. Colors are the same as panels a and b. FAMOUS and CESM2 4x have nonlinear warming factors greater than 1.8.

to its longwave clear-sky component. The forcing per CO_2 doubling decreases with CO_2 concentration for as many models as it increases. Nonlinear forcing contributes less to the sensitivity increase than either other term, although it can be important for individual models. Forcing adjustments play a first-order role in determining the nonlinear forcing.

The substantial uncertainties in some of our findings could be greatly decreased with additional simulations. Longer simulations give better estimates of equilibrium warming (Rugenstein et al., 2020; Dai et al., 2020; Dunne et al., 2020); fixed-SST experiments give better radiative forcing estimates (Forster et al., 2016; Pincus et al., 2016); and simulations at multiple CO_2 levels allow for an assessment of nonlinearities (Good et al., 2016). Simulations that behave in surprising or anomalous ways may be exhibiting nonlinear dynamics, and should not be neglected (Valdes, 2011). Even if a loss of stability causes models to warm outside the range for which they were calibrated, the increase in sensitivity may still be physical. Exploring and documenting the nonlinear frontiers of warming in climate models is essential to assessing the risk of extreme warming for the real world.

Acknowledgments

We thank Tim Andrews for making the HadGEM3-GC3.1-LL abrupt-2xCO2 simulation available at `https://github.com/timothyandrews/abrupt-2xCO2`. CMIP6 data is at `https://pcmdi.llnl.gov/CMIP6/`. LongRunMIP data access is at `http://www.longrunmip`

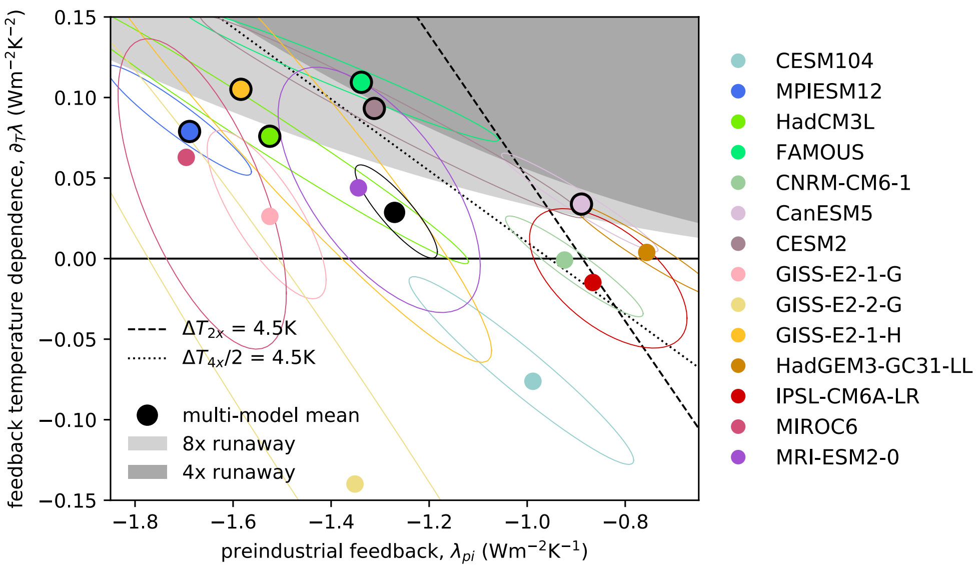

Figure 2. Preindustrial feedback vs. feedback temperature dependence (colored dots; colored ellipsoids give the 75^{th} percentile of uncertainty). Values in the dark (light) gray region imply runaway warming under $4xCO_2$ ($8xCO_2$) and values above the dashed (dotted) black line have a sensitivity estimated from abrupt$2xCO_2$ (abrupt$4xCO_2$) above 4.5K. All thresholds are calculated assuming forcing from Byrne and Goldblatt (2014) and model-mean feedback CO_2 dependence. Colored dots with black outlines experience an additional 3K of equilibrium warming under $8xCO_2$ given our estimate of that model's forcing and $\partial_C\lambda$.

