Climate Dynamics (2020) 54:4515–4543
https://doi.org/10.1007/s00382-020-05242-1

Improved methods for estimating equilibrium climate sensitivity from transient warming simulations

Aiguo Dai[1] · Danqing Huang[2] · Brian E. J. Rose[1] · Jian Zhu[3] · Xiangjun Tian[4]

Received: 23 August 2019 / Accepted: 10 April 2020 / Published online: 19 April 2020

Abstract

Equilibrium climate sensitivity (ECS) refers to the total global warming caused by an instantaneous doubling of atmospheric CO_2 from the pre-industrial level in a climate system. ECS is commonly used to measure how sensitive a climate system is to CO_2 forcing; but it is difficult to estimate for the real world and for fully coupled climate models because of the long response time in such a system. Earlier studies used a slab ocean coupled to an atmospheric general circulation model to estimate ECS, but such a setup is not the same as the fully coupled system. More recent studies used a linear fit between changes in global-mean surface air temperature (ΔT) and top-of-atmosphere net radiation (ΔN) to estimate ECS from relatively short simulations. Here we analyze 1000 years of simulation with abrupt quadrupling ($4\times CO_2$) and another 500-year simulation with doubling ($2\times CO_2$) of pre-industrial CO_2 using the CESM1 model, and three other multi-millennium (~5000 year) abrupt $4\times CO_2$ simulations to show that the linear-fit method considerably underestimates ECS due to the flattening of the $-$dN/dT slope, as noticed previously. We develop and evaluate three other methods, and propose a new method that makes use of the realized warming near the end of the simulations and applies the $-$dN/dT slope calculated from a best fit of the ΔT and ΔN data series to a simple two-layer model to estimate the unrealized warming. Using synthetic data and the long model simulations, we show that the new method consistently outperforms the linear-fit method with small biases in the estimated ECS using $4\times CO_2$ simulations with at least 180 years of simulation. The new method was applied to $4\times CO_2$ experiments from 20 CMIP5 and 19 CMIP6 models, and the resulting ECS estimates are about 10% higher on average and up to 25% higher for models with medium–high ECS (> 3 K) than those reported in the IPCC AR5. Our new estimates suggest an ECS range of about 1.78–5.45 K with a mean of 3.61 K among the CMIP5 models and about 1.85–6.25 K with a mean of 3.60 K for the CMIP6 models. Furthermore, stable ECS estimates require at least 240 (180) years of simulation for using $2\times CO_2$ ($4\times CO_2$) experiments, and using shorter simulations may underestimate the ECS substantially. Our results also suggest that it is the forced $-$dN/dT slope after year 40, not the internally-generated $-$dN/dT slope, that is crucial for an accurate estimate of the ECS, and this forced slope may be fairly stable.

Keywords Climate sensitivity · Equilibrium response · Climate feedback · Climate models · CMIP5 · Global warming

1 Introduction

Equilibrium climate sensitivity (ECS) refers to the increase in global-mean surface (air) temperature in response to a doubling of atmospheric CO_2 (often from the pre-industrial level) after the system reaches a new equilibrium (Hansen et al. 1984; Randall et al. 2007; Flato et al. 2013; Stevens et al. 2016; Knutti et al. 2017). ECS has been a fundamental metric used to quantify how sensitive a climate system is to a given forcing caused by changes in atmospheric greenhouse gases (GHGs), especially in model comparisons (Charney et al. 1979; Cubasch et al. 2001; Randall et al. 2007; Flato et al. 2013), although different forcing agents may cause

✉ Aiguo Dai
adai@albany.edu

[1] Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, NY 12222, USA

[2] School of Atmospheric Sciences, Nanjing University, Nanjing, China

[3] College of Hydrology and Water Resources, Hohai University, Nanjing, China

[4] ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

different surface warming per unit forcing, leading to different efficacy (Hansen et al. 2005). Furthermore, different characteristics of the same type of forcing may cause varying response in surface warming (Hansen et al. 1997). In addition, near-future climate projections are usually more related to the transient climate response rather than ECS (Grose et al. 2018). Thus, ECS is a simplified metric; nevertheless, it still provides a useful measure of the sensitivity of a climate system to CO_2 forcing, which is the primary forcing agent in the current climate change.

However, due to the long response time of Earth's climate system to any forcing perturbation, it is difficult to estimate ECS accurately. Until about 10 years ago, ECS of a fully coupled climate model was commonly estimated using relatively short simulations of its atmospheric component coupled with a slab ocean (e.g., Kiehl et al. 2006), which lacks ocean dynamics but allows the system to reach a new equilibrium quickly. Clearly, the slab-ocean setup is very different from the fully coupled version of the model; thus one could argue that ECS estimated from the slab-ocean experiments may not be the same as that in fully coupled model runs, although some long simulations by coarse-resolution models showed that the slab-ocean approach may work well with only a small bias (Danabasoglu and Gent 2009; Li et al. 2013; Jonko et al. 2013). Partly because of this, the climate community quickly adopted a simple method proposed by Gregory et al. (2004), who observed a fairly linear relationship between the changes (relative to a control run) in the global-mean and annual-mean surface air temperature (ΔT) and top-of-atmosphere (TOA) net radiation (ΔN) (i.e., $\Delta N \approx F - \lambda \Delta T$) in a $4\times CO_2$ simulation and used the $\lambda = -dN/dT$ slope, referred to as the feedback parameter (Andrews et al. 2012), to extrapolate the ΔT when ΔN approaches zero (i.e., when the system reaches a new equilibrium) as twice ECS. This linear-fit method was used to estimate ECS of CMIP5 (Andrews et al. 2012; Forster et al. 2013; Flato et al. 2013) and CMIP6 (Zelinka et al. 2020) models and has become the standard method for estimating ECS in fully coupled climate models, although there have been efforts to improve the Gregory et al.'s method. For example, Geoffroy et al. (2013a, b) estimated ECS using their two-layer energy balance solutions fitted to model data from CMIP5 $4\times CO_2$ experiments; Andrews et al. (2015) and Armour (2017) estimated ECS using linear regressions between ΔT and ΔN data over years 21–150, and Proistosescu and Huybers (2017) estimated ECS using a more sophisticated 3-mode model fit. Recently, Paynter et al. (2018) claim that ECS can be estimated from extended 1%/year CO_2 ramping simulations. These all produce higher ECS values than those reported in Flato et al. (2013) for the reasons discussed below, and a recent analysis of multi-millennium long simulations suggests an underestimate bias of 10–30% for the Gregory et al. (2004)-based ECS estimates (Rugenstein et al. 2019a).

It is well known, however, that Earth's surface temperature responds very rapidly to a doubling or quadrupling of CO_2 during the first few decades as the deep ocean has yet to take up energy; thereafter, the response slows down gradually as the heat exchange with the deep ocean becomes important (Held et al. 2010; Geoffroy et al. 2013a, b; Gregory et al. 2015; Garuba et al. 2018). Thus, the relationship between surface temperature and TOA net radiation differs during these periods of fast and slow adjustment, such that the $\lambda = (-dN/dT)$ slope decreases over time (Senior and Mitchell 2000; Armour et al. 2013; Rose and Rayborn 2016), or is nonlinear and changes with the state of the climate system (Feldl and Roe 2013; Jonko et al. 2013; Meraner et al. 2013; Gregory et al. 2015; Knutti and Rugenstein 2015; He et al. 2017; Ceppi and Gregory 2019), in response to an instantaneous CO_2 change. In particular, changes in the feedback parameter λ over time have been interpreted in terms of regional climate feedbacks (Armour et al. 2013), spatial patterns of ocean heat uptake (Rose et al. 2014; Rose and Rayborn 2016; Rugenstein et al. 2016), a non-unit efficacy of ocean heat uptake (Held et al. 2010; Winton et al. 2010; Geoffroy et al. 2013b; Yoshimori et al. 2016), or changes with the mean state (He et al. 2017; Ceppi and Gregory 2019), although these are not mutually exclusive interpretations (e.g. Haugstad et al. 2017). A recent analysis of multi-millennium simulations by coupled climate models (Rugenstein et al. 2019a) suggests that the feedback parameter may change over time, especially during the first 150 years. All these studies suggest that the $(-dN/dT)$ slope should change (more specifically decrease) with time, rather than being constant, at least during the first 100 years in a $2\times CO_2$ or $4\times CO_2$ experiment. The increasing climate sensitivity ($=1/\lambda$) over time (Senior and Mitchell 2000; Armour et al. 2013; Rose and Rayborn 2016) would imply that the Gregory et al. (2004)'s method should underestimate ECS in a fully coupled model, which is confirmed recently by Paynter et al. (2018) using multi-millennium simulations of two fully coupled models. This raises the following questions: How big is this underestimate for other models? Is the Gregory et al. (2004)'s method of using the $(-dN/dT)$ slope estimated over the whole simulation period appropriate for estimating ECS? Can we improve this extrapolation method for estimating ECS from relatively short simulations to fill this practical need as multi-millennium long simulations are still expensive and impractical for most modeling groups? While the underestimation bias of the Gregory et al. (2004)'s method has been noticed previously (e.g., Andrews et al. 2015; Armor 2017; Proistosescu and Huybers 2017; Paynter et al. 2018; Rugenstein et al. 2019a) and Geoffroy et al. (2013b) proposed a new method to estimate ECS based on their fitted two-layer model, the focus of these previous studies was not on developing and evaluating different methods for estimating ECS from a relatively short simulation, and

these questions have not been systematically investigated. Given that the Gregory et al. (2004) method is still used to estimate ECS in the latest CMIP6 models (Zelinka et al. 2020), there is an urgent need to address the above questions.

Here we first discuss some of the key issues associated with the linear-fit method for estimating ECS, then analyze a 1000-year $4\times CO_2$ simulation using the CESM1 from NCAR (Hurrell et al. 2013), a comprehensive fully coupled climate model, to explore and evaluate several other methods for estimating ECS using varying lengths of simulation. The $\Delta T(t)$ and $\Delta N(t)$ time series from this simulation are fitted with the analytic solutions from the two-layer energy balance model of Geoffroy et al. (2013a, b), and realistic random noise is then added to the fitted $\Delta T(t)$ and $\Delta N(t)$ time series to generate synthetic samples of the noisy $\Delta T(t)$ and $\Delta N(t)$ time series. Four different methods (including that of Gregory et al. 2004) are then used to estimate the predefined ECS in these synthetic data series. This allows us to quantitatively evaluate the performance of these different methods and quantify the bias of the Gregory et al. (2004)'s method under this idealized framework. The stability of the estimated ECS using CESM1-simulated data over varying lengths of simulation is also examined as a measure of the performance. The performance of the four methods is further evaluated using multi-millennium (~5000 year) long simulations with abrupt $4\times CO2$ forcing from three other fully coupled models, in which case the ECS can be estimated approximately as the warming near the end of the simulation. The results suggest that ECS can be estimated reliably from relatively short (e.g., 180 year) simulations using our improved methods.

