Note that all line numbers in our responses refer to the version with tracked comments included in this document. Line numbers from the reviewers refer to the original manuscript.

Reviewer #1

We thank the reviewer for their comments. Below, we detail our responses.

64-65. As a single number to quantify the spread, the standard deviation would also be helpful.

We have added 5-95% confidence intervals throughout the paper.

66. Why do you use only a single decade, rather than all the data, for instance by dividing the dataset into two or using regression (cf Barnes and Barnes, 2015, 10.1175/JCLI-D-15-0032.1)? A single decade would be less precise. You could estimate the statistical uncertainty incurred from the control run.

We calculate ECS using this approach because this is the way most ECS calculations based on the 20th-century observational record are done. Thus, our results can therefore directly provide insight into the impact of variability in the observational estimates of ECS.

The reviewer is correct that using more than a decade might affect the results. If one used the difference between the averages of the first and last 20 years, the range in λ declines from 0.87 $W/m^2/K$ to 0.48 $W/m^2/K$. Using longer averaging periods does not further decrease the range. We now mention this in the paper (line 67).

118. It would be useful to remark here that 16 years is chosen to match the CERES dataset, because that was mentioned some lines above (103-104), where it appears actually to be 17 years and 5 months long.

We have added a statement that the segmentation of the data is done to match the CERES record (line 134). We have also updated the paper to segment the data into 17-year segments to more closely match CERES.

119, 196. Why are monthly anomalies used here, rather than annual? Does it make a difference?

We do this to facilitate the comparison with the CERES regressions, which also uses monthly data. The reason most analyses with CERES data are done with monthly data is because using annual data means there's only 17 data points, and the uncertainties end up being very large. Issues involved in annual vs. monthly regressions are discussed in some detail in Forster (2016, 10.1146/annurev-earth-060614-105156).

167. Again, the standard deviation would be helpful, and could be compared with lines 64-65.

Added.

173, 175. You could give standard errors of the mean for each of these two numbers, and judge the significance of their difference.

We have added the 5-95% confidence intervals to all of these numbers.

174, 175. "analysis" and "calculated" - by what method? From the slope of R against Delta T?

We have clarified the text that we use the method of Gregory et al. (2004), where annual average R is regressed against T, and the slope of the curve is an estimate of λ or Θ (line 194)

204. "agrees" in what sense?

We have changed the sentence to read: "We find that the 15 models whose average short-term Θ falls within the uncertainty of Θ estimated from CERES observations have ECS values ranging from 2.0-3.9 K, with an average of 2.9 K." (line 247)

218. I would say that this is "one source" of the spread, which is not eliminated, but only reduced, by using Theta instead.

We believe that this sentence is phrased correctly. The spread in our estimate from the ensemble is due to the construction of the energy-balance equation. Unlike observational analyses, we know everything else perfectly. Using our revised energy balance equation does not completely solve the problem, but it is an improvement.

233. Why is this material an appendix, rather than being incorporated in the main text?

We felt that this material would not be interesting to most readers, so we put it in the appendix. In retrospect, perhaps that was a bad decision. At this stage in the paper's review cycle, we hesitate to move material around. We can, however, if the reviewer or editor insists.

<u>Reviewer #2</u>

We thank the reviewer for their comments. In this document, we detail our responses.

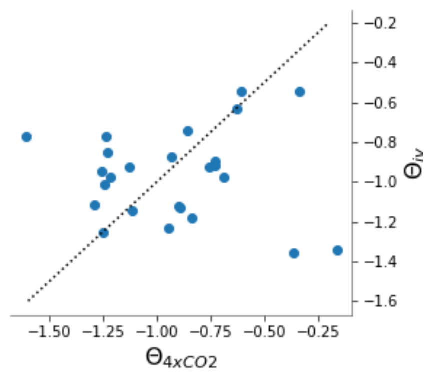
1) It would be helpful to provide a little more physical motivation for the choice of tropical 500 hPa temperature. I see some good reasons why mid-tropospheric temperature should work better (e.g., it should scale better with LR, WV and LW cloud feedbacks), but I don't think this was discussed anywhere. Why use tropical temperature rather than global-mean? Is there a physical rationale, or did this simply work better in MPI-ESM?

Also, although mid-tropospheric temperature clearly works better for the overall feedback, I expect the scaling with Ta might actually be a worse choice for some individual feedback processes (e.g. surface albedo, marine low cloud). This might be worth discussing briefly.

A: To address this, we have added a paragraph to the paper beginning on line 221.

2) A key result is that the revised feedback parameter theta more accurately estimates the "true" feedback strength under CO2 forcing. This is shown to be the case in MPI-ESM (L172-176). However, does this hold for CMIP5 models in general? I.e., do the values of theta estimated in control runs correlate well with those in 4xCO2?

A: This is not a claim we make in the paper, although one might infer it from the MPI model. Indeed, there is *some* correlation between short-term and long-term theta in the CMIP5 ensemble, as seen here:



Caption: Scatter plot of Θ_{4xCO2} vs. $\Theta_{control}$ from the CMIP5 ensemble. Each point represents values from model.

However, because of the outlier models, the relation is hard to interpret and we have not pursued this "emergent constraint" approach in our estimate of ECS using our revised framework [Dessler and Forster (2018, February 6). An estimate of equilibrium climate sensitivity from interannual variability. Retrieved from eartharxiv.org/4et67].

We have added a short statement to the paper to reflect this on line 260: "It may also be possible to use the relation between short-term and long-term Θ as an emergent constraint to convert short-term observations to the long-term response. There is some scatter in the relation in the CMIP5 ensemble, however, so more analysis of how these relate is likely required before ECS can be constrained in this way."