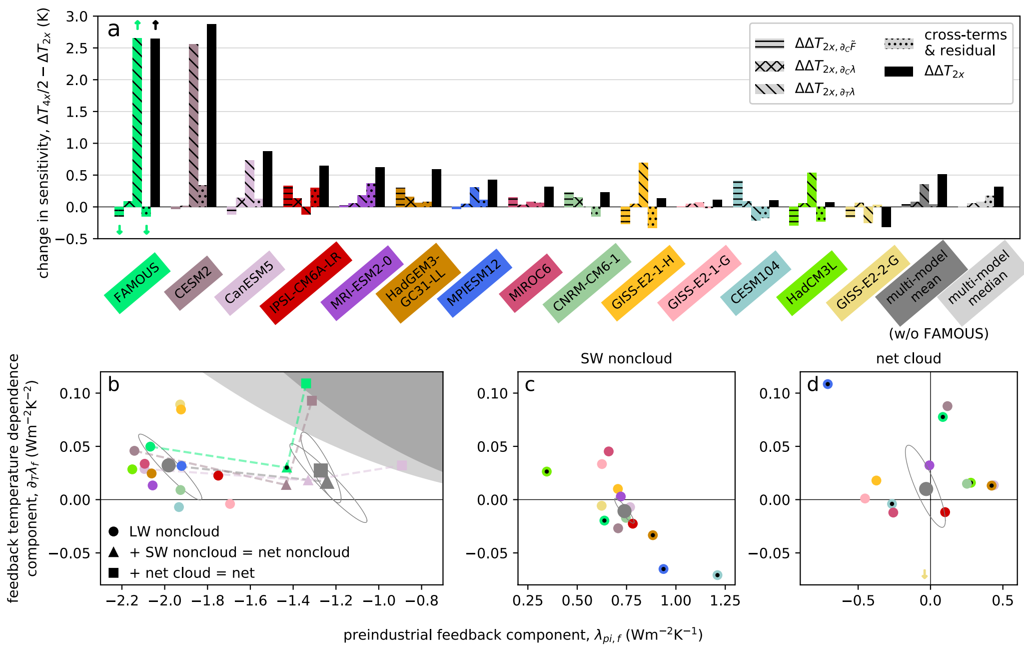

Figure 3. a. Contributions to the change in sensitivity from $2xCO_2$ to $4xCO_2$ (black bars) from nonlinear forcing ($\partial_C \tilde{F}$, horizontally-hatched bars), feedback CO_2 dependence ($\partial_C \lambda$, crossed-hatched bars), and feedback temperature dependence ($\partial_T \lambda$, diagonally-hatched bars). Dotted bars represent cross-terms, higher-order nonlinearities, and errors in our estimates. FAMOUS is not included in the mean and median as the quadratic model suggests it is in a state of runaway under $4xCO_2$.
b., c., and d. Colored circles give estimates of the longwave noncloud, shortwave noncloud, and net cloud components respectively of the preindustrial feedback and feedback temperature dependence. Models with dotted circles use clear-sky fluxes instead of approximate partial radiative perturbation to partition the shortwave flux into noncloud and cloud components. Colors are given by the model names in panel a. Gray circles give the multi-model mean and gray ellipsoids give the estimated 75^{th} percentile of uncertainty. The shaded regions in panel b are as in Fig. 2. Triangles in panel b show the result of adding the shortwave noncloud component to longwave noncloud components. Squares show the result for additionally adding the net cloud component.

.org/. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 786427, project "Couplet" and grant agreement No. 820829, project "CONSTRAIN").

References

Anagnostou, E., John, E. H., Babila, T. L., Sexton, P. F., Ridgwell, A., Lunt, D. J., ... Foster, G. L. (2020, September). Proxy evidence for state-dependence of climate sensitivity in the Eocene greenhouse. *Nature Communications*, *11*(1), 4436. Retrieved 2020-09-11, from `https://www.nature.com/articles/s41467-020-17887-x` (Number: 1 Publisher: Nature Publishing Group) 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 CO 2 concentration was the primary driver of early Cenozoic climate. *Nature*, *533*(7603), 380–384. Retrieved 2019-12-21, from `https://www.nature.com/articles/nature17423` doi: 10.1038/nature17423

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models. *Journal of Climate*, *28*(4), 1630–1648. Retrieved 2019-11-27, from `https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-14-00545.1` doi: 10.1175/JCLI-D-14-00545.1

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-Varying Climate Sensitivity from Regional Feedbacks. *Journal of Climate*, *26*(13), 4518–4534. Retrieved 2018-05-12, from `https://journals.ametsoc.org/doi/10.1175/JCLI-D-12-00544.1` doi: 10.1175/JCLI-D-12-00544.1

