ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL088852

Key Points:
- Equilibrium climate sensitivity (ECS) estimates for a single coupled model can vary by more than 1°C (20%) depending on analysis method
- ECS estimates from ≥300-year coupled simulations from current U.S. models range from 3.1°C to 7.0°C, another method giving 2.7°C to 5.3°C
- Analysis of years 21–150 agrees with slab ocean ECS, but pentadal analysis of years 51–150 reduces bias against long, coupled simulations

Supporting Information:
- Supporting Information S1
- Data Set S1

Correspondence to:
J. P. Dunne,
john.dunne@noaa.gov

Citation:
Dunne, J. P., Winton, M., Bacmeister, J., Danabasoglu, G., Gettelman, A., Golaz, J.-C., et al. (2020). Comparison of equilibrium climate sensitivity estimates from slab ocean, 150-year, and longer simulations. *Geophysical Research Letters*, *47*, e2020GL088852. https://doi.org/10.1029/2020GL088852

Received 12 MAY 2020
Accepted 11 JUL 2020
Accepted article online 23 JUL 2020

Comparison of Equilibrium Climate Sensitivity Estimates From Slab Ocean, 150-Year, and Longer Simulations

John P. Dunne[1], Michael Winton[1], Julio Bacmeister[2], Gokhan Danabasoglu[2], Andrew Gettelman[2], Jean-Christophe Golaz[3], Cecile Hannay[2], Gavin A. Schmidt[4], John P. Krasting[1], L. Ruby Leung[5], Larissa Nazarenko[4], Lori T. Sentman[1], Ronald J. Stouffer[6], and Jonathan D. Wolfe[7]

[1]NOAA/OAR Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [2]National Center for Atmospheric Research, Boulder, CO, USA, [3]Lawrence Livermore National Laboratory, Livermore, CA, USA, [4]NASA Goddard Institute for Space Studies, New York, NY, USA, [5]Pacific Northwest National Laboratory, Richland, WA, USA, [6]Department of Geosciences, University of Arizona, Tucson, AZ, USA, [7]Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract We compare equilibrium climate sensitivity (ECS) estimates from pairs of long (≥800-year) control and abruptly quadrupled CO_2 simulations with shorter (150- and 300-year) coupled atmosphere-ocean simulations and slab ocean models (SOMs). Consistent with previous work, ECS estimates from shorter coupled simulations based on annual averages for years 1–150 underestimate those from SOM (−8% ± 13%) and long (−14% ± 8%) simulations. Analysis of only years 21–150 improved agreement with SOM (−2% ± 14%) and long (−8% ± 10%) estimates. Use of pentadal averages for years 51–150 results in improved agreement with long simulations (−4% ± 11%). While ECS estimates from current generation U.S. models based on SOM and coupled annual averages of years 1–150 range from 2.6°C to 5.3°C, estimates based longer simulations of the same models range from 3.2°C to 7.0°C. Such variations between methods argues for caution in comparison and interpretation of ECS estimates across models.

Plain Language Summary Precise definition and estimation of equilibrium climate sensitivity (ECS) continues to challenge model intercomparison. While annual analyses of years 1–150 of coupled atmosphere-ocean models agree with slab ocean model simulations, they underestimate coupled ECS estimates from multicentennial to millennial scale simulations. However, long-term ECS estimates can be largely recovered through a combination of (1) ignoring the first 50 years of abrupt 4 times preindustrial CO_2 simulation dominated by early timescales of ocean response and (2) using pentadal (5-year) averages instead of annual ones for years 51–150. This variation between methods argues for reconsideration of ECS estimation and application acknowledging that slab ocean estimates systematically ignore potential sources of enhanced sensitivity and simulations longer than 150 years are necessary for precise estimation of the long-term trend.

1. Introduction

Two primary metrics of idealized global climate model 2-m air temperature response of CO_2 greenhouse radiative forcing are the Transient Climate Response (TCR) to 1%CO_2 year^{-1} increase at doubling, and the equilibrium climate sensitivity (ECS) to CO_2 increase to a long term equilibrium doubling. ECS was originally estimated at 3°C ± 1.5°C (Charney et al., 1979) and has continued to serve as a fundamental metric of climate model behavior over the last four decades as estimated by the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment with "*high confidence* that ECS is *extremely unlikely* less than 1°C and *medium confidence* that the ECS is *likely* between 1.5°C and 4.5°C and *very unlikely* greater than 6°C." (Bindoff et al., 2013). While both TCR and ECS are defined as a temperature change from CO_2 doubling, the TCR is easily calculable in an idealized model framework as the global warming at the time of doubling (average of years 61–80), while the ECS of a given model is only fully known after that model has simulated control and doubled or quadrupled CO_2 over the long timescales of ocean heat uptake and after the sea surface temperature (SST) response has come to equilibrium—usually after several millennia of simulation (Krasting et al., 2018; Paynter et al., 2018; Rugenstein et al., 2019, 2020). While TCR is generally considered to more closely resemble the incremental (rather than abrupt) historical and projected CO_2 increase, ECS has been found to display a more robust representation of regional temperature change patterns than TCR (Grose et al., 2018) and is also highly useful both as a fundamental metric of model response and for the calibration of integrated assessment models (Calel & Stainforth, 2017). The relationship between TCR and ECS depends on several factors including the rate and pattern of both surface and interior ocean warming. Further, ECS has a long history of use, through all the IPCC reports and back to the late 1960s (e.g., Möller, 1963) and serves as an integrated high-level metric of the climate system that spans multiple generations of climate model development. A full contextual analysis of the value of the TCR and ECS concepts both historically and in the ongoing IPCC Sixth Assessment is provided in Meehl et al. (2020).

