



UNIVERSITY OF LEEDS

This is a repository copy of *An estimate of equilibrium climate sensitivity from interannual variability*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/159979/>

Version: Preprint

Preprint:

Dessler, AE orcid.org/0000-0003-3939-4820 and Forster, PM orcid.org/0000-0002-6078-0171 (2018) *An estimate of equilibrium climate sensitivity from interannual variability*.

[Preprint - EarthArXiv]

<https://doi.org/10.31223/osf.io/4et67>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

1 **An estimate of equilibrium climate sensitivity from**
2 **interannual variability**

3
4 A.E. Dessler^{1*}, P.M. Forster²
5

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

7 ²School of Earth and Environment, University of Leeds, UK p.m.forster@leeds.ac.uk
8
9

10
11 Main points:

- 12 1. We use interannual variability to estimate equilibrium climate sensitivity (ECS). We
13 estimate ECS is *likely* 2.4-4.6 K (17-83% confidence interval), with a mode and median
14 value of 2.9 and 3.3 K, respectively.
- 15 2. We see no evidence to support low ECS (values less than 2K) suggested by other
16 analyses.
- 17 3. This work shows the value of alternate energy balance frameworks for understanding
18 climate change.
19

20 **Abstract**

21 Estimating the equilibrium climate sensitivity (ECS; the equilibrium warming in response to a
22 doubling of CO₂) from observations is one of the big problems in climate science. Using
23 observations of interannual climate variations covering the period 2000 to 2017 and a model-
24 derived relationship between interannual variations and forced climate change, we estimate
25 ECS is *likely* 2.4-4.6 K (17-83% confidence interval), with a mode and median value of 2.9 and
26 3.3 K, respectively. This analysis provides no support for low values of ECS (below 2 K)
27 suggested by other analyses. The main uncertainty in our estimate is not observational
28 uncertainty, but rather uncertainty in converting observations of short-term, mainly unforced
29 climate variability to an estimate of the response of the climate system to long-term forced
30 warming.

31 **Plain language summary**

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

39

40 **Introduction**

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

$$43 \quad R = F + \lambda T_s \quad (1)$$

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

$$49 \quad \text{ECS} = -F_{2\times\text{CO}_2}/\lambda \quad (2)$$

50 where $F_{2\times\text{CO}_2}$ is the forcing from doubled CO_2 .

51 Equation 1 is a workhorse of climate science and it has been used many times to estimate λ and
52 ECS. Many of these [e.g., Gregory *et al.*, 2002; Annan and Hargreaves, 2006; Otto *et al.*, 2013;
53 Lewis and Curry, 2015; Aldrin *et al.*, 2012; Skeie *et al.*, 2014; Forster, 2016] combine Eq. 1 with
54 estimates of R , F , and T_s over the 19th and 20th centuries to infer λ and ECS. These calculations
55 suggest λ is near $-2 \text{ W/m}^2/\text{K}$ and appear to rule out an ECS larger than $\sim 4 \text{ K}$ [Stevens *et al.*,
56 2016]. The increased likelihood of an ECS below 2 K implied by these calculations led the IPCC
57 Fifth Assessment Report (AR5) to extend their *likely* ECS range downward to include 1.5 K
58 [Collins *et al.*, 2013].

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

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

76 **Results**

77 Traditional energy-balance framework

78 Per Eq. 2, ECS requires estimates of $F_{2\times\text{CO}_2}$ and λ . We use estimates of $F_{2\times\text{CO}_2}$ from fixed sea
79 surface temperature and sea-ice experiments from ten global climate models that submitted
80 output to the Precipitation Driver Response Model Intercomparison Project [Myhre *et al.*,
81 2017b]. They estimate $F_{2\times\text{CO}_2}$ to be normally distributed with a mean of 3.69 W/m² and a
82 standard deviation of 0.13 W/m².

83 We estimate λ from observations of R and T_s . Observations of R come from the Clouds and the
84 Earth's Radiant Energy System (CERES) Energy Balanced and Filled product (ed. 4) [Loeb *et al.*,
85 2018] and cover the period March 2000 to July 2017. Estimates of T_s come from the European
86 Centre for Medium Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERAi) [Dee *et al.*,
87 2011]. In these calculations, monthly and globally averaged anomalies are used, where
88 anomalies are deviations from the mean annual cycle of the data.

89 Given these data, we calculate λ two ways, both based on Eq. 1. First, we use estimates of
90 effective radiative forcing F over the CERES period and calculate λ as the slope of the regression
91 of R-F vs. T_s . We use standard regressions in this paper — an ordinary least-squares fit, with R-F
92 as the dependent variable and T_s as the independent variable [Murphy *et al.*, 2009]. The
93 forcing is based on the IPCC AR5 forcing time series, revised and extended in the following

94 ways. Forcing from CO₂, N₂O and CH₄ have been replaced by calculating new forcing timeseries
95 using concentrations from NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/) with updated
96 formula to convert mixing ratios to forcing [Etminan *et al.*, 2016]. Other forcing components
97 match IPCC AR5 through 2011 and have been extended to July 2017. For aerosols and ozone,
98 the multi-model mean forcing from Myhre *et al.* [2017a] is used. For volcanoes, the forcing
99 from Andersson *et al.* [2015] is taken from their Figure 4, beginning in 2008. Solar forcing after
100 2011 is derived from SORCE data [Lean *et al.*, 2005]. Other minor forcing terms are estimated
101 using the relative change in forcing from 2011-2017 from the RCP4.5 scenario [Meinshausen *et*
102 *al.*, 2011].

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

110 We then estimate a distribution of λ using Monte Carlo sampling. We start by subtracting the
111 10,000 forcing time series from the observed R time series to generate 10,000 estimates of R-F.
112 Then we repeat the following process 500,000 times: 1) randomly select an R-F time series, 2)
113 randomly subsample it and the observed T_s time series, with replacement, 3) regress the
114 sampled R-F and T_s data sets to obtain an estimate of λ . The number of samples taken is set by
115 the number of independent pieces of information in the time series, as estimated by Eq. 6 of
116 Santer *et al.* [2000] (the original data set contains 209 months; we estimate there are ~100-120
117 independent samples due to autocorrelation in the time series).

118 In the second approach, we assume forcing changes linearly over the CERES time period and
119 account for it by detrending R and T_s time series. We do this by subtracting off the linear trend
120 of each time series estimated using a least-squares regression. We then assume that
121 $R_{\text{detrended}} = \lambda T_{s,\text{detrended}}$ and we calculate λ by regression. The distribution of λ is estimated by

122 randomly sampling 500,000 times (with replacement) the detrended R and T_s time series, with
123 each resampled data set providing one estimate λ . As with the previous estimate, we account
124 for autocorrelation by reducing the number of samples taken, using Eq. 6 of Santer et al.
125 [2000]. Plots of R, T_s , and F can be found in Section S1 of the supplement.

126 Distributions of λ for the two approaches are both quite wide (Fig. 1a), with values of
127 -0.51 ± 0.64 and -0.81 ± 0.65 W/m²/K for the R-F and detrended calculations, respectively
128 (uncertainties are 5-95% confidence intervals). The two estimates of λ reflect different ways of
129 handling forcing and they show that different approaches yield similar distributions for λ . These
130 distributions are similar to those estimated as the uncertainty of ordinary least-squares
131 regressions of R-F vs. T_s (-0.52 ± 0.56 W/m²) and detrended R vs. detrended T_s (-0.82 ± 0.64
132 W/m²). Our sign convention is that fluxes are downward positive, so a negative λ means that a
133 warmer planet radiates more energy to space, a necessary requirement for a stable climate.