Relatedly, I would also suggest adding the correlation between R and Ta in CMIP5 piControl to Fig. 4, as additional bars in a different color.

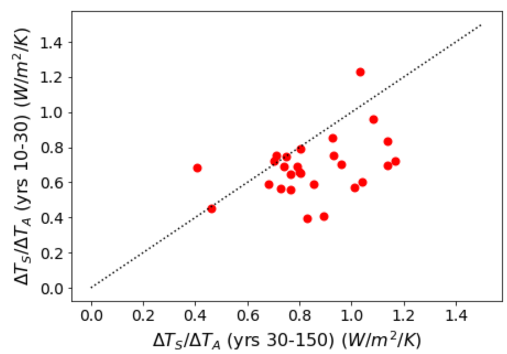
A: We have done that.

3) One important issue that isn't discussed in the paper is that the "pattern effect" doesn't simply go away with the improved relationship; rather, it shifts from the feedback parameter to the Ts/Ta term. This isn't a problem, but the way the paper is currently written, some readers might get that impression.

A: We have added a sentence discussing this: "Thus, the pattern effect's impact on ECS calculations shifts from λ in the traditional framework to the $\Delta T_S/\Delta T_A$ term in Eq. 4." (line 217)

So if most of the curvature in the relationship between radiative response and temperature goes away with the revised framework (Fig. 6), I expect there must be some curvature in the Ta versus Ts relationship in 4xCO2 runs. Can the authors confirm this?

Confirmed.



Caption. Scatterplot of slope of ΔT_S vs. ΔT_A in CMIP5 abrupt4xCO2 runs. Each point represents one model. The dotted line is the 1:1 line. The subscripts (10-30, 30-150) indicate the years of the run from which the slopes are calculated.

We've added a sentence to the paper mentioning that there is curvature in T_A vs T_S relation: "The lack of curvature in the Θ calculations means there is curvature in the relation between T_A and T_S in the models." (line 216)

4) I expect the Ts/Ta ratio cannot be reliably estimated from historical runs in the presence of large variability (for the same reason that lambda cannot be reliably estimated - because of the pattern effect). So we must rely on models to estimate this ratio under future global warming, meaning that it will be important to understand how future patterns of surface warming will develop. I suggest the authors discuss this briefly, for example in the conclusions.

We have added a sentence to the paper mentioning this point: "This also emphasizes the need to improve our understanding of the factors that control $\Delta T_S/\Delta T_A$, as well as how future patterns of surface warming will evolve." (line 218)

Other minor comments:

I suggest using colors in Fig. 6, rather than dark grey and black.

Done

L223: Cite Andrews and Webb 2018 - For future reference, it would be useful to mention the value of theta estimated from observations (horizontal dashed bar in Fig. 7a).

Done.

The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity

Andrew E. Dessler[1]*, Thorsten Mauritsen[2], Bjorn Stevens[2]

[1] Dept. of Atmospheric Sciences, Texas A&M University, College Station, TX 77843

[2] Max Planck Institute for Meteorology, Bundesstraße 53, 20146 Hamburg, Germany

*Correspondence to: adessler@tamu.edu, 979-862-1427

Keywords: Climate sensitivity, climate variability, energy balance

Abstract: Our climate is constrained by the balance between solar energy absorbed by the Earth and terrestrial energy radiated to space. This energy balance has been widely used to infer equilibrium climate sensitivity (ECS) from observations of 20th-century warming. Such estimates yield lower values than other methods and these have been influential in pushing down the consensus ECS range in recent assessments. Here we test the method using a 100-member ensemble of the MPI-ESM1.1 climate model simulations of the period 1850-2005 with known forcing. We calculate ECS in each ensemble member using energy balance, yielding values ranging from 2.1 to 3.9 K. The spread in the ensemble is related to the central hypothesis in the energy budget framework: that global average surface temperature anomalies are indicative of anomalies in outgoing energy (either of terrestrial origin or reflected solar energy). We find that assumption is not well supported over the historical temperature record in the model ensemble or more recent satellite observations. We find that framing energy balance in terms of 500-hPa tropical temperature better describes the planet's energy balance.

The problem

When an energy imbalance is imposed, such as by adding a greenhouse gas to the atmosphere, the climate will shift in such a way to eliminate the energy imbalance. This process is embodied in the traditional linearized energy balance equation:

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

where the forcing F is an imposed energy imbalance, T_S is the global average surface temperature, and λ relates changes in T_S to a change in net top-of-atmosphere (TOA) flux (Gregory et al., 2002; Dessler and Zelinka, 2014). R is the resulting TOA flux imbalance from the combined forcing and response. All quantities are deviations from an equilibrium base state, usually the pre-industrial climate. Equilibrium climate sensitivity (hereafter ECS, the equilibrium warming in response to a doubling of CO_2) is equal to $-F_{2xCO2}/\lambda$, where F_{2xCO2} is the forcing from doubled CO_2.

Many investigators (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) have used Eq. 1 combined with estimates of R, F, and T_s to estimate λ:

$$\lambda = \Delta(R-F)/\Delta T_s \quad (2)$$

where Δ indicates the change between the start of the historical period (usually the mid to late nineteenth century) and a recent period. These calculations result in values of λ near -2 $W/m^2/K$ and appear to rule out ECS larger than ~4 K (Stevens et al., 2016). The substantial likelihood of an ECS below 2 K implied by these calculations led the IPCC Fifth Assessment Report to extend their lower bound on *likely* values of ECS to 1.5 K (Collins et al., 2013).