To estimate ECS without the computational expenditure of multimillennial simulations, several strategies have been used. The first was the "slab" ocean model (SOM) or "mixed layer" approach (Manabe & Stouffer, 1979, 1980) where the atmospheric model is coupled to a simple mixed layer ocean-sea ice model. Early SOM heat flux patterns were derived from observational climatologies while later SOMs where constructed from equilibrated coupled atmosphere-ocean simulations to more adequately reflect coupled model behavior (Bitz et al., 2012). As the SOM with specified lateral and deep ocean heat flux pattern comes rapidly to equilibrium, this approach requires far shorter simulations but has the uncertainty of estimating with a fundamentally different ocean component (Danabasoglu & Gent, 2009; Hansen et al., 1985, 1997) and it assumes no changes in heat transport by the world oceans. An alternative approach that uses shorter runs and extrapolates to equilibrium was put forth by Gregory et al. (2004) and applied to Coupled Model Intercomparison Project Phase 5 (CMIP5) models by Andrews et al. (2012) and is alternatively referred to as the "Effective Climate Sensitivity". In this approach, one conducts two simulations of at least 150 years —a control run and an abrupt quadrupling of CO_2—and regresses the difference in net radiative flux at the top of the atmosphere (ΔF) versus the change in global surface air temperature (ΔT) to extrapolate to the hypothetical radiative balance at equilibrium. Danabasoglu and Gent (2009) estimated the one sigma uncertainty in ECS estimates of approximately 0.18°C (8%) for CCSM3. Several studies have demonstrated the limitations of this approach highlighting the multiple timescales of ocean adjustment (Frölicher et al., 2014; Paynter et al., 2018) and the need to run models out longer than 150 years to achieve a robust estimate of ECS (Gregory et al., 2004). Nonlinearity of the relationship between ΔT and temporal and spatial variation in ocean heat uptake causes extrapolation methods to underestimate the ECS but with decreasing error as the integration lengthens (Armour, 2017; Armour et al., 2013; Ceppi & Gregory, 2017; Senior & Mitchell, 2000; Winton et al., 2010). Ocean heat uptake influences the pattern of surface temperature (Haugstad et al., 2017), which in turn determines the strength of climate feedback due to the spatially heterogeneous nature of these feedbacks (Armour et al., 2013). Specifically, the increase in feedback with time appears to be in large part due to the movement of the pattern of warming away from regions of tropical convection, regions that tends to induce particularly negative climate feedbacks (Bloch-Johnson et al., 2020; Dong et al., 2019; Zhou et al., 2017). Feedback temperature dependence, as mentioned above, can also change the slope of ΔF against ΔT.

Geoffroy et al. (2013) emulate the CMIP5 model nonlinearity of global temperature/heat uptake response to step forcing with a two box (two timescale) model. The kink in this adjustment occurs after the fast timescale adjustment with an *e*-folding time of about 4 years. Including the initial fast timescale adjustment with its steeper slope in the regression biases the ECS estimate low. Geoffroy et al. (2013) find an average long timescale *e*-folding time of 290 years for the CMIP5 models but is limited by the analysis having been based on only the first 150 years. Andrews et al. (2015) demonstrated that linearly fitting only years 21–150 increased the ECS estimate. Alternative methods include fitting functions with two or three exponentials (Proistosescu & Huybers, 2017), specific simulation setups (Saint-Martin et al., 2019), and the local tangent approach (Rugenstein et al., 2016). Recently, Rugenstein et al. (2020) showed through a 15-model intercomparison that the Gregory et al. (2004) method underestimated the long-term estimate by a median 17%.

Another weakness of the Gregory et al. (2004) and Andrews et al. (2012) methods relates to enhanced regression uncertainty in CMIP6 models as they increasingly capture climate modes of variability and their teleconnections. While only a few CMIP5 models were capable of accurately representing the role of ElNiño–Southern Oscillation (ENSO) on global temperature variability, modeling centers have since successfully represented not only ENSO but other modes of variability including Madden-Julian Oscillation and Pacific Decadal Oscillation (Eyring et al., 2019), and multidecadal to centennial modes (e.g., Zhang et al., 2019). Preprocessing the data by taking long averages before performing the regression filters out some of this low-frequency variability. For example, the method of Winton et al. (2020) uses 50-year-binned

10.1029/2020GL088852

averages of ΔF and ΔT before the regression is applied to better capture the forced response and avoid biasing the result with the different relationships between ΔF, Ocean heat uptake and ΔT relationship from natural internal variability such as ENSO. The first heat uptake/temperature pair is discarded and the remaining five that are available in the 300-year simulation are used in the regression.

The first 50 years are discarded to remove a period of SST adjustment during which a pattern of relatively reduced warming emerges in the subpolar North Atlantic and Southern Ocean (Winton et al., 2010). Although the fast mode of global surface temperature adjustment takes place with an e-folding timescale of about 4 years (Geoffroy et al., 2013), high-latitude adjustments—including changes in deep water circulation—are multidecadal or longer as the evolving SST response pattern changes the relationship between surface warming and high-latitude ocean heat uptake (Winton, Griffies et al., 2013).

One of the central experiments for the sixth phase of the CMIP6 Diagnostic, Evaluation and characterization of Klima (DECK) (Eyring et al., 2016) experiments is an abrupt quadrupling of atmospheric CO_2 run out for 150 years to estimate ECS, precluding the Winton et al. (2020) approach. The current approach used in ESMvalTool (Eyring et al., 2016) to estimate ECS is that of Gregory et al. (2004) and Andrews et al. (2012) in which least squares regression is conducted on the full 150 years using annual values of ΔF and ΔT. Making use of several previous generation models that have been run out to equilibrium and more recent ones run out 300 years, we are able to provide both a quantitative multimodel assessment of the Gregory et al. (2004) and Andrews et al. (2012) methods and provide an alternative approach for an improved estimate of the derived ECS among current generation U.S. models. However, we also note that the Andrews et al. (2012) method remains superior compared to SOM estimates in this analysis.

Building on previous work (Krasting et al., 2018; Paynter et al., 2018; Rugenstein et al., 2019, 2020; Winton, Adcroft, et al., 2013), the central factors of concern in the present study with respect to abrupt 4xCO_2 changes in radiative forcing are as follows: (1) Do models return to radiative balance and, if so, how long does it take? (2) Over what timescale (if ever) does the approach to equilibrium to become quasilinear? And (3) how long of a temporal average is required to remove the confounding role of model internal variability. While the first question on decadal scales is largely answered in Winton, Adcroft, et al. (2013) and on millennial scales in Paynter et al. (2018) and Krasting et al. (2018), the present study takes the practical step of translating that understanding into improved analysis of the current short (150-year) CMIP6 DECK simulations to estimate the ECS achieved from multiple century to multimillennial simulations.

