An estimate of equilibrium climate sensitivity from interannual variability

A.E. Dessler[1]*, P.M. Forster[2]

[1] Dept. of Atmospheric Sciences, Texas A&M University. adessler@tamu.edu

[2] School of Earth and Environment, University of Leeds, UK p.m.forster@leeds.ac.uk

Main points:

1. We use interannual variability to estimate equilibrium climate sensitivity (ECS). We estimate ECS is *likely* 2.4-4.6 K (17-83% confidence interval), with a mode and median value of 2.9 and 3.3 K, respectively.
2. We see no evidence to support low ECS (values less than 2K) suggested by other analyses.
3. This work shows the value of alternate energy balance frameworks for understanding climate change.

Abstract

Estimating the equilibrium climate sensitivity (ECS; the equilibrium warming in response to a doubling of CO_2) from observations is one of the big problems in climate science. Using observations of interannual climate variations covering the period 2000 to 2017 and a model-derived relationship between interannual variations and forced climate change, we estimate ECS is *likely* 2.4-4.6 K (17-83% confidence interval), with a mode and median value of 2.9 and 3.3 K, respectively. This analysis provides no support for low values of ECS (below 2 K) suggested by other analyses. The main uncertainty in our estimate is not observational uncertainty, but rather uncertainty in converting observations of short-term, mainly unforced climate variability to an estimate of the response of the climate system to long-term forced warming.

Plain language summary

Equilibrium climate sensitivity (ECS) is the amount of warming resulting from doubling carbon dioxide. It is one of the important metrics in climate science because it is a primary determinant of how much warming we will experience in the future. Despite decades of work, this quantity remains uncertain: the last IPCC report stated a range for ECS of 1.5-4.5 deg. Celsius. Using observations of interannual climate variations covering the period 2000 to 2017, we estimate ECS is *likely* 2.4-4.6 K. Thus, our analysis provides no support for the bottom of the IPCC's range.

Introduction

The response of the climate system to the imposition of a climate forcing is frequently described using the linearized energy balance equation:

$$R = F + \lambda\, T_s \qquad (1)$$

where forcing F is an imposed top-of-atmosphere (TOA) energy imbalance, T_S is the global average surface temperature, and λ is the change in TOA flux per unit change in T_S [Sherwood *et al.*, 2014]. R is the resulting TOA flux imbalance from the combined forcing and response. All quantities are anomalies, i.e., departures from a base state. Equilibrium climate sensitivity (hereafter ECS, the equilibrium warming in response to a doubling of CO_2) can be calculated as:

$$ECS = -F_{2xCO2}/\lambda \qquad (2)$$

where F_{2xCO2} is the forcing from doubled CO_2.

Equation 1 is a workhorse of climate science and it has been used many times to estimate λ and ECS. Many of these [e.g., Gregory *et al.*, 2002; Annan and Hargreaves, 2006; Otto *et al.*, 2013; Lewis and Curry, 2015; Aldrin *et al.*, 2012; Skeie *et al.*, 2014; Forster, 2016] combine Eq. 1 with estimates of R, F, and T_s over the 19th and 20th centuries to infer λ and ECS. These calculations suggest λ is near -2 $W/m^2/K$ and appear to rule out an ECS larger than ~4 K [Stevens *et al.*, 2016]. The increased likelihood of an ECS below 2 K implied by these calculations led the IPCC Fifth Assessment Report (AR5) to extend their *likely* ECS range downward to include 1.5 K [Collins *et al.*, 2013].

However, since AR5 a number of problems with this approach have been identified. These include questions about the impact of internal variability [e.g., Dessler *et al.*, 2018], arguments that ECS inferred from historical energy budget produces an underestimate of the true value [e.g., Armour, 2017; Gregory and Andrews, 2016; Zhou *et al.*, 2016; Andrews and Webb, 2018; Proistosescu and Huybers, 2017; Marvel *et al.*, 2018], the large and evolving uncertainty in forcing over the 20th century [e.g., Forster, 2016], different forcing efficacies of greenhouse gases and aerosols [Shindell, 2014; Kummer and Dessler, 2014], and geographically incomplete or inhomogeneous observations [Richardson *et al.*, 2016].

For robust estimates of ECS, multiple lines of evidence are needed and care needs to be taken in relating the inferred ECS from any method to other estimates. Thus, there is great value in finding alternate ways to approach the problem. Relatively few papers have attempted use short-term interannual variability to estimate ECS [e.g., Forster, 2016; Tsushima *et al.*, 2005; Forster and Gregory, 2006; Chung *et al.*, 2010; Tsushima and Manabe, 2013; Dessler, 2013; Donohoe *et al.*, 2014]. Papers that do typically yield estimates of ECS consistent with the IPCC's canonical ECS range of 1.5-4.5°C, but their uncertainty is so large as to provide no meaningful constraint of the range. In this paper, we present a new methodology that uses interannual fluctuations to help constrain the ECS range.

Results

Traditional energy-balance framework

Per Eq. 2, ECS requires estimates of F_{2xCO2} and λ. We use estimates of F_{2xCO2} from fixed sea surface temperature and sea-ice experiments from ten global climate models that submitted output to the Precipitation Driver Response Model Intercomparison Project [Myhre *et al.*, 2017b]. They estimate F_{2xCO2} to be normally distributed with a mean of 3.69 W/m^2 and a standard deviation of 0.13 W/m^2.

We estimate λ from observations of R and T_S. Observations of R come from the Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled product (ed. 4) [Loeb *et al.*, 2018] and cover the period March 2000 to July 2017. Estimates of T_S come from the European Centre for Medium Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERAi) [Dee *et al.*, 2011]. In these calculations, monthly and globally averaged anomalies are used, where anomalies are deviations from the mean annual cycle of the data.

Given these data, we calculate λ two ways, both based on Eq. 1. First, we use estimates of effective radiative forcing F over the CERES period and calculate λ as the slope of the regression of R-F vs. T_S. We use standard regressions in this paper — an ordinary least-squares fit, with R-F as the dependent variable and T_S as the independent variable [Murphy *et al.*, 2009]. The forcing is based on the IPCC AR5 forcing time series, revised and extended in the following ways. Forcing from CO_2, N_2O and CH_4 have been replaced by calculating new forcing timeseries using concentrations from NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/) with updated formula to convert mixing ratios to forcing [Etminan *et al.*, 2016]. Other forcing components match IPCC AR5 through 2011 and have been extended to July 2017. For aerosols and ozone, the multi-model mean forcing from Myhre et al. [2017a] is used. For volcanoes, the forcing from Andersson et al. [2015] is taken from their Figure 4, beginning in 2008. Solar forcing after 2011 is derived from SORCE data [Lean *et al.*, 2005]. Other minor forcing terms are estimated using the relative change in forcing from 2011-2017 from the RCP4.5 scenario [Meinshausen *et al.*, 2011].

Uncertainty is estimated using radiative forcing uncertainties from 2015. We take the 5%-95% range for each of the 14 different forcing terms in 2015 and turn this into a fractional range by dividing by the median 1750-2015 forcing estimate. This fractional uncertainty is Monte Carlo sampled for each forcing term independently. These fractions are then multiplied by the relevant forcing time series and summed to create 10,000 different realizations of the time series of total radiative forcing. The average forcing time series during the CERES period is plotted in Fig. S1.

We then estimate a distribution of λ using Monte Carlo sampling. We start by subtracting the 10,000 forcing time series from the observed R time series to generate 10,000 estimates of R-F. Then we repeat the following process 500,000 times: 1) randomly select an R-F time series, 2) randomly subsample it and the observed T_S time series, with replacement, 3) regress the sampled R-F and T_S data sets to obtain an estimate of λ. The number of samples taken is set by the number of independent pieces of information in the time series, as estimated by Eq. 6 of Santer et al. [2000] (the original data set contains 209 months; we estimate there are ~100-120 independent samples due to autocorrelation in the time series).

In the second approach, we assume forcing changes linearly over the CERES time period and account for it by detrending R and T_S time series. We do this by subtracting off the linear trend of each time series estimated using a least-squares regression. We then assume that $R_{detrended} = \lambda\ T_{S,detrended}$ and we calculate λ by regression. The distribution of λ is estimated by randomly sampling 500,000 times (with replacement) the detrended R and T_S time series, with each resampled data set providing one estimate λ. As with the previous estimate, we account for autocorrelation by reducing the number of samples taken, using Eq. 6 of Santer et al. [2000]. Plots of R, T_S, and F can be found in Section S1 of the supplement.

Distributions of λ for the two approaches are both quite wide (Fig. 1a), with values of -0.51±0.64 and -0.81±0.65 $W/m^2/K$ for the R-F and detrended calculations, respectively (uncertainties are 5-95% confidence intervals). The two estimates of λ reflect different ways of handing forcing and they show that different approaches yield similar distributions for λ. These distributions are similar to those estimated as the uncertainty of ordinary least-squares regressions of R-F vs. T_S (-0.52±0.56 W/m^2) and detrended R vs. detrended T_S (-0.82±0.64 W/m^2). Our sign convention is that fluxes are downward positive, so a negative λ means that a warmer planet radiates more energy to space, a necessary requirement for a stable climate.

The extreme width of the λ distributions is a consequence of scatter in the relationship between R-F and T_S (Fig. 1b) [Spencer and Braswell, 2010; Xie *et al.*, 2016], which is due to both weak coupling between the surface and ΔR [Dessler *et al.*, 2018] and weather noise. This means that our observational estimate of λ is quite uncertain, with almost all of the uncertainty coming from month-to-month variability in the R time series. Switching to another temperature data set, such as MERRA2 [Gelaro *et al.*, 2017], or using only the median forcing, yields very similar distributions. Systematic errors in the CERES time series are small; the data are stable to better than 0.5 W/m^2/decade (stability of the shortwave is 0.3 W/m^2/decade [Loeb *et al.*, 2007], and longwave is 0.15 W/m^2/decade [Susskind *et al.*, 2012]). Because we are regressing R vs. temperature, spurious trends in the data have little impact on our analysis [Dessler, 2010].