We test this energy balance methodology through a perfect model experiment consisting of an analysis of a 100-member ensemble of runs of the MPI Earth System Model, MPI-ESM1.1. This is the latest coupled climate model from the Max Planck Institute for Meteorology and consists of the ECHAM6.3 atmosphere and land model coupled to the MPI-OM ocean model. The atmospheric resolution is T63 spectral truncation, corresponding to about 200 km, with 47 vertical levels, whereas the ocean has a nominal resolution of about 1.5 degrees and 40 vertical levels. MPI-ESM1.1 is a bug-fixed and improved version of the MPI-ESM used during CMIP5 (Giorgetta et al., 2013) and nearly identical to the MPI-ESM1.2 (Mauritsen et al., 2018) model being used to provide output to CMIP6, except that the historical forcings are from the MPI-ESM.

Each of the 100 members simulates the years 1850-2005 (Fig. 1) and use the same evolution of historical natural and anthropogenic forcings. The members differ only in their initial conditions —each starts from a different state sampled from a 2000-year control simulation. We calculate effective radiative forcing F for the ensemble by subtracting top-of-atmosphere flux R in a run with climatological sea surface temperatures (SSTs) and a constant pre-industrial atmosphere from average R from an ensemble of three runs using the same SSTs but the time-varying atmospheric composition used in the historical runs (Hansen et al., 2005; Forster et al., 2016). The three-member ensemble begins with perturbed atmospheric states. We estimate F_{2xCO2} using the same approach in a set of fixed SST runs in which CO_2 increases at 1% per year, which yields a F_{2xCO2} value of 3.9 W/m^2.

We calculate λ using Eq. 2 for each ensemble member, producing values ranging from -1.88 to -1.01 $W/m^2/K$ (5-95% range -1.63 to -1.17 $W/m^2/K$), with an ensemble median of -1.43 $W/m^2/K$ (Fig. 2a). In this calculation, $\Delta(R\text{-}F)$ and ΔT_S are the average difference between the first and last decade of each run. The spread in λ depends to some extent on how the calculation is set up — if one used the difference between the averages of the first and last 20 years, for example, the range in λ declines from 0.87 $W/m^2/K$ to 0.48 $W/m^2/K$. Using longer averaging periods does not further decrease the range.

Deleted: average

We also calculate ECS = $-F_{2xCO2}/\lambda$ for each ensemble member, producing values ranging from 2.08 to 3.87 K (5-95% range 2.39 to 3.34 K) (Fig. 2b), with an ensemble median of 2.72 K. Thus, our analysis shows that λ and ECS estimated from the historical record can vary widely simply due to internal variability. Given that we have only a single realization of the 20th century, we should not consider estimates based on the historical period to be precise — even with perfect observations. This supports previous work that also emphasized the impact of internal

Deleted: average

Deleted: 6

Deleted: ¶
With respect to precision of the estimates

variability on estimates of λ and ECS (Huber et al., 2014; Andrews et al., 2015; Zhou et al., 2016; Gregory and Andrews, 2016).

Previous researchers have questioned whether the historical record provides an accurate measure of λ and ECS, and we can check this by comparing the ensemble values to ECS estimates from a 2xCO_2 run of the MPI-ESM1.2, which is physically very close to MPI-ESM1.1. An abrupt 2xCO_2 run yields an ECS of 2.93 K in response to an abrupt doubling of CO_2 (estimated by regressing years 100-1000 of a 1000-year run) — 8% larger than the ensemble median. This is in line with the 10% difference in ECS estimated by Mauritsen and Pincus (2017) to arise from the average CMIP5 model time-dependent feedback, but smaller than suggested in other recent studies of ECS in transient climate runs (e.g., Armour, 2017; Proistosescu and Huybers, 2017).

Deleted: average

Deleted: and the changes between the MPI-ESM1.1 and MPI-ESM1.2 are not believed to be important for its climate sensitivity

Deleted: 6

Deleted: average

Thus, there are a number of issues that need to be considered when interpreting estimates of λ and ECS derived from the historical period. In addition to the precision and accuracy issues discussed above, it also includes the large and evolving uncertainty in forcing over the 20th century (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).

Why are estimates using the traditional energy balance approach imprecise?

In this section, we explain the physical process by which internal variability leads to the large spread in λ and ECS estimated from the ensemble. We begin by observing that Eqs. 1 and 2 parameterize R-F in terms of global average surface temperature, T_S. In model runs with strong forcing driving large warming, such as abrupt 4xCO_2 simulations, there is indeed a strong correlation between these variables (e.g., Gregory et al., 2004). However, because R-F in such runs is dominated by a monotonic trend, correlations will exist with any geophysical field that also exhibits a monotonic trend, regardless of whether there is a physical connection between the fields. Thus, one should not take the correlation between R-F and T_S in these runs as proving causality.

If T_S is a good proxy for the response R-F, we would expect to also see a correlation in measurements dominated by interannual variations. Observational data allow us to test this hypothesis. We use observations of R from the Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled product (ed. 4) (Loeb et al., 2009), which cover the period March 2000 to July. 2017. Our sign convention throughout the paper is that downward fluxes are positive. Temperatures come from the European Centre for Medium Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERAi) (Dee et al., 2011). We assume forcing changes linearly over this time period and account for it by detrending ΔR and ΔT anomaly time series using a linear least-squares fit to remove the long-term trend.