2. Methods

We take advantage of the combination of previous generation models that were contributed to LongRunMIP (Rugenstein et al., 2019, 2020) along with the University of Arizona implementation of the Manabe Climate Model (MCM_UA) based on the GFDL R30c and MOM1 component models described in Delworth et al. (2002) with air-sea flux adjustment to reproduce SSTs and a fixed 40-m mixed layer, GFDL's second generation of climate model $\mathrm{GFDL_{ESM2G}}$ (Dunne et al., 2012) based on GFDL's CM2.1 (Delworth et al., 2006). We also take advantage of GFDL's fourth-generation model development products, $\mathrm{GFDL_{CM4}}$ (Held et al., 2019) and $\mathrm{GFDL_{ESM4.1}}$ (Dunne, Horowitz, et al., 2020). The climate sensitivity in CM4 has been previously documented in Winton et al. (2020). Additionally, long simulations with Goddard Institute for Space Studies $\mathrm{GISS_{E2.1\text{-}G}}$ (2091 years) (Kelley et al., 2020), the National Center for Atmospheric Research $\mathrm{NCAR_{CESM2(CAM6)}}$ (898 years) (Danabasoglu et al., 2020; Gettelman et al., 2019), and the Department of Energy $\mathrm{DOE_{E3SMv1}}$ (300 years) (Golaz et al., 2019) are used.

3. Data

Data from CMIP6 models can be found on the Earth System Grid Federation (https://esgf-node.llnl.gov/projects/esgf-llnl/). All global annual values for temperature and net radiation at the top of the atmosphere for control and 4xCO_2 runs used in this study as well as Matlab™ scripts to analyze the data are supplied as supporting information. LongRunMIP data are from Rugenstein et al. (2020).

Table 1

Equilibrium Climate Sensitivity (ECS; °C doubled CO_2^{-1}) Estimates From Slab Ocean Models (SOM) (e.g., Manabe & Stouffer, 1979) Under Atmosphere-Land-Sea Ice Simulations, Long Equilibration Runs (e.g., Senior & Mitchell, 2000; See Footnotes), and 150- and 300-year Runs With 1- and 50-year Averaging Periods for ΔF and ΔT (Column Heading Notation: Averaging-Period$_{data\text{-}span}$) Where Preindustrial Reference F and T Were Estimated From Least Squares Regression Over the First 300 years of the Control Run

Model		SOM (Manabe & Stouffer, 1979)	>800 years (Senior & Mitchell, 2000)	50yr$_{51\text{-}300}$ (Winton et al., 2020)	1yr$_{1\text{-}150}$ (Gregory et al., 2004)	1yr$_{21\text{-}150}$ (Andrews et al., 2015)	5yr$_{51\text{-}150}$ (this study)
			Latest U.S. CMIP6 models				
DOE$_{E3SMv1}$				**7.02**	5.31 ± 0.29	5.68	5.99
GFDL$_{CM4}$		4.1		**4.93**	3.91 ± 0.29	4.45	4.88
GFDL$_{ESM4.1}$ five-member ensemble of 4xCO_2 simulations	101	3.25	3.37	**3.06**	2.66 ± 0.14	2.63	2.93
	126				2.65	2.67	2.84
	151				2.68	2.65	2.75
	176				2.65	2.63	2.89
	201				2.64	2.68	2.90
	Ave				2.66	2.65	2.90
	Std. Dev				0.02	0.02	0.13
GISS$_{E2.1G}$		3.0	3.21	**3.23**	2.72 ± 0.10	2.83	3.10
NCAR$_{CESM2(CAM6)}$		5.3[e]	6.58	**6.53**	5.26 ± 0.29	6.24	6.60
			Previous GFDL Models				
GFDL$_{ESM2G}$		3.4[d]	3.27[a]	**2.92**	2.34 ± 0.14	2.68	3.04
MCM_UA		3.4[c]	3.45	**3.60**	3.76 ± 0.17	3.97	4.13
			LongRunMIP Models (Rugenstein et al., 2020)				
CCSM3		2.32[f]	2.46[k] (2.73)	**2.60**	2.50	2.66	2.81
CESM$_{1.0.4}$		3.20[i]	3.57 (3.38)	**3.29**	2.88	3.18	3.44
CNRM$_{CM61}$			5.47 (5.7)	**4.91**	4.94	4.88	4.70
ECHAM5$_{MPIOM}$		5.55[g]	6.0, 5.4[g] (5.83)	**5.84**	5.26	4.93	4.67
FAMOUS			7.42 (8.55)	**6.28**	5.56	5.85	6.15
GFDL$_{CM3}$			4.84 (4.67)	**4.38**	3.96	4.21	4.18
GFDL$_{ESM2M}$		3.4[d]	3.34[b] (3.25)	**2.97**	2.45	2.63	2.87
GISS$_{E2R}$		2.4[j]	2.40 (2.44)	**2.28**	2.16	2.30	2.28
HADCM3L		3.3[k]	3.40 (3.45)	**3.30**	2.90	3.27	3.44
HADGEM2			4.71 (4.77)	**4.72**	4.52	5.65	5.80
IPSL$_{CM5A}$			4.27 (4.76)	**4.05**	4.04	4.19	4.27
MIROC32		4.0[h]	(4.49)		4.12	4.11	4.27
MPIESM11			3.46 (3.35)	**3.51**	3.02	3.22	3.49
MPIESM12			3.47 (3.42)	**3.44**	3.04	3.23	3.37

Note. The 1σ uncertainties associated with the regression are provided for models not already contributed to LongRunMIP. Columns that represent the long term (≥800 years regressing 50-year-binned averages; hereafter *long*) with values from LongRunMIP provided in parentheses are provided along with 1-year averages over years 1–150 (1yr$_{1\text{-}150}$) (Gregory et al., 2004), 1-year averages over years 21–150 (1yr$_{21\text{-}150}$) (Andrews et al., 2015), regressing 50-year-binned averages over years 51–300 (50yr$_{51\text{-}300}$) (Winton et al., 2020), and regressing 5-year-binned averages over years 51–150 (5yr$_{51\text{-}150}$) methods.
[a]Gettelman et al. (2019). [b]Delworth et al. (2002). [c]Krasting et al. (2018) based 4 times simulation, [d]Stouffer et al. (2006). [e]Kiehl et al. (2006). [f]Danabasoglu and Gent (2009). [g]Meehl et al. (2013). [h]Li et al. (2013). [i]Paynter et al. (2018) based 2 times simulation. [j]Schmidt et al. (2014). [k]Randall et al. (2007).