134 The extreme width of the λ distributions is a consequence of scatter in the relationship
135 between R-F and T_s (Fig. 1b) [Spencer and Braswell, 2010; Xie *et al.*, 2016], which is due to both
136 weak coupling between the surface and ΔR [Dessler *et al.*, 2018] and weather noise. This
137 means that our observational estimate of λ is quite uncertain, with almost all of the uncertainty
138 coming from month-to-month variability in the R time series. Switching to another
139 temperature data set, such as MERRA2 [Gelaro *et al.*, 2017], or using only the median forcing,
140 yields very similar distributions. Systematic errors in the CERES time series are small; the data
141 are stable to better than 0.5 W/m²/decade (stability of the shortwave is 0.3 W/m²/decade
142 [Loeb *et al.*, 2007], and longwave is 0.15 W/m²/decade [Susskind *et al.*, 2012]). Because we are
143 regressing R vs. temperature, spurious trends in the data have little impact on our analysis
144 [Dessler, 2010].

145 The distributions of λ plotted in Fig. 1a are derived mainly from the response to interannual
146 variability (Fig. S3), so we will refer to them hereafter as λ_{iv} . The λ in Eq. 2, however, is the
147 climate system's response to forcing from doubled CO₂ (hereafter $\lambda_{2 \times CO_2}$), so we cannot simply
148 plug λ_{iv} into Eq. 2 to derive ECS. In fact, this disconnect between what we can measure (λ_{iv})

149 and what is required to calculate ECS ($\lambda_{2\times\text{CO}_2}$) is one reason scientists have largely avoided using
150 interannual variability to infer ECS.

151 We therefore modify Eq. 2 to account for this:

$$152 \quad \text{ECS} = - \frac{F_{2\times\text{CO}_2}}{\lambda_{iv,obs}} \frac{\lambda_{iv}}{\lambda_{2\times\text{CO}_2}} \quad (3)$$

153 where $\lambda_{iv,obs}$ is the observed value (from Fig. 1a), mainly the response to interannual variability,
154 while the ratio $\lambda_{iv}/\lambda_{2\times\text{CO}_2}$ is a transfer function that converts $\lambda_{iv,obs}$ into the required value $\lambda_{2\times\text{CO}_2}$.
155 We estimate this transfer function using models that submitted required output to the 5th
156 phase of the Coupled Model Intercomparison Project (CMIP5) [Taylor *et al.*, 2012]. The
157 numerator λ_{iv} is derived from the models' control runs, in which climate variations arise
158 naturally from internal variability. To facilitate comparison with the observations, as well as
159 avoid any issues with long-term drift, we first break each control run into 16-year segments and
160 calculate monthly anomalies of ΔR and ΔT_s during each segment, where anomalies are
161 deviations from the average annual cycle of each 16-year period. We expect these model
162 segments to contain the same types of climate variations that are in the observations (e.g.,
163 weather noise, ENSO). Then, we calculate λ_{iv} for each segment as the slope of the regression of
164 ΔR vs. ΔT_s for that segment. Finally, we average the segments' values of λ_{iv} to come up with a
165 single value of λ_{iv} for each model (Table S1).

166 The CMIP5 archive does not include doubled CO_2 runs, but it does have abrupt $4\times\text{CO}_2$ runs from
167 which we can estimate $\lambda_{4\times\text{CO}_2}$. $\lambda_{4\times\text{CO}_2}$ is calculated from these runs using the Gregory *et al.*
168 [Gregory *et al.*, 2004] method: we regress all 150 years of annual R vs. annual average T_s , and
169 take the resulting slope as an estimate of $\lambda_{4\times\text{CO}_2}$, where R and T_s are deviations from the pre-
170 industrial control run.

171 If we assume that $\lambda_{2\times\text{CO}_2} \approx \lambda_{4\times\text{CO}_2}$, so we can re-write Eq. 3 as:

$$172 \quad \text{ECS} \approx - \frac{F_{2\times\text{CO}_2}}{\lambda_{iv,obs}} \frac{\lambda_{iv}}{\lambda_{4\times\text{CO}_2}} \quad (4)$$

173 Recent work suggests that λ_{4xCO_2} is less negative (i.e., implying a higher ECS) than λ_{2xCO_2}
174 [Armour, 2017; Proistosescu and Huybers, 2017]. On the other hand, we use all 150 years of the
175 4xCO₂ runs to estimate λ_{4xCO_2} , which tends to produce values that are too negative [Andrews *et al.*,
176 2015; Rugenstein *et al.*, 2016; Rose and Rayborn, 2016; Armour, 2017]. These two errors
177 tend to cancel, but how much of a bias is left — and in which direction — remains an
178 uncertainty in this analysis. The CMIP5 ensemble’s distribution of $\lambda_{iv}/\lambda_{4xCO_2}$ is plotted in Fig. 2;
179 it has an average of 0.81 and a standard deviation of 0.34.

180 We then use a Monte Carlo approach to estimate ECS. We produce 500,000 estimates of ECS
181 by randomly sampling the distributions of F_{2xCO_2} , $\lambda_{iv,obs}$ (Fig. 1a), and $\lambda_{iv}/\lambda_{4xCO_2}$ (Fig. 2) and
182 plugging them into Eq. 3; negative ECS values or values greater than 10 K are viewed as
183 physically implausible and thrown out (sensitivity to the 10-K threshold is shown in Table 1). We
184 produce two ECS distributions — one using $\lambda_{iv,obs}$ from the R-F calculation and one using $\lambda_{iv,obs}$
185 from the detrended calculation. The ECS distributions (Fig. 3) have 17-83% confidence intervals
186 (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
187 calculations, respectively. The modes are 3.0 and 2.4 K, while the medians are 4.2 and 3.3 K.

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

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

198 Modified energy-balance framework

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

$$201 \quad R = F + \Theta T_A \quad (5)$$

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

206 In this framework, the equilibrium warming of the tropical atmosphere ΔT_A in response to
 207 doubled CO_2 is equal to $-F_{2\times\text{CO}_2}/\Theta_{2\times\text{CO}_2}$. ECS can therefore be written

$$208 \quad \text{ECS} = -\frac{F_{2\times\text{CO}_2}}{\Theta_{iv,obs}} \frac{\Theta_{iv}}{\Theta_{2\times\text{CO}_2}} \frac{\Delta T_S}{\Delta T_A} \approx -\frac{F_{2\times\text{CO}_2}}{\Theta_{iv,obs}} \frac{\Theta_{iv}}{\Theta_{4\times\text{CO}_2}} \frac{\Delta T_S}{\Delta T_A} \quad (6)$$

209 where $\Theta_{iv,obs}$ is the analog to $\lambda_{iv,obs}$, $\Theta_{iv}/\Theta_{2\times\text{CO}_2}$ is the transfer function that allows us to use
 210 short-term variability to estimate ECS, and $\Delta T_S/\Delta T_A$ is the ratio of the temperature changes at
 211 equilibrium in response to doubled CO_2 . As we did previously, we will further assume that
 212 $\Theta_{4\times\text{CO}_2} \approx \Theta_{2\times\text{CO}_2}$.

213 We use the same forcing $F_{2\times\text{CO}_2}$ that was used in the previous section. The distributions of the
 214 scaling factor $\Theta_{iv}/\Theta_{4\times\text{CO}_2}$ (Fig. 4a) come from the CMIP5 ensemble. These are calculated the
 215 same way as the $\lambda_{iv}/\lambda_{4\times\text{CO}_2}$ ratios were, except atmospheric temperatures are substituted for
 216 surface temperatures. Just as we did for $\lambda_{iv,obs}$, we calculate $\Theta_{iv,obs}$ two ways: by regressing $R-F$
 217 vs. T_A and by regressing detrended R vs. detrended T_A . Distributions of $\Theta_{iv,obs}$ for the two
 218 approaches are similar (Fig. 1a), with values of -0.98 ± 0.32 and -1.09 ± 0.29 $\text{W/m}^2/\text{K}$ for the $R-F$
 219 and detrended calculations, respectively (uncertainties are 5-95% confidence intervals).
 220 Because of their similarities, in the rest of this section we will show results using the detrended
 221 calculation, although results for both distributions can be found in Table 2.

222 Finally, the distribution of the temperature ratio $\Delta T_S/\Delta T_A$ is also estimated from the CMIP5
 223 ensemble. For each model, ΔT_S and ΔT_A are estimated as the average difference of the first and

224 last decades of the abrupt 4xCO₂ runs; we then take the ratio of these values. Comparisons of
225 the models to observations show that models do well at simulating this ratio (Sect. S3). The
226 resulting distribution of $\Delta T_S/\Delta T_A$ constructed by the CMIP5 models (Fig. 5a) has an ensemble
227 average and standard deviation of 0.86 ± 0.10 .