Bitz, C. M., Shell, K. M., Gent, P. R., Bailey, D. A., Danabasoglu, G., Armour, K. C., ... Kiehl, J. T. (2012). Climate Sensitivity of the Community Climate System Model, Version 4. *Journal of Climate*, *25*(9), 3053–3070. Retrieved 2019-11-27, from `https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-11-00290.1` doi: 10.1175/JCLI-D-11-00290.1

Bjordal, J., Storelvmo, T., Alterskjr, K., & Carlsen, T. (2020, November). Equilibrium climate sensitivity above 5 C plausible due to state-dependent cloud feedback. *Nature Geoscience*, *13*(11), 718–721. Retrieved 2020-11-20, from `https://www.nature.com/articles/s41561-020-00649-1` (Number: 11 Publisher: Nature Publishing Group) doi: 10.1038/s41561-020-00649-1

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, *42*(12), 4973–4980. Retrieved 2019-11-27, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL064240` doi: 10.1002/2015GL064240

Bloch-Johnson, J., Rugenstein, M., & Abbot, D. S. (2020). Spatial Radiative Feedbacks from Internal Variability Using Multiple Regression. *Journal of Climate*, *33*(10), 4121–4140. Retrieved 2020-04-22, from `https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-19-0396.1` (Publisher: American Meteorological Society) doi: 10.1175/JCLI-D-19-0396.1

Block, K., & Mauritsen, T. (2013). Forcing and feedback in the MPI-ESM-LR coupled model under abruptly quadrupled CO2. *Journal of Advances in Modeling Earth Systems*, *5*(4), 676–691. Retrieved 2019-11-27, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jame.20041` doi: 10.1002/jame.20041

Boer, G. J., Hamilton, K., & Zhu, W. (2005). Climate sensitivity and climate change under strong forcing. *Climate Dynamics*, *24*(7), 685–700. Retrieved 2018-10-31, from `https://doi.org/10.1007/s00382-004-0500-3` doi: 10.1007/s00382-004-0500-3

Byrne, B., & Goldblatt, C. (2014). Radiative forcing at high concentrations of well-mixed greenhouse gases. *Geophysical Research Letters*, *41*(1), 152–160. Retrieved 2019-11-27, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058456` 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 2018-06-28, from http://www.pnas.org/content/110/35/14162 doi: 10.1073/pnas.1303365110

Ceppi, P., & Gregory, J. M. (2017, December). Relationship of tropospheric stability to climate sensitivity and Earths observed radiation budget. *Proceedings of the National Academy of Sciences*, *114*(50), 13126–13131. Retrieved 2018-06-28, from http://www.pnas.org/content/114/50/13126 doi: 10.1073/pnas.1714308114

Colman, R., & McAvaney, B. (2009). Climate feedbacks under a very broad range of forcing. *Geophysical Research Letters*, *36*(1). Retrieved 2019-11-27, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GL036268 doi: 10.1029/2008GL036268

Crucifix, M. (2006). Does the Last Glacial Maximum constrain climate sensitivity? *Geophysical Research Letters*, *33*(18). Retrieved 2019-12-20, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL027137 doi: 10.1029/2006GL027137

Dai, A., Huang, D., Rose, B. E. J., Zhu, J., & Tian, X. (2020, April). Improved methods for estimating equilibrium climate sensitivity from transient warming simulations. *Climate Dynamics*. Retrieved 2020-04-23, from https://doi.org/10.1007/s00382-020-05242-1 doi: 10.1007/s00382-020-05242-1

Dong, Y., Proistosescu, C., Armour, K. C., Battisti, D. S., Dong, Y., Proistosescu, C., … Battisti, D. S. (2019). Attributing Historical and Future Evolution of Radiative Feedbacks to Regional Warming Patterns using a Greens Function Approach: The Preeminence of the Western Pacific. *Journal of Climate*. Retrieved 2019-09-28, from https://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-18-0843.1 doi: 10.1175/JCLI-D-18-0843.1

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 2019-12-21, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029093 doi: 10.1029/2018JD029093