4. Results

A comparison of different methods of estimating ECS for a suite of climate models is provided in Table 1. Among models for which both slab ocean model (SOM; Column 2) and long (>800-year regressing 50-year-binned averages; hereafter *long*; Column 3) coupled atmosphere-ocean estimates are available, SOM estimates tend to be 6% ± 7% lower than *long* estimates (Figure 2, upper left). One example of significant disagreement is NCAR$_{CESM2(CAM6)}$ for which the *long* estimate is 1.4°C higher than the SOM based estimate. Early analysis suggests this is a result of the cloud response to the warming surface in NCAR$_{CESM2(CAM6)}$ (Danabasoglu et al., 2020; Gettelman et al., 2019). Several studies have demonstrated a strong increase in ECS estimates with warming climate with 4xCO_2 perturbations often giving a higher ECS than 2xCO_2 experiments (e.g., Bloch-Johnson et al., 2015; Meraner et al., 2013; Rohrschneider et al., 2019). As many of the SOM estimates come from 2xCO_2 experiments whereas all of the fully coupled simulations come from 4xCO_2 experiments, this nonlinearity could explain this result.

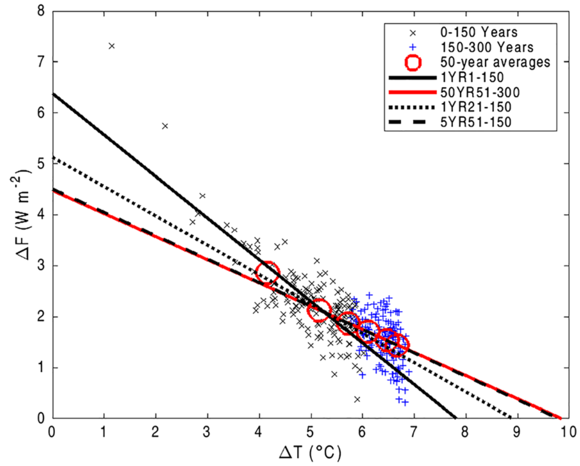

Figure 1. Example estimation of ECS from different methods described in Table 1 for $GFDL_{CM4}$ model from the regression of the difference in net radiation at the top of the atmosphere in the $4xCO_2$ simulation from the control (ΔF; W m^{-2}) versus the difference in 2-m air temperature as annual values from those simulations for the first 150 years (black x), and next 150 years (blue +) along with 50-year averages (red o) and regression using $1yr_{1\text{-}150}$ (black solid), $50yr_{51\text{-}300}$ (red solid), $1yr_{21\text{-}150}$ (black dot), and $5yr_{51\text{-}150}$ (black dash) methods.

When we compute $50yr_{51\text{-}300}$ (Column 4 in Table 1) following Winton et al. (2020) from 300-year simulations in which the first 50 years is ignored and the slope/intercept is calculated by regressing 50-year-binned averages ($50yr_{51\text{-}300}$, fourth column), we find good correspondence with *long* ECS with $50yr_{51\text{-}300}$ tending to underestimate *long* ECS by approximately 5% ± 5% with the exception of FAMOUS, which gave a 21% lower $50yr_{51\text{-}300}$ estimate than its *long* estimate. The lack of convergence of ECS in FAMOUS is discussed in Rugenstein et al. (2020) but could not be explained. As this model displays a fundamentally different and unexplained behavior than the other models, it is excluded from subsequent analysis in this study. Similar to $NCAR_{CESM2(CAM6)}$, $GFDL_{CM4}$ also exhibits a much lower ECS (0.8°C) based on SOM than $50yr_{51\text{-}300}$. We also note that this analysis—using a linear regression of the entire 1- to 300-year period to reference the control—gets a slightly lower estimate for $GFDL_{CM4}$ (4.93°C) than the value of 5.0°C provided in Winton et al. (2020). This is due to their having referenced individual 50-year time differences from the control and ignoring data corresponding to the first 50 years of perturbation. Overall, we found that differences in treatment of the control drift resulted in relatively small ECS estimates (<0.1°C) using year 51–300 analyses.

Because the standard simulation time of the abrupt $4xCO_2$ simulations in the CMIP6 experimental design is only 150 years, we next turn our attention to comparability of alternative methods for performing such calculations from the CMIP6 multimodel ensemble. For all estimates of ECS based on the first 150 years alone, we applied a single linear estimate of the 0–300 year control drift for all ECS calculations as we found the uncertainty control drift to inflate markedly when restricted to year 1–150 and 51–150 analysis. We find the annual-150 year method (1 $yr_{1\text{-}150}$; Column 5 in Table 1) (Andrews et al., 2012; Gregory et al., 2004) slightly overestimates *long* ECS for the first generation MCM_UA but strongly underestimates by approximately 0.4–1.7°C, or −18% ± 5% of 50 $yr_{51\text{-}300}$ among the more recent models analyzed here. Important to note is that MCM_UA differs from all of the other models considered here in not having an explicit mixed layer but rather a fixed 40 m mixed layer depth. As such, it is unable to represent the immediate surface-warming based shoaling and reduction of ventilation of the upper ocean that, in all the other models, leads to a latitudinal shift in sea surface warming away from the tropics toward higher latitudes and results in a strong initial ΔF:ΔT slope that subsides as warming propagates into the ocean interior and surface stratification subsides (Cubasch et al., 1992). Held et al. (2010) found that surface ocean warming initially was focused in the tropics while higher-latitude warming occurred later except for the North Atlantic subpolar region which cooling.