The distributions of λ plotted in Fig. 1a are derived mainly from the response to interannual variability (Fig. S3), so we will refer to them hereafter as λ_{iv}. The λ in Eq. 2, however, is the climate system's response to forcing from doubled CO_2 (hereafter λ_{2xCO2}), so we cannot simply plug λ_{iv} into Eq. 2 to derive ECS. In fact, this disconnect between what we can measure (λ_{iv})

and what is required to calculate ECS (λ_{2xCO2}) is one reason scientists have largely avoided using interannual variability to infer ECS.

We therefore modify Eq. 2 to account for this:

$$\text{ECS} = -\frac{F_{2\times CO2}}{\lambda_{iv,obs}}\frac{\lambda_{iv}}{\lambda_{2\times CO2}} \quad (3)$$

where $\lambda_{iv,obs}$ is the observed value (from Fig. 1a), mainly the response to interannual variability, while the ratio $\lambda_{iv}/\lambda_{2xCO2}$ is a transfer function that converts $\lambda_{iv,obs}$ into the required value λ_{2xCO2}. We estimate this transfer function using models that submitted required output to the 5th phase of the Coupled Model Intercomparison Project (CMIP5) [Taylor *et al.*, 2012]. The numerator λ_{iv} is derived from the models' control runs, in which climate variations arise naturally from internal variability. To facilitate comparison with the observations, as well as avoid any issues with long-term drift, we first break each control run into 16-year segments and calculate monthly anomalies of ΔR and ΔT_S during each segment, where anomalies are deviations from the average annual cycle of each 16-year period. We expect these model segments to contain the same types of climate variations that are in the observations (e.g., weather noise, ENSO). Then, we calculate λ_{iv} for each segment as the slope of the regression of ΔR vs. ΔT_S for that segment. Finally, we average the segments' values of λ_{iv} to come up with a single value of λ_{iv} for each model (Table S1).

The CMIP5 archive does not include doubled CO_2 runs, but it does have abrupt 4xCO_2 runs from which we can estimate λ_{4xCO2}. λ_{4xCO2} is calculated from these runs using the Gregory et al. [Gregory *et al.*, 2004] method: we regress all 150 years of annual R vs. annual average T_S, and take the resulting slope as an estimate of λ_{4xCO2}, where R and T_S are deviations from the pre-industrial control run.

If we assume that $\lambda_{2xCO2} \approx \lambda_{4xCO2}$, so we can re-write Eq. 3 as:

$$\text{ECS} \approx -\frac{F_{2\times CO2}}{\lambda_{iv,obs}}\frac{\lambda_{iv}}{\lambda_{4\times CO2}} \quad (4)$$

Recent work suggests that λ_{4xCO2} is less negative (i.e., implying a higher ECS) than λ_{2xCO2} [Armour, 2017; Proistosescu and Huybers, 2017]. On the other hand, we use all 150 years of the $4xCO_2$ runs to estimate λ_{4xCO2}, which tends to produce values that are too negative [Andrews *et al.*, 2015; Rugenstein *et al.*, 2016; Rose and Rayborn, 2016; Armour, 2017]. These two errors tend to cancel, but how much of a bias is left — and in which direction — remains an uncertainty in this analysis. The CMIP5 ensemble's distribution of $\lambda_{iv}/\lambda_{4xCO2}$ is plotted in Fig. 2; it has an average of 0.81 and a standard deviation of 0.34.

We then use a Monte Carlo approach to estimate ECS. We produce 500,000 estimates of ECS by randomly sampling the distributions of F_{2xCO2}, $\lambda_{iv,obs}$ (Fig. 1a), and $\lambda_{iv}/\lambda_{4xCO2}$ (Fig. 2) and plugging them into Eq. 3; negative ECS values or values greater than 10 K are viewed as physically implausible and thrown out (sensitivity to the 10-K threshold is shown in Table 1). We produce two ECS distributions — one using $\lambda_{iv,obs}$ from the R-F calculation and one using $\lambda_{iv,obs}$ from the detrended calculation. The ECS distributions (Fig. 3) have 17-83% confidence intervals (corresponding to the IPCC's *likely* range) of 2.5-7.0 K and 2.0-5.7 K for the R-F and detrended calculations, respectively. The modes are 3.0 and 2.4 K, while the medians are 4.2 and 3.3 K.

Overall, our calculated ECS distributions overlap substantially with the IPCC's range, although our distributions are shifted to higher values: we see a ~30% chance that ECS exceeds 4.5 K, while the IPCC assigns that a 17% chance. And we see less support for low values of ECS: the chance of an ECS below 2 K is 6-15%, while the IPCC assigns a 17% chance it is below 1.5 K.

Table 1 lists the statistics of these distributions, as well as a number of sensitivity tests to determine the robustness of the calculation. For example, we have done ECS calculations using a F_{2xCO2} distribution derived from the CMIP5 abrupt $4xCO_2$ runs instead of the distribution from the PDRMIP (see Sect. S2 for more about this). All of the ECS distributions are similar to those shown in Fig. 3, leading us to conclude that our conclusions are robust with respect to the many choices in how the calculation is done.

Modified energy-balance framework

Recently, Dessler et al. [2018] suggested a revision of Eq. 1, where the TOA flux is parameterized in terms of tropical atmospheric temperature, not global surface temperature:

$$R = F + \Theta T_A \tag{5}$$

where T_A is the tropical average (30°N-30°S) 500-hPa temperature and Θ converts this quantity to TOA flux. R and F are the same global average quantities they were in equation 1. They demonstrated that T_A correlated better with R-F than T_S does (Fig. 1c), thereby providing a superior way to describe global energy balance.

In this framework, the equilibrium warming of the tropical atmosphere ΔT_A in response to doubled CO_2 is equal to $-F_{2xCO2}/\Theta_{2xCO2}$. ECS can therefore be written

$$\text{ECS} = -\frac{F_{2\times \text{CO2}}}{\Theta_{iv,obs}} \frac{\Theta_{iv}}{\Theta_{2\times CO2}} \frac{\Delta T_S}{\Delta T_A} \approx -\frac{F_{2\times \text{CO2}}}{\Theta_{iv,obs}} \frac{\Theta_{iv}}{\Theta_{4\times CO2}} \frac{\Delta T_S}{\Delta T_A} \tag{6}$$

where $\Theta_{iv,obs}$ is the analog to $\lambda_{iv,obs}$, $\Theta_{iv}/\Theta_{2xCO2}$ is the transfer function that allows us to use short-term variability to estimate ECS, and $\Delta T_S/\Delta T_A$ is the ratio of the temperature changes at equilibrium in response to doubled CO_2. As we did previously, we will further assume that $\Theta_{4xCO2} \approx \Theta_{2xCO2}$.

We use the same forcing F_{2xCO2} that was used in the previous section. The distributions of the scaling factor $\Theta_{iv}/\Theta_{4xCO2}$ (Fig. 4a) come from the CMIP5 ensemble. These are calculated the same way as the $\lambda_{iv}/\lambda_{4xCO2}$ ratios were, except atmospheric temperatures are substituted for surface temperatures. Just as we did for $\lambda_{iv,obs}$, we calculate $\Theta_{iv,obs}$ two ways: by regressing R-F vs. T_A and by regressing detrended R vs. detrended T_A . Distributions of $\Theta_{iv,obs}$ for the two approaches are similar (Fig. 1a), with values of -0.98±0.32 and -1.09±0.29 W/m^2/K for the R-F and detrended calculations, respectively (uncertainties are 5-95% confidence intervals). Because of their similarities, in the rest of this section we will show results using the detrended calculation, although results for both distributions can be found in Table 2.

Finally, the distribution of the temperature ratio $\Delta T_S/\Delta T_A$ is also estimated from the CMIP5 ensemble. For each model, ΔT_S and ΔT_A are estimated as the average difference of the first and last decades of the abrupt 4xCO_2 runs; we then take the ratio of these values. Comparisons of the models to observations show that models do well at simulating this ratio (Sect. S3). The resulting distribution of $\Delta T_S/\Delta T_A$ constructed by the CMIP5 models (Fig. 5a) has an ensemble average and standard deviation of 0.86±0.10.

Long forced runs of the MPI-ESM1.1, GFDL CM3, and ESM2M models all show this ratio increases as the climate continues to warm beyond year 150. In runs of the GFDL CM3 and ESM2M, in which CO_2 increases at 1% per year until doubling and then remains fixed, the ratio increases from 0.79 and 0.70, 300 years after CO_2 doubles, to 0.86 and 0.76 at equilibrium (GFDL values are personal communication, David Paynter, 2018, based on runs described in [Paynter *et al.*, 2018]). The ratio in an abrupt 4xCO_2 run of the MPI model increases from 0.79 in years 140-150 to 0.87 in years 2400-2500. Thus, we conclude that values of this ratio obtained from the 150-year CMIP5 4xCO_2 simulations may be low biased, which would lead our ECS to also be low biased.

As in the previous section, we use a Monte Carlo approach and produce 500,000 estimates of ECS by randomly sampling the distributions of F_{2xCO2}, $\Theta_{iv,obs}$, $\Theta_{iv}/\Theta_{4xCO2}$, and $\Delta T_S/\Delta T_A$, and then plugging the values into Eq. 6. The resulting ECS distribution (Fig. 6a) shows a similar structure to the λ-based distributions in Fig. 3: a broad maximum between 2 and 3 K and a tail towards higher ECS values.

There is also a puzzling peak below 1°C. The only way for an ECS estimate to be close to zero is if $\Theta_{iv,obs}$ is very large or one of the other factors in Eq. 6 is close to zero. Analysis of the terms in Eq. 6 suggests that the term causing the low ECS values is $\Theta_{iv}/\Theta_{4xCO2}$, whose distribution approaches zero (Fig. 4a). These low values come from the GISS models (Fig. 7a, Table S1) and if they are removed from the ensemble, the bump below 1 K disappears (Fig. 6b), although the statistics of the distribution do not change much.

This result emphasizes that the scaling factor $\Theta_{iv}/\Theta_{4xCO2}$ is unconstrained by observations and has not been previously studied. That doesn't mean, however, that we know nothing about it — we do have observations of Θ_{iv} and can compare those to each model's value of Θ_{iv}. We find that 15 of the 25 CMIP5 models produce estimates of Θ_{iv} in agreement with the CERES observations (Fig. 7b). If we construct distributions of $\Theta_{iv}/\Theta_{4xCO2}$ and $\Delta T_S/\Delta T_A$ from just those models (Figs. 4b and 5b), we obtain the ECS distribution in Fig. 6c (hereafter referred to as the "good-Θ" distribution).