Based on these analyses, we recommend a new method and apply it to the $4\times CO_2$ simulations from 20 CMIP5 and 19 CMIP6 models to produce new ECS estimates for these models. Our results show that the previously reported ECS values (Andrews et al. 2012; Forster et al. 2013; Flato et al. 2013) are substantially underestimated, as noticed previously (Geoffroy et al. 2013b; Andrews et al. 2015; Armour 2017; Proistosescu and Huybers 2017; Paynter et al. 2018; Rugenstein et al. 2019a, b). This implies that Earth's ECS may be higher than our estimates, which have remained in the range of 1.5–4.5 °C since 1979 despite the tremendous progresses in climate modeling and paleoclimate research made in the last 4 decades (Charney et al. 1979; Randall et al. 2007, PALEOSENS 2012; Flato et al. 2013; Stevens et al. 2016; Knutti et al. 2017; Marvel et al. 2018a). However, the latest CMIP6 models show a slightly higher ECS range of 1.8–5.6 K based on the same Gregory et al. (2004) method (Zelinka et al. 2020), while our new estimates suggest a range of 1.85–6.25 K.

We emphasize that our results of the performance of the various methods, which are confirmed by our analyses of the multi-millennium simulations, are derived from the synthetic data that resemble the noise level (including some temporal variations in dT, dN, and dN/dT) and the long-term evolution of the dT and dN seen in the CESM1 $4\times CO_2$ simulation. Thus, they are likely to be only indicative for the performance when applied to climate model simulations, and the biases may change with data from individual models. Nevertheless, the *relative* performance found here is likely applicable to real model data because the synthetic data used here capture the overall long-term evolution of the data series seen in the published multi-millennium simulations (Danabasoglu and Gent 2009; Li et al. 2013; Jonko et al. 2013; Paynter et al. 2018), which show that the dT gradually approaches an upper limit with only small fluctuations on decadal to centennial time scales after the first few hundred years. These short-term fluctuations, which are represented by the noise in our synthetic data, may result from internal variations (e.g., due to warming- and thus time-dependent feedbacks) that may change the short-term climate sensitivity (or the dN/dT slope), but these short-term variations do not appear to affect the long-term evolution of dT and thus the final ECS as suggested by the published multi-millennium time series (Danabasoglu and Gent 2009; Li et al. 2013; Jonko et al. 2013; Paynter et al. 2018). Our analyses of available multi-millennium simulations also show a stable long-term dN/dT slope that allows the use of a constant dN/dT slope for estimating ECS using a relatively short simulation. From this perspective, we believe our test results on the relative performance of the various methods are likely applicable to real model simulations.

Clearly, for models that continue to exhibit wild fluctuations over extended periods of time (e.g., > 100 years) in the dT and dN series even after the first few hundred years, no simple method can reliably estimate their ECS from a short simulation over a few hundred years, and the only way to estimate their ECS is to make very long simulations. However, for most climate models, as shown below with the CMIP5 and CMIP6 models, our improved method appears to work reasonably well for estimating ECS from a relatively short simulation. Furthermore, our analyses of available multi-millennium simulations revealed that the (−dN/dT) slope does not show any long-term trends, although it varies considerably on decadal to centennial time scales due to internal variability. This long-term stability provides a physical basis for using a constant slope to estimate ECS based on a short simulation. Thus, our study fills a practical need for such a method, as most modeling groups cannot afford to make multi-millennium simulations for each new version of their models.

2 Model experiments and analysis method

2.1 Model experiments

We ran the Community Earth System Model version 1.2.1 (CESM1, with CAM4 as the atmospheric model) (Hurrell et al. 2013) for 1000 years after an instantaneous quadrupling of the pre-industrial CO_2 (at 284.7 ppmv) ($4\times CO_2$) with a 2.5° lon × ~1.875 lat atmospheric grid and about 1.12° lon × 0.47° lat for the oceans. The CESM1 is a widely used fully coupled climate model, with an ECS of about 3.2 °C when coupled to a slab ocean model (SOM) (Gettelman et al. 2012). This 1000-year simulation from the $4\times CO_2$ experiment was used to test and evaluate various methods for estimating ECS. The same model (with the same resolution) was also run for 500 years after an instantaneous doubling of the pre-industrial CO_2 ($2\times CO_2$). This $2\times CO_2$ experiment was analyzed to verify whether the ECS estimate is similar to that derived from the $4\times CO_2$ experiment after accounting for the initial forcing difference. A pre-industrial control (piControl) run with the same model configuration was also done to provide the baseline for calculating the change in the global-mean and annual-mean surface air temperature (ΔT) and TOA net radiation (ΔN, positive downward) for the $4\times CO_2$ and $2\times CO_2$ runs. The CESM1 control run did not show noticeable long-term drifts in the ΔT and ΔN time series.

To evaluate the performance of the four methods for estimating ECS, we also obtained and examined ten multi-millennium simulations from the long run project (https://data.iac.ethz.ch/longrunmip/; Rugenstein et al. 2019b), and from Dr. Cao Li of MPI and Dr. David Paynter of GFDL. Many of these simulations only had ≤ 3000 years of simulation and their TOA forcing is still positive near the end of the run. As a result, here we only included three long abrupt $4\times CO_2$ simulations from CESM 1.0.4 (for 5900 years, on a 2.5° lon × ~ 1.875° lat atmospheric grid with CAM4), GISS-E2-R (for 5000 years, on a 2.5° lon × 2.0° lat grid), and MPI-ESM-1.1 (for 4458 years, on a 1.875° lon × ~ 1.875° grid) in our analyses. These three long simulations show steady global-mean temperature toward the end of the simulation with TOA forcing less than 0.1 W/m^2, and thus the warming near the end can be used as a fairly accurate estimate of the true ECS, against which the ECS estimates from different methods and using different length of simulation can be evaluated.

We also analyzed the $4\times CO_2$ experiment (only ~ 150 years available after the quadrupling of pre-industrial CO_2), 1% CO_2 increase per year experiment (1pctCO_2) and a pre-industrial control run provided by 20 fully coupled climate models participated in the Coupled Model Inter-comparison Project phase 5 (CMIP5) (Table 1; Taylor et al. 2012) and 19 models from the phase 6 of the project (CMIP6) (Table 2, Eyring et al. 2016). The ECS estimates using Gregory et al. (2004)'s method and our new method for these models were compared, together with the estimates of their transient climate response (TCR), defined as the ΔT (relative to the climatology of the pre-industrial control run) around the time of CO_2 doubling (i.e., from years 61 to 80) in the 1pctCO_2 run.

2.2 Methods for estimating ECS from short-term simulations

After many tests, we applied the following four methods to estimate ECS using ΔT and ΔN data from a $4\times CO_2$ (and $2\times CO_2$) experiment. **Method 1** is the linear-fit method of Gregory et al. (2004), who utilized $\Delta N = F - \lambda\,\Delta T$ and used all annual data points from a $4\times CO_2$ experiment to estimate the $\lambda = (-dN/dT)$ slope using least-squares fitting and calculated ECS as $\theta\times F/\lambda$, where F is the global-mean TOA *effective* forcing for a quadrupling of pre-industrial CO_2, which is estimated as the intercept at $\Delta T = 0$ from the fitted equation as in Gregory et al. (2004) and Andrew et al. (2012), and θ is the $2\times CO_2$ to $4\times CO_2$ forcing ratio for converting the equilibrium ΔT change for $4\times CO_2$ to that for $2\times CO_2$. In this study, we use $\theta = 3.8749/8.1246 = 0.4769$ (based on the radiative forcing for $2\times CO_2$ and $4\times CO_2$ from Byrne and Goldblatt 2014), instead of 0.5 as in Gregory et al. (2004). Note $\theta = 1$ if data from a $2\times CO_2$ experiment are used. **Method 2** is similar to method 1, except that we exclude the first 40 years of data (for reasons discussed below) in the linear fitting: $\Delta N = a$—$b\,\Delta T$, so that $ECS = \theta\times a/b$. Note that parameter a is no longer equal to F. In **method 3**, we use the mean ΔT averaged over the last 50 years of the simulation (ΔT_{mean}) as the realized warming and ($\Delta N_{mean}/b$) as the unrealized warming, so that $ECS = \theta\times(\Delta T_{mean} + \Delta N_{mean}/b)$, where ΔN_{mean} is the ΔN averaged over the last 50 years of the simulation (referred to as the remaining forcing), and b is the same slope as in method 2. **Method 4** is similar to method 3, except that the slope b is not estimated directly from the model data using least squares fitting; instead, we first fit the $\Delta T(t)$ and $\Delta N(t)$ time series to analytic functions based on a simple two-layer climate model (described below), and then use these fitted equations to estimate the $b = (-dN/dT)$ slope for estimating ECS as $\theta\times(\Delta T_{mean} + \Delta N_{mean}/b)$. In methods 3–4, we separate ECS into the large realized component and a small unrealized component, and the estimated $(-dN/dT)$ slope only affects the estimation of the unrealized part; in contrast, the estimated ECS is entirely determined by the estimated $(-dN/dT)$ slope in methods 1–2.

For methods 1–3, the procedure to estimate ECS is fairly straightforward; however, for method 4 it requires

Table 1 The transient climate response (TCR, in K), and the equilibrium climate sensitivity (ECS, in K), climate feedback parameter (λ), and climate sensitivity parameter (1/λ) estimated using Gregory et al. (2004)'s method (with subscript "$_g$") and our new method (without subscript), and the fast and slow response time (τ_f and τ_s in Eqs. (1, 2)) for the 20 CMIP5 models with 140 or 150 years of $4\times CO_2$ simulations

Model	Transient climate response (TCR)	Climate feedback ParamGreg (λ_g)	Climate sensitivity ParamGreg ($1/\lambda_g$)	Eq. climate sensitivity gregory (ECS_g)	Climate feedback parameter (λ)	Climate sensitivity parameter (1/λ)	Eq. climate sensitivity method 4 (ECS)	Eq. climate sensitivity method 2 (ECS2)	ECS[1] (Geoffroy et al. 2013b)	Unrealized warming×0.4769	t_f (year)	t_s (year)
ACCESS1-0	1.881	0.814	1.229	3.598	0.534	1.873	4.285	4.083		1.696	3.079	306.929
ACCESS1-3	1.623	0.865	1.156	3.324	0.517	1.934	4.145	3.857		1.825	4.044	433.687
CanESM2	2.278	1.028	0.973	3.515	0.891	1.122	3.660	3.617	3.720	0.836	3.673	184.737
CCSM4	1.676	1.225	0.816	2.784	0.989	1.011	2.966	2.883	2.861	0.731	2.902	187.738
CNRM-CM5-1	2.034	1.134	0.882	3.091	1.657	0.604	2.861	3.072	3.052	0.443	5.353	262.749
CNRM-CM5-2	1.756	1.101	0.908	3.262	1.028	0.973	3.366	3.383		0.916	2.559	260.197
GISS-E2-H	1.734	1.671	0.598	2.309	1.433	0.698	2.396	2.359		0.426	2.263	160.475
GISS-E2-R	1.618	1.605	0.623	2.281	1.385	0.722	2.382	2.375	2.146	0.559	1.479	198.037
HadGEM2-ES	2.495	0.63	1.587	4.391	0.392	2.551	5.358	4.634	5.294	2.342	4.862	561.014
inmcm4	1.251	1.389	0.72	1.982	2.295	0.436	1.746	1.739	1.860	0.330	4.015	851.142
IPSL-CM5A-LR	2.008	0.741	1.35	3.979	0.610	1.639	4.280	4.192	4.054	1.500	5.652	311.191
IPSL-CM5A-MR	2.024	0.788	1.269	3.972	0.591	1.692	4.432	4.144		1.671	6.380	423.518
IPSL-CM5B-LR	1.511	1.006	0.994	2.514	0.692	1.444	2.848	2.596		0.839	3.459	221.880
MIROC5	1.476	1.503	0.665	2.604	1.411	0.709	2.660	2.662	2.671	0.619	2.593	256.216
MIROC-ESM	2.256	0.885	1.13	4.511	0.838	1.193	4.615	5.045		1.499	5.594	364.833
MPI-ESM-LR	2.000	1.107	0.903	3.488	0.826	1.211	3.830	3.808	3.720	1.021	2.941	215.571
MPI-ESM-MR	2.025	1.165	0.858	3.315	0.895	1.117	3.590	3.624		0.887	3.392	209.547
MPI-ESM-P	2.022	1.214	0.824	3.32	0.945	1.059	3.578	3.553		0.838	2.764	188.001
MRI-CGCM3	1.561	1.26	0.794	2.478	1.066	0.939	2.575	2.605	2.575	0.506	4.086	155.856
NorESM1-M	1.431	1.091	0.917	2.754	0.682	1.466	3.326	3.041	3.100	1.319	3.437	344.930
Multi-model Mean	1.833	1.111	0.960	3.174	0.984	1.220	3.445	3.364	3.187	1.040	3.726	304.912
Standard deviation	0.321	0.284	0.258	0.710	0.461	0.524	0.905	0.845	0.968	0.554	1.282	166.433