228 Long forced runs of the MPI-ESM1.1, GFDL CM3, and ESM2M models all show this ratio
229 increases as the climate continues to warm beyond year 150. In runs of the GFDL CM3 and
230 ESM2M, in which CO₂ increases at 1% per year until doubling and then remains fixed, the ratio
231 increases from 0.79 and 0.70, 300 years after CO₂ doubles, to 0.86 and 0.76 at equilibrium
232 (GFDL values are personal communication, David Paynter, 2018, based on runs described in
233 [Paynter *et al.*, 2018]). The ratio in an abrupt 4xCO₂ run of the MPI model increases from 0.79
234 in years 140-150 to 0.87 in years 2400-2500. Thus, we conclude that values of this ratio
235 obtained from the 150-year CMIP5 4xCO₂ simulations may be low biased, which would lead our
236 ECS to also be low biased.

237 As in the previous section, we use a Monte Carlo approach and produce 500,000 estimates of
238 ECS by randomly sampling the distributions of F_{2xCO_2} , $\Theta_{iv,obs}$, $\Theta_{iv}/\Theta_{4xCO_2}$, and $\Delta T_S/\Delta T_A$, and then
239 plugging the values into Eq. 6. The resulting ECS distribution (Fig. 6a) shows a similar structure
240 to the λ -based distributions in Fig. 3: a broad maximum between 2 and 3 K and a tail towards
241 higher ECS values.

242 There is also a puzzling peak below 1°C. The only way for an ECS estimate to be close to zero is
243 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
244 Eq. 6 suggests that the term causing the low ECS values is $\Theta_{iv}/\Theta_{4xCO_2}$, whose distribution
245 approaches zero (Fig. 4a). These low values come from the GISS models (Fig. 7a, Table S1) and if
246 they are removed from the ensemble, the bump below 1 K disappears (Fig. 6b), although the
247 statistics of the distribution do not change much.

248 This result emphasizes that the scaling factor $\Theta_{iv}/\Theta_{4xCO_2}$ is unconstrained by observations and
249 has not been previously studied. That doesn't mean, however, that we know nothing about it
250 — we do have observations of Θ_{iv} and can compare those to each model's value of Θ_{iv} . We find

251 that 15 of the 25 CMIP5 models produce estimates of Θ_{iv} in agreement with the CERES
252 observations (Fig. 7b). If we construct distributions of $\Theta_{iv}/\Theta_{4\times CO_2}$ and $\Delta T_S/\Delta T_A$ from just those
253 models (Figs. 4b and 5b), we obtain the ECS distribution in Fig. 6c (hereafter referred to as the
254 “good- Θ ” distribution).

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

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

264 We have also performed corresponding “good- λ ” ECS calculations in which the $\lambda_{iv}/\lambda_{4\times CO_2}$
265 distribution in Eq. 4 is constructed using only those models whose λ_{iv} agrees with $\lambda_{iv,obs}$. The
266 ECS distributions obtained from these calculations (Table 1) are similar to distributions from the
267 λ calculations using all models.

268 **Discussion**

269 There are several reasons why ECS estimated from the revised energy balance framework (Eq.
270 6) should be considered more reliable than that estimated from the traditional framework (Eq.
271 4) used in previous papers [e.g., Forster, 2016; Tsushima *et al.*, 2005; Forster and Gregory,
272 2006; Chung *et al.*, 2010; Tsushima and Manabe, 2013; Dessler, 2013; Donohoe *et al.*, 2014].
273 Fig. 1 shows the main advantage — that $\Theta_{iv,obs}$ is better constrained than $\lambda_{iv,obs}$. This is what
274 leads to the narrower distributions of ECS in Fig. 6 than in Fig. 3. Of particular note, the $\lambda_{iv,obs}$
275 distributions have non-zero probabilities of values close to zero; since ECS is proportional to
276 $1/\lambda_{iv,obs}$, this generates a large tail towards unrealistically large ECS values.

277 There are additional reasons that lead us to conclude that the estimates from the revised
278 framework are superior. It has been suggested that $\lambda_{iv,obs}$ exhibits significant decadal variability
279 in models [Andrews *et al.*, 2015; Gregory and Andrews, 2016; Zhou *et al.*, 2016; Dessler *et al.*,
280 2018]. This opens the possibility that the observed $\lambda_{iv,obs}$, based on 16 years of data, is biased
281 with respect to the long-term average; if so, then ECS estimated from these observations would
282 also be biased. Model simulations suggest that $\Theta_{iv,obs}$ exhibits smaller decadal variability
283 [Dessler *et al.*, 2018], making Θ_{iv} estimated from CERES data a more robust estimate of the
284 climate system's actual long-term value. There is also evidence that Θ changes less than λ
285 during transient climate change [Dessler *et al.*, 2018], making the assumption that $\Theta_{2xCO_2} \approx$
286 Θ_{4xCO_2} a far better one than the assumption that $\lambda_{4xCO_2} \approx \lambda_{2xCO_2}$.

287 It is also worth stepping back and asking what could cause our calculation to be seriously in
288 error. It seems unlikely that forcing from doubled CO₂ is wrong given our good understanding
289 of the physics of CO₂ forcing [e.g., Feldman *et al.*, 2015]. Estimates of $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$ are
290 derived from observations we view to be reliable, so our judgment is that they are also unlikely
291 to be significantly wrong. The $\Delta T_s/\Delta T_A$ factor comes from climate model simulations, but
292 models have long been able to accurately reproduce the observed pattern of surface warming
293 [e.g., Stouffer and Manabe, 2017], and we have simple physical arguments explaining how the
294 atmospheric and surface temperature should be connected [Xu and Emanuel, 1989]. Finally,
295 we can compare the models to data [Compo *et al.*, 2011; Poli *et al.*, 2016] to validate their
296 simulation of this ratio (Sect. S3).

297 Thus, the transfer function $\Theta_{iv}/\Theta_{4xCO_2}$ seems the most probable place for a significant error to
298 occur. That said, there are reasons to believe the models' estimates of this ratio. As mentioned
299 above, we can directly compare Θ_{iv} in the models to observations, and find agreement in the
300 majority of models (Fig. 7). We also argue that while errors may exist in a model (i.e., in the
301 cloud feedback), this will affect both the numerator and denominator and such errors will tend
302 to cancel out. As a preliminary test of this, we have analyzed three different versions of the
303 MPI-ESM 1.2 model that have had their cloud feedbacks modified to produce different ECS
304 [Thorsten Mauritsen and Diego Jimenez, personal communication, 2018]. The three versions

305 are the standard model (ECS calculated from an abrupt 4xCO₂ run using the Gregory method =
306 3.0 K), an “iris” version [described in Mauritsen and Stevens, 2015] (ECS = 2.6 K), and a “high
307 ECS” version, in which the convective parameterization has been tweaked to generate a large,
308 positive cloud feedback (ECS = 5.2 K). Despite large differences in the ECS, these three versions
309 have similar values of $\lambda_{iv}/\lambda_{4xCO_2}$ of 1.17, 1.15, and 1.11 for the standard, iris, and high ECS
310 versions, respectively. The corresponding values of $\Theta_{iv}/\Theta_{4xCO_2}$ are 1.06, 0.96, and 1.10. While
311 one must be careful about conclusions based on a single model, this nevertheless provides
312 some support for the hypothesis that errors in Θ_{4xCO_2} will cancel errors in Θ_{iv} when the ratio is
313 taken and that the ratio $\Theta_{4xCO_2}/\Theta_{iv}$ may well be more accurate than either Θ_{4xCO_2} or Θ_{iv} are
314 individually.