We consider the "good-Θ" ECS distributions to be the best estimates of ECS from this analysis. Those ECS distributions have 17-83% confidence intervals (corresponding to the IPCC's *likely* range) of 2.4-4.7 K and 2.4-4.4 K for the R-F and detrended calculations, respectively. Averaging these gives us our single best estimate for the *likely* range, 2.4-4.6 K, and 5-95% range, 1.9-5.7 K. The modes are 2.6 and 3.1 K (average 2.9 K), and the medians of both are 3.3 K.

These distributions suggest a 15-20% chance ECS exceeds 4.5 K and a 6% chance of an ECS below 2 K. We therefore conclude that the IPCC's upper end of the *likely* ECS range is about right, but that the low end is too low. We would conclude that, in the parlance of the IPCC, ECS is *very unlikely* to be below 2 K.

We have also performed corresponding "good-λ" ECS calculations in which the $\lambda_{iv}/\lambda_{4xCO2}$ distribution in Eq. 4 is constructed using only those models whose λ_{iv} agrees with $\lambda_{iv,obs}$. The ECS distributions obtained from these calculations (Table 1) are similar to distributions from the λ calculations using all models.

Discussion

There are several reasons why ECS estimated from the revised energy balance framework (Eq. 6) should be considered more reliable than that estimated from the traditional framework (Eq. 4) used in previous papers [e.g., Forster, 2016; Tsushima *et al.*, 2005; Forster and Gregory, 2006; Chung *et al.*, 2010; Tsushima and Manabe, 2013; Dessler, 2013; Donohoe *et al.*, 2014]. Fig. 1 shows the main advantage — that $\Theta_{iv,obs}$ is better constrained than $\lambda_{iv,obs}$ This is what leads to the narrower distributions of ECS in Fig. 6 than in Fig. 3. Of particular note, the $\lambda_{iv,obs}$ distributions have non-zero probabilities of values close to zero; since ECS is proportional to $1/\lambda_{iv,obs}$, this generates a large tail towards unrealistically large ECS values.

There are additional reasons that lead us to conclude that the estimates from the revised framework are superior. It has been suggested that $\lambda_{iv,obs}$ exhibits significant decadal variability in models [Andrews *et al.*, 2015; Gregory and Andrews, 2016; Zhou *et al.*, 2016; Dessler *et al.*, 2018]. This opens the possibility that the observed $\lambda_{iv,obs}$, based on 16 years of data, is biased with respect to the long-term average; if so, then ECS estimated from these observations would also be biased. Model simulations suggest that $\Theta_{iv,obs}$ exhibits smaller decadal variability [Dessler *et al.*, 2018], making Θ_{iv} estimated from CERES data a more robust estimate of the climate system's actual long-term value. There is also evidence that Θ changes less than λ during transient climate change [Dessler *et al.*, 2018], making the assumption that $\Theta_{2xCO2} \approx \Theta_{4xCO2}$ a far better one than the assumption that $\lambda_{4xCO2} \approx \lambda_{2xCO2}$.

It is also worth stepping back and asking what could cause our calculation to be seriously in error. It seems unlikely that forcing from doubled CO_2 is wrong given our good understanding of the physics of CO_2 forcing [e.g., Feldman *et al.*, 2015]. Estimates of $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$ are derived from observations we view to be reliable, so our judgment is that they are also unlikely to be significantly wrong. The $\Delta T_S/\Delta T_A$ factor comes from climate model simulations, but models have long been able to accurately reproduce the observed pattern of surface warming [e.g., Stouffer and Manabe, 2017], and we have simple physical arguments explaining how the atmospheric and surface temperature should be connected [Xu and Emanuel, 1989]. Finally, we can compare the models to data [Compo *et al.*, 2011; Poli *et al.*, 2016] to validate their simulation of this ratio (Sect. S3).

Thus, the transfer function $\Theta_{iv}/\Theta_{4xCO2}$ seems the most probable place for a significant error to occur. That said, there are reasons to believe the models' estimates of this ratio. As mentioned above, we can directly compare Θ_{iv} in the models to observations, and find agreement in the majority of models (Fig. 7). We also argue that while errors may exist in a model (i.e., in the cloud feedback), this will affect both the numerator and denominator and such errors will tend to cancel out. As a preliminary test of this, we have analyzed three different versions of the MPI-ESM 1.2 model that have had their cloud feedbacks modified to produce different ECS [Thorsten Mauritsen and Diego Jimenez, personal communication, 2018]. The three versions are the standard model (ECS calculated from an abrupt 4xCO_2 run using the Gregory method = 3.0 K), an "iris" version [described in Mauritsen and Stevens, 2015] (ECS = 2.6 K), and a "high ECS" version, in which the convective parameterization has been tweaked to generate a large, positive cloud feedback (ECS = 5.2 K). Despite large differences in the ECS, these three versions have similar values of $\lambda_{iv}/\lambda_{4xCO2}$ of 1.17, 1.15, and 1.11 for the standard, iris, and high ECS versions, respectively. The corresponding values of $\Theta_{iv}/\Theta_{4xCO2}$ are 1.06, 0.96, and 1.10. While one must be careful about conclusions based on a single model, this nevertheless provides some support for the hypothesis that errors in Θ_{4xCO2} will cancel errors in Θ_{iv} when the ratio is taken and that the ratio $\Theta_{4xCO2}/\Theta_{iv}$ may well be more accurate than either Θ_{4xCO2} or Θ_{iv} are individually.

We have also constructed an error budget to determine which term contributes most to the width of the distributions in Fig. 6. We do this by sequentially setting each term to have zero uncertainty by replacing that term's distribution in the Monte Carlo calculation with a single number, the ensemble average. This has little effect on the mean, median, or mode, but does change the width of the distribution (Table 3). By comparing the widths of the resulting distributions (defined as the distance between the 17th and 83rd percentiles), Fig. 8 shows that the biggest contributor to ECS uncertainty is the uncertainty in $\Theta_{iv}/\Theta_{4xCO2}$. Eliminating the uncertainty in that reduces the 17-83% confidence interval to 2.8-4.0 K. *Thus, developing a theoretical argument for the value of this ratio would be a key advance in climate science.* The next most important uncertainty is uncertainty in $\Theta_{iv,obs}$, followed by the uncertainty in $\Delta T_S/\Delta T_A$ and then the uncertainty in F_{2xCO2}.

Conclusions

Estimating ECS from observations remains one of the big problems in climate science. Despite several decades of intense investigations, the uncertainty in this parameter remains stubbornly large, with the last IPCC assessment reporting a *likely* range of 1.5-4.5 K (17-83% confidence interval). Because of this, there is great value in finding alternate ways to approach the problem.

In this paper, we have used observations of interannual climate variations covering the period 2000 to 2017 along with a model-derived relationship between interannual variations and forced climate change to estimate ECS. We interpret the observations using a modified energy balance framework (Eq. 5) in which the response of TOA flux is proportional to the atmospheric temperature. We conclude ECS is *likely* 2.4-4.6 K (17-83% confidence interval), with a mode and median value of 2.9 and 3.3 K, respectively. Overall, our analysis suggests that the upper end of the IPCC's range is set about right, but this analysis provides little evidence to support estimates of ECS in the bottom third of the IPCC's *likely* range.

One of the key parts of our calculations is the use of CMIP5 climate models to convert the observations of interannual variability into an estimate of the response of the system to doubled CO_2. This is the main uncertainty in our analysis and future efforts to pin this transfer function down would be extremely valuable.

References

Aldrin, M., M. Holden, P. Guttorp, R. B. Skeie, G. Myhre, & T. K. Berntsen (2012), Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content, Environmetrics, 23, 253-271, doi: 10.1002/env.2140.

Andersson, S. M., B. G. Martinsson, J.-P. Vernier, J. Friberg, C. A. M. Brenninkmeijer, M. Hermann, et al. (2015), Significant radiative impact of volcanic aerosol in the lowermost stratosphere, Nature Communications, 6, 7692, doi: 10.1038/ncomms8692.

Andrews, T., & M. J. Webb (2018), The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical Pacific warming, J. Climate, 31, 641-654, doi: 10.1175/jcli-d-17-0087.1.

Andrews, T., J. M. Gregory, & M. J. Webb (2015), The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models, J. Climate, 28, 1630-1648, doi: 10.1175/JCLI-D-14-00545.1.

Annan, J. D., & J. C. Hargreaves (2006), Using multiple observationally-based constraints to estimate climate sensitivity, Geophys. Res. Lett., 33, doi: 10.1029/2005gl025259.

Armour, K. C. (2017), Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks, Nature Clim. Change, 7, 331-335, doi: 10.1038/nclimate3278.

Chung, E. S., B. J. Soden, & B. J. Sohn (2010), Revisiting the determination of climate sensitivity from relationships between surface temperature and radiative fluxes, Geophys. Res. Lett., 37, doi: 10.1029/2010gl043051.

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, et al. (2013), Long-term climate change: Projections, commitments and irreversibility, in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, et al. (2011), The Twentieth Century Reanalysis Project, Q. J. R. Meteor. Soc., 137, 1-28, doi: 10.1002/qj.776.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteor. Soc., 137, 553-597, doi: 10.1002/qj.828.

Dessler, A. E. (2010), A Determination of the Cloud Feedback from Climate Variations over the Past Decade, Science, 330, 1523-1527, doi: 10.1126/science.1192546.

Dessler, A. E. (2013), Observations of climate feedbacks over 2000-10 and comparisons to climate models, J. Climate, 26, 333-342, doi: 10.1175/jcli-d-11-00640.1.

Dessler, A. E., T. Mauritsen, & B. Stevens (2018), The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity, Atmos. Chem. Phys., 18, 5147-5155, doi: 10.5194/acp-18-5147-2018.

Donohoe, A., K. C. Armour, A. G. Pendergrass, & D. S. Battisti (2014), Shortwave and longwave radiative contributions to global warming under increasing CO2, Proc. Natl. Acad. Sci., 111, 16700-16705, doi: 10.1073/pnas.1412190111.

Etminan, M., G. Myhre, E. J. Highwood, & K. P. Shine (2016), Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing, Geophys. Res. Lett., 43, 12,614-612,623, doi: 10.1002/2016GL071930.

