Supporting Information for "Feedback Temperature Dependence Determines the Risk of High Warming"

Jonah Bloch-Johnson[1], Raymond T. Pierrehumbert[1], and Dorian S. Abbot[1]

Contents of this file

Corresponding author: J. Bloch-Johnson, Department of the Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637, USA. (jsbj@uchicago.edu)

[1]Department of the Geophysical Sciences,

University of Chicago, Chicago, Illinois,

USA.

Introduction

This supporting information includes two method descriptions and two additional figures. Section S1 describes our zero-dimensional energy balance model. Section S2 describes how we estimate a for a given GCM, and Section S3 describes how including a CO_2-dependent feedback influences this estimate, and our estimates of warming generally. Figure S1 demonstrates that higher-order terms become essential to determining warming when the Earth is quadratically unstable, suggesting that in these cases, the Earth's sensitivity may be incredibly difficult to estimate using GCMs. Figure S2 demonstrates what happens when the Earth jumps to a warmer state. Figure S3 shows the effect of feedback CO_2 dependence on estimates of equilibrium warming. Table S1 lists estimates of a for various GCMs with and without accounting for feedback CO_2 dependence.

Section S1. Model details

Zero-dimensional energy balance models of the Earth used in the climate feedback literature are variations of the equation

$$c\frac{dT}{dt} = N(T, C)$$

where c is the thermal inertia of the Earth system, T is the global annual mean surface temperature, N is the global annual mean top-of-atmosphere energy flux, and C is the CO_2 concentration. Specifically, this model is used to understand how changing C changes the equilibrium value(s) of T, which we can find by solving $N(T, C) = 0W/m^2$ for T under different values of C.

Suppose we refer to a given seasonal and spatial (including vertical) pattern of atmospheric and surface temperature, water vapor, and clouds as an "atmospheric climatology." For the fully-dimensional Earth, knowing T and C are not enough to determine N because there are many atmospheric climatologies associated with a given T, and for a given CO_2 concentration C, they might have very different average net top-of-atmosphere energy fluxes N. As a result, $N(T, C)$ is not a well-defined function for the Earth. Moreover, there are atmospheric climatologies and CO_2 concentrations for which N is 0, but for which the system is not in equilibrium, due to continued internal transfers of energy (e.g., continued lateral heat transport, or ocean heat uptake). These two problems break our ability to use Equation 1 to estimate changes in the equilibrium value of T, i.e. the sensitivity.

We will give some examples to illustrate these two problems, and then propose a way of defining $N(T, C)$ such that these problems are avoided.

Suppose that the preindustrial Earth had $T = 287K$ and $C = 270ppm$. Since the preindustrial Earth was roughly in equilibrium, this would imply $N(287K, 270ppm) = 0W/m^2$. However, suppose we were to instantaneously change the surface and atmosphere to be uniformly $287K$. T would still equal $287K$, but N would be greatly changed — namely, it would be $L_*(1-\alpha)/4-\sigma(287K)^4 \approx -145W/m^2$, implying $N(287K, 270ppm) = -145W/m^2$.

More realistically, imagine that we subjected a GCM to two experiments: one in which CO_2 was abruptly doubled from preindustrial conditions (i.e., increased to $540ppm$), and one in which CO_2 was abruptly quadrupled (i.e., increased to $1080ppm$). Suppose the abrupt doubling resulted in an equilibrium warming of $3K$, implying $N(290K, 540ppm) = 0W/m^2$. Suppose we paused the abrupt-quadrupling run when its average surface temperature T reached $290K$. Given the differential warming of different regions of the Earth, the atmospheric climatology associated with the paused abrupt-quadrupling model would look different than the atmospheric climatology of the finished abrupt-doubling model, even though they both have the same average surface temperature T.

Suppose we were to decrease the CO_2 concentration of the paused abrupt-quadrupling run from $1080ppm$ to $540ppm$. If the resulting TOA energy flux N was nonzero, it would imply $N(290K, 540ppm) \neq 0W/m^2$, giving an example of our first problem — that the same pair of T and C could imply different values of N. If not, $N(290K, 540ppm) = 0W/m^2$. However, we would expect that the paused run should not be in a state of internal equilibrium, so that if we were to unpause it, energy would still move around the Earth system, e.g. through ocean heat uptake, which would in turn continue to affect T.

This illustrates the second problem: knowing that $N = 0W/m^2$ for a given T and C does not guarantee that the Earth is in equilibrium.