These data show that ΔR is poorly correlated with ΔT_S in response to interannual variability (Fig. 3a), as has been noted many times in the literature; see, e.g., Sect. 5 of Forster (2016). In particular, the low correlation coefficient tells us that ΔT_S explains little of the variance in ΔR. Using explicit estimates of forcing or other temperature datasets (e.g., MERRA-2) yield the same result.

GCMs that submitted output to the 5th phase of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al., 2012) also show this poor correlation. To demonstrate this, we have calculated the correlation coefficient between ΔT_S and ΔR in CMIP5 pre-industrial control runs (these are runs for which forcing F = 0). To facilitate comparison with the CERES data, as well as avoid any issues with long-term drift in the control runs, we break each run into 17-year segments to match the length of the CERES data and calculate the correlation coefficient of monthly anomalies of ΔR and ΔT_S for each segment. Fig. 4 shows that the correlation between ΔR and ΔT_S in the models is similar to that from the CERES analysis.

Recent work provides an explanation: the response of Δ(R-F) to a particular ΔT_S is determined not only by the global average magnitude, but also by the pattern of warming (Armour et al., 2013; Andrews et al., 2015; Gregory and Andrews, 2016; Zhou et al., 2016, 2017; Andrews and Webb, 2018). During El Nino cycles that dominate the observations in Fig. 3, the spatial pattern of warm and cool regions changes, leading to responses in Δ(R-F) that do not scale cleanly with ΔT_S — something Stevens et al. (2016) refer to as "pattern effects"

Deleted: 6

Deleted: (2016)

To demonstrate how this also generates the spread in λ in the model ensemble (Fig. 2a), we calculate the local response λ_r in three equal-area regions (90°S-19.4°S, 19.4°S-19.4°N, 19.4°N-90°N). We define λ_r as the regional analog to λ (Eq. 2):

$$\lambda_r = \Delta(R\text{-}F)_r/\Delta T_{S,r} \qquad (3)$$

where the "r" subscript indicates a regional average value.

We find that λ_r varies between the regions (Fig. 5). This means that different ensemble members with similar global average ΔT_S but different patterns of surface warming produce different values of global average Δ(R-F), thereby leading to spread in the estimated λ among the ensemble members. We also see strong variability in λ_r within each region, suggesting that how the warming is distributed within the region also drives some of the spread in estimated λ in the ensemble.

This explanation is consistent with analyses showing that λ changes during transient runs as the pattern of surface temperature evolves (Senior and Mitchell, 2000; Armour et al., 2013; Andrews et al., 2015; Gregory and Andrews, 2016; Stevens et al., 2016). In our model ensemble, however, the pattern changes are caused by internal variability rather than differing regional heat capacities that cause some regions to warm more slowly than others during forced warming.

A better way to describe energy balance

Our analysis demonstrates limitations of the conventional energy balance framework (Eq. 1). It has been previously noted that ΔR correlates better with tropospheric temperatures than ΔT_S (Murphy, 2010; Spencer and Braswell, 2010; Trenberth et al., 2015). Recent analyses have also stressed the importance of atmospheric temperatures — through its influence on lapse rate — as providing a fundamental control on the planet's energy budget (Zhou et al., 2016; Ceppi and Gregory, 2017). Based on this, we test a new energy balance framework constructed using the temperature of the tropical atmosphere:

$$R - F = \Theta T_A \qquad (4)$$

where T_A is the tropical average (30°N-30°S) 500-hPa temperature and Θ relates this quantity to R-F. R and F are the same global average quantities they were in equation 1. ECS can be expressed in terms of Θ:

$$\mathrm{ECS} = -\frac{\Delta F_{2\times CO2}}{\Theta}\frac{\Delta T_S}{\Delta T_A} \qquad (5)$$

where ΔT_S and ΔT_A are the equilibrium changes in these quantities in response to doubled CO_2. The CMIP5 ensemble average ratio $\Delta T_S/\Delta T_A$ is 0.86±0.10 (±1σ), where Δ represents the average difference between the first and last decades of the abrupt $4xCO_2$ runs.

Deleted: ;

Deleted: t

Support for Eq. 4 can be found in the observations: ΔR shows a tighter correlation with ΔT_A than with ΔT_S in observations (Figs. 3a vs. 3b). CMIP5 models also show this (Fig. 4). Given that the slope of these plots can be taken as estimates of Θ and λ, the tighter correlation leads to more accurate estimates of Θ than λ, both in absolute and relative terms.

Turning to the model ensemble, we next demonstrate that Θ is a more precise metric than λ. We do this by calculating Θ [= $\Delta(R\text{-}F)/\Delta T_A$] in each ensemble member, yielding values ranging from -1.18 to -0.89 W/m^2/K (5-95% range -1.16 to -0.92 W/m^2/K), with an ensemble median of -1.04 W/m^2/K (Fig. 2a). There is clearly less variability in Θ among the ensemble members than for λ. This reflects less variability in the regional response Θ_r (= $\Delta(R\text{-}F)_r/\Delta T_{A,r}$) than in λ_r (Fig. 5), as well as less variability within the regions. We therefore conclude that interannual variability has less of an impact on Θ than λ. We show additional evidence for the superior precision of Θ in the Appendix.