Feldman, D. R., W. D. Collins, P. J. Gero, M. S. Torn, E. J. Mlawer, & T. R. Shippert (2015), Observational determination of surface radiative forcing by CO2 from 2000 to 2010, Nature, 519, 339, doi: 10.1038/nature14240.

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

Forster, P. M. D., & J. M. Gregory (2006), The climate sensitivity and its components diagnosed from Earth Radiation Budget data, J. Climate, 19, 39-52.

Gelaro, R., W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, et al. (2017), The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419-5454, doi: 10.1175/jcli-d-16-0758.1.

Gregory, J. M., & T. Andrews (2016), Variation in climate sensitivity and feedback parameters during the historical period, Geophys. Res. Lett., 43, 3911-3920, doi: 10.1002/2016GL068406.

Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, & N. A. Rayner (2002), An observationally based estimate of the climate sensitivity, J. Climate, 15, 3117-3121, doi: 10.1175/1520-0442(2002)015<3117:aobeot>2.0.co;2.

Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, et al. (2004), A new method for diagnosing radiative forcing and climate sensitivity, Geophys. Res. Lett., 31, doi: 10.1029/2003gl018747.

Kummer, J. R., & A. E. Dessler (2014), The impact of forcing efficacy on the equilibrium climate sensitivity, Geophys. Res. Lett., 41, 3565-3568, doi: 10.1002/2014gl060046.

Lean, J., G. Rottman, J. Harder, & G. Kopp (2005), SORCE contributions to new understanding of global change and solar variability, Solar Physics, 230, 27-53, doi: 10.1007/s11207-005-1527-2.

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

Loeb, N. G., S. Kato, K. Loukachine, N. Manalo-Smith, & D. R. Doelling (2007), Angular distribution models for top-of-atmosphere radiative flux estimation from the Clouds and the Earth's Radiant Energy System instrument on the Terra satellite. Part II: Validation, Journal of Atmospheric and Oceanic Technology, 24, 564-584.

Loeb, N. G., D. R. Doelling, H. Wang, W. Su, C. Nguyen, J. G. Corbett, et al. (2018), Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product, J. Climate, 31, 895-918, doi: 10.1175/jcli-d-17-0208.1.

Marvel, K., R. Pincus, G. A. Schmidt, & R. L. Miller (2018), Internal variability and disequilibrium confound estimates of climate sensitivity from observations, Geophys. Res. Lett., doi: 10.1002/2017GL076468.

Mauritsen, T., & B. Stevens (2015), Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models, Nature Geosci, 8, 346-351, doi: 10.1038/ngeo2414.

Meinshausen, M., S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, J.-F. Lamarque, et al. (2011), The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213, doi: 10.1007/s10584-011-0156-z.

Murphy, D. M., S. Solomon, R. W. Portmann, K. H. Rosenlof, P. M. Forster, & T. Wong (2009), An observationally based energy balance for the Earth since 1950, J. Geophys. Res., 114, 14, doi: 10.1029/2009jd012105.

Myhre, G., W. Aas, R. Cherian, W. Collins, G. Faluvegi, M. Flanner, et al. (2017a), Multi-model simulations of aerosol and ozone radiative forcing due to anthropogenic emission changes during the period 1990–2015, Atmos. Chem. Phys., 17, 2709-2720, doi: 10.5194/acp-17-2709-2017.

Myhre, G., P. M. Forster, B. H. Samset, Ø. Hodnebrog, J. Sillmann, S. G. Aalbergsjø, et al. (2017b), PDRMIP: A Precipitation Driver and Response Model Intercomparison Project—

Protocol and Preliminary Results, Bull. Am. Met. Soc., 98, 1185-1198, doi: 10.1175/bams-d-16-0019.1.

Otto, A., F. E. L. Otto, O. Boucher, J. Church, G. Hegerl, P. M. Forster, et al. (2013), Energy budget constraints on climate response, Nature Geoscience, 6, 415-416, doi: 10.1038/ngeo1836.

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

Poli, P., H. Hersbach, D. P. Dee, P. Berrisford, A. J. Simmons, F. Vitart, et al. (2016), ERA-20C: An Atmospheric Reanalysis of the Twentieth Century, J. Climate, 29, 4083-4097, doi: 10.1175/jcli-d-15-0556.1.

Proistosescu, C., & P. J. Huybers (2017), Slow climate mode reconciles historical and model-based estimates of climate sensitivity, Science Advances, 3, doi: 10.1126/sciadv.1602821.

Richardson, M., K. Cowtan, E. Hawkins, & M. B. Stolpe (2016), Reconciled climate response estimates from climate models and the energy budget of Earth, Nature Clim. Change, 6, 931-935, doi: 10.1038/nclimate3066.

Rose, B. E. J., & L. Rayborn (2016), The effects of ocean heat uptake on transient climate sensitivity, Current Climate Change Reports, 2, 190-201, doi: 10.1007/s40641-016-0048-4.

Rugenstein, M. A. A., K. Caldeira, & R. Knutti (2016), Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes, Geophys. Res. Lett., 43, 9877-9885, doi: 10.1002/2016GL070907.

Santer, B. D., T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, et al. (2000), Statistical significance of trends and trend differences in layer-average atmospheric temperature time series, J. Geophys. Res., 105, 7337-7356, doi: 10.1029/1999jd901105.

Sherwood, S. C., S. Bony, O. Boucher, C. Bretherton, P. M. Forster, J. M. Gregory, & B. Stevens (2014), Adjustments in the forcing-feedback framework for understanding climate change, Bull. Am. Met. Soc., 96, 217-228, doi: 10.1175/BAMS-D-13-00167.1.

Shindell, D. T. (2014), Inhomogeneous forcing and transient climate sensitivity, Nature Climate Change, 4, 274, doi: 10.1038/nclimate2136.

Skeie, R. B., T. Berntsen, M. Aldrin, M. Holden, & G. Myhre (2014), A lower and more constrained estimate of climate sensitivity using updated observations and detailed radiative forcing time series, Earth System Dynamics, 5, 139-175, doi: 10.5194/esd-5-139-2014.

Spencer, R. W., & W. D. Braswell (2010), On the diagnosis of radiative feedback in the presence of unknown radiative forcing, J. Geophys. Res., 115, doi: 10.1029/2009JD013371.

Stevens, B., S. C. Sherwood, S. Bony, & M. J. Webb (2016), Prospects for narrowing bounds on Earth's equilibrium climate sensitivity, Earth's Future, 4, 512-522, doi: 10.1002/2016EF000376.

Stouffer, R. J., & S. Manabe (2017), Assessing temperature pattern projections made in 1989, Nature Clim. Change, 7, 163-165, doi: 10.1038/nclimate3224.

Susskind, J., G. Molnar, L. Iredell, & N. G. Loeb (2012), Interannual variability of outgoing longwave radiation as observed by AIRS and CERES, J. Geophys. Res., 117, doi: 10.1029/2012jd017997.
Taylor, K. E., R. J. Stouffer, & G. A. Meehl (2012), An overview of CMIP5 and the experiment design, Bull. Am. Met. Soc., 93, 485-498, doi: 10.1175/bams-d-11-00094.1.
Tsushima, Y., & S. Manabe (2013), Assessment of radiative feedback in climate models using satellite observations of annual flux variation, Proc. Natl. Acad. Sci., 110, 7568-7573, doi: 10.1073/pnas.1216174110.
Tsushima, Y., A. Abe-Ouchi, & S. Manabe (2005), Radiative damping of annual variation in global mean surface temperature: comparison between observed and simulated feedback, Climate Dynamics, 24, 591-597, doi: 10.1007/s00382-005-0002-y.
Xie, S.-P., Y. Kosaka, & Y. M. Okumura (2016), Distinct energy budgets for anthropogenic and natural changes during global warming hiatus, Nature Geoscience, 9, 29-33, doi: 10.1038/ngeo2581.
Xu, K. M., & K. A. Emanuel (1989), Is the tropical atmosphere conditionally unstable?, Mon. Wea. Rev., 117, 1471-1479.
Zhou, C., M. D. Zelinka, & S. A. Klein (2016), Impact of decadal cloud variations on the Earth's energy budget, Nature Geosci, 9, 871-874, doi: 10.1038/ngeo2828.

Acknowledgments: A.E.D. acknowledges support from NSF grant AGS-1661861 to Texas A&M University. P.M.F. acknowledges support from the Natural Environment Research Council project NE/P014844/1. We thank Bjorn Stevens and Thorsten Mauritsen for their insight into this analysis. We also acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison, and the WCRP's Working Group on Coupled Modeling for their roles in making available the WCRP CMIP5 multimodel dataset. CERES data were downloaded from ceres.larc.nasa.gov, CMIP5 data were downloaded from cmip.llnl.gov, and ECMWF reanalysis were downloaded from www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. Code and data can be found here: https://zenodo.org/record/1323162.

Table 1. ECS values from the λ runs

Summary of the statistics of the ECS distributions derived using Eq. 4. "%<2" and "%>4.5" gives the percent of ECS values that are below 2 K or above 4.5 K. Units are in K, except for "%<2" and "%>4.5", which are in percent.

run	mean	mode	median	5-95%	17-83%	%<2	%>4.5
all-Lambda-1	4.63	2.98	4.22	1.7-8.8	2.5-7.0	6	32
all-Lambda-1-f	4.43	2.85	3.99	1.6-8.7	2.3-6.8	8	30
all-Lambda-1-f_20-150	4.59	2.98	4.19	1.6-8.9	2.4-7.0	7	31
all-Lambda-1_8K	4.17	2.98	3.95	1.7-7.3	2.4-6.1	6	25
all-Lambda-1_12K	5.00	2.98	4.41	1.8-10.2	2.6-7.7	6	36
all-Lambda-2	3.78	2.44	3.29	1.4-8.0	2.0-5.7	15	26
all-Lambda-2_8K	3.52	2.44	3.19	1.4-6.8	2.0-5.2	15	21
all-Lambda-2_12K	3.97	2.44	3.35	1.4-8.8	2.0-6.0	15	28
good-Lambda-1	4.20	2.71	3.73	1.6-8.4	2.3-6.3	9	28
good-Lambda-2	3.66	2.31	3.19	1.4-7.7	1.9-5.4	16	24
good-Lambda-1-f_20-150	4.18	2.71	3.72	1.4-8.5	2.2-6.4	10	28
good-Lambda-1-f	3.98	2.44	3.48	1.4-8.3	2.1-6.0	12	25

Names containing "all" or "good" include all models or just the ones whose λ_{iv} agrees with the CERES observations, respectively. The names with "-1" or "-2" use $\lambda_{iv,obs}$ derived using estimates of forcing (the R-F calculations) and the detrended calculations, respectively. The names including "-f" use forcing from the CMIP5 abrupt 4x CO_2 runs (see Sect. S2). The names including "-f_20-150" calculate F_{2xCO2} and λ_{4xCO2} from years 20-150 of the abrupt 4xCO_2 runs (see Sect. S2). Names with "-8K" and "-12K" change the plausibility threshold above which ECS values are considered non-physical and are thrown out.