If we associate with each T a specific atmospheric climatology, and if we choose these atmospheric climatologies such that when $N = 0W/m^2$, the Earth is in equilibrium, we can avoid both of these problems. Therefore, we define $N(T, C)$ in the following way: for each T choose an atmospheric climatology such that there is *some* CO_2 concentration C for which that atmospheric climatology would be in equilibrium, both internal and externally. Then, calculate the net top-of-atmosphere radiative fluxes that this atmospheric climatology would have if you were to set the CO_2 level to C, but *held the atmospheric climatology fixed*. Let N be the global annual mean value of those fluxes.

As an example, the preindustrial climate was roughly in equilibrium with $C = 270ppm$, and had $T = 287K$. Therefore, to evaluate $N(287K, 540ppm)$, for example, you would take the atmospheric climatology of the preindustrial Earth, set the CO_2 concentration to $540ppm$, and then, *without letting the system evolve*, measure the radiative transfer at the top of the atmosphere over the course of a year. The global annual average of this transfer is our desired N. Presumably, in this case, the value of $N(287K, 540ppm)$ would be close to the forcing F_{2x} without any adjustments (stratospheric or otherwise).

Suppose we define the "existence and uniqueness condition" as the condition that every T has one and only one atmospheric climatology such that there is some C for which that planet would be in equilibrium. If the existence and uniqueness condition holds, then our definition of $N(T, C)$ creates a well-defined function (each T implies a unique atmospheric climatology, implying a unique N for a given C), solving the first problem. Further, since

for a given T, our definition ensures that $N(T, C)$ is strictly monotonically increasing (e.g., $\frac{\partial N}{\partial C}|_T > 0$), $N(T, C)$ is zero only when C is the CO_2 value for which the atmospheric climatology associated with T is in equilibrium. This solves the second problem, since $N(T, C) = 0W/m^2$ therefore implies the system is in equilibrium.

Generally speaking, the existence and uniqueness condition might not hold, in which case our definition might not solve these two problems. However, since our goal in this paper is to understand the limitations of the linear model of equilibrium warming, i.e. $\Delta T = -F/\lambda$, our definition needs only hold when the linear model can work, and the linear model can only work when the existence and uniqueness condition holds:

• Since a given forcing $F(C)$ is simply a monotonic function of C, its inverse function F^{-1} is well-defined. If the Earth is linear, then for a given T, the CO_2 concentration $C = F^{-1}(\lambda(T-287K))+270ppm$ will be in equilibrium with that T. Therefore, the linear model implies that for every T, there is a C such that that T will be in equilibrium with that C, implying the existence of some atmospheric climatology that could be associated with that T. This implies that if there is no atmospheric climatology associated with a given T for which the Earth could be in equilibrium for some C, the linear model must not hold.

• If there are two atmospheric climatologies associated with a specific T such that there is some C for which these climatologies would be in equilibrium, then either these climatologies would be in equilibrium for the same C, or for different C's. If the C's are different, then it would be possible to change from one C to the other without changing T, i.e. $\Delta T = 0K$ for a nonzero F, implying an infinite λ and a broken linear model. If the C's

are the same, but the climatologies are different, then either there is some different CO_2 concentration C_{diff} such that measuring the net top-of-atmosphere flux from changing to C_{diff} without letting the system evolve (i.e. the forcing) would be different for the two models, or there would be no such C_{diff}.

If there is such a C_{diff}, these two different climatologies would have different forcings for the same CO_2 increase. For the linear model to hold, the warming caused by the increase to C_{diff} would have to be the same, because otherwise you could move between the two new Ts without changing C, suggesting an infinite sensitivity. As a result, we have different forcings giving us the same temperature increase, once more breaking the linear model.

If there was no such C_{diff} for which this was the case, the two climatologies would be indistinguishable from each other, and we could just choose one; such a definition of $N(T, C)$ would also solve the two problems presented at the beginning of this section.

To conclude, whenever we would expect the linear model to work, our definition ensures a well-defined $N(T, C)$ for which a zero-valued N implies equilibrium. Therefore, we can use it in our investigation.

Section S2. Adding a CO_2-dependent feedback

We start by adding a quadratic term representing the CO_2 dependence of the feedback to the quadratic model:

$$-F = \lambda \Delta T + a\Delta T^2 + bD\Delta T \tag{S1}$$

where D is the number of CO_2 doublings (i.e., $D \equiv \Delta log_2(C)$) and b represents the CO_2 dependence of the feedback (i.e., $b \equiv \partial^2 N / \partial T \partial log_2(C)$). Note that we use $log_2(C)$ as the

dependent variable rather than C, as we expect N to scale roughly linearly with $log_2(C)$ rather than C.