Deleted: average

As far as accuracy goes, we can compare Θ in the ensemble over the historical period to Θ in response to much larger warming. The ensemble median of Θ from the historic period (Fig. 2), -1.04±0.01 W/m^2/K (5-95% confidence interval), is close to the value obtained from an analysis of the first 150 years of an abrupt $4xCO_2$ run of the same model, Θ = -1.03±0.04 W/m^2/K, as well as Θ calculated from all 2600 years of this run, Θ = -1.00±0.01 W/m^2/K (values from the $4xCO_2$ runs are all obtained using the Gregory method (Gregory et al., 2004) using annual

Deleted: average

Deleted: 4

average R and temperatures). On the other hand, λ changes substantially in the $4xCO_2$ run as the climate warms: λ = -1.36±0.07 $W/m^2/K$ when calculated from the first 150 years, but λ = -0.95±0.01 $W/m^2/K$ from all 2600 years of that run.

Deleted: .

We can verify this result in the CMIP5 abrupt $4xCO_2$ ensemble. It has been previously demonstrated that plots of R-F vs. T_S do not trace straight lines as the climate warms (Andrews et al., 2015; Rugenstein et al., 2016; Rose and Rayborn, 2016; Armour, 2017), so λ and ECS calculated in a single model run may depend on the portion of the run selected. In the CMIP5 abrupt $4xCO_2$ ensemble, for example, average λ calculated by regressing years 10-30 (λ_{10-30}) is more negative than λ calculated from years 30-150 (λ_{30-150}) by 0.49 $W/m^2/K$ (Fig. 6).

Deleted: 50

Several explanations for this have been advanced, most prominently that λ is function of the pattern of surface warming (Senior and Mitchell, 2000; Armour et al., 2013; Andrews et al., 2015; Gregory and Andrews, 2016; Zhou et al., 2016; Stevens et al., 2016). Using Θ largely eliminates this pattern effect: Θ_{10-30} and Θ_{30-150} have an average difference of 0.13 $W/m^2/K$ for the CMIP5 ensemble (Fig. 6). Thus, we find additional evidence that Θ tends to be similar for different amounts and patterns of warming.

Deleted: 6

The lack of curvature in the Θ calculations means there is curvature in the relation between T_A and T_S in the models. Thus, the pattern effect's impact on ECS calculations shifts from λ in the traditional framework to the $\Delta T_S/\Delta T_A$ term in Eq. 4. This also emphasizes the need to improve our understanding of the factors that control $\Delta T_S/\Delta T_A$, as well as how future patterns of surface warming will evolve.

There are several plausible reasons why T_A may control R better than T_S. It seems likely that several of the feedbacks — e.g., lapse rate, water vapor, longwave cloud — should be more strongly influenced by atmospheric temperatures than T_S. More recently, it has been shown that atmospheric temperatures also play a key role in regulating low clouds (Zhou et al., 2016, 2017), thereby influencing the shortwave cloud feedback. This is also consistent with Ceppi et al. (2017), who identified a dependence of ECS on atmospheric stability in models. We have not further investigated this — ultimately, our use of T_A in Eq. 4 is based on observations

(Murphy, 2010; Spencer and Braswell, 2010; Trenberth et al., 2015) that it correlates well with R. Other metrics, such as global average atmospheric temperature work almost as well. Clearly, further investigations on how to best describe the Earth's energy balance are warranted.

Finally, one of our ultimate goals for this revised framework is to help produce better estimates of ECS. We are working on a detailed analysis of ECS based on this framework and will publish that in a follow-on paper, but we briefly show here how the advantages of the revised energy balance framework may be leveraged to do this. Fig. 7a shows Θ calculated from control runs of 25 CMIP5 models. To calculate Θ in the control runs, we break each control run into 17-year segments and calculate monthly anomalies of ΔR and ΔT_A during each segment. Then, we calculate Θ for each segment as the slope of the regression of ΔR vs. ΔT_A for that segment. Thus, for each control run, we generate a large number of estimates of Θ. The value in Fig. 7a is the average of these individual values.

Deleted: 6

Fig. 7b shows the ECS of these models, calculated from the first 150 years of the abrupt 4xCO_2 runs using the Gregory method. If we assume that models with more accurate simulation of short-term Θ produce more accurate estimates of ECS (Brown and Caldeira, 2017; Wu and North, 2002), then we can use Figs. 7a and 7b to constrain ECS. We find that the 15 models whose average short-term Θ falls within the uncertainty of Θ estimated from CERES observations have ECS values ranging from 2.0-3.9 K, with an average of 2.9 K. This excludes many of the highest ECS models, a result consistent with other analyses (Cox et al., 2018; Lewis and Curry, 2015).

Deleted: (Gregory et al., 2004)

Deleted: agrees with the

It would not have been possible to draw this conclusion with the conventional energy balance framework. Fig. 7c shows the comparison between λ from the control runs (calculated the same way Θ was calculated) and CERES observations. Because of the much larger uncertainty in the observational estimate of short-term λ, almost all models fall within the observational range, thereby prohibiting any constraint on the ECS range.

It may also be possible to use the relation between short-term and long-term Θ as an emergent constraint to convert short-term observations to the long-term response. There is some scatter in the relation in the CMIP5 ensemble, however, so more analysis of how these relate is likely required before ECS can be constrained in this way.

Conclusions

We have estimated ECS in each of a 100-member climate model ensemble using the same energy-balance constraint used by many investigators to estimate ECS from 20th-century historical observations. We find that the method is imprecise — the estimates of ECS range from 2.1 to 3.9 K (Fig. 2), with some ensemble members far from the model's true value of 2.9 K. Given that we only have a single ensemble of reality, one should recognize that estimates of ECS derived from the historical record may not be a good estimate of our climate system's true value.