Dunne, J. P., Winton, M., Bacmeister, J., Danabasoglu, G., Gettelman, A., Golaz, J.-C., … Wolfe, J. D. (2020). Comparison of equilibrium climate sensitivity estimates from slab ocean, 150-year, and longer simulations. *Geophysical Research Letters*, *n/a*(n/a), e2020GL088852. Retrieved 2020-07-30, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL088852 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL088852) doi: 10.1029/2020GL088852

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*(1), 1–26. Retrieved 2020-04-23, from https://www.jstor.org/stable/2958830 (Publisher: Institute of Mathematical Statistics)

Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, *43*(24), 12,614–12,623. Retrieved 2019-11-27, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL071930 doi: 10.1002/2016GL071930

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved 2020-01-10, from https://www.geosci-model-dev.net/9/1937/2016/gmd-9-1937-2016.html doi: https://doi.org/10.5194/gmd-9-1937-2016

Farnsworth, A., Lunt, D. J., O'Brien, C. L., Foster, G. L., Inglis, G. N., Markwick, P., … Robinson, S. A. (2019). Climate Sensitivity on Geological Timescales Controlled by Nonlinear Feedbacks and Ocean Circulation. *Geophysical Research Letters*, *46*(16), 9880–9889. Retrieved 2019-11-27, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083574 doi: 10.1029/2019GL083574

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., … Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *Journal of Geophysical Research: Atmospheres*, *121*(20), 12,460–12,475. Retrieved 2019-12-24, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025320 doi: 10.1002/2016JD025320

Frey, W. R., & Kay, J. E. (2018, April). The influence of extratropical cloud phase and amount feedbacks on climate sensitivity. *Climate Dynamics*, *50*(7), 3097–3116. Retrieved 2020-11-23, from https://doi.org/10.1007/s00382-017-3796-5 doi: 10.1007/s00382-017-3796-5

Friedrich, T., Timmermann, A., Tigchelaar, M., Timm, O. E., & Ganopolski, A. (2016, November). Nonlinear climate sensitivity and its implications for future greenhouse warming. *Science Advances*, *2*(11), e1501923. Retrieved 2019-12-20, from https://advances.sciencemag.org/content/2/11/e1501923 doi: 10.1126/sciadv.1501923

Good, P., Andrews, T., Chadwick, R., Dufresne, J.-L., Gregory, J. M., Lowe, J. A., … Shiogama, H. (2016). nonlinMIP contribution to CMIP6: model intercomparison project for non-linear mechanisms: physical basis, experimental design and analysis principles (v1.0). *Geoscientific Model Development*, *9*(11), 4019–4028. Retrieved 2019-12-19, from https://www.geosci-model-dev.net/9/4019/2016/gmd-9-4019-2016-discussion.html doi: https://doi.org/10.5194/gmd-9-4019-2016

Gregory, J. M., Andrews, T., & Good, P. (2015, November). The inconstancy of the transient climate response parameter under increasing CO2. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2054), 20140417. Retrieved 2020-02-27, from https://royalsocietypublishing.org/doi/full/10.1098/rsta.2014.0417 doi: 10.1098/rsta.2014.0417

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., … Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3), L03205. Retrieved 2018-01-19, from http://onlinelibrary.wiley.com/doi/10.1029/2003GL018747/abstract doi: 10.1029/2003GL018747

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., & Stone, P. (1985). Climate Response Times: Dependence on Climate Sensitivity and Ocean Mixing. *Science*, *229*(4716), 857–859. Retrieved 2018-05-15, from http://science.sciencemag.org/content/229/4716/857 doi: 10.1126/science.229.4716.857

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., … Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, *110*(D18). Retrieved 2019-11-27, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD005776 doi: 10.1029/2005JD005776

Haugstad, A. D., Armour, K. C., Battisti, D. S., & Rose, B. E. J. (2017). Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, *44*(14), 7455–7463. Retrieved 2020-04-22, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074372 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL074372) doi: 10.1002/2017GL074372

Heydt, A. S. v. d., Dijkstra, H. A., Wal, R. S. W. v. d., Caballero, R., Crucifix, M., Foster, G. L., … Ziegler, M. (2016). Lessons on Climate Sensitivity From Past Climate Changes. *Current Climate Change Reports*, *2*(4), 148–158. Retrieved 2019-12-19, from https://link.springer.com/article/10.1007/s40641-016-0049-3 doi: 10.1007/s40641-016-0049-3