Efforts to remove the role of the initial response of the 150-year runs have been proposed. We find that the revised 150 year method using annual averages but ignoring the first 20 years proposed by Andrews et al. (2015; Column 6) increases the ECS estimate on average by 8% ± 6% and removing the first 50 years in the annual analysis ($1yr_{51\text{-}150}$; Column 6 in Table 1) leads to further increase of 3% ± 5%. We found that ignoring more than 50 years led to considerable degradation in the reliability of the regression. Ignoring initial slopes associated with ocean equilibration timescales is not the only challenge, however, as current generation climate models include representation of a complex combination of interannual, decadal and centennial scale modes of variability. We find that the ECS underestimate can be further reduced when both the first 50 years are ignored and the annual data is averaged into pentads. An example visual comparison of these methods is provided in Figure 1 for the case of $GFDL_{CM4}$.

Excluding FAMOUS, we find that the original Gregory et al. (2004) method ($1yr_{1\text{-}150}$) underestimates the *long* estimate by an average of −14% ± 8% (Figure 2, middle left panel), while the Andrews et al. (2012) method ($1yr_{21\text{-}150}$) is less biased relative to the *long* estimate but with a larger uncertainty (−8% ± 10%; Figure 2, middle right panel). We find an improved correspondence to the *long* estimate when 5-year averages are first calculated before linear regression is performed ($5yr_{51\text{-}150}$; Table 1, Column 7; Figure 2,

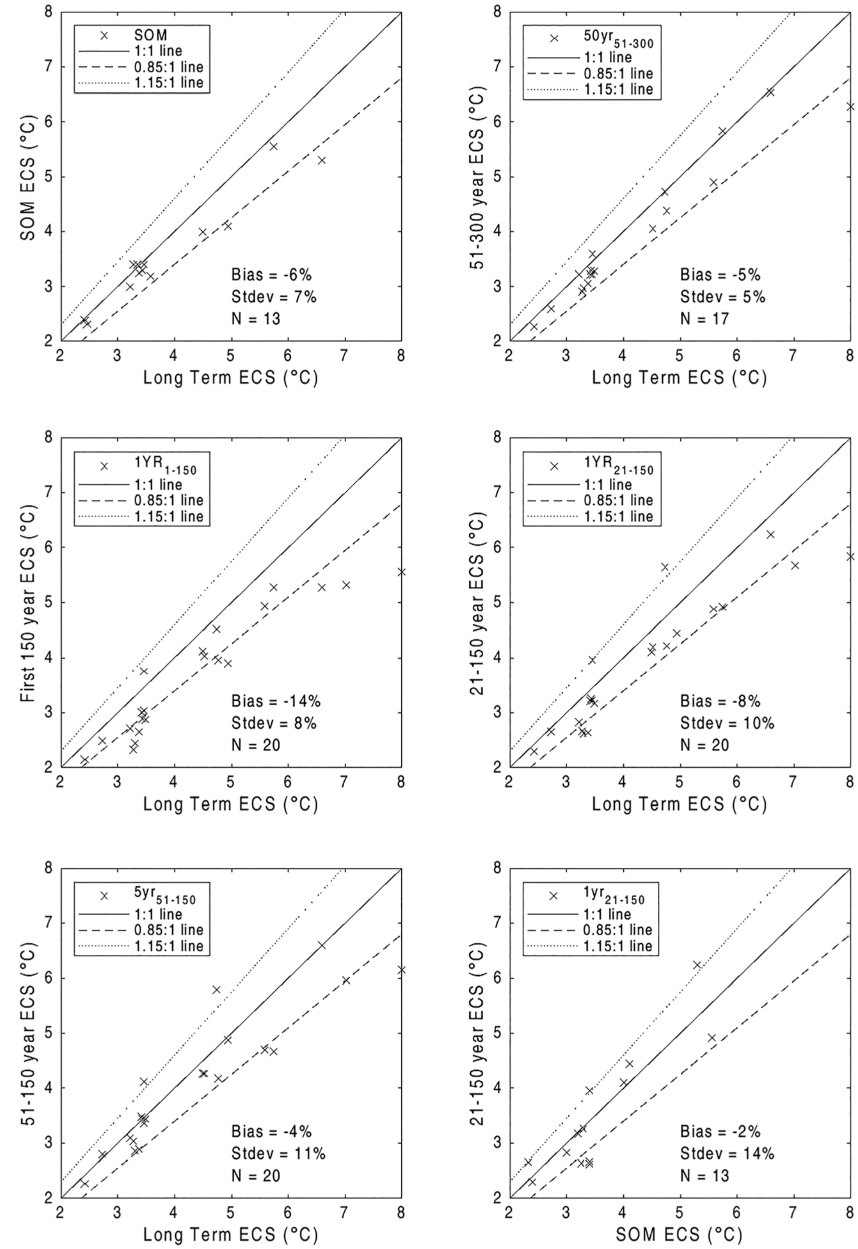

Figure 2. Visual comparison of ECS (°C) based on long (≥800 years) abrupt $4xCO_2$ runs versus slab ocean model (SOM) estimates (upper left), 300 year estimates (upper right), and 150 year estimates using the $1yr_{1\text{-}150}$ (middle left) $1yr_{21\text{-}150}$ (middle right), and $50yr_{51\text{-}150}$ methods (lower left) and SOM estimate versus the $1yr_{21\text{-}150}$ method (lower right) where the 1:1 (solid), 1.0:1.15 (dashed), and 1.15:1 (dotted) lines are also shown on each plot for reference.

lower left panel) with underestimation of the *long* estimate but slightly larger uncertainty of $-4\% \pm 11\%$. The overall results are shown in Figure 2 that illustrates that the $5yr_{51\text{-}150}$ tends to follow the 1:1 line when compared to long simulations (Figure 2, lower left panel), whereas the $1yr_{1\text{-}150}$ approach follows more closely to the 0.85:1 line (middle left panel). It is important to note that this analysis suggests that some of the higher ECS estimated from 300-year simulations is due to processes that begin to manifest within the first 150 years but are potentially masked by the early response and that the value of running the simulations out to 300 years is not to uncover a differing response from the 51- to 150-year period but rather to boost the signal to noise in the result. However, it is also clear that some long term processes are also at work in some models such as FAMOUS that serve to further elevate ECS and may result in a biased underestimate in the final equilibrium value (e.g., Bloch-Johnson et al., 2015; Meraner et al., 2013; Rohrschneider et al., 2019).