315 We have also constructed an error budget to determine which term contributes most to the
316 width of the distributions in Fig. 6. We do this by sequentially setting each term to have zero
317 uncertainty by replacing that term’s distribution in the Monte Carlo calculation with a single
318 number, the ensemble average. This has little effect on the mean, median, or mode, but does
319 change the width of the distribution (Table 3). By comparing the widths of the resulting
320 distributions (defined as the distance between the 17th and 83rd percentiles), Fig. 8 shows that
321 the biggest contributor to ECS uncertainty is the uncertainty in $\Theta_{iv}/\Theta_{4xCO_2}$. Eliminating the
322 uncertainty in that reduces the 17-83% confidence interval to 2.8-4.0 K. *Thus, developing a*
323 *theoretical argument for the value of this ratio would be a key advance in climate science.* The
324 next most important uncertainty is uncertainty in $\Theta_{iv,obs}$, followed by the uncertainty in $\Delta T_S/\Delta T_A$
325 and then the uncertainty in F_{2xCO_2} .

326 **Conclusions**

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

332 In this paper, we have used observations of interannual climate variations covering the period
333 2000 to 2017 along with a model-derived relationship between interannual variations and
334 forced climate change to estimate ECS. We interpret the observations using a modified energy
335 balance framework (Eq. 5) in which the response of TOA flux is proportional to the atmospheric
336 temperature. We conclude ECS is *likely* 2.4-4.6 K (17-83% confidence interval), with a mode and
337 median value of 2.9 and 3.3 K, respectively. Overall, our analysis suggests that the upper end of
338 the IPCC's range is set about right, but this analysis provides little evidence to support estimates
339 of ECS in the bottom third of the IPCC's *likely* range.

340 One of the key parts of our calculations is the use of CMIP5 climate models to convert the
341 observations of interannual variability into an estimate of the response of the system to
342 doubled CO₂. This is the main uncertainty in our analysis and future efforts to pin this transfer
343 function down would be extremely valuable.

344 **References**

- 345 Aldrin, M., M. Holden, P. Guttorp, R. B. Skeie, G. Myhre, & T. K. Berntsen (2012), Bayesian
346 estimation of climate sensitivity based on a simple climate model fitted to observations of
347 hemispheric temperatures and global ocean heat content, *Environmetrics*, 23, 253-271, doi:
348 10.1002/env.2140.
- 349 Andersson, S. M., B. G. Martinsson, J.-P. Vernier, J. Friberg, C. A. M. Brenninkmeijer, M.
350 Hermann, et al. (2015), Significant radiative impact of volcanic aerosol in the lowermost
351 stratosphere, *Nature Communications*, 6, 7692, doi: 10.1038/ncomms8692.
- 352 Andrews, T., & M. J. Webb (2018), The dependence of global cloud and lapse rate feedbacks on
353 the spatial structure of tropical Pacific warming, *J. Climate*, 31, 641-654, doi: 10.1175/jcli-
354 d-17-0087.1.
- 355 Andrews, T., J. M. Gregory, & M. J. Webb (2015), The dependence of radiative forcing and
356 feedback on evolving patterns of surface temperature change in climate models, *J. Climate*,
357 28, 1630-1648, doi: 10.1175/JCLI-D-14-00545.1.
- 358 Annan, J. D., & J. C. Hargreaves (2006), Using multiple observationally-based constraints to
359 estimate climate sensitivity, *Geophys. Res. Lett.*, 33, doi: 10.1029/2005gl025259.
- 360 Armour, K. C. (2017), Energy budget constraints on climate sensitivity in light of inconstant
361 climate feedbacks, *Nature Clim. Change*, 7, 331-335, doi: 10.1038/nclimate3278.
- 362 Chung, E. S., B. J. Soden, & B. J. Sohn (2010), Revisiting the determination of climate
363 sensitivity from relationships between surface temperature and radiative fluxes, *Geophys.*
364 *Res. Lett.*, 37, doi: 10.1029/2010gl043051.

365 Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, et al. (2013),
366 Long-term climate change: Projections, commitments and irreversibility, in *Climate Change*
367 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment
368 Report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin,
369 G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M.
370 Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY,
371 USA.

372 Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, et al. (2011),
373 The Twentieth Century Reanalysis Project, *Q. J. R. Meteor. Soc.*, 137, 1-28, doi:
374 10.1002/qj.776.

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

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

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

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

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

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

391 Feldman, D. R., W. D. Collins, P. J. Gero, M. S. Torn, E. J. Mlawer, & T. R. Shippert (2015),
392 Observational determination of surface radiative forcing by CO₂ from 2000 to 2010, *Nature*,
393 519, 339, doi: 10.1038/nature14240.

394 Forster, P. M. (2016), Inference of climate sensitivity from analysis of Earth's energy budget,
395 *Annual Review of Earth and Planetary Sciences*, 44, 85-106, doi: 10.1146/annurev-earth-
396 060614-105156.

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

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

402 Gregory, J. M., & T. Andrews (2016), Variation in climate sensitivity and feedback parameters
403 during the historical period, *Geophys. Res. Lett.*, 43, 3911-3920, doi:
404 10.1002/2016GL068406.

405 Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, & N. A. Rayner (2002), An
406 observationally based estimate of the climate sensitivity, *J. Climate*, 15, 3117-3121, doi:
407 10.1175/1520-0442(2002)015<3117:aobeot>2.0.co;2.

408 Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, et al. (2004),
409 A new method for diagnosing radiative forcing and climate sensitivity, *Geophys. Res. Lett.*,
410 31, doi: 10.1029/2003gl018747.

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

413 Lean, J., G. Rottman, J. Harder, & G. Kopp (2005), *SORCE contributions to new understanding*
414 *of global change and solar variability*, *Solar Physics*, 230, 27-53, doi: 10.1007/s11207-005-
415 1527-2.

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

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

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

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

429 Mauritsen, T., & B. Stevens (2015), Missing iris effect as a possible cause of muted hydrological
430 change and high climate sensitivity in models, *Nature Geosci*, 8, 346-351, doi:
431 10.1038/ngeo2414.

432 Meinshausen, M., S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, J.-F. Lamarque, et al.
433 (2011), The RCP greenhouse gas concentrations and their extensions from 1765 to 2300,
434 *Climatic Change*, 109, 213, doi: 10.1007/s10584-011-0156-z.

435 Murphy, D. M., S. Solomon, R. W. Portmann, K. H. Rosenlof, P. M. Forster, & T. Wong (2009),
436 An observationally based energy balance for the Earth since 1950, *J. Geophys. Res.*, 114,
437 14, doi: 10.1029/2009jd012105.

438 Myhre, G., W. Aas, R. Cherian, W. Collins, G. Faluvegi, M. Flanner, et al. (2017a), Multi-model
439 simulations of aerosol and ozone radiative forcing due to anthropogenic emission changes
440 during the period 1990–2015, *Atmos. Chem. Phys.*, 17, 2709-2720, doi: 10.5194/acp-17-
441 2709-2017.

442 Myhre, G., P. M. Forster, B. H. Samset, Ø. Hodnebrog, J. Sillmann, S. G. Aalbergsjø, et al.
443 (2017b), PDRMIP: A Precipitation Driver and Response Model Intercomparison Project—

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

446 Otto, A., F. E. L. Otto, O. Boucher, J. Church, G. Hegerl, P. M. Forster, et al. (2013), Energy
447 budget constraints on climate response, *Nature Geoscience*, 6, 415-416, doi:
448 10.1038/ngeo1836.

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

452 Poli, P., H. Hersbach, D. P. Dee, P. Berrisford, A. J. Simmons, F. Vitart, et al. (2016), ERA-20C:
453 An Atmospheric Reanalysis of the Twentieth Century, *J. Climate*, 29, 4083-4097, doi:
454 10.1175/jcli-d-15-0556.1.

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

457 Richardson, M., K. Cowtan, E. Hawkins, & M. B. Stolpe (2016), Reconciled climate response
458 estimates from climate models and the energy budget of Earth, *Nature Clim. Change*, 6,
459 931-935, doi: 10.1038/nclimate3066.

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

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

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

468 Sherwood, S. C., S. Bony, O. Boucher, C. Bretherton, P. M. Forster, J. M. Gregory, & B. Stevens
469 (2014), Adjustments in the forcing-feedback framework for understanding climate change,
470 *Bull. Am. Met. Soc.*, 96, 217-228, doi: 10.1175/BAMS-D-13-00167.1.