Hoffman, P. F., Kaufman, A. J., Halverson, G. P., & Schrag, D. P. (1998). A Neoproterozoic Snowball Earth. *Science*, *281*(5381), 1342–1346. Retrieved 2019-12-19, from https://science.sciencemag.org/content/281/5381/1342 doi: 10.1126/science.281.5381.1342

Ingersoll, A. P. (1969). The Runaway Greenhouse: A History of Water on Venus. *Journal of the Atmospheric Sciences*, *26*(6), 1191–1198. Retrieved 2019-12-19, from https://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281969%29026%3C1191%3ATRGAHO%3E2.0.CO%3B2 doi: 10.1175/1520-0469(1969)026⟨1191:TRGAHO⟩2.0.CO;2

Jonko, A. K., Shell, K. M., Sanderson, B. M., & Danabasoglu, G. (2013). Climate Feedbacks in CCSM3 under Changing CO2 Forcing. Part II: Variation of Climate Feedbacks and Sensitivity with Forcing. *Journal of Climate*, *26*(9), 2784–2795. Retrieved 2019-11-27, from https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-12-00479.1 doi: 10.1175/JCLI-D-12-00479.1

Kamae, Y., Watanabe, M., Ogura, T., Yoshimori, M., & Shiogama, H. (2015). Rapid Adjustments of Cloud and Hydrological Cycle to Increasing CO<Subscript>2</Subscript>: a Review. *Current Climate Change Reports*, *1*(2), 103–113. Retrieved 2018-08-01, from https://link.springer.com/article/10.1007/s40641-015-0007-5 doi: 10.1007/s40641-015-0007-5

Knutti, R., Rugenstein, M. A. A., & Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, *10*(10), 727–736. Retrieved 2018-11-01, from https://www.nature.com/articles/ngeo3017 doi: 10.1038/ngeo3017

Köhler, 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*, 1102–1114. Retrieved 2018-06-28, from http://doi.org/10.1002/2017PA003190 doi: 10.1002/2017PA003190

Koll, D. D. B., & Cronin, T. W. (2018, October). Earths outgoing longwave radiation linear due to H2O greenhouse effect. *Proceedings of the National Academy of Sciences*, *115*(41), 10293–10298. Retrieved 2019-09-30, from https://www.pnas.org/content/115/41/10293 doi: 10.1073/pnas.1809868115

Komabayasi, M. (1967). Discrete Equilibrium Temperatures of a Hypothetical Planet with the Atmosphere and the Hydrosphere of One Component-Two Phase System under Constant Solar Radiation. *Journal of the Meteorological Society of Japan. Ser. II*, *45*(1), 137–139. doi: 10.2151/jmsj1965.45.1_137

Kutzbach, J. E., He, F., Vavrus, S. J., & Ruddiman, W. F. (2013). The dependence of equilibrium climate sensitivity on climate state: Applications to studies of climates colder than present. *Geophysical Research Letters*, *40*(14), 3721–3726. Retrieved 2019-12-19, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50724 doi: 10.1002/grl.50724

Martínez-Botí, M. A., Foster, G. L., Chalk, T. B., Rohling, E. J., Sexton, P. F., Lunt, D. J., … Schmidt, D. N. (2015). Plio-Pleistocene climate sensitivity evaluated using high-resolution CO 2 records. *Nature*, *518*(7537), 49–54. Retrieved 2019-12-20, from https://www.nature.com/articles/nature14145 doi: 10.1038/nature14145

Meraner, K., Mauritsen, T., & Voigt, A. (2013). Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, *40*(22), 5944–5948. Retrieved 2019-11-27, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058118 doi: 10.1002/2013GL058118

Myhre, G., Highwood, E. J., Shine, K. P., & Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophysical Research Letters*, *25*(14), 2715–2718. Retrieved 2020-01-10, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98GL01908 doi: 10.1029/98GL01908

Pierrehumbert, R. T. (2013, August). Hot climates, high sensitivity. *Proceedings of the National Academy of Sciences*, *110*(35), 14118–14119. Retrieved 2018-10-31, from http://www.pnas.org/content/110/35/14118 doi: 10.1073/pnas.1313417110

Pincus, R., Forster, P. M., & Stevens, B. (2016, September). The Radiative Forcing Model Intercomparison Project (RFMIP): experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(9), 3447–3460. Retrieved 2020-04-23, from https://www.geosci-model-dev.net/9/3447/2016/ (Publisher: Copernicus GmbH) doi: https://doi.org/10.5194/gmd-9-3447-2016