Alternatively, when the SOM estimate is used as the true value (Figure 2, lower right panel), it is the $1yr_{21\text{-}150}$ method that follows more closely the 1:1 line with a low bias of −2% ± 14%, whereas the $5yr_{51\text{-}150}$ method gives a high bias of 4% ± 15%. While the $NCAR_{CESM2(CAM6)}$ and DOE_{E3SMv1} models give similar ECS of 5.3°C with the $1yr_{1\text{-}150}$ method, $NCAR_{CESM2(CAM6)}$ gives a significantly higher ECS with both $5yr_{51\text{-}150}$ and $50yr_{51\text{-}300}$ methods (6.66°C and 6.53°C, respectively) while DOE_{E3SMv1} gives a much higher ECS with $50yr_{51\text{-}300}$ (7.02°C) than $5yr_{51\text{-}150}$ (5.96°C) methods, highlighting the potential for significant differences in results between methods across models. Further, while the $5yr_{51\text{-}150}$ method appears superior to the $1yr_{1\text{-}150}$ approach in 16 of 21 cases, the $5yr_{51\text{-}150}$ approach strongly overestimates the long estimate in the case of MCM_UA, $ECHAM5_{MPIOM}$, and HAD_{GEM2}. As such, the above diversity in model behavior should serve as caution in interpreting the uncertainty associated with each method and the potential role of a suite of factors including the nonlinearity of CO_2 response, lack of equilibration of initial and final states, and long-term feedbacks associated with adjustments in ocean circulation.

We also conducted a suite of sensitivity studies varying both the averaging windows from 1 to 50 years, and length of initial simulation ignored from 0 to 50 years. As discussed in the supporting information, we found little advantage to increasing the averaging window without ignoring an initial segment. In contrast, we found that including an averaging window of 5 years filtered out most of the interannual variability when the first 50 years of simulation was excluded. Further, we found that fidelity declined slightly as the averaging window increased beyond 5 years. We attribute this slight loss in fidelity as an effective decrease in the span along the x axis from 95 years with the 5-year window to 50 years with the 50-year window. Overall, we found that analysis of 150-year simulation results largely converged with analysis of 300-year simulation when the first 50 years were excluded both with an averaging window of 1 year, and slightly more so with an averaging window of 5 years.

5. Conclusions

We find that much of the character of the long-term behavior of the ECS estimates can be captured with exclusion of the first third of a 150-year simulation and calculating 5-year averages before least squares regression for an ECS estimate with only slight underestimation (−4% ± 11%). With the original method of Gregory et al. (2004), however, we find an underestimation of −14% ± 8% of ECS compared to those estimated from *long* (>800 year) runs. Using the modified method of Andrews et al. (2012), we find a smaller underestimation of −8% ± 10% compared to those estimated from *long* runs but good agreement (−2% ± 14%) with SOM-based estimates. One interpretation of these results is that the CMIP6 experimental design significantly underestimates the *long* ECS with CMIP6 class models, but that this deficiency can be largely addressed using the modified $5yr_{51\text{-}150}$ method excluding the initial part of the simulation and taking pentadal averages of years 51–150 to calculate the temperature at radiative balance. As such, we find a large range of ECS estimates among current generation U.S. models from 2.6°C using the $1yr_{21\text{-}150}$ method for $GFDL_{ESM4.1}$ to 7.0°C using the $50yr_{51\text{-}300}$ method for DOE_{E3SM1}—well outside the IPCC assessment that ECS is very unlikely greater than 6°C (Bindoff et al., 2013). However, we also find evidence that estimates from *long* abrupt $4xCO_2$ simulations are significantly higher than SOM estimates as well as considerable divergence in the relationships between different methods for different models. There might also be an additional discrepancy due to the typical use of $2xCO_2$ simulations for SOM estimates but $4xCO_2$ simulations for coupled estimates. Under the assumption that the global surface air temperature responds linearly to an increase in atmospheric CO_2, the $4xCO_2$ and the $2xCO_2$ should give the same climate sensitivity. In practice the linear assumption is not strictly satisfied (Jonko et al., 2013) as forcing has been shown to be supralogarithmically dependent on the CO_2 concentration (Byrne & Goldblatt, 2014; Etminan et al., 2016; Gregory et al., 2015) along with other arguments for nonlinearity in the temperature dependence of radiative feedbacks (e.g., Bloch-Johnson et al., 2015; Meraner et al., 2013; Rohrschneider et al., 2019). As such, we argue that more research should be done to standardize methods to estimate ECS through a more comprehensive comparison of ECS through both multimillennial climate perturbation simulations such as conducted in LongRunMIP (Rugenstein et al., 2019, 2020) and slab ocean model comparisons to better understand the causes of these differences and derive a more robust estimate of climate sensitivity from current generation models.

Note also that the concept of ECS is predicated on the initial and final states both being in equilibrium. With computationally intensive models such as those used in CMIP6, models have typically not spun up the ocean to equilibrium for practical considerations. As such, while the surface temperature was stable, the deep ocean thermal status was continuing to evolve. When such a system is perturbed, it may respond differently, at least in transient (e.g., He et al., 2017), to the equilibrium of a slab ocean model.

Whereas ECS began as a convenient idealized model construct (e.g., Charney et al., 1979), it has emerged as a routine test of models as if ECS could be measured and interpreted precisely and accurately. We argue that the concept of ECS should be considered more notional than absolute and useful more in idealized studies of relative sensitivity with the understanding absolute value of this metric will depend on the state of the system and the nature of the imposed forcing with different methods of estimating ECS accessing different feedbacks.

Data Availability Statement

Data used in this analysis as well as analysis scripts are provided in the supporting information and is available at Dunne, Winton, et al. (2020; http://doi.org/10.5281/zenodo.3932257).