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

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

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

478 Stevens, B., S. C. Sherwood, S. Bony, & M. J. Webb (2016), Prospects for narrowing bounds on
479 Earth's equilibrium climate sensitivity, *Earth's Future*, 4, 512-522, doi:
480 10.1002/2016EF000376.

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

483 Susskind, J., G. Molnar, L. Iredell, & N. G. Loeb (2012), Interannual variability of outgoing
484 longwave radiation as observed by AIRS and CERES, *J. Geophys. Res.*, 117, doi:
485 10.1029/2012jd017997.

486 Taylor, K. E., R. J. Stouffer, & G. A. Meehl (2012), An overview of CMIP5 and the experiment
487 design, *Bull. Am. Met. Soc.*, 93, 485-498, doi: 10.1175/bams-d-11-00094.1.

488 Tsushima, Y., & S. Manabe (2013), Assessment of radiative feedback in climate models using
489 satellite observations of annual flux variation, *Proc. Natl. Acad. Sci.*, 110, 7568-7573, doi:
490 10.1073/pnas.1216174110.

491 Tsushima, Y., A. Abe-Ouchi, & S. Manabe (2005), Radiative damping of annual variation in
492 global mean surface temperature: comparison between observed and simulated feedback,
493 *Climate Dynamics*, 24, 591-597, doi: 10.1007/s00382-005-0002-y.

494 Xie, S.-P., Y. Kosaka, & Y. M. Okumura (2016), Distinct energy budgets for anthropogenic and
495 natural changes during global warming hiatus, *Nature Geoscience*, 9, 29-33, doi:
496 10.1038/ngeo2581.

497 Xu, K. M., & K. A. Emanuel (1989), Is the tropical atmosphere conditionally unstable?, *Mon.*
498 *Wea. Rev.*, 117, 1471-1479.

499 Zhou, C., M. D. Zelinka, & S. A. Klein (2016), Impact of decadal cloud variations on the Earth's
500 energy budget, *Nature Geosci*, 9, 871-874, doi: 10.1038/ngeo2828.

501

502

503 **Acknowledgments:** A.E.D. acknowledges support from NSF grant AGS-1661861 to Texas
504 A&M University. P.M.F. acknowledges support from the Natural Environment Research
505 Council project NE/P014844/1. We thank Bjorn Stevens and Thorsten Mauritsen for their
506 insight into this analysis. We also acknowledge the modeling groups, the Program for Climate
507 Model Diagnosis and Intercomparison, and the WCRP's Working Group on Coupled Modeling
508 for their roles in making available the WCRP CMIP5 multimodel dataset. CERES data were
509 downloaded from ceres.larc.nasa.gov, CMIP5 data were downloaded from cmip.llnl.gov, and
510 ECMWF reanalysis were downloaded from [www.ecmwf.int/en/forecasts/datasets/reanalysis-](http://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim)
511 [datasets/era-interim](http://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim). Code and data can be found here: <https://zenodo.org/record/1323162>.
512

513
514
515
516
517
518
519

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

520 Names containing “all” or “good” include all models or just the ones whose λ_{iv} agrees with the
521 CERES observations, respectively. The names with “-1” or “-2” use $\lambda_{iv,obs}$ derived using
522 estimates of forcing (the R-F calculations) and the detrended calculations, respectively. The
523 names including “-f” use forcing from the CMIP5 abrupt 4x CO₂ runs (see Sect. S2). The names
524 including “-f_20-150” calculate F_{2xCO_2} and λ_{4xCO_2} from years 20-150 of the abrupt 4xCO₂ runs
525 (see Sect. S2). Names with “-8K” and “-12K” change the plausibility threshold above which
526 ECS values are considered non-physical and are thrown out.

527
528
529

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

530 Names follow the same convention as Table 1. The names including “noGISS-” include all
531 models except the two GISS models. In the “-corr” calculations, each Monte Carlo value of ECS
532 uses values of $\Delta T_S/\Delta T_A$ and $\Theta_{iv}/\Theta_{4xCO_2}$ from the same model.
533
534

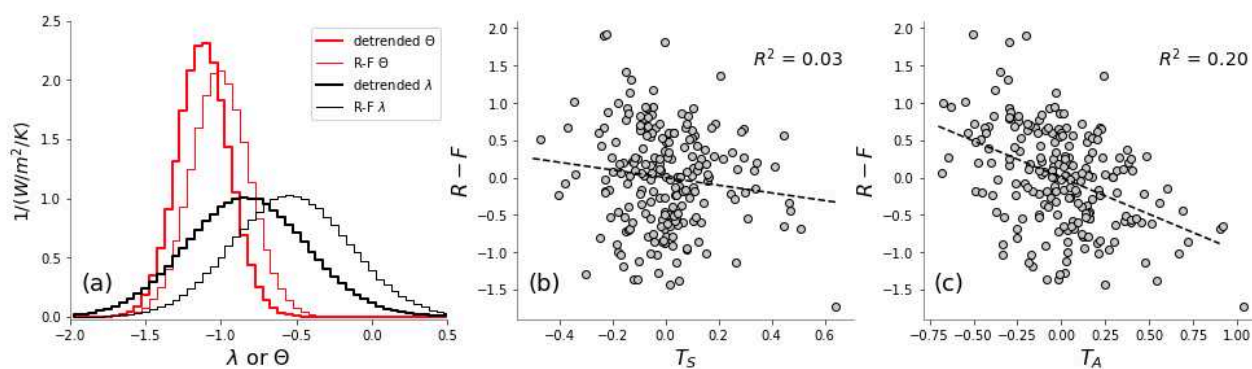
535 Table 3. Error budget calculations
 536 Summary of the statistics of the ECS distribution when one of the input distributions has no
 537 uncertainty.

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

538 Most name conventions Table 1. For these calculations, we take the “all-Theta-2” or “good-
 539 Theta-2” calculation and sequentially set the uncertainty in one term to zero. The “-noF”, “-
 540 noRat”, “-nodtdt”, and “-noTheta” correspond to no uncertainty in F_{2xCO_2} , $\Theta_{iv}/\Theta_{4xCO_2}$, $\Delta T_s/\Delta T_A$,
 541 and Θ_{iv} , respectively.

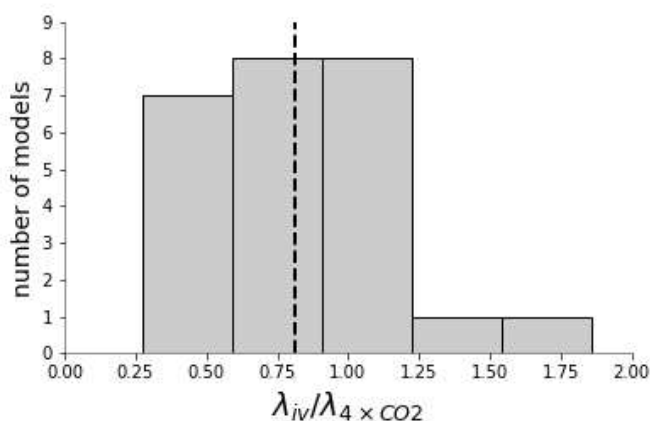
542

543



544
545
546
547
548
549
550
551
552

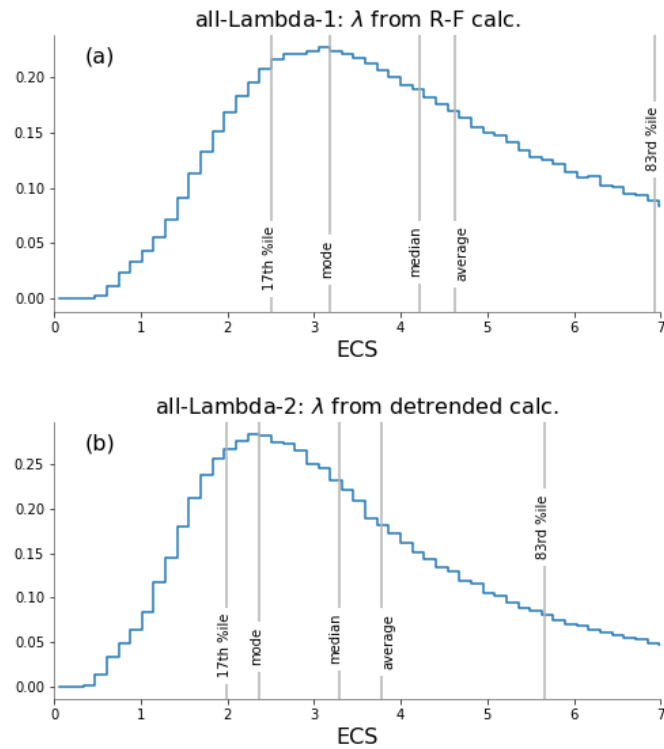
Figure 1. (a) Distribution of $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$ (W/m^2); (b) scatter plot of $R-F$ (W/m^2) vs. T_S (K), the dashed line is a least-squares fit; (c) same as panel (b), but the regression is against T_A (K).