The CNRM-CM5-2, HadGEM2-ES, IPSL-CM5A-MR and MRI-CGCM3 had 140 years of simulations, while all others had 150 years of simulations. The model data were downloaded from https://cmip.llnl.gov/cmip5/data_portal.html. TCR = global-mean surface air T averaged years 61–80 of a 1%/year CO_2 run minus the control run climatology. λ_g (= −dN/dT) and ECS_g (= 0.4769*F/λ_g) are based on Gregory et al. (2004), where F is the TOA effective forcing for $4\times CO_2$ and is estimated as the intercept at T = 0. The use of 0.4769, instead of 0.5, led to slightly lower ECS_g than that in Table 9.5 of Flato et al. (2013). λ (= −dN/dT) and ECS are based on our new method 4

Table 2 The transient climate response (TCR, in K), and the equilibrium climate sensitivity (ECS, in K), climate feedback parameter (λ), and climate sensitivity parameter (1/λ) estimated using Gregory et al. (2004)'s method (with subscript "g") and our new method (without subscript, right three columns)) for the 19 CMIP6 models using their abrupt 4×CO_2 simulations

Model	Transient climate response (TCR)	Climate feedback ParamGreg (λ_g)	Climate sensitivity ParamGreg ($1/\lambda_g$)	Eq. climate sensitivity gregory (ECS_g)	Climate feedback parameter (λ)	Climate sensitivity parameter (1/λ)	Eq. climate sensitivity method 4 (ECS)	Eq. climate sensitivity method 2 (ECS2)	Unrealized warming ×0.4769	t_f (year)	t_s (year)
AWI-CM-1-1-MR	2.124	1.192	0.839	2.997	0.965	1.036	3.138	3.035	0.533	3.261	150.116
BCC-CSM2-MR	1.654	1.164	0.859	2.726	0.775	1.290	3.070	2.996	0.807	3.905	196.261
BCC-ESM1	1.794	1.110	0.901	2.933	0.700	1.429	3.394	3.299	1.025	4.122	252.480
CAMS-CSM1-0	1.760	1.653	0.605	2.220	1.400	0.714	2.287	2.223	0.289	2.677	127.626
CanESM5	2.826	0.702	1.425	5.232	0.586	1.706	5.579	5.452	1.758	6.047	251.293
CESM2-WACCM	1.968	0.699	1.431	4.475	0.472	2.117	5.307	5.259	2.244	4.056	314.745
E3SM-1–0	3.068	0.694	1.440	4.919	0.384	2.605	5.933	5.487	1.874	7.694	252.095
EC-Earth3-Veg	2.639	0.838	1.193	4.042	0.667	1.499	4.274	4.150	0.780	4.149	123.343
FGOALS-f3-L	2.034	1.188	0.842	2.962	0.953	1.050	3.150	3.017	0.706	3.024	197.586
GISS-E2-1-G	1.751	1.427	0.701	2.593	1.174	0.852	2.763	2.621	0.806	2.318	355.296
GISS-E2-1-H	1.948	1.245	0.803	2.896	0.979	1.021	3.094	2.931	0.728	3.611	203.177
GISS-E2-2-G	1.738	0.441	2.266	2.207	1.005	0.995	2.083	2.155	0.310	3.334	427.284
INM-CM4-8	1.354	1.718	0.582	1.714	1.148	0.871	1.792	1.774	0.113	2.371	74.495
INM-CM5-0	1.428	1.789	0.559	1.784	1.051	0.952	1.955	1.900	0.296	2.916	156.578
MIROC6	1.564	1.491	0.671	2.453	1.121	0.892	2.647	2.461	0.622	3.361	470.750
MPI-ESM1-2-h	1.091	1.200	0.833	2.873	0.939	1.065	3.162	3.080	1.122	3.714	319.938
MRI-ESM2-0	1.605	0.732	1.367	2.978	0.755	1.324	3.330	3.186	1.091	3.039	254.455
NESM3	2.804	0.837	1.195	4.447	0.805	1.242	4.500	4.541	1.123	2.471	219.039
SAM0-UNICON	2.280	1.011	0.989	3.598	0.873	1.146	3.798	3.817	1.185	3.393	314.899
Multi-model mean	1.970	1.112	1.026	3.161	0.882	1.253	3.435	3.336	0.916	3.656	245.340
Standard deviation	0.539	0.385	0.424	1.023	0.255	0.469	1.200	1.158	0.564	1.296	103.675

The FGOALS-f3-L, MPI-ESM1-2-h had 160 years of simulations, MIROC6 had 250 years of simulations, and all others had 150 years of simulations. The model data were downloaded from https://esgf-node.llnl.gov/search/cmip6/. TCR = global-mean surface air T averaged years 61–80 of a 1%/year CO_2 run minus the control run climatology. λ_g (= −dN/dT) and ECS_g (=0.4769*F/λ_g) are based on Gregory et al. (2004), where F is the TOA effective forcing for 4×CO_2 and is estimated as the intercept at ΔT = 0. The use of 0.4769, instead of 0.5, led to slightly lower ECS_g than that in Table S1 of Zelinka et al. (2020). λ (= −dN/dT) and ECS are based on our new method 4

some iterations. First, we derive a first ECS guess using method 3, then we use this ECS/θ as the α parameter in the fitting function for ΔT(t) (i.e., Eq. 1 in Sect. 2.3) and start the fitting (with α being specified). The fitted functions will allow us to derive a new α = ECS/θ using method 4 (which is usually larger than the previous estimate). If this new α estimate differs from the previous estimate by more than a threshold of 0.0001 K, then we replace the α parameter in the ΔT(t) fitting function with this new estimate and repeat the fitting (again with α being specified) and the ECS estimate (and α) again, until the new estimate stabilizes (usually within 50 iterations).

It becomes apparent that how one estimates the $b = (-dN/dT)$ slope is critical for deriving ECS. To explore this issue more, we re-visit some basics of the least squares linear fitting in Appendix. It is clear that for noisy data series $X_i = x_i + \varepsilon_{xi}$ and $Y_i = y_i + \varepsilon_{yi}$, where ε_{xi} and ε_{yi} are random noise added to the signal x_i and y_i, which are linearly related: $y_i = a + b\, x_i$, the estimated slope between X_i and Y_i with X_i as the predictor would be $b/(1 + \sigma^2_{\varepsilon x}/\sigma^2_x)$, where $\sigma^2_{\varepsilon x}$ and σ^2_x are the variance of the noise (ε_{xi}) and signal (x_i) of the predictor. Note here ε_x and ε_y are assumed to be uncorrelated, which may not be the case for certain climate data (Proistosescu et al. 2018). Thus, the magnitude of the estimated slope from noisy data will always be smaller than that of the true slope between the two variables! And this underestimation increases with the noise-to-signal ratio ($\sigma_{\varepsilon x}/\sigma_x$). For the data without noise (i.e., x_i and y_i), the dy/dx and dx/dy slopes are related: $dy/dx = b = 1/(1/b) = 1/(dx/dy)$. However, when they contain noise, such a relationship no long exists: $dY/dX = b/(1 + \sigma^2_{\varepsilon x}/\sigma^2_{\varepsilon x})$, while $dY/dX = b/(1 + \sigma^2_{\varepsilon y}/\sigma^2_{\varepsilon y})$. Because the noise-to-signal ratio is generally not the same for X and Y, the estimated dY/dX slope will not be equal to the inverse of the estimated dX/dY slope!

These results have important implications for our analysis of the ΔT(t) and ΔN(t) time series from the CESM1 or CMIP models, as these data series (Fig. 1) contain considerable noise generated from internal variability (Dai and Bloecker 2019) superimposed on the CO_2 forcing-induced long-term change (the signal). The noise-to-signal ratio is particularly large for the ΔN(t) data series, thus the underestimation will be more pronounced for the (−dT/dN) slope (i.e., when ΔN(t) is used as the predictor) than for the (−dN/dT) slope. Because of this, one should avoid using the estimated (−dT/dN) slope to derive ECS. Similarly, since the noise level is similar while the forced signal is almost doubled in $4\times CO_2$ experiments compared with $2\times CO_2$ experiments, the ECS estimate based on $4\times CO_2$ experiments should contain less uncertainty than that based on $2\times CO_2$ experiments. However, it is unclear whether the simple rescaling using θ first induced by Gregory et al. (2004) and used previously (e.g., Andrews et al. 2012) would lead to the same ECS estimates from the two types of experiments. We will apply our methods to both the CESM1 $4\times CO_2$ and $2\times CO_2$ experiments to examine this issue in Sect. 3.3.