Deleted: this suggests that some skepticism is appropriate when considering

The source of the imprecision relates to the construction of the traditional energy balance equation (Eq. 1). In it, the response of TOA net flux (R-F) is parameterized in terms of global average surface temperature (T_S). Recent research has suggested that the response is not just determined by the magnitude of T_S, but includes other factors, such as the pattern of T_S (e.g., Armour et al., 2013; Andrews et al., 2015; Gregory and Andrews, 2016; Zhou et al., 2017) or the lapse rate (e.g., Zhou et al., 2017; Ceppi and Gregory, 2017; Andrews and Webb, 2018). As a result, two ensemble members with the same ΔT_S can have different climate responses, Δ(R-F), leading to spread in the inferred λ.

The lack of a direct relationship between T_S and radiation balance suggests that it may be profitable to investigate alternative formulations. We test parameterizing the response in terms of 500-hPa tropical temperature (Eq. 4) and find that it is superior in many ways. Ultimately, how investigators describe the energy balance of the planet will depend on the problem and the available data. The surface temperature is indeed special, so the traditional framework may be preferred for some problems. But investigators may find that the alternatives are superior for certain problems, for instance constraining Earth's climate sensitivity.

Appendix

It has been previously noted in analyses of the historical record that λ exhibits significant interdecadal variability (Andrews et al., 2015; Gregory and Andrews, 2016; Zhou et al., 2016). We can reproduce this in a 2000-year control run (a run with fixed pre-industrial boundary conditions) of the MPI-ESM1.1 model. Fig. 8 shows λ calculated in a sliding 17-year window and confirms significant temporal variability in λ. We can similarly calculate Θ and find that temporal variability in Θ is substantially smaller (Fig. 8).

Deleted: 6

This result is reproduced in the CMIP5 control models. Fig. 9 plots the standard deviation of each CMIP5 model's set of short-term λ divided by the standard deviation of that model's set of short-term Θ (as described previously, we calculate time series of short-term λ and Θ values for each model by regressing anomalies in a 17-year sliding window of the control runs). All of the models fall above 1, demonstrating that there is less variability in the Θ time series than in the λ time series in every climate model. This confirms that Θ is more robust with respect to internal variability than λ. It also suggests that Θ estimated from the satellite data (Fig. 3) should be considered a better estimate of the climate system's long-term value than λ estimated from the same data set.

Deleted: 6

Deleted: Page Break

Acknowledgements: This work was supported by NSF grant AGS-1661861 to Texas A&M University. This work was initiated while AED was on Faculty Development Leave from Texas A&M during the Fall of 2016; he thanks Texas A&M and the Max Planck Institut für Meteorologie for supporting this research. Computational resources were made available by Deutsches Klimarechenzentrum (DKRZ) through support from German Federal Ministry of Education and Research (BMBF), and by the Swiss National Supercomputing Centre (CSCS).

Deleted: completed

Code and links to data can be found at
https://github.com/aedessler/DesslerMauritsenStevens18

References

Aldrin, M., Holden, M., Guttorp, P., Skeie, R. B., Myhre, G., and Berntsen, T. K.: 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, 10.1002/env.2140, 2012.

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

Andrews, T., and Webb, M. J.: The Dependence of Global Cloud and Lapse Rate Feedbacks on the Spatial Structure of Tropical Pacific Warming, J. Climate, 31, 641-654, 10.1175/jcli-d-17-0087.1, 2018.

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

Armour, K. C., Bitz, C. M., and Roe, G. H.: Time-varying climate sensitivity from regional feedbacks, J. Climate, 26, 4518-4534, 10.1175/jcli-d-12-00544.1, 2013.

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

Brown, P. T., and Caldeira, K.: Greater future global warming inferred from Earth’s recent energy budget, Nature, 552, 10.1038/nature24672, 2017.

Ceppi, P., and Gregory, J. M.: Relationship of tropospheric stability to climate sensitivity and Earth’s observed radiation budget, Proc. Natl. Acad. Sci., 10.1073/pnas.1714308114, 2017.

Collins, M., et al.: 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: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Cox, P. M., Huntingford, C., and Williamson, M. S.: Emergent constraint on equilibrium climate sensitivity from global temperature variability, Nature, 553, 319-322, 10.1038/nature25450, 2018.

Dee, D. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteor. Soc., 137, 553-597, 10.1002/qj.828, 2011.

Dessler, A. E., and Zelinka, M. D.: Climate feedbacks, in: Encyclopedia of Atmospheric Sciences, edited by: North, G. R., Pyle, J., and Zhang, F., Elsevier, 18–25, 2014.

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

Forster, P. M., et al.: Recommendations for diagnosing effective radiative forcing from climate models for CMIP6, J. Geophys. Res., 121, 12460-12475, 10.1002/2016jd025320, 2016.

Giorgetta, M. A., et al.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572-597, 10.1002/jame.20038, 2013.

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

Gregory, J. M., et al.: A new method for diagnosing radiative forcing and climate sensitivity, Geophys. Res. Lett., 31, 10.1029/2003gl018747, 2004.

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

Hansen, J., et al.: Efficacy of climate forcings, Journal of Geophysical Research: Atmospheres, 110, n/a-n/a, 10.1029/2005JD005776, 2005.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Rev. Geophys., 48, 10.1029/2010rg000345, 2010.

Huber, M., Beyerle, U., and Knutti, R.: Estimating climate sensitivity and future temperature in the presence of natural climate variability, Geophys. Res. Lett., 41, 2086-2092, 10.1002/2013GL058532, 2014.

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

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

Loeb, N. G., et al.: Toward optimal closure of the Earth's top-of-atmosphere radiation budget, J. Climate, 22, 748-766, 10.1175/2008jcli2637.1, 2009.

Mauritsen, T., et al.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2), in preparation, 2018.