Table 2. ECS values from the Θ runs

Same as Table 1, but derived using Eq. 6.

run	mean	mode	median	5-95%	17-83%	%<2	%>4.5
all-Theta-1	3.33	2.58	3.14	0.7-6.2	2.1-4.6	15	19
all-Theta-2	2.96	2.31	2.82	0.7-5.4	1.9-4.1	20	11
all-Theta-1-corr	3.36	2.58	3.13	0.8-6.5	2.0-4.8	16	20
all-Theta-1-f	3.11	2.44	2.91	0.7-6.0	1.9-4.4	21	16
all-Theta-1-f_20_150	2.98	2.31	2.75	0.6-5.8	1.8-4.3	24	14
good-Theta-1	3.56	2.58	3.33	2.0-5.9	2.4-4.7	6	20
good-Theta-2	3.43	3.12	3.33	1.9-5.3	2.4-4.4	6	15
good-Theta-1-corr	3.58	2.44	3.33	1.9-6.2	2.3-4.8	7	21
good-Theta-1-f	2.81	2.17	2.65	0.5-5.1	1.8-3.9	25	10
good-Theta-1-f_20-150	2.71	2.03	2.51	0.4-5.0	1.7-3.8	30	9
noGISS-Theta-1	3.56	2.58	3.28	1.9-6.3	2.3-4.8	8	21
noGISS-Theta-2	3.18	2.31	2.94	1.7-5.5	2.1-4.2	13	12

Names follow the same convention as Table 1. The names including "noGISS-" include all models except the two GISS models. In the "-corr" calculations, each Monte Carlo value of ECS uses values of $\Delta T_S/\Delta T_A$ and $\Theta_{iv}/\Theta_{4xCO2}$ from the same model.

Table 3. Error budget calculations

Summary of the statistics of the ECS distribution when one of the input distributions has no uncertainty.

run	mean	mode	median	5-95%	17-83%	%<2	%>4.5
error-all-Theta-2-noF	2.97	2.31	2.82	0.7-5.4	1.9-4.1	20	11
error-all-Theta-2-noRat	2.97	2.71	2.90	2.1-4.1	2.4-3.5	3	2
error-all-Theta-2-nodtdt	2.97	2.31	2.85	0.7-5.3	1.9-4.0	19	11
error-all-Theta-2-noTheta	2.89	2.31	2.78	0.7-5.0	2.0-3.9	18	8
error-good-Theta-2-noF	3.43	3.25	3.32	1.9-5.3	2.4-4.4	6	15
error-good-Theta-2-noRat	3.43	3.25	3.35	2.4-4.7	2.8-4.0	0	7
error-good-Theta-2-nodtdt	3.43	3.25	3.35	2.0-5.2	2.4-4.3	5	14
error-good-Theta-2-noTheta	3.34	3.53	3.35	2.1-4.8	2.4-4.2	4	10

Most name conventions Table 1. For these calculations, we take the “all-Theta-2” or “good-Theta-2” calculation and sequentially set the uncertainty in one term to zero. The “-noF”, “-noRat”, “-nodtdt”, and “-noTheta” correspond to no uncertainty in F_{2xCO2}, $\Theta_{iv}/\Theta_{4xCO2}$, $\Delta T_S/\Delta T_A$, and Θ_{iv}, respectively.

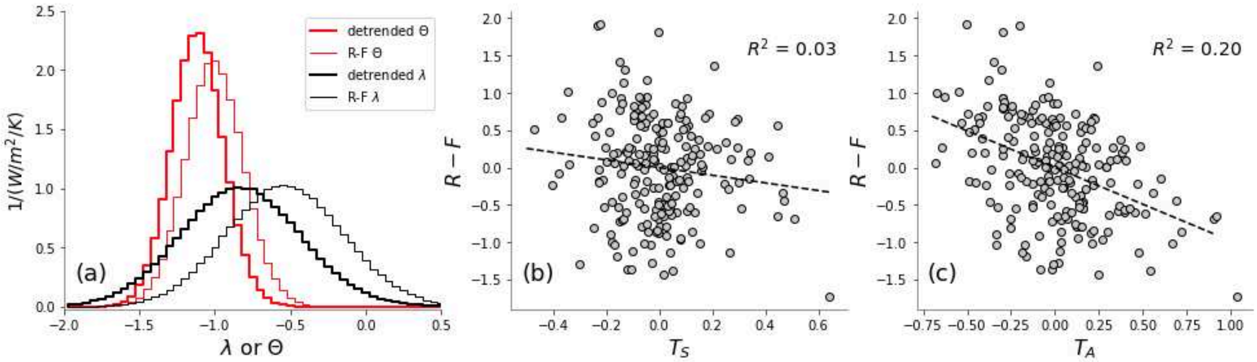

Figure 1. (a) Distribution of $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$ (W/m^2); (b) scatter plot of R-F (W/m^2) vs. T_S (K), the dashed line is a least-squares fit; (c) same as panel (b), but the regression is against T_A (K).

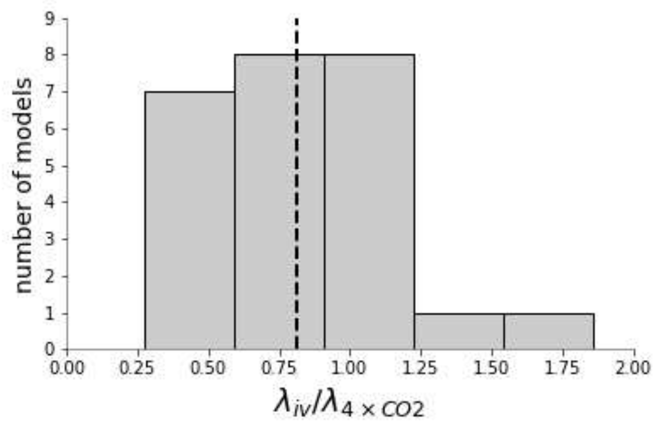

Figure 2. Distribution of $\lambda_{iv}/\lambda_{4xCO2}$ from 25 CMIP5 models; the black dashed line is the mean of the distribution. See methods for description of how the value is calculated in each model.

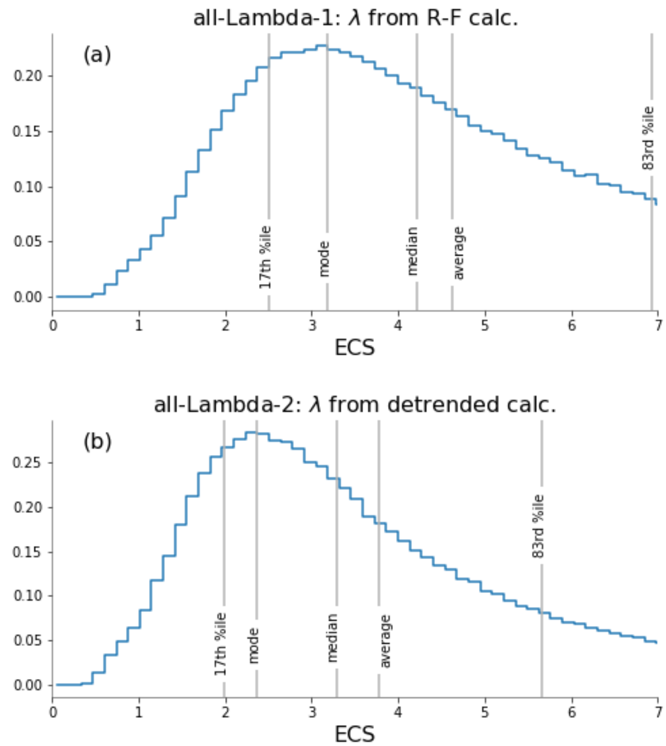

Figure 3. Distributions of ECS using the traditional energy balance framework (Eq. 4). (a) Calculated using $\lambda_{iv,obs}$ from the R-F regression, (b) Calculated using $\lambda_{iv,obs}$ from the detrended regression. "17th %ile" and "83rd %ile" are 17^{th} and 83^{rd} percentile, corresponding to the IPCC's *likely* range.

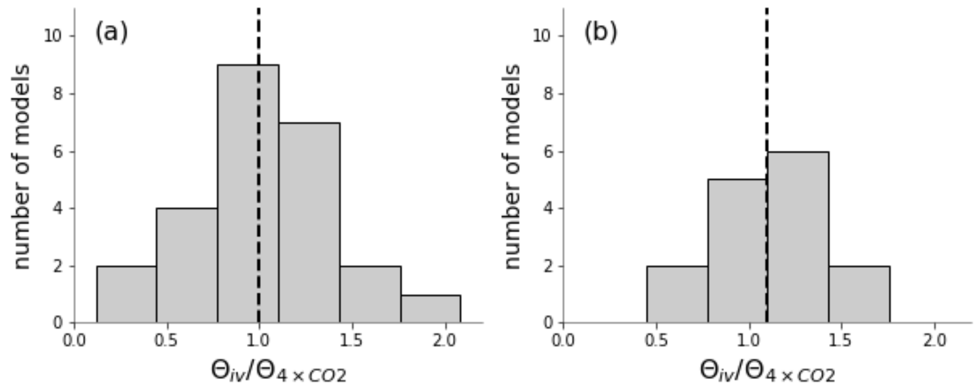

Figure 4. Distribution of $\Theta_{iv}/\Theta_{4xCO2}$ from (a) 25 CMIP5 models and (b) from those 15 models whose Θ_{iv} agrees with observations. The black dashed lines are the means of the distributions.