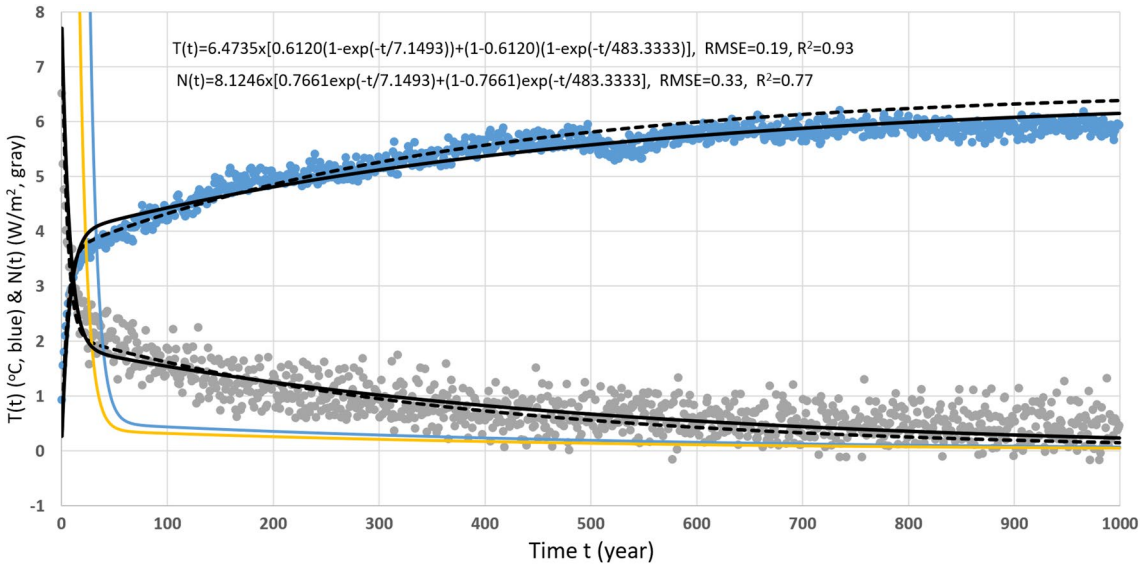

Fig. 1 Time series of anomalies (relative to the control climate) of global-mean and annual-mean surface air temperature (T(t), blue dots) and TOA net radiation flux (N(t), gray dots, positive downward) from the CESM1 $4\times CO_2$ experiment for 1000 years. The solid (dashed) black line is the best fit to the 1000 (first 500) years of data points using the two-layer model solutions of Geoffroy et al. (2013b), with the fitting results for the 1000 year case shown on the figure. The blue (orange) line is the dT/dt (−dN/dt) slope based on the fitted functions (solid black lines) in unit of K/100 year (W/m^2 per 100 years) on the left y-axis

Because the estimated ECS is proportional to $(1+\sigma_{\varepsilon x}^2/\sigma_x^2)/b$ in our methods 1–2, the underestimation of the (−dN/dT) slope should lead to an overestimation of the ECS in these methods. However, as shown below, this bias is smaller than the overestimation bias of the (−dN/dT) slope due to the use of the data points from the early part of the simulation. As a result, ECS is actually underestimated by these methods.

2.3 Fitting the model data

To describe the *long-term* behavior of the ΔT(t) and ΔN(t) data series from a model experiment and also to help devise a framework for testing and evaluating the performance of the four methods for estimating ECS, we need to fit the ΔT(t) and ΔN(t) time series from the CESM1 or the CMIP models with certain analytic equations. We explored a few different forms of the fitting functions based on visual examinations of the graphic shapes of the ΔT(t) and ΔN(t) time series. They include the solutions from the two-layer energy balance model with a fast (τ_f) and a slow (τ_s) response time (Geoffroy et al. 2013a, b):

$$\Delta T(t) = \alpha\left[\beta\left(1-e^{-\frac{t}{\tau_f}}\right)+(1-\beta)\left(1-e^{-\frac{t}{\tau_s}}\right)\right], \tag{1}$$

$$\Delta N(t) = F_o\left[Be^{-\frac{t}{\tau_f}}+(1-B)e^{-\frac{t}{\tau_s}}\right], \tag{2}$$

where F_o is the initial TOA forcing and is set to 8.1246 W/m^2 for 4×CO_2 and 3.8749 W/m^2 for 2×CO_2 according to Byrne and Goldblatt (2014). Note that F_o, which differs from the effective forcing F in Gregory et al. (2004)'s method, cannot be easily estimated from the model output from the experiments because the TOA forcing changes rapidly at the start of the simulations. Hansen et al. (2005) proposed a 2×CO_2 experiment with fixed SSTs and sea ice from a control run to estimate the TOA forcing as a measure of the direct radiative forcing (without feedbacks) from a doubling of CO_2. We performed two 40-year simulations using the CESM1 with fixed SSTs and sea ice from the piControl run, one with an instantaneous doubling and one with an instantaneous quadrupling of atmospheric CO_2. Results from these runs (not shown) revealed steady evolutions for the global-mean ΔT and ΔN, with a slight increase in ΔT by 0.19 K and 0.55 K for the 2×CO_2 and 4×CO_2 cases, respectively. The TOA forcing estimated from these runs is 4.1384 W m^{-2} and 8.0257 W m^{-2} for 2×CO_2 and 4×CO_2, respectively, which are reasonably close to the generic estimates from Byrne and Goldblatt (2014). Given the uncertainties associated with the Hansen experiment approach (e.g., increased water vapor, snow-ice feedback over land and other processes may still contribute to the TOA forcing in such experiments), practical difficulties to make such experiments for other CMIP models, and the insensitivity of the fitted $\Delta N(t)$ to F_o for large t (e.g., for $t>100$ years), we chose to use the F_o estimates from Byrne and Goldblatt (2014) for CESM1 and all other CMIP models. Our tests showed that using slightly different F_o (e.g., $F_o=8.1246$ or 8.0257 W m^{-2}) did not change the ECS estimate or the fitted $\Delta N(t)$ for large t values. This is because Eqs. (1, 2) are not used in methods 1–3, and used in method 4 only for estimating the small component of the unrealized warming.

The other parameters α, β, τ_f, τ_s, and B in Eqs. (1, 2) need to be estimated during the fitting to the model data. Note that the use of a different B in Eq. (2) allows for a non-unit efficacy of deep ocean heat uptake in the two-layer model (Geoffroy et al. 2013b); otherwise, the B would be the same as the β in Eq. (1) and the fitting to the ΔN data would be poor for the first ~ 40 years. This suggests that a non-unit efficacy of deep ocean heat uptake is necessary for the two-layer model to work well. The parameter α in Eq. (1) is the equilibrium temperature change (thus ECS = θ×α). Our tests showed that it is better to estimate this parameter first (e.g., using the methods of Sect. 2.2) before the optimization-based fitting described below; otherwise, this parameter (and thus ECS) will vary considerably with different lengths of simulation, which is an undesirable property for any method.

The two-layer model has been shown by Geoffroy et al. (2013b) to fit well the global-mean ΔT(t) and ΔN(t) time series from CMIP5 models. This is also confirmed by our own analyses discussed in Sect. 3.4. However, a new study by Rohrschneider et al. (2019) suggests that the two-layer model may not be able to accurately describe the behavior of complex climate models, as it only contains two response time scales. Thus, some uncertainties may exist in using a fitted two-layer model to represent the long-term evolution of the ΔT(t) and ΔN(t) time series in fully coupled models, although our addition to the two-layer model of the short-term variations derived from the CESM1 data series should improve the representation in our benchmark tests discussed below.

From Eqs. (1, 2), we can derive the λ = (−dN(t)/dT(t)) slope as

$$\lambda = -\frac{dN(t)}{dT(t)} = \frac{F_o\left(\frac{B}{\tau_f}e^{-t/\tau_f}+\frac{1-B}{\tau_s}e^{-t/\tau_s}\right)}{\alpha\left(\frac{\beta}{\tau_f}e^{-t/\tau_f}+\frac{1-\beta}{\tau_s}e^{-t/\tau_s}\right)} \approx \frac{F_o(1-B)}{\alpha(1-\beta)} \text{ for } t \gg \tau_f \text{ and } \tau_s \gg \tau_f. \tag{3}$$

This implies that the feedback parameter (λ) becomes constant for $t \gg \tau_f$ (e.g., for t > 40–50 years, see Fig. 3a) for the two-layer model, which excludes short-term variations generated by internal variability.

The following cost function is used to determine the parameters in the fitting functions:

$$\ell = \frac{1}{2}\sum_{t=1}^{N}\left(\mathbf{y}(t) - \mathbf{y}_{obs,t}\right)^2, \tag{4}$$

where $y(t) = \Delta T(t)$ or $\Delta N(t)$ computed from the fitting function at time t, and $\mathbf{y}_{obs,t}$ is the data time series (with N data points) from the CESM1 or CMIP model experiments. Thus, the optimized parameters can be computed iteratively by performing evaluations of the cost function and its gradient using a suitable descent algorithm (Liu and Nocedal 1989). This allows us to derive the *optimal combination* of the parameters in any given form of a fitting function that minimizes the cost function (i.e., in a least-square sense).

2.4 A framework for evaluating ECS estimates

An ideal method for estimating ECS using the ΔT(t) and ΔN(t) data from a limited length of simulation should produce stable estimates as the length of simulation (L) varies above a minimum number of years. Thus, we will apply the four methods to the ΔT(t) and ΔN(t) data over varying lengths of simulation (i.e., from year 1 to year L only), with $L = 100, 110, \ldots, 1000$ years. The stability of the ECS estimates over varying L will provide one measure of the performance for the methods.

The best way to quantitatively evaluate the performance of the methods for estimating ECS using short simulations is to compare their estimates with the realized warming near the end of a multi-millennium simulation, as the latter is likely very close to the true ECS. We were able to obtain three long (~5000 year) abrupt $4\times CO_2$ simulations from three different coupled models in this testing. This provides additional verification of the test results using synthetic data described below.

Another quantitative evaluation is to examine how well a method can recover a pre-defined ECS (e.g., as represented by Eq. 1) in noisy data series that resemble model data. Please note that this specified ECS may slightly differ from the true ECS of the model that was used to produce the model data used in the fitting. To do so, we notice that the two-layer model solutions (Eqs. 1, 2) provide a reasonable fit to the noisy ΔT(t) and ΔN(t) data from the CESM1 (Fig. 1) and other CMIP5 $4\times CO_2$ experiments (see Sect. 3.5). The residuals from this fitting to all years of data looked like noise (not shown), with some lag-1 autocorrelations ($r_1 = 0.63$ and 0.17 for ΔT and ΔN, respectively, over years 1–500). Thus, we generated 10,000 synthetic time series for ΔT and ΔN (for each given $L = 100, 110, \ldots, 1000$ years) by adding random noise to the smooth time series calculated using the fitted functions (fitted to all years of simulation, i.e., for $L = 1000$ years), which define ECS (as $\theta \times a$) and the initial forcing (F_o) in the noisy synthetic data. We emphasize that this pre-defined ECS is not intended to represent the true ECS of the CESM1 (which is unknown); rather, it serves as the target ECS for a method to recover from a noisy data series that closely *resembles* the data series from a fully coupled climate model. The random noise was generated by randomly sampling the pairs of the ΔT and ΔN residual time series (so that their correlation of about −0.53 was preserved) either for one year each time or for a 10-year block each time. The 10-year block sampling largely preserves the lag-1 autocorrelation in the residual time series and the results presented below use this sampling; however, the results are very similar for the yearly sampling case. For a short length of simulations (L = 100–200 years), the inter-ensemble spread in the estimated ECS is considerable (around 0.1–0.2 K). The ensemble-mean results from the 10,000 samples provide a more stable estimate of the performance of the methods than the results based on a single realization from the CESM1 or the other three models. Thus, the ensemble-mean results based the synthetic data are useful for assessing the mean performance of the methods, provided that the synthetic data resemble real model data.

For method 4, the fitting procedure of Sect. 2.3 with the same fitting function forms as those used to generate the synthetic data was applied to each of the noisy synthetic time series (with the iterations described in Sect. 2.2) to estimate the (−dN/dT) slope needed for calculating ECS in this method. For example, when we use Eqs. (1, 2) to fit the data over all available years of simulation (i.e., 1000 years) and then use these fitted functions to generate 10,000 pairs of synthetic ΔT and ΔN time series for each length of simulation ($L = 100, 110, \ldots, 1000$ years), we would use Eqs. (1, 2) to fit each of these synthetic time series at the given L iteratively until parameter α in (1) stabilizes. The newly fitted ΔT(t) and ΔN(t) functions (rather than the functions used in generating the synthetic data series) were then used to estimate the (−dN/dT) slope using Eq. (3) that was used in method 4 to estimate ECS for the given L.