553
554
555
556
557

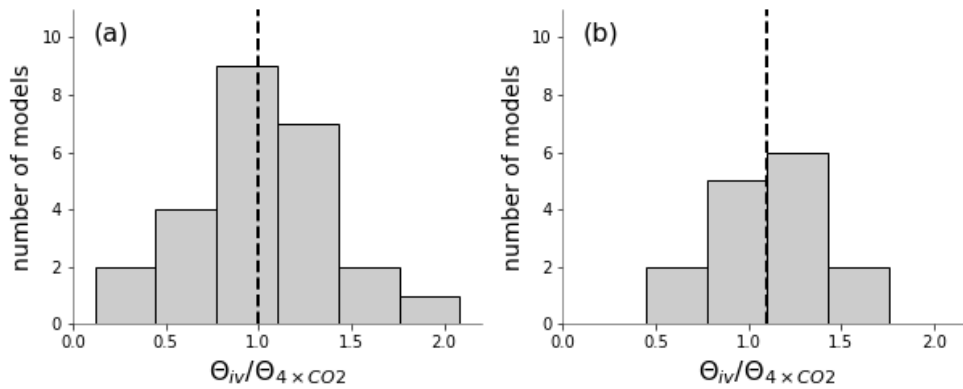
Figure 2. Distribution of $\lambda_{iv}/\lambda_{4 \times CO_2}$ from 25 CMIP5 models; the black dashed line is the mean of the distribution. See methods for description of how the value is calculated in each model.

558
559
560



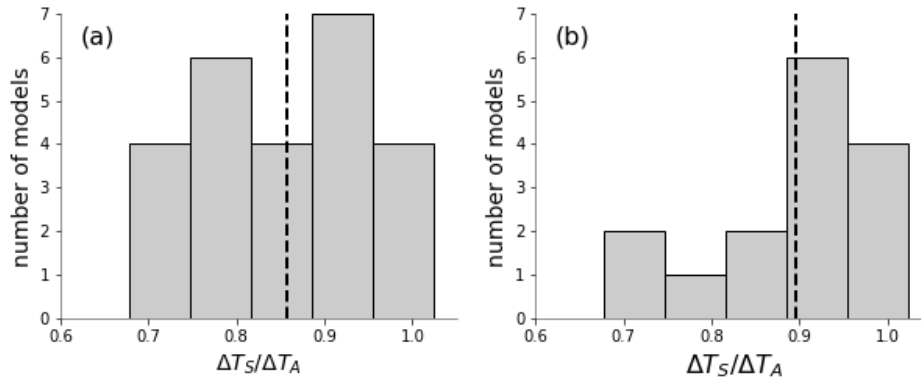
561
562
563
564
565
566

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 17th and 83rd percentile, corresponding to the IPCC’s *likely* range.



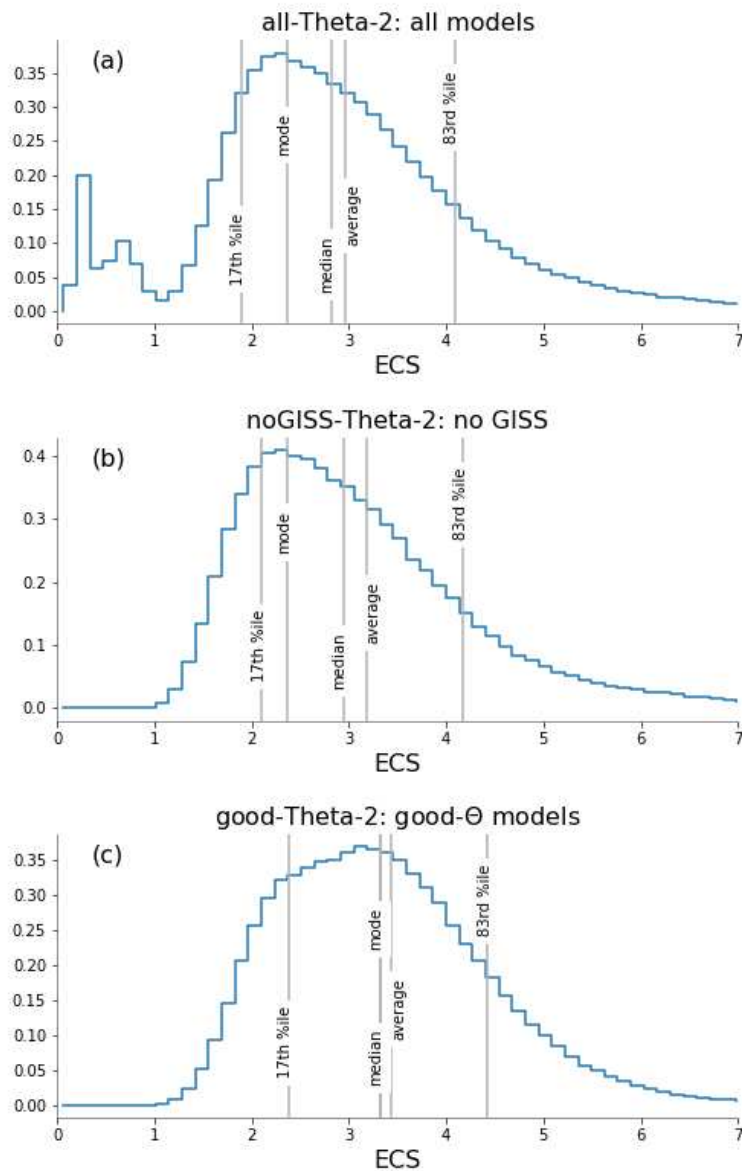
567
568
569
570
571
572

Figure 4. Distribution of $\Theta_{iv}/\Theta_{4xCO_2}$ 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.



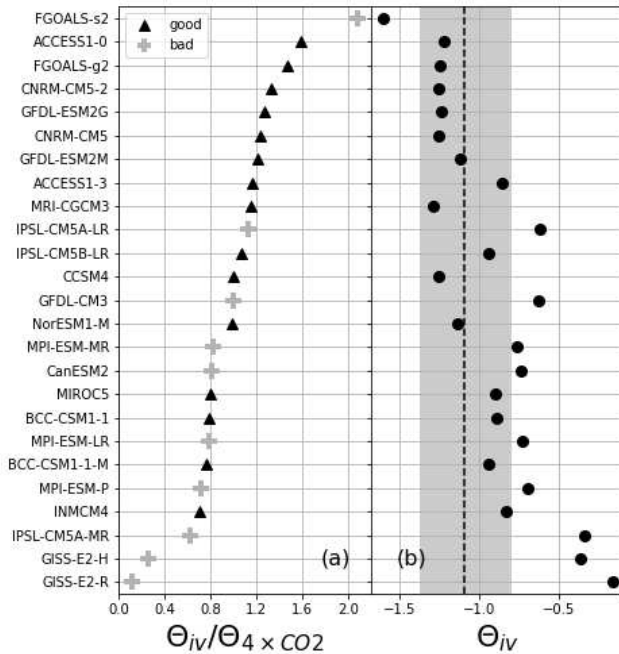
573
574
575
576
577

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.