Rieger, V. S., Dietmller, S., & Ponater, M. (2017). Can feedback analysis be used to uncover the physical origin of climate sensitivity and efficacy differences? *Climate Dynamics*, *49*(7), 2831–2844. Retrieved 2019-11-27, from https://doi.org/10.1007/s00382-016-3476-x doi: 10.1007/s00382-016-3476-x

Roe, G. H., & Armour, K. C. (2011, July). How sensitive is climate sensitivity? *Geophysical Research Letters*, *38*(14). Retrieved 2018-10-31, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GL047913 doi: 10.1029/2011GL047913

Rohrschneider, T., Stevens, B., & Mauritsen, T. (2019). On simple representations of the climate response to external radiative forcing. *Climate Dynamics*, *53*(5), 3131–3145. Retrieved 2019-11-27, from https://doi.org/10.1007/s00382-019-04686-4 doi: 10.1007/s00382-019-04686-4

Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N., & Koll, D. D. B. (2014). The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, *41*(3), 1071–1078. Retrieved 2018-06-28, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058955 doi: 10.1002/2013GL058955

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., … Yang, S. (2019). LongRunMIP motivation and design for a large collection of millennial-length AO-GCM simulations. *Bulletin of the American Meteorological Society*. Retrieved 2019-11-27, from https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-19-0068.1 doi: 10.1175/BAMS-D-19-0068.1

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., … Knutti, R. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, *n/a*(n/a). Retrieved 2019-11-27, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083898 doi: 10.1029/2019GL083898

Rugenstein, M., Gregory, J. M., Schaller, N., Sedlek, J., & Knutti, R. (2016). Multiannual OceanAtmosphere Adjustments to Radiative Forcing. *Journal of Climate*, *29*(15), 5643–5659. Retrieved 2018-08-01, from https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-16-0312.1 doi: 10.1175/JCLI-D-16-0312.1

Seeley, J., & Jeevanjee, N. (2020, August). *H2O windows and CO2 radiator fins: a clear-sky explanation for the peak in ECS* [preprint]. Retrieved 2020-08-28, from http://www.essoar.org/doi/10.1002/essoar.10503539.1 (Archive Location: world Publisher: Earth and Space Science Open Archive Section: Atmospheric Sciences) doi: 10.1002/essoar.10503539.1

Senior, C. A., & Mitchell, J. F. B. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, *27*(17), 2685–2688. Retrieved 2018-06-28, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL011373 doi: 10.1029/2000GL011373

Shaffer, G., Huber, M., Rondanelli, R., & Pedersen, J. O. P. (2016). Deep time evidence for climate sensitivity increase with warming. *Geophysical Research Letters*, *43*(12), 6538–6545. Retrieved 2018-05-15, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL069243 doi: 10.1002/2016GL069243

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., & Stevens, B. (2014). Adjustments in the Forcing-Feedback Framework for Understanding Climate Change. *Bulletin of the American Meteorological Society*, *96*(2), 217–228. Retrieved 2018-08-01, from https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00167.1 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. *Reviews of Geophysics*, *n/a*(n/a), e2019RG000678. Retrieved 2020-09-15, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019RG000678 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019RG000678) doi: 10.1029/2019RG000678

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjr, K., Collins, W., Sima, A., ... Forster, P. M. (2020, January). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chemistry and Physics Discussions*, 1–37. Retrieved 2020-04-23, from https://www.atmos-chem-phys-discuss.net/acp-2019-1212/ (Publisher: Copernicus GmbH) doi: https://doi.org/10.5194/acp-2019-1212

Snyder, C. W. (2019, September). Revised estimates of paleoclimate sensitivity over the past 800,000years. *Climatic Change*, *156*(1), 121–138. Retrieved 2020-04-03, from https://doi.org/10.1007/s10584-019-02536-0 doi: 10.1007/s10584-019-02536-0

Soden, B. J., Broccoli, A. J., & Hemler, R. S. (2004, October). On the Use of Cloud Forcing to Estimate Cloud Feedback. *Journal of Climate*, *17*(19), 3661–3665. Retrieved 2018-06-28, from https://journals.ametsoc.org/doi/full/10.1175/1520-0442%282004%29017%3C3661:OTUOCF%3E2.0.CO%3B2 doi: 10.1175/1520-0442(2004)017⟨3661:OTUOCF⟩2.0.CO;2