We emphasize that the ECS biases for the methods may vary with data from individual model simulations, and they will likely differ from the biases estimated here using the synthetic data. However, we think the relative performance shown in our tests is likely applicable to actual model data because our synthetic data closely resemble the noise level and the long-term evolution of the ΔT and ΔN in a real model simulation by the CESM1. Although the fitted functions imply a constant slope of λ = (−dN/dT) after ~100 years (cf. Eq. 3), the noise added to the analytic solution could still generate some short-term variations in λ in our synthetic data. Thus, our synthetic data, which include the solution from the two-layer model and a realistic noise derived from CESM1, are

not necessarily inconsistent with the notion that the dN/dT slope may vary over time due to time- or warming-dependent feedbacks.

3 Results

3.1 The varying relationship between ΔT and ΔN

Since one needs to use the slope between ΔT and ΔN to extrapolate the ΔT when ΔN = 0 for estimating ECS, it is helpful to examine the various estimates of the (−dN/dT) and (−dT/dN) slopes using the CESM1 simulations, although their influence is smaller in methods 3–4 than in methods 1–2 as the (−dN/dT) slope is used only to estimate the small unrealized warming in methods 3–4. Figure 2 shows that the (−dN/dT) slope is steepest during the first several decades when the fast response (i.e., the first term in Eq. 1) dominates; the slope then flattens as the simulation continues. Previous studies have shown that changing oceanic heat uptake patterns, atmospheric stability, cloud feedback, and other processes can lead to this increased climate sensitivity over time (e.g., Jonko et al. 2013; Rose and Rayborn 2016; Stevens et al. 2016; Armour 2017). Clearly, including the data points from the early decades (green line in Fig. 1) would steepen the (−dN/dT) slope, thus leading to a smaller intercept on the x-axis and therefore a lower ECS estimate. Furthermore, we noticed that adding additional years of data (e.g., adding years 501–1000) did not greatly change the slope of the green line in Fig. 2. This suggests that this slope is determined mainly by the data points from

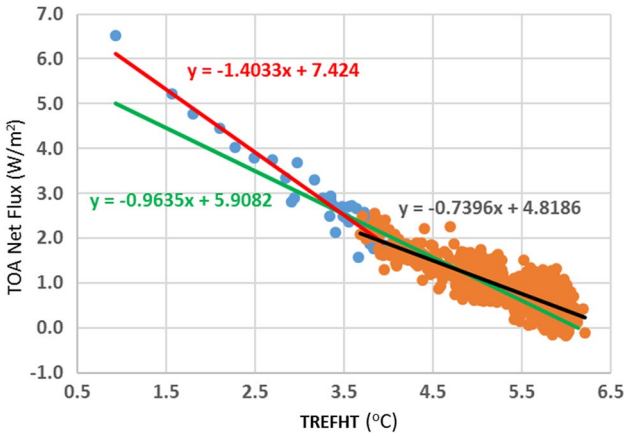

Fig. 2 Scatter plot of the anomalies (relative to the control-run climate) of the global-mean and annual-mean surface air temperature (TREFHT, x axis) and top-of-atmosphere (TOA) net radiation flux (y axis, positive downward) from year 1 to year 40 (blue dots, red line = linear fit) and year 41 to year 1000 (red dots, black line = linear fit) from the CESM1 4×CO_2 simulation. The green line is the linear fit to the data from year 1 to year 1000. The regression equation (in the same color as the line) is also shown

the first few hundred years, especially the first few decades. Such a change in the (−dN/dT) slope over time has been noticed previously (e.g., Andrews et al. 2015; Gregory et al. 2015; Rose and Rayborn 2016; Rugenstein et al. 2019b), while Armour (2017) excluded the first 20 years in his regression.

To explore the temporal evolution of the (−dN/dT) and the (−dT/dN) slopes in more detail, in Fig. 3 we show the time series of the (−dN/dT) and (−dT/dN) slopes estimated using local 50 years of data (blue line) from the CESM1 4×CO_2 experiment, using the data from year 1 to the plotted year (green line), and using the data from year 41 to the plotted year (red line). They are compared with the slopes calculated using the two-layer model solutions fitted to the CESM1 output (black line), which may be considered as the slopes between the CO_2-induced long-term ΔT and ΔN changes (referred to as the signals here). It is clear that the local estimates of the slopes are highly variable and differ greatly from the slopes between the long-term ΔT and ΔN signals, presumably due to the changing effects of the internal climate variability (referred to as the noise, Dai and Bloecker 2019), such as the decadal variations in Arctic sea–ice cover (Fig. 4) and the associated amplification of Arctic warming (Dai et al. 2019).

Figure 4 shows that Arctic sea ice, whose melting amplifies surface warming (Screen and Simmonds 2010; Dai et al. 2019), exhibits large fluctuations on 5- to 200-year time scales. These internally-generated sea–ice fluctuations cause similar variations not only in Arctic surface air temperature, but also in global-mean surface air temperature; while the TOA net radiation (N) does not vary closely with the sea ice fluctuations. Thus, the sea–ice induced T variations will likely affect the feedback parameter (λ) on those time scales. This differs from the impacts from the cloud, lapse rate and water vapor feedbacks that may change over time in response to external forcing or among different climate models (Andrews et al. 2015). However, the sea–ice's effect appears to be small on the long-term relationship between global-mean ΔT and ΔN, as the λ = (−dN/dT) slope in the 2-layer model solutions becomes nearly constant after year ~ 40–50 (Fig. 3), and the λ in multi-millennium simulations shows little trend (see Fig. 10 below). In other words, internal variability may induce short-term (1–10^2 year) variations in the feedback parameter λ = (−dN/dT) when it is estimated using local ΔT and ΔN data. However, these large short-term variations may not be important for the eventual warming reached by the system. This is because the overall long-term evolution is reasonably characterized by the fitted two-layer model solutions (Fig. 1), and they imply a near constant λ after about year 40–50 (Fig. 3). The (−dN/dT) slope estimated using all data points after year 40, which effectively filters out the temporal variations, also shows little change for simulations of 150 years or longer (red line

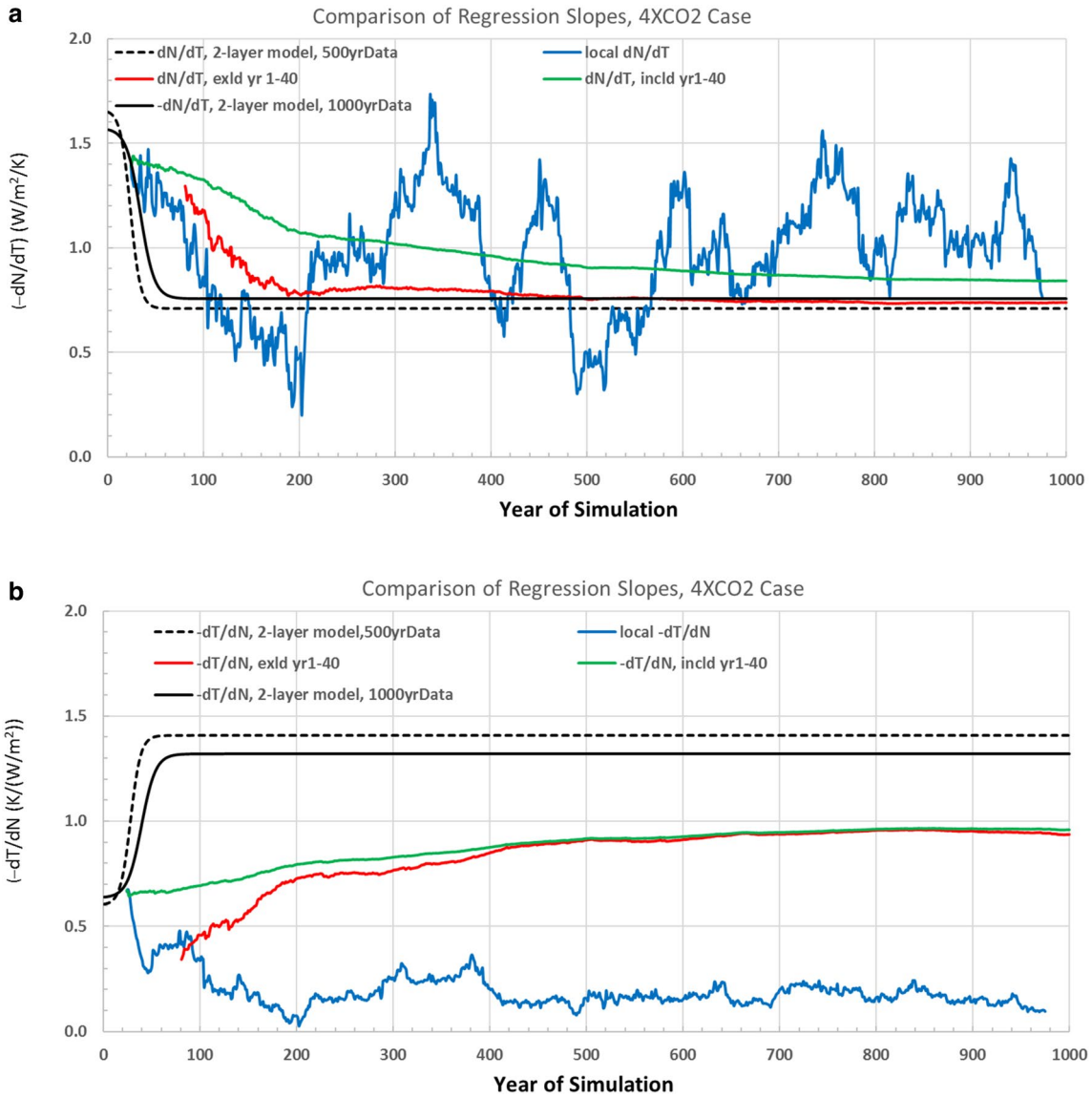

Fig. 3 Comparisons of the least-squares estimates of the **a** (−dN/dT) and **b** (−dT/dN) slopes using annual and global-mean TOA net flux (N, W/m^2, positive downward) and surface air temperature (T, K) data from year 1 to the plotted year of the CESM1 4×CO_2 simulation (labeled as "incld yr1-40") and using data from year 41 to the plotted year of simulation (labeled as "exld yr1-40"). Also shown is the (−dN/dT) or (−dT/dN) slope (blue line) estimated using the local 50 years of data (centered at the plotted year). The black solid (dashed) line is the (−dN/dT) or (−dT/dN) slope calculated using the fitted 2-layer model functions based on the 1000 (first 500) years of data shown in Fig. 1. The difference between the green and red lines reflects the effect of the first 40 years. The black lines stabilize approximately after about year 60

in Fig. 3a). Thus, the estimated local −dN/dT slopes are not suitable for estimating ECS; instead, the −dN/dT slope estimated using all data points after year 40 is a better choice for estimating ECS as it mainly reflects the forced long-term relationship between dT and dN. This conclusion is confirmed by our analyses of the multi-millennium simulations (Sect. 3.3).