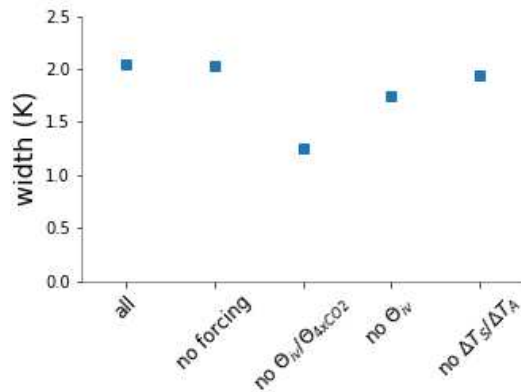
578
 579
 580
 581
 582
 583
 584
 585

Figure 6. Distributions of ECS using the revised energy balance framework (Eq. 6). Panel (a) uses all models for the distributions of $\Theta_{iv}/\Theta_{4xCO_2}$ 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 17th and 83rd percentile, corresponding to the IPCC’s *likely* range.



586
587
588
589
590
591
592

Figure 7. CMIP5 model estimates of (a) $\Theta_{iv}/\Theta_{4xCO_2}$ 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.



593
594
595
596
597
598
599

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, $\Theta_{iv}/\Theta_{4xCO_2}$, $\Theta_{iv,obs}$, 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. Dessler¹, P.M. Forster²

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

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

Contents of this file

Sect. S1: additional plots of data going into the calculations of $\lambda_{iv,obs}$ and $\Theta_{iv,obs}$

Sect. S2: alternate ways to calculate $F_{2\times CO_2}$, $\lambda_{iv,obs}$, and $\Theta_{iv,obs}$

Sect. S3: testing models' ability to estimate $\Delta T_S/\Delta T_A$

Sect. S4: estimating the distribution of $\lambda_{4\times CO_2}$

Table S1: summary statistics of CMIP5 models

Table S2 and S3: summary statistics of $\lambda_{4\times CO_2}$

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 $1.1\text{-}3.1 \text{ W/m}^2$. 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 $0.62 \text{ W/m}^2/\text{K}$ and 5-95% confidence interval of $\pm 0.16 \text{ W/m}^2/\text{K}$.

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.

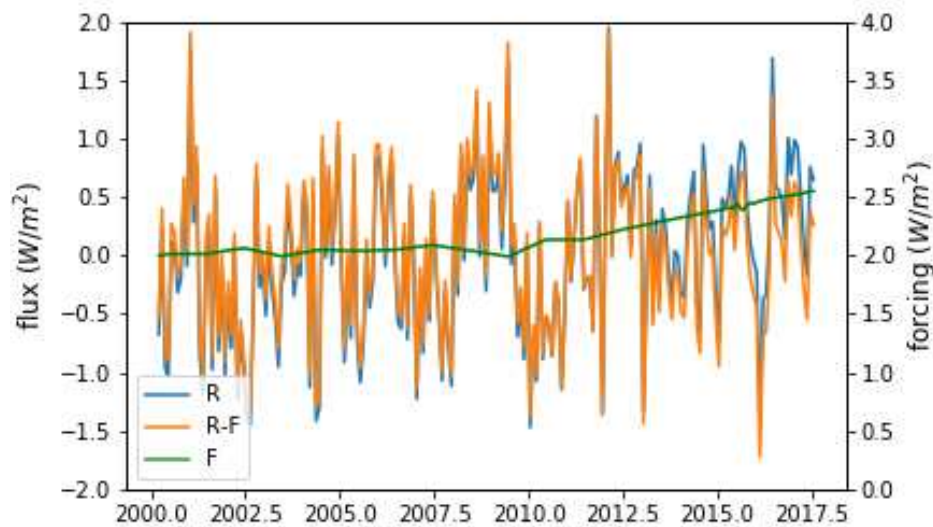


Fig. S1. Time series of global average, monthly anomalies of CERES R (blue), median forcing F (green), and R-F (orange).

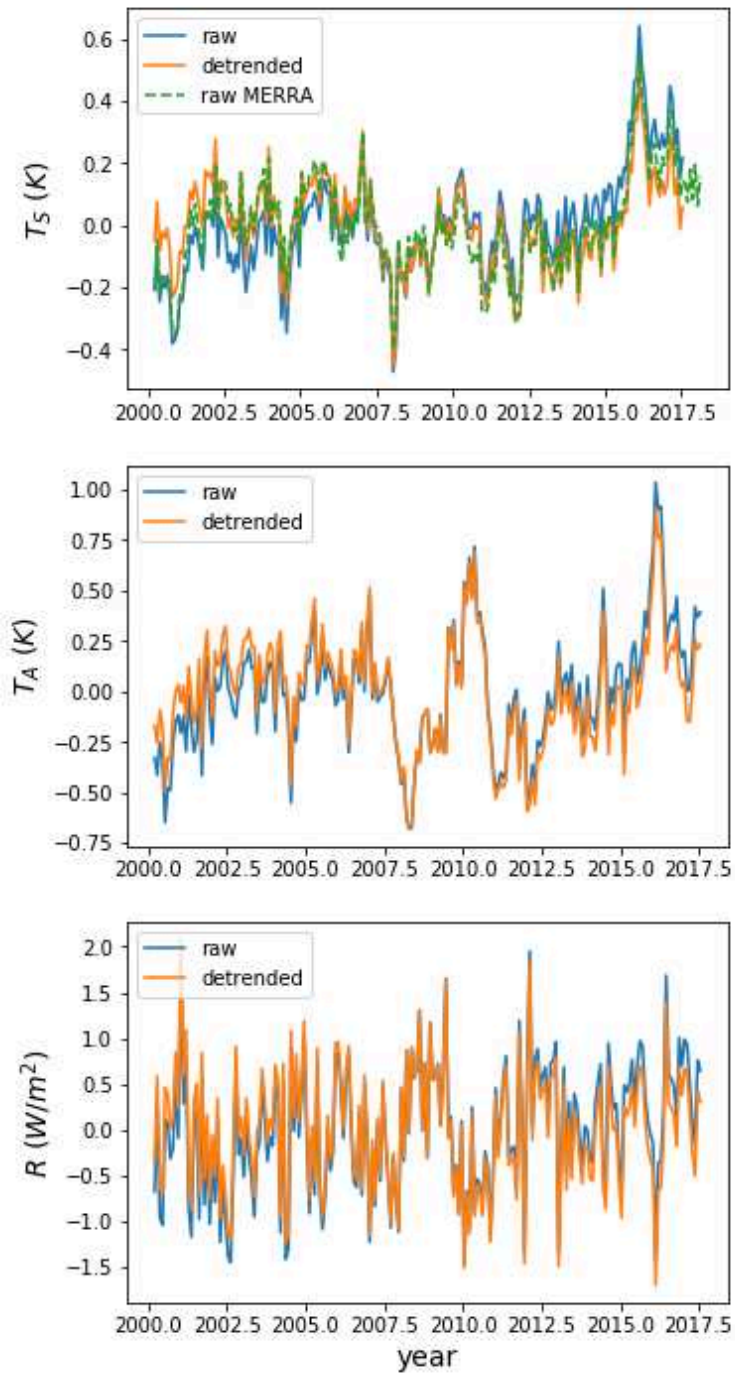


Fig. S2. Time series of global average, monthly anomalies of CERES R and ERAi global average surface temperature and 500-hPa tropical average (30°N-30°S) temperature. The raw time series is before detrending; the detrended time series has the linear trend, estimated using a least-squares fit, removed.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

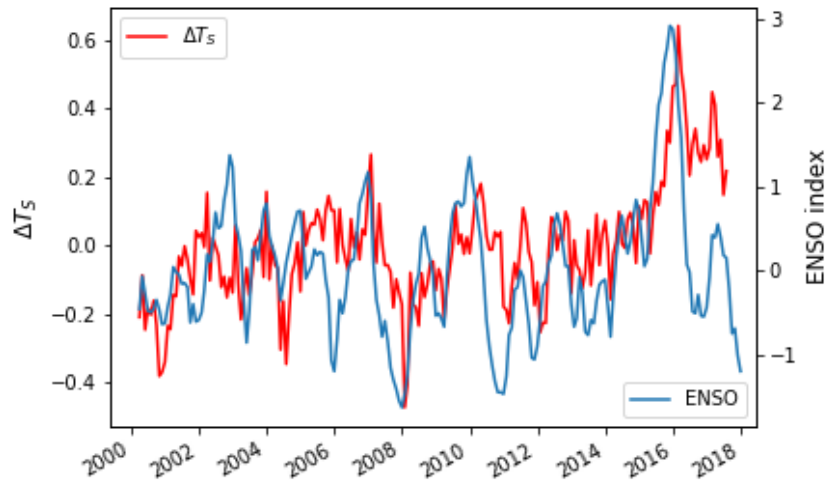


Fig. S3. Time series of global average surface temperature anomaly ΔT_s (K; left-hand axis) and Nino3 ENSO index (right-hand axis). ENSO index downloaded from <https://www.esrl.noaa.gov/psd/data/timeseries/monthly/NINO3/>.