Acknowledgments

We thank the efforts at NOAA, NASA, NCAR, and DoE for supporting U.S. supercomputing for climate studies through the Earth System Grid Federation (https://esgf-node.llnl.gov/projects/esgf-llnl/). The National Center for Atmospheric Research is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement 1852977. The Energy Exascale Earth System Model (E3SM) project is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. Work at LLNL was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. PNNL is operated for U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC05-76RL01830. LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). Climate modeling at GISS is supported by the NASA Modeling, Analysis and Prediction program, and resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. We are also very appreciative of the four anonymous reviewers that contributed comments to substantially improve this manuscript.

References

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

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*, L09712. https://doi.org/10.1029/2012GL051607

Armour, K. (2017). Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, *7*, 331–335. https://doi.org/10.1038/nclimate3278

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, *26*(13), 4518–4534. https://doi.org/10.1175/JCLI-D-12-00544.1

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., et al. (2013). Detection and attribution of climate change: From global to regional. In T. F. Stocker, et al., (Eds.), *Climate Change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Chap. 10, pp. 867–952). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Bitz, C. M., Shell, K. M., Gent, P. R., Bailey, D. A., Danabasoglu, G., Armour, K. C., et al. (2012). Climate sensitivity of the community climate system model, version 4. *Journal of Climate*, *25*(9), 3053–3070. https://doi.org/10.1175/JCLI-D-11-00290.1

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, *42*, 4973–4980.

Bloch-Johnson, J., Rugenstein, M., & Abbot, D. S. (2020). Spatial radiative feedbacks from internal variability using multiple regression. *Journal of Climate*, *33*(10), 4121–4140. https://doi.org/10.1175/JCLI-D-19-0396.1

Byrne, B., & Goldblatt, C. (2014). Radiative forcing at high concentrations of well-mixed greenhouse gases. *GRL*, *41*(1), 152–160. https://doi.org/10.1002/2013GL058456

Calel, R., & Stainforth, D. A. (2017). On the physics of three integrated assessment models. *Bulletin of the American Meteorological Society*, *98*(6), 1199–1216. https://doi.org/10.1175/BAMS-D-16-0034.1

Ceppi, P., & Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proceedings of the National Academy of Sciences*, *114*(50), 13126–13131. https://doi.org/10.1073/pnas.1714308114

Charney, J. G., Arakawa, A., Baker, D. J., Bolin, B., Dickinson, R. E., Goody, R. M., et al. (1979). *Carbon dioxide and climate: A scientific assessment* (pp. 2030–2050). Washington, DC: National Academy of Sciences.

Cubasch, U., Hasselmann, K., Höck, H., Maier-Reimer, E., Mikolajewicz, U., Santer, B. D., & Sausen, R. (1992). Time-dependent greenhouse warming computations with a coupled ocean-atmosphere model. *Climate Dynamics*, *8*(2), 55–69. https://doi.org/10.1007/BF00209163

Danabasoglu, G., & Gent, P. R. (2009). Equilibrium climate sensitivity: Is it accurate to use a slab ocean model? *Journal of Climate*, *22*(9), 2494–2499.

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916.

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., et al. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643–674.

Delworth, T. L., Stouffer, R., Dixon, K., Spelman, M., Knutson, T., Broccoli, A., et al. (2002). Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *Climate Dynamics*, *19*(7), 555–574.

Dong, Y., Proistosescu, C., Armour, K. C., & Battisti, D. S. (2019). Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a Green's function approach: The preeminence of the Western Pacific. *Journal of Climate*, *32*(17), 5471–5491. https://doi.org/10.1175/JCLI-D-18-0843.1

Dunne, J. P., Horowitz, L. W., Adcroft, A., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS002015. https://doi.org/10.1029/2019MS002015

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2012). GFDL's ESM 2 global coupled climate–carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, *25*(19), 6646–6665. https://doi.org/10.1175/JCLI-D-11-00560.1

Dunne, J. P., Winton, M., Bacmeister, J., Danabasoglu, G., Gettelman, A., Golaz, J.-C., et al. (2020). Supplementary material to "Comparison of equilibrium climate sensitivity estimates from slab ocean, 150-year, and longer simulations" (version 1.0.0) [data set]. *Zenodo*. http://doi.org/10.5281/zenodo.3932257

Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *GRL*, *43*(24), 12,614–12,623. https://doi.org/10.1002/2016GL071930

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development (Online)*, *9*. LLNL-JRNL-736881

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., et al. (2016). ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *9*, 1747–1802. https://doi.org/10.5194/gmd-9-1747-2016

Frölicher, T. L., Winton, M., & Sarmiento, J. L. (2014). Continued global warming after CO_2 emissions stoppage. *Nature Climate Change*, *4*(1), 40–44. https://doi.org/10.1038/nclimate2060

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., & Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *Journal of Climate*, *26*(6), 1859–1876. https://doi.org/10.1175/JCLI-D-12-00196.1

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High climate sensitivity in the Community Earth System Model version 2 (CESM2). *Geophysical Research Letters*, *46*, 8329–8337. https://doi.org/10.1029/2019GL083978

Golaz, J. C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129. https://doi.org/10.1029/2018MS001603

Gregory, J. M., Andrews, T., & Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO_2. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2054). https://doi.org/10.1098/rsta.2014.0417

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*, L03205. https://doi.org/10.1029/2003GL018747

Grose, M. R., Gregory, J., Colman, R., & Andrews, T. (2018). What climate sensitivity index is most useful for projections? *Geophysical Research Letters*, *45*, 1559–1566. https://doi.org/10.1002/2017GL075742

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., & Stone, P. (1985). Climate response times: Dependence on climate sensitivity and ocean mixing. *Science*, *229*(4716), 857–859. https://doi.org/10.1126/science.229.4716.857

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. *Journal of Geophysical Research*, *102*(D6), 6831–6864. https://doi.org/10.1029/96JD03436

Haugstad, A. D., Armour, K. C., Battisti, D. S., & Rose, B. E. J. (2017). Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, *44*, 7455–7463. https://doi.org/10.1002/2017GL074372

He, J., Winton, M., Vecchi, G., Jia, L., & Rugenstein, M. (2017). Transient climate sensitivity depends on base climate ocean circulation. *Journal of Climate*, *30*(4), 1493–1504. https://doi.org/10.1175/JCLI-D-16-0581.1

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. https://doi.org/10.1029/2019MS001829

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), 2418–2427. https://doi.org/10.1175/2009JCLI3466.1