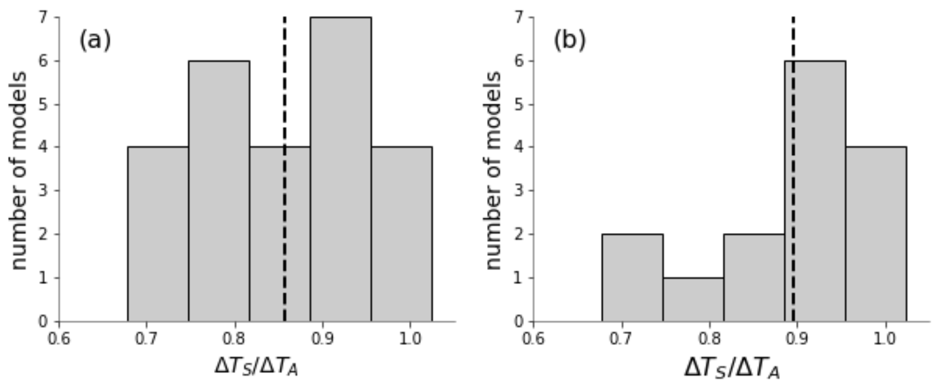

Figure 5. Distribution of $\Delta T_S/\Delta T_A$ from (a) 25 CMIP5 models and (b) from those 15 models whose Θ_{iv} agrees with observations. The black dashed lines are the means of the distributions.

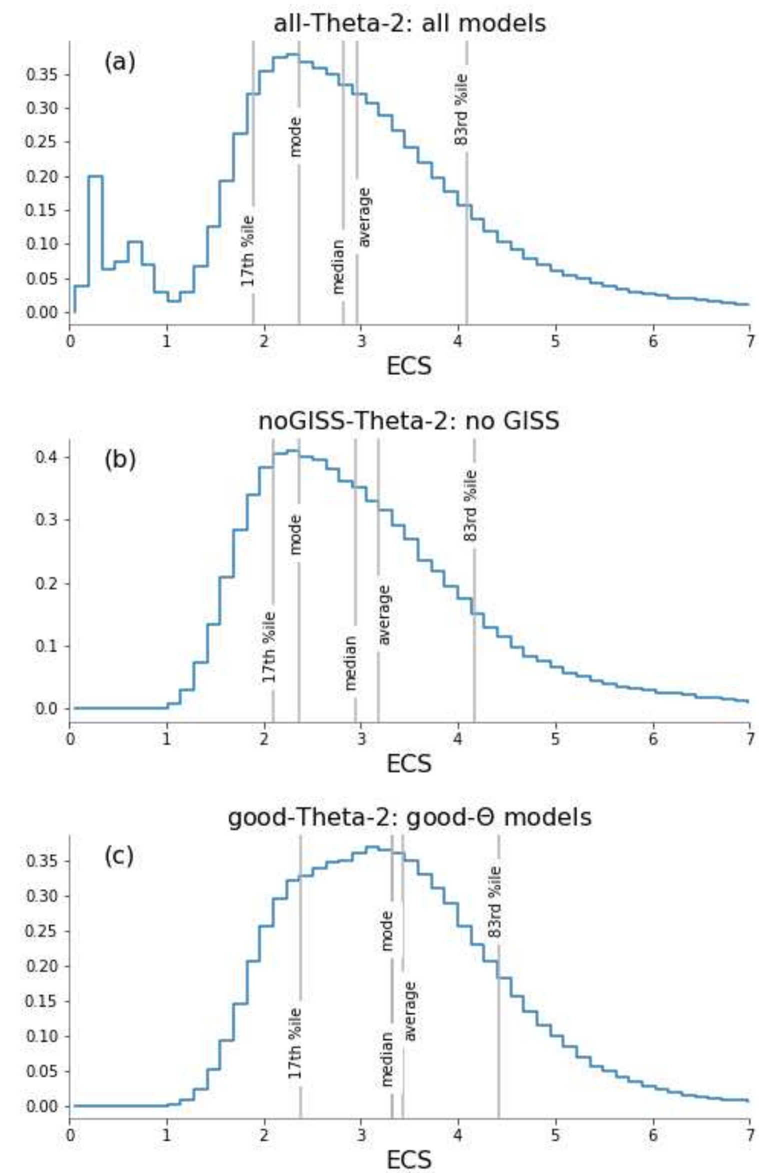

Figure 6. Distributions of ECS using the revised energy balance framework (Eq. 6). Panel (a) uses all models for the distributions of $\Theta_{iv}/\Theta_{4xCO2}$ and $\Delta T_S/\Delta T_A$, (b) uses all models except for the two GISS models, (c) uses 15 models whose Θ_{iv} agrees with the value estimated from observations. All calculations use $\Theta_{iv,obs}$ from the detrended calculation. "17th %ile" and "83rd %ile" are 17^{th} and 83^{rd} percentile, corresponding to the IPCC's *likely* range.

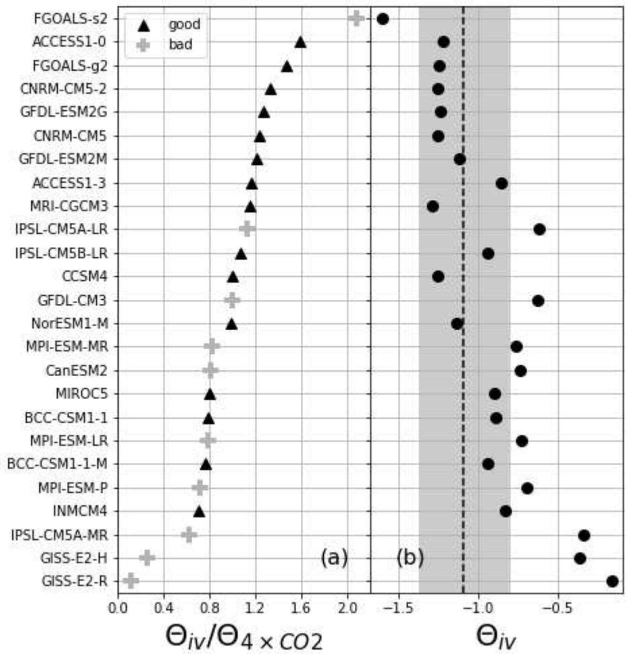

Figure 7. CMIP5 model estimates of (a) $\Theta_{iv}/\Theta_{4xCO2}$ and (b) Θ_{iv} (W/m^2). The gray region in panel (b) shows the observational range (from the detrended calculation). The black triangle symbols in panel a) indicate that the model's Θ_{iv} agrees with observations; the gray cross symbols indicate that it does not.

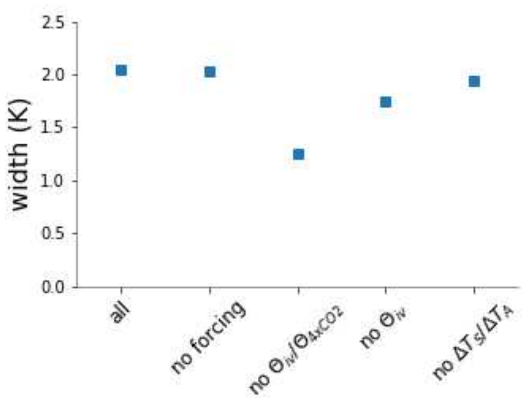

Figure 8. Error budget analysis of ECS estimates. The "all" point is the width of the ECS distribution from the good-Theta-2 calculation (Table 3). Then, from left to right, is the width when the uncertainty in forcing, $\Theta_{iv}/\Theta_{4xCO2}$, $\Theta_{iv,obs}$, and $\Delta T_S/\Delta T_A$ distributions are sequentially set to zero. For all points, "width" is the difference between the 17th and 83rd percentile of the ECS distribution.

Journal of Geophysical Research

Supporting Information for

An estimate of equilibrium climate sensitivity from interannual variability

A.E. Dessler[1], P.M. Forster[2]

[1] Dept. of Atmospheric Sciences, Texas A&M University

[2] School of Earth and Environment, University of Leeds, UK

Contents of this file

S1. Data going into the calculations of $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$

This section shows additional plots of the CERES, temperature, and forcing data. Fig. S1 shows the CERES R time series, the median forcing F time series, and the R-F time series. The CERES data are anomalies (deviations from the mean annual cycle); the forcing data are referenced to pre-industrial. These data go into the R-F estimates of $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$. Median forcing over the period analyzed in this paper, relative to preindustrial, is 2.2 W/m^2, with 5-95% confidence interval of 1.1-3.1 W/m^2. While the forcing uncertainty is large, what's important for this analysis is the uncertainty of the *slope* of the regression of forcing vs. temperature. Regressing all 10,000 forcing time series vs. T_S yields a median value of 0.62 W/m^2/K and 5-95% confidence interval of ±0.16 W/m^2/K.

Fig. S2 shows the raw and detrended CERES and ERAi temperature data. The detrended time series are used to estimate the detrended $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$. These two plots show that both forcing and detrending are minor adjustments to the data. The top panel in Fig. S2 also shows good agreement between ERAi and MERRA2. This supports our analysis that most of the uncertainty in $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$ comes from the scatter in CERES R measurements. Fig. S3 shows the correspondence between ΔT_S and the Nino3 index, which demonstrates that most of the variability in ΔT_S is due to interannual variability and not long-term climate change.

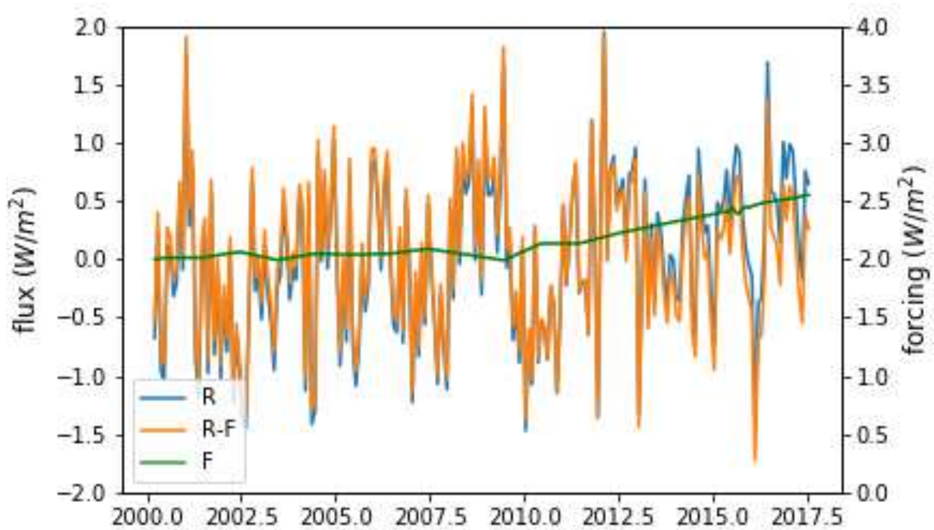

Fig. S1. Time series of global average, monthly anomalies of CERES R (blue), median forcing F (green), and R-F (orange).