We first show the effects of adding this term on our GCM estimates of a. Instead of regressing model results against Equation 1 from the paper, we regress them against our altered version Equation S1 above. Results for a and b are shown in Table S1. Only one of the GCMs is affected by more than $0.02 W/m^2/K^2$ (CAM3). However, this GCM has only three runs, and has a value of b with a much larger magnitude than any of the other models, suggesting that the small sample size may limit the usefulness of this estimate of a and b.

Our regressions suggest a reasonable range of the feedback CO_2 dependence (b) of roughly $\pm 0.1\ W/m^2/K$ per doubling, disregarding the outlier associated with the under-sampled model. The effect of feedback CO_2 dependence b is qualitatively different than feedback temperature dependence a, because the nonlinearity associated with a positive a is self-amplifying (warming makes you more sensitive, and being more sensitive makes you warm more) in a way that can lead to jumps to warmer states, or extreme increases in sensitivity. The positive CO_2 dependence has some capacity for self-amplifying if C increases as a function of T (e.g. through the melting of methane hydrates, or the increased release of soil carbon). However, this effect is limited (e.g., there is only so much carbon to be released, and a certain threshold must be reached to release it), while the capacity for self-amplification due to positive temperature dependence does not have a threshold or limited capacity.

The qualitative difference in the two effects can be seen by comparing Figure 2 in the paper with Figure S3. While Figure 2 shows how equilibrium warming (ΔT) changes with feedback temperature dependence (a) for a fixed preindustrial feedback (λ), Figure S3 shows how equilibrium warming changes with feedback CO_2 dependence (b) for a fixed preindustrial feedback. While CO_2 dependence clearly affects the exact value of warming associated with a given CO_2 increase, it does not cause the same extreme behavior, such as loss of stability, or greatly heightened warming, caused by positive feedback temperature dependence.

Section S3. Estimating a for a GCM

For each GCM, we have a collection of runs where CO_2 was abruptly increased to $n \times C_0$ (where C_0 is the preindustrial CO_2 level) and then allowed to equilibrate. The initial radiative imbalance is the forcing ($F_{n\times} = N(T_0, nC_0)$) which we pair with the resulting warming, $\Delta T_{n\times}$. We assume that there is no CO_2-dependent feedback; the previous section discusses the impact of this assumption. Assuming no CO_2-dependent feedback is the same as assuming that lines of N are parallel as we change C, so that $N(T, nC_0) - N(T, C_0)$ is independent of T. Specifically, $N(T, nC_0) - N(T, C_0) = N(T_0, nC_0) - N(T_0, C_0) = N(T_0, nC_0) = F_{n\times}$. So, $F_{n\times} = N(T_0 + \Delta T_{n\times}, nC_0) - N(T_0 + \Delta T_{n\times}, C_0) = -N(T_0 + \Delta T_{n\times}, C_0)$. If we have m different runs, this gives us $m + 1$ points on the curve $N(T, C_0)$, including $N(T_0, C_0) = 0$. We can then fit a curve of the form $N = \lambda T + aT^2$ to these points. The resulting values are plotted in Figure 2 and listed in the first numerical column of Table S1.