Murphy, D. M.: Constraining climate sensitivity with linear fits to outgoing radiation, Geophys. Res. Lett., 37, 10.1029/2010GL042911, 2010.

Otto, A., et al.: Energy budget constraints on climate response, Nature Geoscience, 6, 415-416, 10.1038/ngeo1836, 2013.

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

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

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

Santer, B. D., et al.: Statistical significance of trends and trend differences in layer-average atmospheric temperature time series, J. Geophys. Res., 105, 7337-7356, 10.1029/1999jd901105, 2000.

Senior, C. A., and Mitchell, J. F. B.: The time-dependence of climate sensitivity, Geophys. Res. Lett., 27, 2685-2688, 10.1029/2000GL011373, 2000.

Shindell, D. T.: Inhomogeneous forcing and transient climate sensitivity, 4, 274, 10.1038/nclimate2136, 2014.

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

Spencer, R. W., and Braswell, W. D.: On the diagnosis of radiative feedback in the presence of unknown radiative forcing, J. Geophys. Res., 115, 10.1029/2009JD013371, 2010.
Stevens, B., Sherwood, S. C., Bony, S., and Webb, M. J.: Prospects for narrowing bounds on Earth's equilibrium climate sensitivity, Earth's Future, 4, 512-522, 10.1002/2016EF000376, 2016.
Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bull. Am. Met. Soc., 93, 485-498, 10.1175/bams-d-11-00094.1, 2012.
Trenberth, K. E., Zhang, Y., Fasullo, J. T., and Taguchi, S.: Climate variability and relationships between top-of-atmosphere radiation and temperatures on Earth, J. Geophys. Res., 10.1002/2014JD022887, 2015.
Wu, Q., and North, G. R.: Climate sensitivity and thermal inertia, Geophys. Res. Lett., 29, 10.1029/2002GL014864, 2002.
Zhou, C., Zelinka, M. D., and Klein, S. A.: Impact of decadal cloud variations on the Earth/'s energy budget, Nature Geosci, 9, 871-874, 10.1038/ngeo2828, 2016.
Zhou, C., Zelinka, M. D., and Klein, S. A.: Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach, Journal of Advances in Modeling Earth Systems, 9, 2174-2189, 10.1002/2017MS001096, 2017.

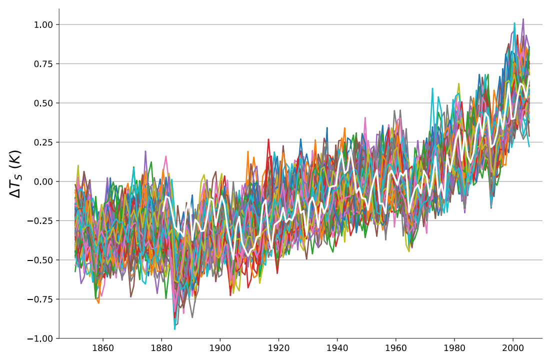


Fig. 1. Plot of annual and global average surface temperature from the 100 members of the MPI-ESM1.1 ensemble (colored lines), along with the GISTEMP measurements (Hansen et al., 2010) (white line). Temperatures are referenced to the 1951-1980 average.

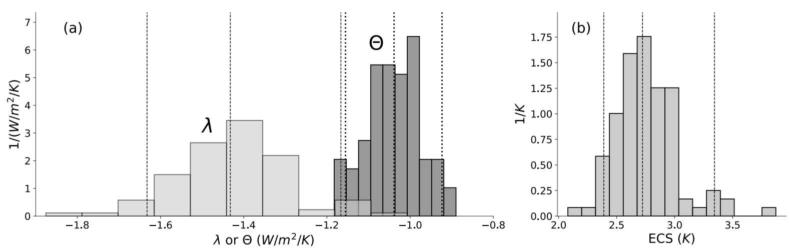


Figure 2. PDFs of (a) λ (lighter) and Θ (darker) and (b) ECS derived from the members of the MPI-ESM1.1 historical ensemble. The vertical lines are the 5th, 50th, and 95th percentile of each distribution.

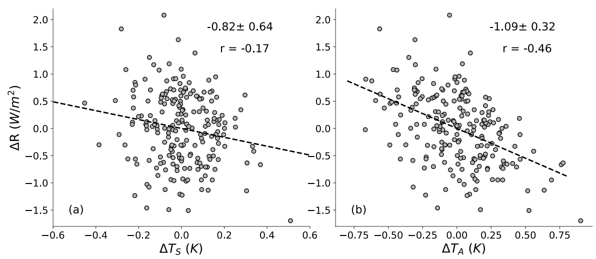


Figure 3. Scatter plot of detrended monthly anomalies of ΔR vs. (a) global average surface temperature ΔT_S, (b) tropical average 500-hPa temperature ΔT_A. Observations cover the period March 2000-July 2017 and anomalies are deviations from the mean annual cycle. The dashed lines are ordinary least-squares fits; the slope, 5-95% confidence interval, and correlation coefficient are shown on each panel. Confidence intervals account for autocorrelation of the time series (Santer et al., 2000).

Deleted: Jan.