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., ... Midgley, P. (2013). *AR5 Climate Change 2013: The Physical Science Basis IPCC* (Tech. Rep.). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Retrieved 2019-12-19, from https://www.ipcc.ch/report/ar5/wg1/

Stolpe, M. B., Medhaug, I., Beyerle, U., & Knutti, R. (2019). Weak dependence of future global mean warming on the background climate state. *Climate Dynamics*, *53*(7), 5079–5099. Retrieved 2019-11-27, from https://doi.org/10.1007/s00382-019-04849-3 doi: 10.1007/s00382-019-04849-3

Stouffer, R. J., & Manabe, S. (2003). Equilibrium response of thermohaline circulation to large changes in atmospheric CO2 concentration. *Climate Dynamics*, *20*(7), 759–773. Retrieved 2019-11-27, from https://doi.org/10.1007/s00382-002-0302-4 doi: 10.1007/s00382-002-0302-4

Tan, I., Storelvmo, T., & Zelinka, M. D. (2016, April). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, *352*(6282), 224–227. Retrieved 2020-04-24, from https://science.sciencemag.org/content/352/6282/224 (Publisher: American Association for the Advancement of Science Section: Report) doi: 10.1126/science.aad5300

Taylor, K. E., Crucifix, M., Braconnot, P., Hewitt, C. D., Doutriaux, C., Broccoli, A. J., ... Webb, M. J. (2007, June). Estimating Shortwave Radiative Forcing and Response in Climate Models. *Journal of Climate*, *20*(11), 2530–2543. Retrieved 2020-03-08, from https://journals.ametsoc.org/doi/

10.1175/JCLI4143.1 (Publisher: American Meteorological Society) doi: 10.1175/JCLI4143.1

Valdes, P. (2011). Built for stability. *Nature Geoscience*, *4*(7), 414–416. Retrieved 2020-01-06, from https://www.nature.com/articles/ngeo1200 doi: 10.1038/ngeo1200

Winton, M., Adcroft, A., Dunne, J. P., Held, I. M., Shevliakova, E., Zhao, M., ... Zhang, R. (2020). Climate Sensitivity of GFDL's CM4.0. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001838. Retrieved 2020-04-22, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001838 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001838) doi: 10.1029/2019MS001838

Wolf, E. T., HaqqMisra, J., & Toon, O. B. (2018). Evaluating Climate Sensitivity to CO2 Across Earth's History. *Journal of Geophysical Research: Atmospheres*, *123*(21), 11,861–11,874. Retrieved 2020-04-21, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029262 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018JD029262) doi: 10.1029/2018JD029262

Yoshimori, M., Yokohata, T., & Abe-Ouchi, A. (2009). A Comparison of Climate Feedback Strength between CO2 Doubling and LGM Experiments. *Journal of Climate*, *22*(12), 3374–3395. Retrieved 2019-09-28, from https://journals.ametsoc.org/doi/full/10.1175/2009JCLI2801.1 doi: 10.1175/2009JCLI2801.1

Zaliapin, I., & Ghil, M. (2010, March). Another Look at Climate Sensitivity. *Nonlinear Processes in Geophysics*, *17*(2), 113–122. Retrieved 2018-10-31, from http://arxiv.org/abs/1003.0253 (arXiv: 1003.0253) doi: 10.5194/npg-17-113-2010

Zhang, Y., Jeevanjee, N., & Fueglistaler, S. (2020). Linearity of Outgoing Longwave Radiation: From an Atmospheric Column to Global Climate Models. *Geophysical Research Letters*, *47*(17), e2020GL089235. Retrieved 2020-11-20, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL089235 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL089235) doi: https://doi.org/10.1029/2020GL089235

Zhou, C., Zelinka, M. D., & Klein, S. A. (2017). Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach. *Journal of Advances in Modeling Earth Systems*, *9*(5), 2174–2189. Retrieved 2018-05-12, from https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017MS001096 doi: 10.1002/2017MS001096

Zhu, J., Poulsen, C. J., & Tierney, J. E. (2019, September). Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks. *Science Advances*, *5*(9), eaax1874. Retrieved 2019-12-21, from https://advances.sciencemag.org/content/5/9/eaax1874 doi: 10.1126/sciadv.aax1874