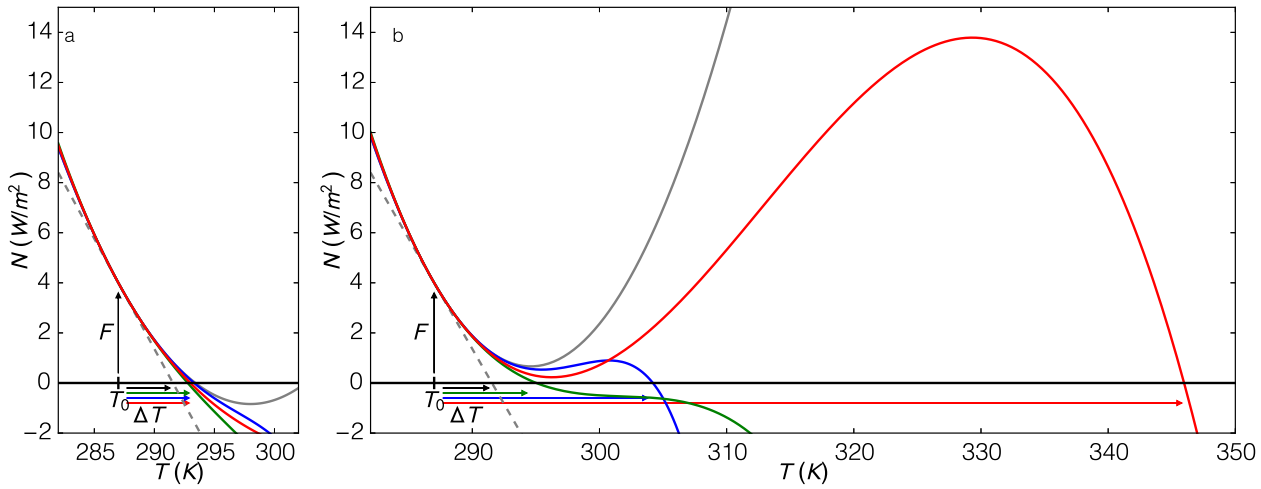

Figure S1. The effect of higher-order terms. Global annual-mean net top-of-atmosphere energy flux N as a function of global annual-mean surface temperature T for a fixed CO_2 concentration, after CO_2 has been increased from the preindustrial value. a) The three colored curves have the same values of preindustrial feedback λ ($-0.88 W/m^2/K$), feedback temperature dependence a ($0.02 W/m^2/K^2$), and CO_2 forcing F ($4\ W/m^2$), but different values of higher-order terms (red, $-7.5 \times 10^{-4} \Delta T^3$; green, $-1.3 \times 10^{-3} \Delta T^3$; blue, $-4 \times 10^{-6} \Delta T^5$, where each term is added to $\lambda \Delta T + a \Delta T^2$ to estimate N). Gray dashed and solid lines show the linear and quadratic approximations of N respectively. For this collection of λ, a, and F, the quadratic model does not run away, and the various higher-order terms do not significantly affect the total warming. b) The same as a), except that now a is ($0.058 W/m^2/K^2$), so that the quadratic model does run away. As a result, higher-order terms *must* come into play. The different higher-order terms cause very different warmings, and different qualitative behaviors (the green curve experiences no jump to a warmer state, while the red and blue curves do).