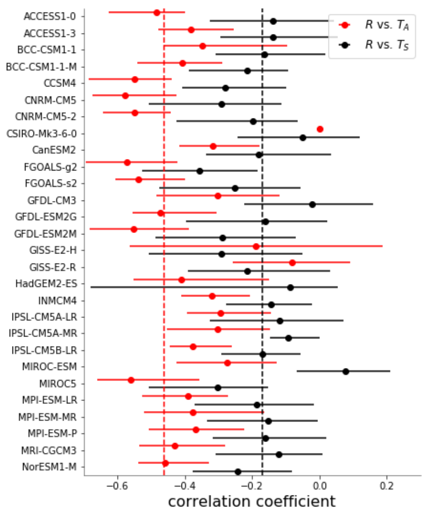


Fig. 4. Correlation coefficients between ΔR and temperature in CMIP5 control runs: black and red symbols represent the correlation with ΔT_S and ΔT_A, respectively. The dot is the average of the correlation coefficients from the 17-year segments of the model run; the bars indicate the maximum and minimum values from the control run. The dashed lines are the corresponding correlation coefficients from the CERES regressions in Fig. 2.

Deleted: ΔT_S

Deleted: 6

Deleted: blue

Deleted: is

Deleted: the average of the CMIP5 models, while the red dashed line is

Deleted: a

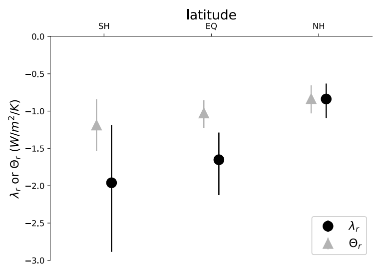


Fig. 5. λ_r and Θ_r calculated as regional average Δ(R-F) divided by regional average temperature (ΔT_S for λ and ΔT_A for Θ). The regions are 90°S-19.4°S (SH), 19.4°S-19.4°N (EQ), and 19.4°N-90°N (NH). The values are calculated for each member of the 100-member ensemble; the solid symbols are the ensemble average while the bars show the 5-95% range.

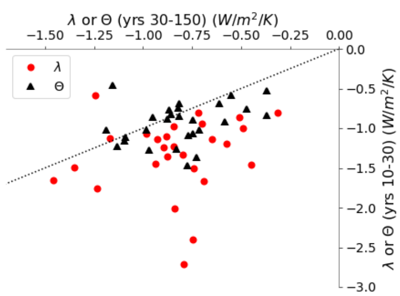


Fig. 6. Scatterplot of $\lambda_{10\text{-}30}$ vs. $\lambda_{30\text{-}150}$ (red circles) in CMIP5 abrupt4xCO_2 runs, as well as $\Theta_{10\text{-}30}$ vs. $\Theta_{30\text{-}150}$ (black triangles) in the same models. Each point represents one model. The dotted line is the 1:1 line. The subscripts (10-30, 30-150) indicate the years of the run from which the values are calculated.

Deleted: gray

Deleted: crosses

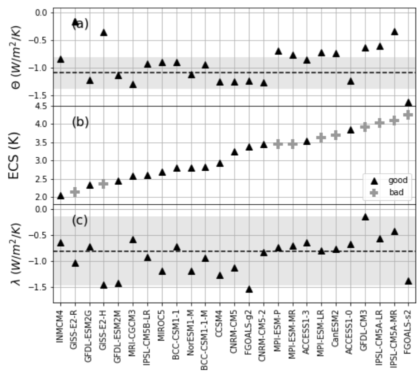


Figure 7. (a) Θ from individual CMIP5 control runs. The dotted line is the estimate from CERES observations; the gray region is the 5-95% confidence band. (b) ECS from each CMIP5 model, estimated from the first 150 years of abrupt 4xCO_2 runs using the Gregory method (Gregory et al., 2004). "Good" models are those whose Θ agrees with observations in panel (a), "bad" models are those that do not. (c) Same as panel (a), but for λ.

Deleted: (calculation described in the Appendix)

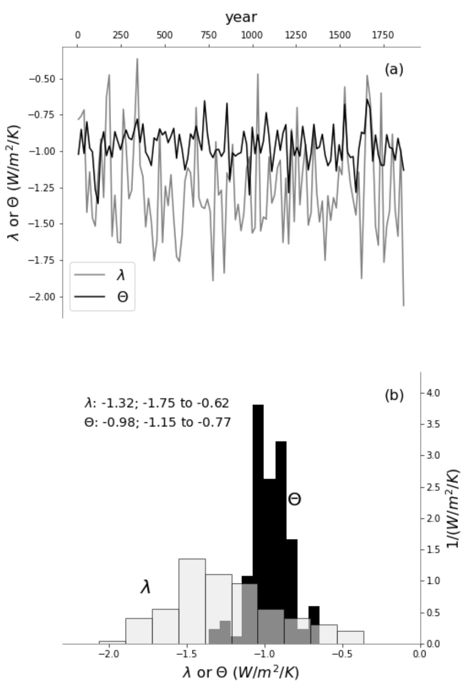


Fig. 8. (a) Time series of λ (gray) and Θ (black) estimated in a 17-year sliding window of a 2000-year control run of the MPI-ESM1.1. (b) PDFs of the time series in panel a. Median and 5-95% confidence interval for each distribution is displayed on the plot.

Deleted: 6

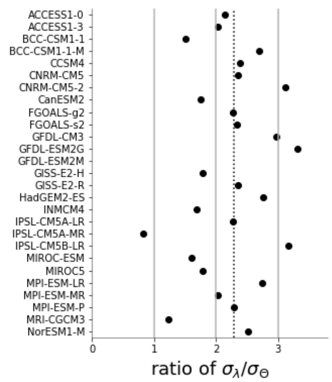


Fig. 9. The standard deviation of the λ time series divided by the standard deviation of the Θ time series. Each time series is calculated from 17-year segments of CMIP5 control runs. The dotted line is the ensemble average.

Deleted: 6