These results seem to suggest that the unrealized warming and thus ECS estimate depend primarily on the long-term (10^2–10^3 year) relationship between ΔT and ΔN that is likely to be determined largely by processes involving the deep oceans, whereas the short-term (1–10^2 year) changes in various feedback processes (e.g., ice-albedo feedback) may not be important for estimating the long-term (−dN/dT) slope and thus ECS. If this is true, then the various factors that have been found to influence the feedback parameter as discussed in the Introduction may not be important for *estimating* the equilibrium warming. One exception is the strengthening of the water–vapor feedback with increasing surface temperature (Meraner et al. 2013), which should

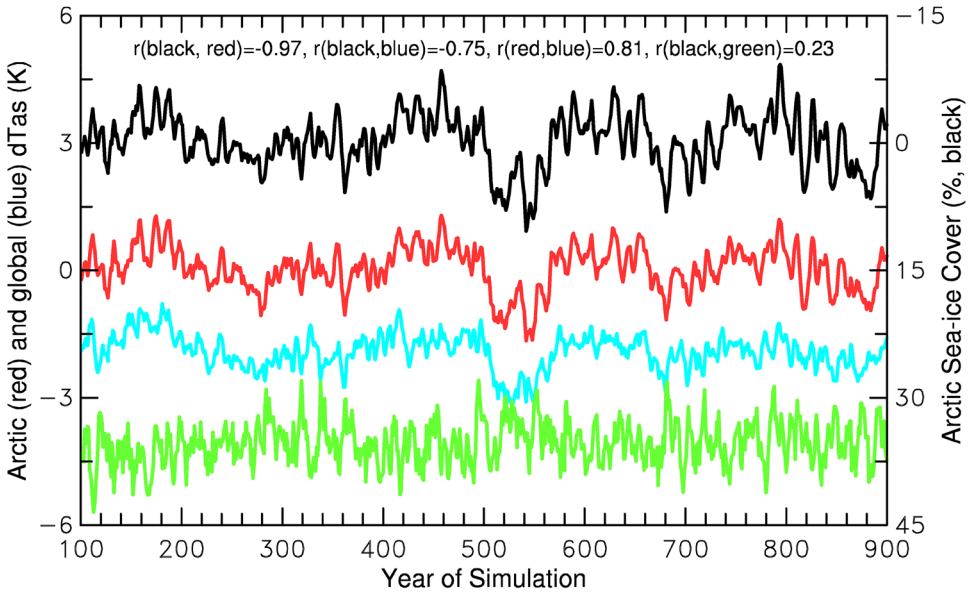

Fig. 4 Time series of the 5–200 year variations in Arctic (north of 67° N, red) and global (blue) annual surface air temperature, Arctic annual sea–ice cover (black line), and global TOA net energy flux (green, W/m^2) from the CESM1 4×CO_2 experiment, with the 5-year and 201-year moving averages being removed from the original data series. The global temperature and TOA energy flux anomalies were multiplied by a factor of five and shifted downward by 2 and 4 units, respectively, in order to use the left y-axis and be separated from the red line. The correlations between a pair of the lines are shown

work on long time scales; however, the unrealized warming after the first 200 years is relatively small (Fig. 1), thus the effect of this strengthening on estimating ECS in our methods 3–4 should be small after for $L > 200$ years. Our analyses of the multi-millennium simulations (Sect. 3.3) also seem to support these arguments.

Figure 3 also shows that including the first 40 years of the model data in estimating the (−dN/dT) or (−dT/dN) slope would considerably increase the estimated slope compared with the case excluding the first 40 years, even for long lengths (> 200 years) of simulation. Furthermore, all the (−dT/dN) estimates are poor approximation of the (−dT/dN) slope between the ΔT and ΔN signals (black lines in Fig. 3b), thus making them a poor choice for estimating the ΔT when ΔN = 0. On the other hand, the (−dN/dT) slope (red line in Fig. 3a) estimated using the data from year 41 to the end of simulation (i.e., the year on the x-axis in Fig. 3) is fairly close to the slope between the signals (black lines in Fig. 3a) when L exceeds about 180 years, making this a comparatively better choice for estimating ECS. Figure 3 also shows that the −dN/dT (−dT/dN) slope from the fitted two-layer model solutions (i.e., between the signals) decreases (increases) rapidly during the first few decades but stabilizes approximately after 40–50 years. Thus, one should exclude the first 40 years or so in estimating the (−dN/dT) slope for extrapolating the ΔT as the ΔN approaches zero.

In Fig. 5, we further evaluate the various estimates of the (−dN/dT) and (−dT/dN) slopes using 10,000 samples of the synthetic data described in Sect. 2.4, in contrast to Fig. 3, in which only one sample of data from the CESM1 4×CO_2 simulation was used. Figure 5a confirms that the Gregory et al. (2004)'s method (green line) considerably overestimates the pre-defined (−dN/dT) slope (and thus underestimates the ECS) embedded in the noisy synthetic data by ~ 100% for $L = 100$ years to about ~ 20% for $L = 1000$ years. Excluding the first 40 years (red line in Fig. 5a) reduces this bias substantially, especially for $L > 150$ years; however, it still overestimates the target (−dN/dT) slope substantially (by about 30–100%) for $L < 200$ years and slightly (< 10%) for $L > 300$ years. On the other hand, fitting the data with the two-layer model (Eqs. 1, 2) first and then estimating the slopes using the fitted functions (as in method 4) reproduce the pre-defined slope in the noisy synthetic data for L > 600 years, although the spread of the estimates is large for $L < 500$ years (Fig. 5a). Similarly, method 4 reproduces the (−dT/dN) slope well while the other two methods substantially underestimate this slope (Fig. 5b). Furthermore, accounting for the underestimation due to the existence of the noise in the data (see Sect. 2.2; i.e., comparing to the blue line in Fig. 5) would increase the mean bias for the (−dN/dT) slope in Fig. 5a, but change the bias to be positive for the (−dT/dN) slope in Fig. 5b for methods 1–3. Since the ECS estimate is proportional to 1/(−dN/dT) or (−dT/dN), an overestimation of (−dN/dT) would lead to an underestimation of ECS. Thus, method 4 is recommended for estimating the (−dN/dT) and (−dT/dN) slopes in the ΔT and ΔN data from climate model simulations.

We noticed that the fitting to the first 500 years of data (dashed black lines in Fig. 1) was better (i.e., with smaller RSME) than the fitting to the 1000 years of data (solid black lines in Fig. 1), partly due to the noticeable drop in the ΔT series several years after year 500 and after year 800 (blue dots in Fig. 1; also evident in Fig. 4), which likely resulted from internal variability. As a result, the residual series from the 500-year fitting looks more random than from the 1000-year fitting, and the slopes

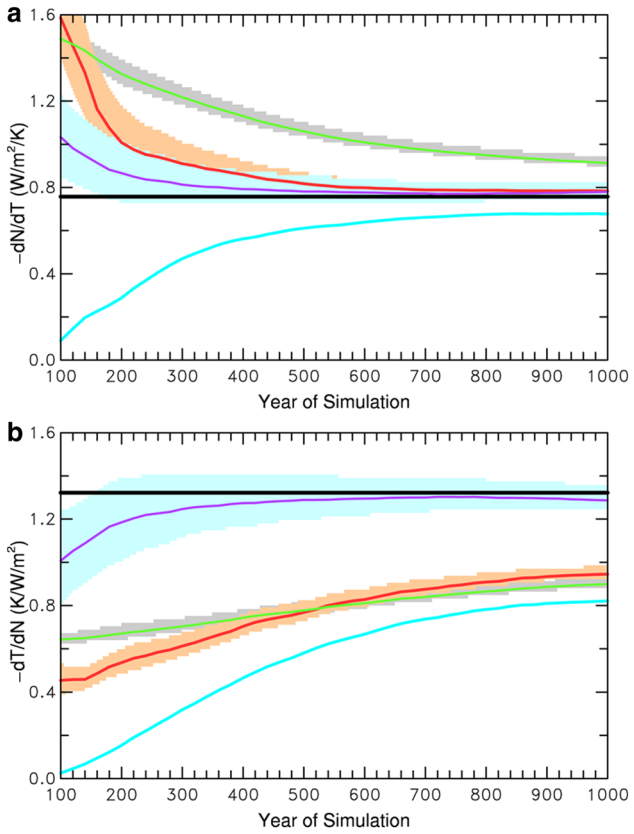

Fig. 5 **a** The −dN/dT slopes calculated using the two-layer model solutions (black line) for the plotted year on the x-axis, and estimated using synthetic data from year 1 to the plotted year (green line) and from year 41 to the plotted year (red line). The synthetic data were derived by combining the two-layer model solutions (fitted to the CESM1 model data over years 1–1000) with random noise sampled from the residual time series from the fitting. The magenta line (very close to the black line after year 600) is the –dN/dT slope estimated using the same synthetic data but with Method 4 (i.e., using the two-layer model solutions fitted to the synthetic data up to the plotted year, see text for details). The black line is the target (i.e., predefined) slope embedded in the noisy synthetic data for each of the methods to recover. Shading around the green, red and magenta lines indicates the ± 1 SD range estimated from 10,000 samples of the synthetic data. The blue line is the black line multiplied by the correction factor of $1/\left(1+{}^{S_\varepsilon}/_{S_s}\right)$, where S_ε and S_s are the variance of the noise and signal (i.e., changes following the two-layer model solution for T). T and N are the global-mean and annual-mean surface air temperature and TOA net radiation, respectively. **b** Same as **a** but for the −dT/dN slope

estimated by method 4 in Fig. 5 would have converged sooner to the target values if only the first 500-years of data were used. This also applies to the ECS estimates shown in Fig. 6; that is, the ECS from method 4 (dashed black line in Fig. 6) would have converged to the target ECS (solid black line in Fig. 6) sooner (around $L = 180$ years instead of 600–700 years) if the first 500 years of data were used. Nevertheless, both cases (using either the first 500 or 1000 years of data) suggest that method 4 recovers the target slopes (and ECS) much better than the other methods.

3.2 Evaluation of the ECS estimates using synthetic data and CESM1 simulation

Figure 6 shows how well the four methods recover the predefined ECS (represented by the solid black line) embedded into the 10,000 samples of the noisy synthetic data for each L of 100, 110, …, 1000 years. Consistent with the biases in the slope estimates shown in Fig. 5, method 4 (dashed black line in Fig. 6) clearly outperforms the other methods, including methods 2–3, especially for $L < 400$ years, with minimum (< 0.03 K) biases for $L > 600$ years (180 years if only the first 500 years of data used). In contrast, Gregory et al. (2004)'s method (red lines in Fig. 6) substantially underestimates ECS by about 0.6 K for $L = 100$ years to 0.08 K for $L = 1000$ years, even for the case without noise in the data! This is simply due to the large positive bias in its estimate of the (−dN/dT) slope due to the use of the data for the first 4 decades. Excluding these 40 years in method 2 improves the ECS estimate substantially (blue line in Fig. 6), and making use of the realized warming in method 3 further improves the ECS estimate slightly (magenta line Fig. 6). However, these two methods still underestimate the target ECS, especially for $L < 400$ years.