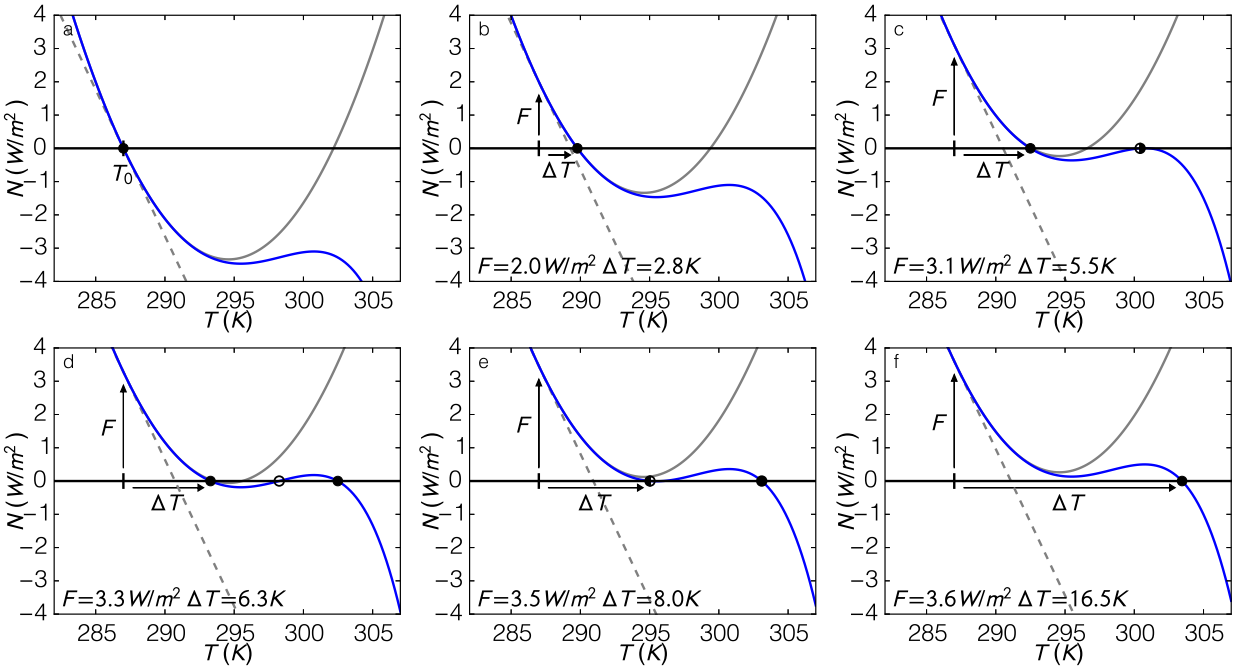

Figure S2. Jumping to a warmer state. Global annual-mean net top-of-atmosphere energy flux N as a function of global annual-mean surface temperature T, where each panel has a successively higher CO_2 concentration. The blue curves have the same λ ($-0.88 W/m^2/K$) and a ($0.058 W/m^2/K^2$) as the red curve in Figure 1a, but with a higher-order term ($-4 \times 10^{-6} \Delta T^5$) added. Gray dashed and solid lines show the linear and quadratic approximations of N respectively. a) The planet is in preindustrial equilibrium. b) After $2 W/m^2$ of forcing, the two approximations estimate the warming well. c) After $3.1 W/m^2$ of forcing, another part of the blue curve intersects the x-axis, so that a pair of new equilibria, one unstable and the other stable, is created in a saddle-node bifurcation. d) As the forcing increases, this pair separates, until in e), the stable equilibrium that the Earth is tracking collides with the unstable equilibrium. f) Under further forcing, these two equilibria disappear in another saddle-node bifurcation, and the Earth warms until it reaches the new stable equilibrium.

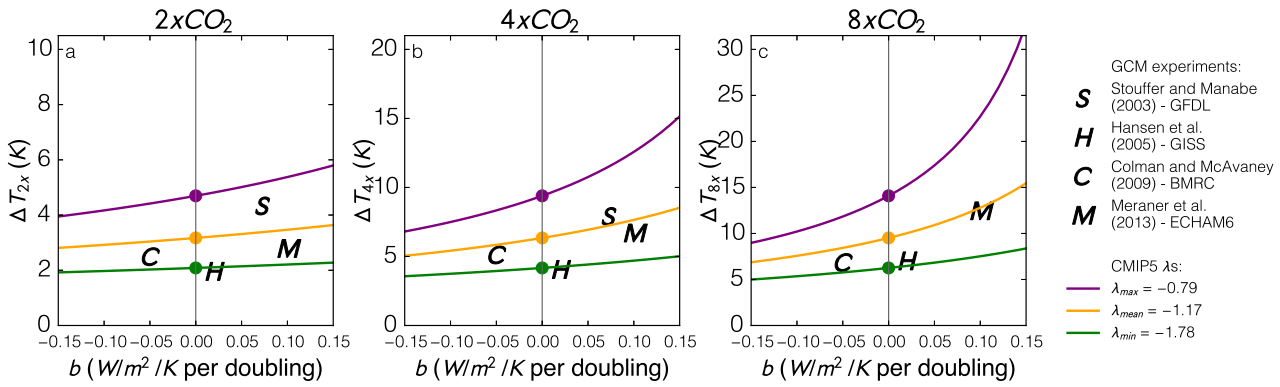

Figure S3. The effect of feedback CO_2 dependence. This figure is analogous to Figure 2 in the paper, except instead of showing how equilibrium warming (ΔT) changes with feedback temperature dependence (a) for a fixed preindustrial feedback (λ), this shows how warming changes with feedback CO_2 dependence (b) for a fixed preindustrial feedback. This effect is shown for one, two, and three doublings of CO_2 in panels a), b), and c) respectively. Letters are centered at values of b and equilibrium warmings for various GCMs. While warming changes with feedback CO_2 dependence (with less negative λ and larger CO_2 increases causing a larger deviation), these changes do not exhibit the same extreme behavior as the positive values of feedback temperature dependence in Figure 2. As an example, there are no shaded areas where the model runs away.

Table S1. Estimating feedback temperature dependence (a) with and without a CO_2-dependent feedback. b is the feedback CO_2 dependence. R^2_{linear} measures the linear fit and R^2_{quad} measures the quadratic fit (without taking into account CO_2 dependence). In both cases, fits assume $N(T_0, C_0) = 0$.

GCM	a^*(w/o CO_2 dep.)	a^*(w/ CO_2 dep.)	b^{**}	# of model runs	R^2_{linear}	R^2_{quad}
GFDL	-0.030	-0.044	0.07	3	0.897	1.000
GISS-E	0.059	0.048	0.02	8	0.985	1.000
BMRC	-0.034	-0.016	-0.05	9	0.953	0.996
CAM3	0.042	0.186	-0.42	3	0.993	1.000
ECHAM6	0.031	0.019	0.10	4	0.909	1.000

* feedback temperature dependence ($W/m^2/K$ per K)

** feedback CO_2 dependence ($W/m^2/K$ per *doubling*)