S2. Alternate ways to calculate F_{2xCO_2} and λ_{4xCO_2} and Θ_{4xCO_2}

One potential issue in our calculation is that the forcing we use is from fixed SST runs while the values of λ_{4xCO_2} and Θ_{4xCO_2} come from abrupt $4xCO_2$ runs. To evaluate the impact of any possibly incompatibility, we have also calculated ECS using a distribution of F_{2xCO_2} 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_{4xCO_2} and Θ_{4xCO_2} . It is well known that removal of the first few decades in the Gregory regression produces a less negative λ_{4xCO_2} [e.g., Andrews *et al.*, 2015], which implies a higher ECS. The effect of this on Θ_{4xCO_2} is smaller [Dessler *et al.*, 2018]. To test the impact of this, we produce ECS estimates where λ_{4xCO_2} 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_{2xCO_2} ” 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.

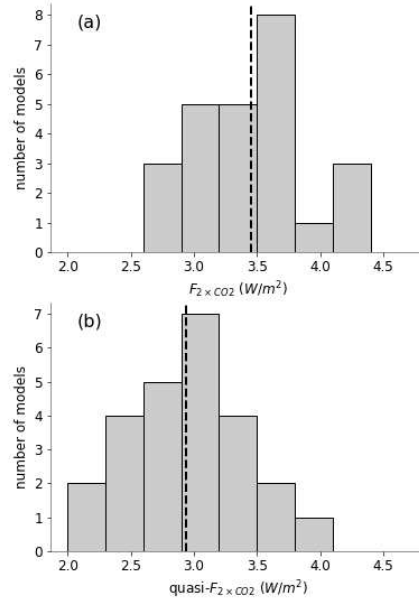


Fig. S4. Distribution of $F_{2 \times CO_2}$ from CMIP5 abrupt 4xCO₂ runs. Panel (a) uses all 150 years of the run, while panel (b) uses years 20-150. The dashed lines are the ensemble averages of 3.45 and 2.94 W/m².

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:

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

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 4xCO₂ 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).

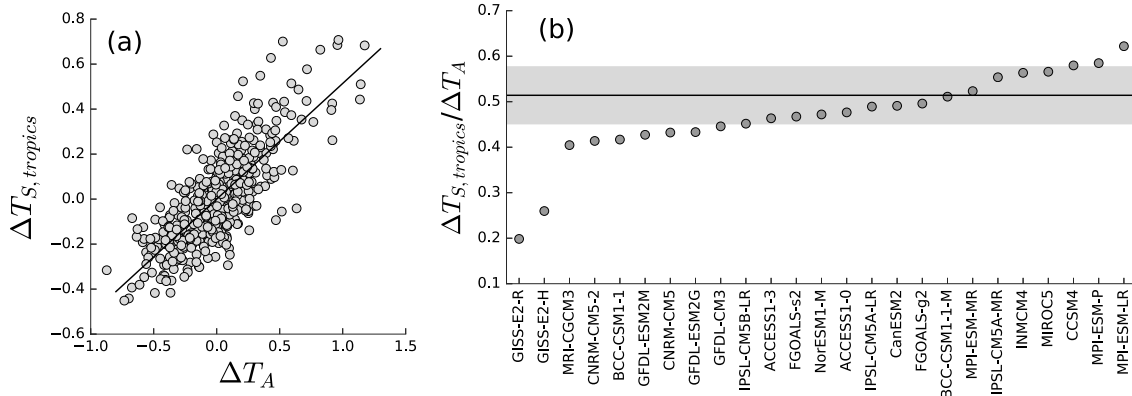


Figure S5. Estimates of $\Delta T_{S,tropics}/\Delta T_A$. (a) Scatter plot of monthly $\Delta T_{S,tropics}$ (K; tropical avg. surface temperature) anomalies vs. ΔT_A (K) anomalies from ERAi reanalysis (1979-2016). The solid line is the best fit line. (b) The slope of the same fit from the last 37 years of the CMIP5 ensemble's abrupt $4\times CO_2$ runs. The black line and gray region shows the slope and uncertainty of the fit to observations in panel a.

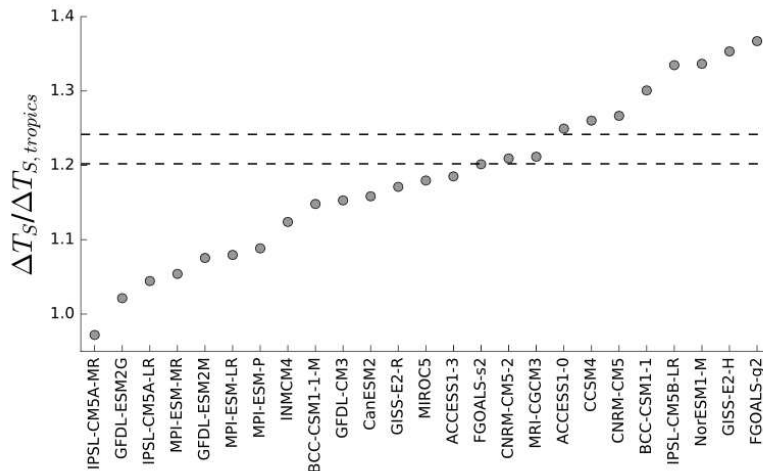


Figure S6. Estimates of polar amplification in the models, $\Delta T_S/\Delta T_{S,tropics}$. For the CMIP5 ensemble, this is calculated by differencing the average of the first and last decades of the CMIP5 ensemble's abrupt $4\times CO_2$ runs. The two dashed lines are observational estimates (see 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 $4\times CO_2$ 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 $\lambda_{4\times CO_2}$

In the main text, we focus on estimating the distributions of ECS. However, we could also produce an observational estimate of the distribution of $\lambda_{4\times CO_2}$. We do this with the following two equations:

$$\lambda_{4\times CO_2} \approx \lambda_{iv,obs} \frac{\lambda_{4\times CO_2}}{\lambda_{iv}} \quad (S2)$$

$$\lambda_{4\times CO_2} \approx \Theta_{iv,obs} \frac{\Theta_{4\times CO_2} \Delta T_A}{\Theta_{iv} \Delta T_S} \quad (S3)$$

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 $\lambda_{iv}/\lambda_{4\times CO_2}$, $\Theta_{iv}/\Theta_{4\times CO_2}$, 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/m²/K (17-83% confidence interval)

Table S1. Values for individual models

Model	λ_{iv}	Θ_{iv}	λ_{4xCO_2}	Θ_{4xCO_2}	$\Delta T_S/\Delta T_A$	F_{2xCO_2}
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_{2xCO_2} is derived from that model's abrupt $4xCO_2$ run and has units of W/m^2 .

Table S2. $\lambda_{4\times\text{CO}_2}$ 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 $\text{W}/\text{m}^2/\text{K}$.

Table S3. $\lambda_{4\times\text{CO}_2}$ 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 $\text{W}/\text{m}^2/\text{K}$.