For the single sample of data from the CESM1 $4\times CO_2$ experiment, the estimated ECS is relatively stable around 3.12 ± 0.06 K for $L > 450$ years for methods 3–4, while it increases with L from about 2.6 K at $L = 100$ years to 3.05 K for $L = 1000$ years for method 1 (red line in Fig. 7). Method 2 (blue line in Fig. 7) produces much improved smooth estimates of ECS compared with method 1, but its ECS estimate still slightly increases with L even after year 450 and is generally below that from methods 3–4 for L = 300–800 years (except for a few decades around year 560 and during year 840–920). The small variations in the ECS from methods 3–4 results from result from variations in the 50-year average for the realized warming (T_{mean}) and remaining forcing (N_{mean}). From Fig. 1, we know that the realized warming by the end of the 1000-year simulation is around 6 K. Using a conversion ratio of 0.485 (see below), this means that the true ECS for the CESM1 should not be less than 2.91 K. Thus, method 1 underestimates this ECS considerably, especially for short L (e.g., < 300 years), as its own ECS estimates increase substantially with L. From this perspective, methods 2–4 are better than method 1 as they produce more stable ECS estimates for $L >$ ~ 300 cases. While methods 2–4 show similar performance for the single sample from the CESM1, the large ensemble-mean results based on the

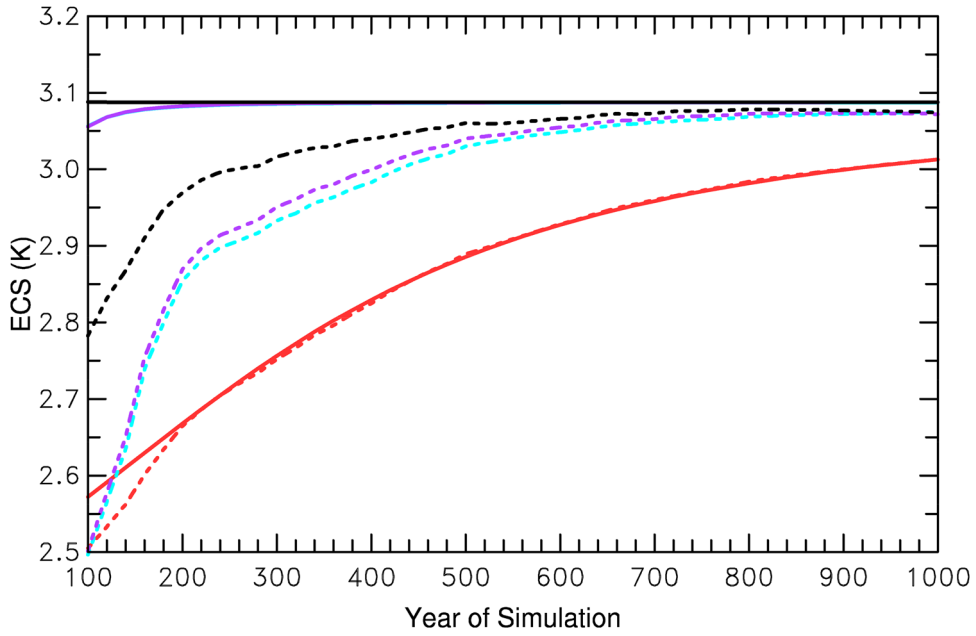

Fig. 6 Ensemble mean of the equilibrium climate sensitivity (ECS) (as a function of the ending year of the synthetic data period) estimated using the Gregory et al. (2004)'s linear fitting method (Method 1, red lines), new Method 2 (blue line), Method 3 (magenta line), and Method 4 (black lines) (see Sect. 2.2 for details) and the synthetic data generated by the fitted two-layer model solutions from Fig. 1 (i.e., data without noise, solid lines) and the data generated by adding random noise to the two-layer model solutions (i.e., data with noise, dashed lines). The black solid line overlaps the pre-defined target ECS contained in the synthetic data. Methods 2 and 3 also recover the target ECS for the no-noise case. The SD of the ECS estimates from the 10,000 synthetic samples is around 0.026 K (for year 100) to 0.014 K (for year 500) for the red dashed line, 0.125–0.025 K for blue dashed line, 0.124–0.033 K for the magenta dashed line, and 0.233–0.049 K for the dashed black line

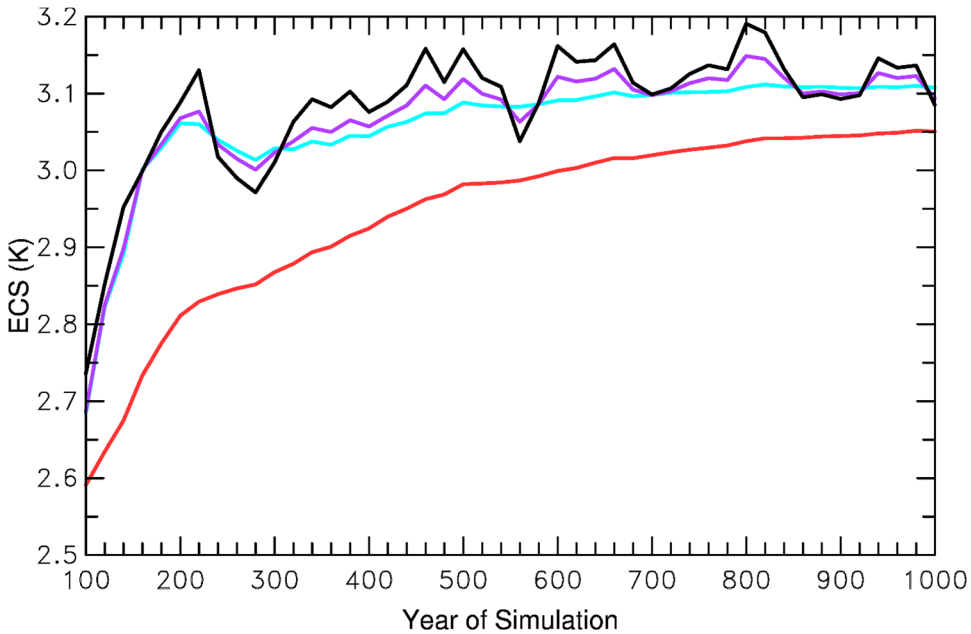

Fig. 7 ECS estimated using Method 1(red), 2 (blue), 3 (magenta) and 4 (black) and the model data up to the plotted year from the CESM1 $4\times CO_2$ experiment

synthetic data (Fig. 6) suggest that method 4 outperforms methods 2–3, especially for $L < 500$ years.

Our best ECS estimate of 3.12 ± 0.06 K for $L = 500$–1000 years based on method 4 is slightly lower than the ECS of 3.2 K estimated from a SOM experiment done by Gettelman et al. (2012). However, if we use a $4\times CO_2$ to $2\times CO_2$ conversion ratio of 0.485 (as discussed below), instead of 0.4769 (the ratio of the initial TOA forcing from

$2\times CO_2$ and $4\times CO_2$ based on Byrne and Goldblatt 2014) used in Fig. 7, then it would yield an ECS of 3.17 K, very close to the SOM estimate. Figure 6 indicates that our method underestimates ECS by about 0.02 K (for $L>600$) based on our tests with the synthetic data. Given all these uncertainties, our estimated ECS should be considered very similar to that based on the SOM experiment from Gettelman et al. (2012). This seems to suggest that the traditional SOM approach provides a reliable ECS estimate for the fully coupled model, at least for the CESM1. This also implies that changes in the ocean circulation over the 1000-year period do not seem to play an important role in determining the model's climate sensitivity. Unfortunately, we do not have the SOM-based estimate for the other CMIP5 models discussed below to verify this conclusion. Previous studies using multi-millennium simulations by fully coupled models with coarse resolutions have shown that slab ocean runs may either slightly overestimate (Li et al. 2013) or underestimate (Danabasoglu and Gent 2009) the ECS of the fully coupled model, but this bias seems to be small, which is consistent with our finding.

3.3 Evaluation of the ECS estimates using multi-millennium simulations

The above evaluation using synthetic data suggests that method 4 outperforms the other methods with an underestimate bias of <0.2 K for $L\geq 150$ years (Fig. 6). The analysis of the CESM1 simulation (Fig. 7) also suggests that method 4 yielded much smaller underestimates of the true ECS. However, one may question the validity of the synthetic data in representing climate model data. For the CESM1 simulation, it is not long enough for us to calculate its final warming although we concluded that its ECS should be not less than 2.91 K based on the already realized warming. Here we further evaluate the four methods using multi-millennium simulations from three other coupled models, in which case the realized warming by the end of such long simulations is likely to be very close to the final equilibrium warming. Thus, we have a better benchmark to evaluate the performance of the methods using actual model data.

Figure 8 shows the time series of the global-mean temperature (ΔT) and TOA net energy flux (ΔN) from these long simulations. Both ΔT and ΔN show rapid changes during the first few hundred years and then gradually approach a steady state until about year 1500–2000; thereafter, their changes are small, although considerable short-term fluctuations and small long-term changes still continue. By the end of the simulations, mean ΔN is very close to zero and mean ΔT is very steady. Thus, we use the mean ΔT of the last 300 years as our best estimate of the true ECS (denoted as ECS_{best}). However, please note that for CESM1.0.4, ΔN is still above zero near the end of its simulation (≈ 0.08 W/m^2, Fig. 8a) and therefore its ESC_{best} still underestimates its true ECS slightly.

Figure 8 also shows that the equilibrium warming (i.e., ECS) represented by the fitted two-layer model is very close to ECS_{best}, and this also true for the fitting to the first 1000 years of data only (short-dashed line in Fig. 8), even though the two fittings may differ noticeably during the early part of the simulation (e.g., for $L<1500$ years). This suggests that the two-layer model fitted to our CESM1 1000 year simulation (Fig. 1) may closely reflect the true ECS of the model, and therefore the pre-defined ECS (from the fitted two-layer model) in our synthetic data should be very close the actual ECS of the CESM1. This further suggests that our synthetic data and the tests based on them are highly relevant to applications to actual model data. On the other hand, the two-layer model fitted only to the first 150 years of data slightly underestimates ECS_{best} for all the three cases (see the long-dashed line in Fig. 8), leading a small underestimate bias in the ECS from method 4 for $L=150$ in Fig. 9. These results are consistent with the test results using the synthetic data as shown in Fig. 6.

The ECS estimates for the three models by the four methods are shown in Fig. 9 as a function of varying simulation length (L) used in the analysis. It is clear that methods 2–4 capture ECS_{best} well with small biases (<0.1 K) even for short L of 150–500 years). In contrast, the Gregory method (method 1) underestimates the ECS by 0.2–0.4 °C for L = 150–500, although the bias decreases to around 0.1–0.2 °C for L = 1000. Due to the use of the warming and ΔN of the last 50 years of the (analyzed) simulation, the ECS from methods 3–4 shows some small fluctuations. Despite this, methods 3–4 show slightly smaller underestimates than method 2, especially for relatively small L (e.g., for $L<2000$). Thus, we conclude that a model's ECS can be estimated fairly reliably by our methods 2–4 (especially method 4) with an error (mainly underestimate) of within about 0.1 K even using simulations of only 150–500 years, while the Gregory method may underestimate the true ECS by 0.2–0.4 °C using such simulations.

The fact that our methods 2–4 can capture ECS_{best} well implies that the feedback parameter λ ($=-N/dT$) does not change greatly in a long simulation. To verify that, in Fig. 10 we show the time series of the λ estimated using the data over a moving window of 101, 201, 301 and 501 years centered around the plotted year. It shows that λ becomes increasingly less variable when more data are used to estimate it, and it is relatively stable after the first 500 years without obvious trends, although there appears to be some oscillations around a period of about 1200 years in the λ from the CESM1.0.4 run (Fig. 10a), which seems to suggest some millennial-scale climate oscillation (e.g. in ocean overturning circulation) that projects onto the radiative feedback parameter ($-dN/dT$).