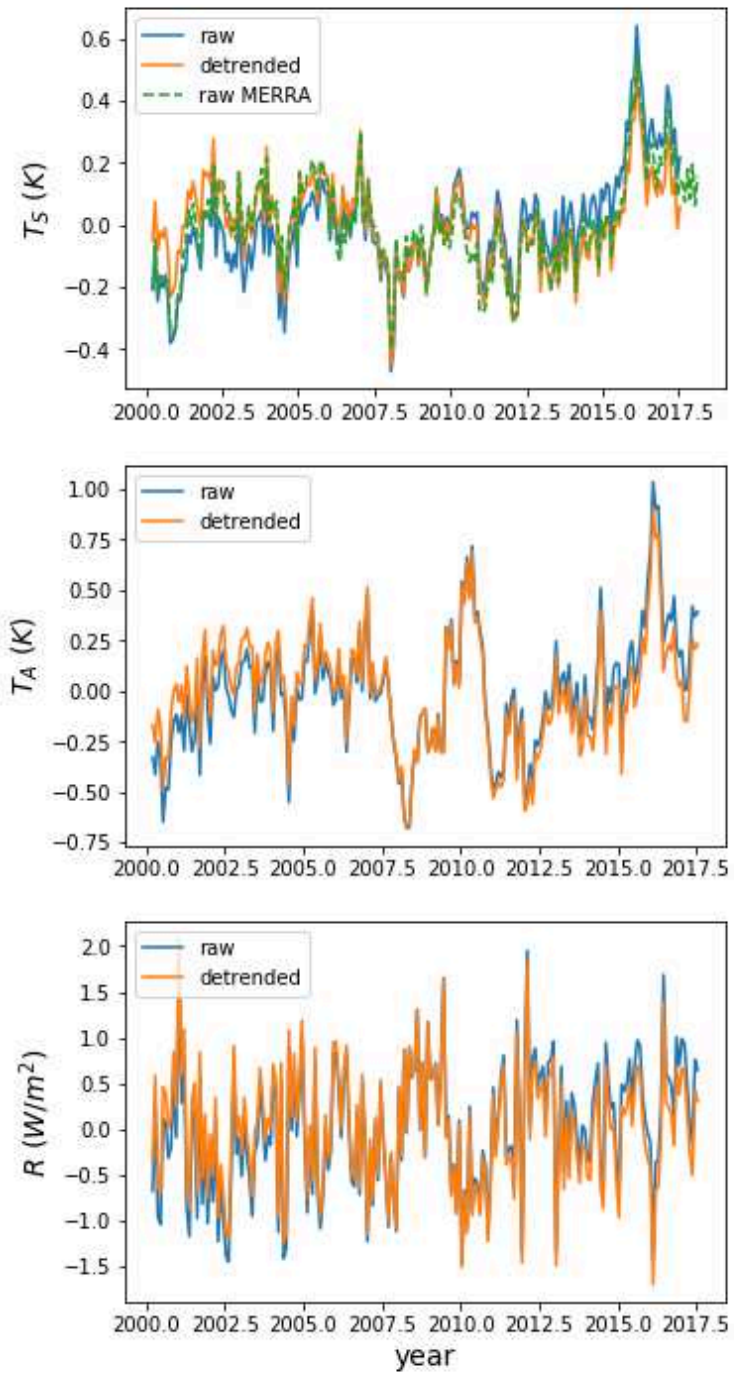

Fig. S2. Time series of global average, monthly anomalies of CERES R and ERAi global average surface temperature and 500-hPa tropical average (30°N-30°S) temperature. The raw time series is before detrending; the detrended time series has the linear trend, estimated using a least-squares fit, removed. The top plot also shows the raw MERRA2 surface temperature for comparison to the ERAi data.

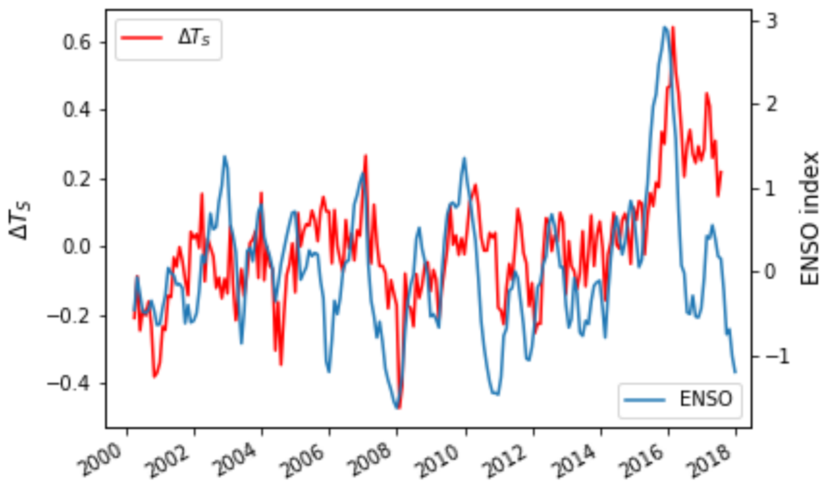

Fig. S3. Time series of global average surface temperature anomaly ΔT_S (K; left-hand axis) and Nino3 ENSO index (right-hand axis). ENSO index downloaded from https://www.esrl.noaa.gov/psd/data/timeseries/monthly/NINO3/.

S2. Alternate ways to calculate F_{2xCO2} and λ_{4xCO2} and Θ_{4xCO2}

One potential issue in our calculation is that the forcing we use is from fixed SST runs while the values of λ_{4xCO2} and Θ_{4xCO2} come from abrupt $4xCO_2$ runs. To evaluate the impact of any possibly incompatibility, we have also calculated ECS using a distribution of F_{2xCO2} obtained from the $4xCO_2$ runs using the Gregory method [*Gregory et al.*, 2004] (Fig. S4a, Table S1). The ECS distributions obtained from this (all-Lambda-1-f, good-Lambda-1-f, all-Theta-1-f, good-Theta-1-f) are summarized in Table 1 and 2. ECS estimated using these forcing distributions are close to those using PDRMIP forcing, so we conclude that this is not a significant uncertainty in our analysis.

Another potential issue is that we use of all 150 years of the CMIP5 abrupt $4xCO_2$ runs to estimate λ_{4xCO2} and Θ_{4xCO2}. It is well known that removal of the first few decades in the Gregory regression produces a less negative λ_{4xCO2} [e.g., *Andrews et al.*, 2015], which implies a higher ECS. The effect of this on Θ_{4xCO2} is smaller [*Dessler et al.*, 2018]. To test the impact of this, we produce ECS estimates where λ_{4xCO2} is calculated from years 20-150 (all-Lambda-1-f_20-150, good-Lambda-1-f_20-150, all-Theta-1-f_20-150, good-Theta-1-f_20-150). For consistency in these calculations, we use a forcing distribution also derived using these years (Fig. S4b). Note that we call this "quasi-F_{2xCO2}" because it should really not be considered a forcing — it is instead just the y-intercept of the Gregory plot for a regression covering years 20-150, which we need to use in order to correctly estimate the x-intercept, the ECS.

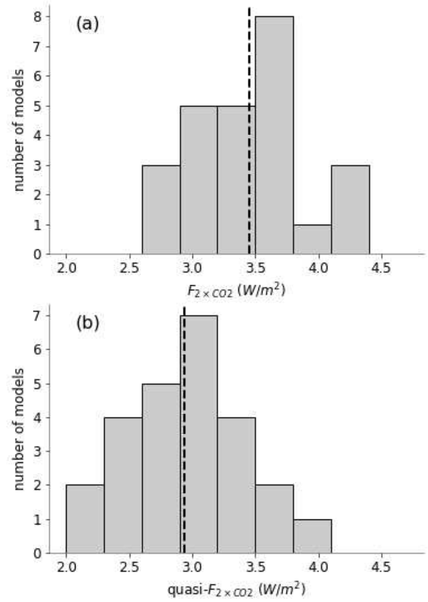

Fig. S4. Distribution of F_{2xCO2} from CMIP5 abrupt $4xCO_2$ runs. Panel (a) uses all 150 years of the run, while panel (b) uses years 20-150. The dashed lines are the ensemble averages of 3.45 and 2.94 W/m^2.

S3. Testing models' ability to estimate $\Delta T_S/\Delta T_A$

To evaluate the accuracy of the CMIP5 ensemble's estimate of $\Delta T_S/\Delta T_A$, we re-write it as the product of two terms:

$$\frac{\Delta T_S}{\Delta T_A} = \frac{\Delta T_{S,tropics}}{\Delta T_A} \frac{\Delta T_S}{\Delta T_{S,tropics}} \quad \text{(S1)}$$

where ΔT_S and ΔT_A are the global average surface temperature and tropical average atmospheric temperature, respectively, and $\Delta T_{S,tropics}$ is the tropical (30°N-30°S) average surface temperature. The term $\Delta T_{S,tropics}/\Delta T_A$ is a measure of the tropical lapse rate, which is understood to be controlled by moist convective adjustment [*Xu and Emanuel*, 1989]. Fig. S5a plots monthly average anomalies of $\Delta T_{S,tropics}$ vs. ΔT_A from the ERAi and, as expected, there is a clear correlation between these variables. The slope derived from this regression is 0.51±0.06 (5-95% confidence interval).