Jonko, A. K., Shell, K. M., Sanderson, B. M., & Danabasoglu, G. (2013). Climate feedbacks in CCSM3 under changing CO_2 forcing. Part II: Variation of climate feedbacks and sensitivity with forcing. *Journal of Climate*, *26*(9), 2784–2795. https://doi.org/10.1175/JCLI-D-12-00479.1

Kelley, M., Schmidt, G. A., Nazarenko, L., Miller, R. L., Bauer, S. E., Ruedy, R., et al. (2020). GISS-E2.1: Configurations and climatology. *Journal of Advances in Modeling Earth Systems*, e2019MS002025. https://doi.org/10.1029/2019MS002025

Kiehl, J. T., Shields, C. A., Hack, J. J., & Collins, W. D. (2006). The climate sensitivity of the Community Climate System Model version 3 (CCSM3). *Journal of Climate*, *19*(11), 2584–2596. https://doi.org/10.1175/JCLI3747.1

Krasting, J. P., Stouffer, R. J., Griffies, S. M., Hallberg, R. W., Malyshev, S. L., Samuels, B. L., & Sentman, L. T. (2018). Role of ocean model formulation in climate response uncertainty. *Journal of Climate*, *31*(22), 9313–9333. https://doi.org/10.1175/JCLI-D-18-0035.1

Li, C., von Storch, J-S., & Marotzke, J. (2013). Deep-ocean heat uptake and equilibrium climate response. *Climate Dynamics*, *40*(5-6), 1071–1071.

Manabe, S., & Stouffer, R. J. (1979). A CO_2-climate sensitivity study with a mathematical model of the global climate. *Nature*, *282*(5738), 491–493. https://doi.org/10.1038/282491a0

Manabe, S., & Stouffer, R. J. (1980). Sensitivity of a global climate model to an increase of CO_2 concentration in the atmosphere. *Journal of Geophysical Research*, *85*(C10), 5529–5554. https://doi.org/10.1029/JC085iC10p05529

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J. F., Stouffer, R. J., et al. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, *6*(26), eaba1981. https://doi.org/10.1126/sciadv.aba1981

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., et al. (2013). Climate change projections in CESM1 (CAM5) compared to CCSM4. *Journal of Climate*, *26*(17), 6287–6308.

Meraner, K., Mauritsen, T., & Voigt, A. (2013). Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, *40*, 5944–5948. https://doi.org/10.1002/2013GL058118

Möller, F. (1963). On the influence of changes in the CO_2 concentration in air on the radiation balance of the Earth's surface and on the climate. *Journal of Geophysical Research*, *68*(13), 3877–3886. https://doi.org/10.1029/JZ068i013p03877

Paynter, D., Frölicher, T. L., Horowitz, L. W., & Silvers, L. G. (2018). Equilibrium climate sensitivity obtained from multimillennial runs of two GFDL climate models. *Journal of Geophysical Research: Atmospheres*, *123*, 1921–1941. https://doi.org/10.1002/2017JD027885

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*, e1602821.

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., et al. (2007). Climate models and their evaluation. In Climate change 2007: The physical science basis. *Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR)* (589–662). Cambridge University Press.

Rohrschneider, T., Stevens, B., & Mauritsen, T. (2019). On simple representations of the climate response to external radiative forcing. *Climate Dynamics*, *53*(5–6), 3131–3145. https://doi.org/10.1007/s00382-019-04686-4

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., et al. (2019). LongRunMIP–motivation and design for a large collection of millennial-length AO-GCM simulations. *Bulletin of the American Meteorological Society*, accepted, *100*(12), 2551–2570. https://doi.org/10.1175/BAMS-D-19-0068.1

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., et al. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, *47*, e2019GL083898.

Rugenstein, M. A., Gregory, J. M., Schaller, N., Sedláček, J., & Knutti, R. (2016). Multiannual ocean-atmosphere adjustments to radiative forcing. *Journal of Climate*, *29*(15), 5643–5659. https://doi.org/10.1175/JCLI-D-16-0312

Saint-Martin, D., Geoffroy, O., Watson, L., Douville, H., Bellon, G., Voldoire, A., et al. (2019). Fast-forward to perturbed equilibrium climate. *GRL*, *46*(15), 8969–8975. https://doi.org/10.1029/2019GL083031

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., et al. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, *6*(1), 141–184.

Senior, C. A., & Mitchell, J. F. B. (2000). Time-dependence of climate sensitivity. *Geophysical Research Letters*, *27*(17), 2685–2688. https://doi.org/10.1029/2000GL011373

Stouffer, R. J., Broccoli, A. J., Delworth, T. L., Dixon, K. W., Gudgel, R., Held, I., et al. (2006). GFDL's CM2 global coupled climate models. Part IV: Idealized climate response. *Journal of Climate*, *19*(5), 723–740. https://doi.org/10.1175/JCLI3632.1

Winton, M., Adcroft, A., Dunne, J. P., Held, I. M., Shevliakova, E., Zhao, M., et al. (2020). Climate sensitivity of GFDL's CM4.0. *Journal of Advances in Modeling Earth Systems*, *12*(1). https://doi.org/10.1029/2019MS001838

Winton, M., Adcroft, A., Griffies, S. M., Hallberg, R., Horowitz, L. W., & Stouffer, R. J. (2013). Influence of ocean and atmosphere components on simulated climate sensitivities. *Journal of Climate*, *26*(1), 231–245. https://doi.org/10.1175/JCLI-D-12-00121.1

Winton, M., Griffies, S. M., Samuels, B. L., Sarmiento, J. L., & Frölicher, T. L. (2013). Connecting changing ocean circulation with changing climate. *Journal of Climate*, *26*(7), 2268–2278. https://doi.org/10.1175/JCLI-D-12-00296.1

Winton, M., Takahashi, K., & Held, I. M. (2010). Importance of ocean heat uptake efficacy to transient climate change. *Journal of Climate*, *23*(9), 2333–2344. https://doi.org/10.1175/2009JCLI3139.1

Zhang, L., Delworth, T. L., Cooke, W., & Yang, X. (2019). Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Climate Change*, *9*(1), 59–65. https://doi.org/10.1038/s41558-018-0350-3

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