The ERAi data set, covering 1979-2016 (37 years), contains both long-term warming and interannual variability. Because of this, we compare the ERAi results to what we consider to be the most analogous model period, the last 37 years of the CMIP5 ensemble's 150-year abrupt $4xCO_2$ runs. Ensemble average $\Delta T_{S,tropics}$ over this period is 1.07 K, similar to the warming in the ERAi from 1979-2016. While a few models appear to have issues with this metric, there is generally good agreement between the models and from observations (Fig. S5b).

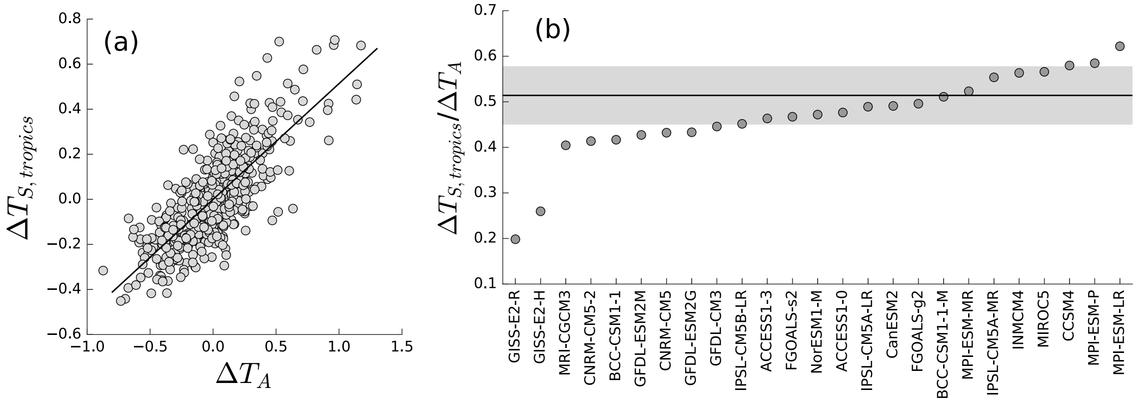

Figure S5. Estimates of $\Delta T_{S,tropics}/\Delta T_A$. (a) Scatter plot of monthly $\Delta T_{S,tropics}$ (K; tropical avg. surface temperature) anomalies vs. ΔT_A (K) anomalies from ERAi reanalysis (1979-2016). The solid line is the best fit line. (b) The slope of the same fit from the last 37 years of the CMIP5 ensemble's abrupt 4xCO_2 runs. The black line and gray region shows the slope and uncertainty of the fit to observations in panel a.

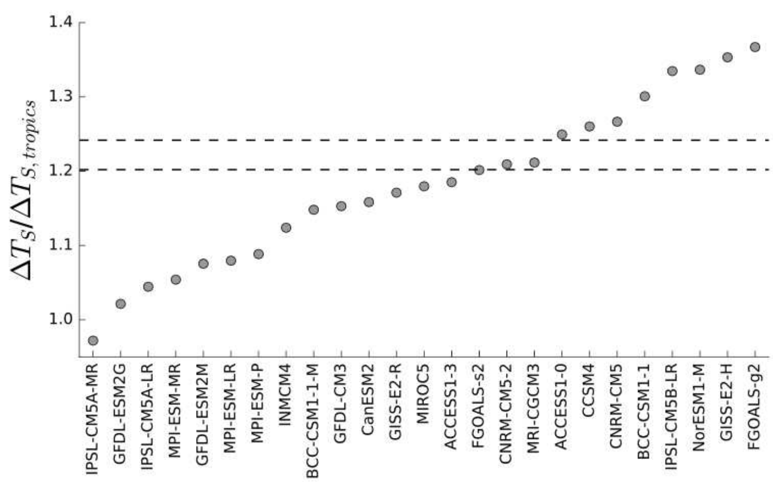

Figure S6. Estimates of polar amplification in the models, $\Delta T_S/\Delta T_{S,tropics}$. For the CMIP5 ensemble, this is calculated by differencing the average of the first and last decades of the CMIP5 ensemble's abrupt 4xCO_2 runs. The two dashed lines are observational estimates (see text).

The second term on the right-hand side of Eq. S1, $\Delta T_S/\Delta T_{S,tropics}$, is a measure of polar amplification in the pattern of surface warming. We estimate this by differencing the averages of the first and last decade of observations or models. The ECMWF 20th century reanalysis [*Poli et al.*, 2016] produces a value of 1.20 over the years 1900-2010 while the NOAA 20th century reanalysis project [*Compo et al.*, 2011] produces a value of 1.23 over the years 1851-2014. We estimate this ratio in each CMIP5 abrupt 4xCO_2 run and the ensemble agrees well with observations (Fig. S6), with a CMIP5 ensemble average of 1.18 and standard deviation of 0.11.

Such good agreement is not surprising — climate models have long demonstrated considerable skill in simulating the large-scale patterns of surface warming [e.g., *Stouffer and Manabe*, 2017].

S4. Estimating the distribution of λ_{4xCO2}

In the main text, we focus on estimating the distributions of ECS. However, we could also produce an observational estimate of the distribution of λ_{4xCO2}. We do this with the following two equations:

$$\lambda_{4\times CO2} \approx \lambda_{iv,obs}\frac{\lambda_{4\times CO2}}{\lambda_{iv}} \tag{S2}$$

$$\lambda_{4\times CO2} \approx \Theta_{iv,obs}\frac{\Theta_{4\times CO2}}{\Theta_{iv}}\frac{\Delta T_A}{\Delta T_S} \tag{S3}$$

We use the same Monte Carlo approach we did in the main text: distributions of $\Theta_{iv,obs}$ and $\lambda_{iv,obs}$ come from the observations and distributions of $\lambda_{iv}/\lambda_{4xCO2}$, $\Theta_{iv}/\Theta_{4xCO2}$, and $\Delta T_S/\Delta T_A$ come from the CMIP5 models. The resulting distributions are summarized in Tables S2 and S3. We note that the Θ calculations provide a consistent bound for λ of -0.7 to -1.5 W/m^2/K (17-83% confidence interval)

Table S1. Values for individual models

Model	λ_{iv}	Θ_{iv}	λ_{4xCO2}	Θ_{4xCO2}	$\Delta T_S/\Delta T_A$	F_{2xCO2}
ACCESS1-0	-0.69	-1.22	-0.75	-0.77	0.96	2.88
ACCESS1-3	-0.66	-0.86	-0.82	-0.74	0.91	2.91
BCC-CSM1-1	-0.74	-0.89	-1.21	-1.12	0.93	3.38
BCC-CSM1-1-M	-0.91	-0.94	-1.31	-1.23	0.92	3.69
CCSM4	-1.26	-1.25	-1.24	-1.26	0.99	3.63
CNRM-CM5	-1.14	-1.25	-1.11	-1.01	0.94	3.63
CNRM-CM5-2	-1.01	-1.25	-1.06	-0.94	0.89	3.64
CanESM2	-0.77	-0.73	-1.03	-0.90	0.88	3.80
FGOALS-g2	-1.55	-1.25	-0.83	-0.85	1.00	2.82
FGOALS-s2	-1.35	-1.60	-0.88	-0.77	0.87	3.75
GFDL-CM3	-0.21	-0.63	-0.75	-0.63	0.80	2.94
GFDL-ESM2G	-0.80	-1.24	-1.42	-0.98	0.68	3.33
GFDL-ESM2M	-1.41	-1.12	-1.34	-0.92	0.74	3.26
GISS-E2-H	-1.48	-0.36	-1.57	-1.36	0.91	3.70
GISS-E2-R	-1.03	-0.16	-1.70	-1.35	0.77	3.64
INMCM4	-0.65	-0.83	-1.51	-1.18	0.80	3.07
IPSL-CM5A-LR	-0.57	-0.61	-0.79	-0.54	0.71	3.19
IPSL-CM5A-MR	-0.46	-0.33	-0.81	-0.54	0.68	3.32
IPSL-CM5B-LR	-0.93	-0.94	-1.00	-0.87	0.91	2.61
MIROC5	-1.18	-0.90	-1.58	-1.13	0.84	4.25
MPI-ESM-LR	-0.78	-0.72	-1.14	-0.91	0.81	4.11
MPI-ESM-MR	-0.69	-0.76	-1.18	-0.93	0.80	4.08
MPI-ESM-P	-0.72	-0.70	-1.25	-0.98	0.80	4.32
MRI-CGCM3	-0.58	-1.29	-1.26	-1.11	0.88	3.27
NorESM1-M	-1.19	-1.13	-1.11	-1.15	1.02	3.10

Units on λ and Θ are $W/m^2/K$, $\Delta T_S/\Delta T_A$ is unitless; F_{2xCO2} is derived from that model's abrupt $4xCO_2$ run and has units of W/m^2.

Table S2. λ_{4xCO2} estimated from Eq. S2

run	mean	mode	median	5-95%	17-83%
all-Lambda-1	-0.73	-0.63	-0.64	-1.9 to +0.2	-1.3 to -0.2
all-Lambda-2	-1.16	-0.95	-1.03	-2.6 to -0.2	-1.8 to -0.5
good-Lambda-1	-0.85	-0.79	-0.78	-2.1 to +0.2	-1.5 to -0.2
good-Lambda-2	-1.20	-0.95	-1.07	-2.6 to -0.2	-1.8 to -0.5

See Table 1 for a description of the runs. Units are $W/m^2/K$.

Table S3. λ_{4xCO2} estimated from Eq. S3

run	mean	mode	median	5-95%	17-83%
all-Theta-1	-1.41	-1.11	-1.00	-4.2 to -0.5	-1.5 to -0.7
all-Theta-2	-1.56	-1.11	-1.11	-4.6 to -0.6	-1.6 to -0.8
all-Theta-1-corr	-1.41	-1.11	-1.00	-4.2 to -0.5	-1.5 to -0.7
good-Theta-1	-1.01	-1.11	-0.96	-1.6 to -0.6	-1.3 to -0.7
good-Theta-2	-1.05	-1.11	-0.99	-1.6 to -0.6	-1.4 to -0.8
good-Theta-1-corr	-1.01	-1.11	-0.96	-1.6 to -0.6	-1.3 to -0.7
noGISS-Theta-1	-1.00	-1.11	-0.96	-1.6 to -0.5	-1.4 to -0.7
noGISS-Theta-2	-1.11	-1.11	-1.07	-1.8 to -0.6	-1.5 to -0.8

See Table 2 for a description of the runs. Units are $W/m^2/K$.