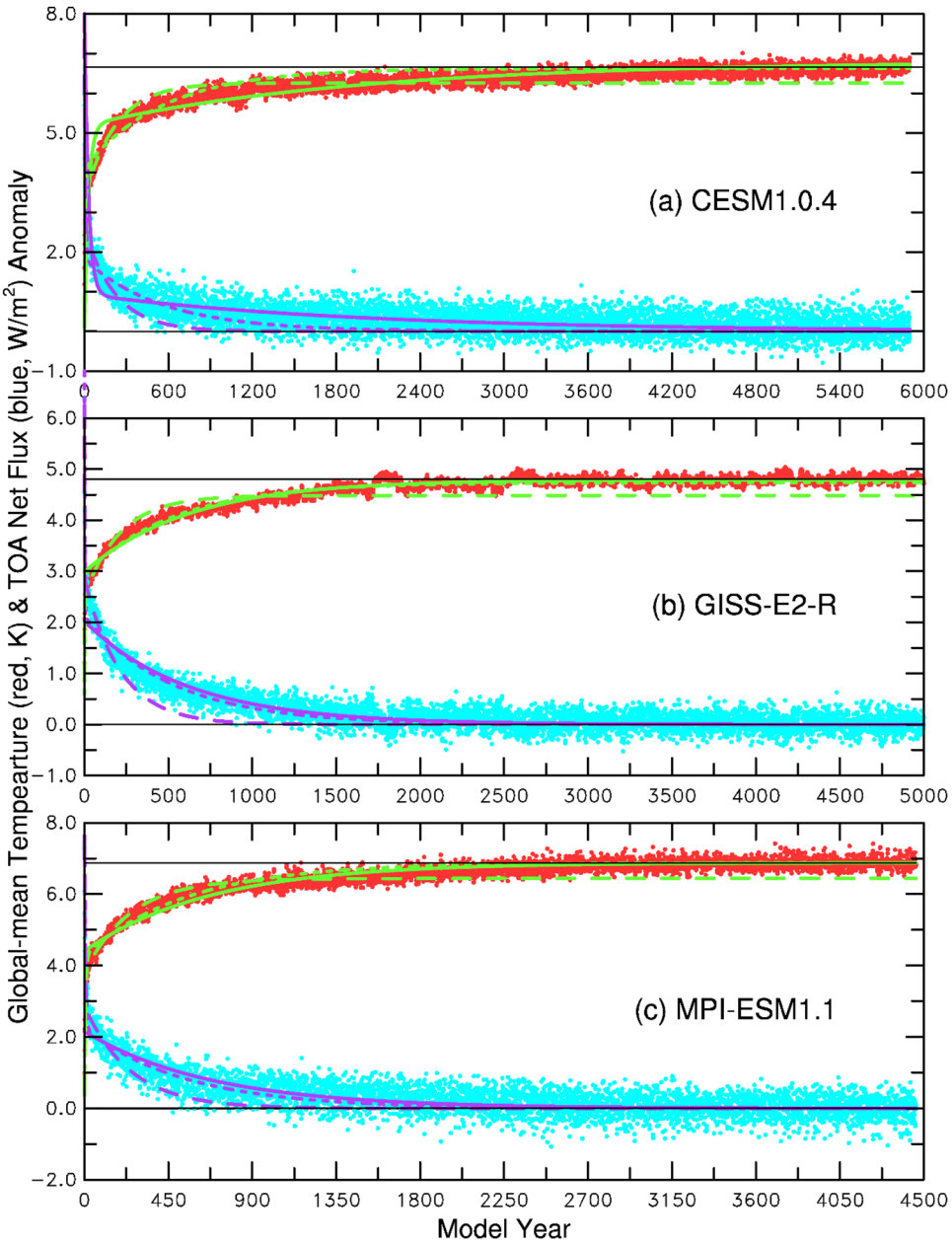

Fig. 8 Time series of annual global-mean surface air temperature (red dots, K) and top-of-atmosphere net energy flux (blue dots, W/m^2, positive downward) anomalies (relative to control run climatology) from the abrupt 4×CO_2 experiment using **a** CESM1.0.4, **b** GISS-E2-R, and **c** MPI-ESM1.1. The top thin black line represents the mean warming of the last 300 years of the simulation while the bottom thin black line is the zero line. The two-layer model fit to the temperature and flux data is represented by the green and magenta lines, respectively, with the solid line for the fitting to all model data, short dashed line for the fitting to the first 1000 years, and long dashed line for the fitting to the first 150 years of data only

These local (−dN/dT) slopes largely reflect the relationship between internally-generated dN and dT fluctuations, rather than between the externally-forced dN and dT changes, especially after the first 500–1000 years. This is because the relationship is dominated by the internal variations after year 1000, as the externally-forced dN and dT changes are small compared with the internally-generated year-to-year fluctuations, which degrade the fitting considerably (Fig. 11).

On the other hand, during the early part of the simulation ($L < 1000$ years), the externally-forced dN and dT changes are large compared with the internally-generated year-to-year fluctuations (Fig. 11), which makes the (−dN/dT) slope over the early period more representative of the forced relationship between dN and dT than that of the later periods. However, the (−dN/dT) slope during the first ~ 40 years is considerably steeper than that during year 41–1000 (Fig. 11), and our ECS results from methods 1 and 2 (Fig. 9) suggest that the (−dN/dT) slope from year 41–1000 is a better choice than that of year 1–40 for estimating ECS through extrapolation.

The (−dN/dT) slope estimated using the data from year 41 to the plotted year, which is used to estimate ECS in our methods 2–3, shows little variation for $L \geq 150$ years, although it decreases slightly over time, mainly over the first millennium (Fig. 10). This contributes to the small

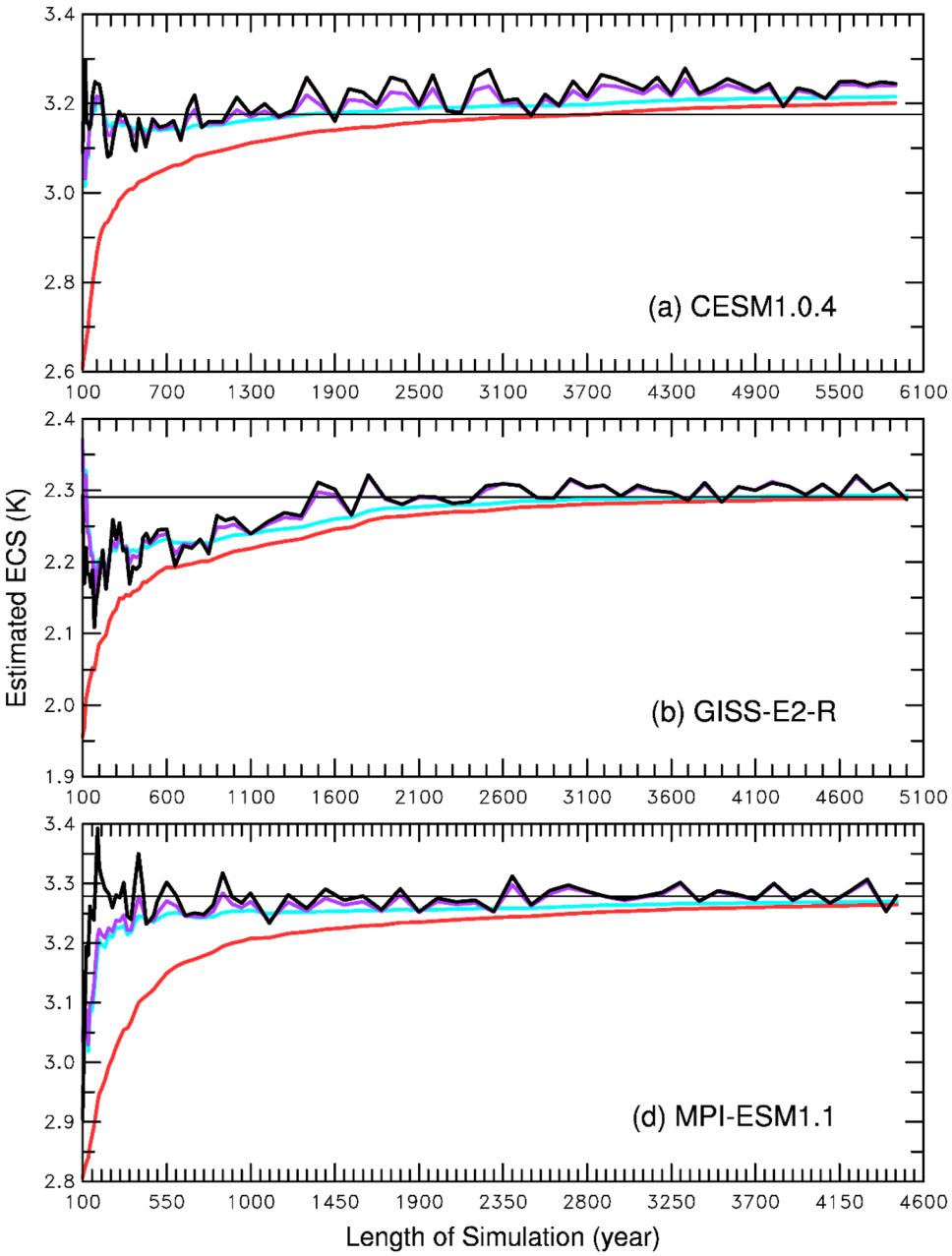

Fig. 9 ECS estimated using Method 1(red), 2 (blue), 3 (magenta) and 4 (black) and the model data up to the plotted year from the abrupt $4\times CO_2$ experiment using **a** CESM1.0.4, **b** GISS-E2-R, and **c** MPI-ESM1.1 shown in Fig. 8. The thin black line represents the mean warming of the last 300 years of the multi-millennium simulation. For the CESM1.0.4, the warming of the last 300 years slightly underestimates the ECS as the TOA net flux is still positive (≈ 0.08 W/m^2) near the end of the simulation (cf. Fig. 8a). Note that a factor of 0.4769, the TOA forcing ratio between 2×CO2 and 4×CO2 based on Byrne and Goldblatt (2014), was used to convert the temperature change from the $4\times CO_2$ experiment to that for the $2\times CO_2$ case

underestimate biases during the early part of the simulations shown in Fig. 9, especially for methods 2–3. This estimated slope represents the dN and dT relationship mainly during the early part of the simulation (i.e., from year 41 up to year 1000), as the later years contribute relatively little to the forced changes in the scatter plot (Fig. 11). The estimated slope (dashed green line in Fig. 10) using data from year 1001 to the plotted year is much closer to the locally estimated slope than that using data after year 41, although noticeably differences still exist. This further suggests that the externally-forced (−dN/dT) slope, which is reflected mainly in the data of the first 1000 years, may differ from the internally-generated (−dN/dT) slope, which is reflected mainly in the data after year 1000. Clearly, one should not use the internally-generated (−dN/dT) slope for estimating ECS.

The above analyses suggest that (1) the (−dN/dT) slope (i.e., the feedback parameter λ) during year 41–1000 mainly reflect the forced relationship, which does not change a lot over time and can be used to estimate ECS through extrapolation; (2) locally estimated (−dN/dT) slope may fluctuate greatly on short (10–10^2 year) time scales and it represents mainly internally-generated relationship, especially after year 1000, that may differ from the forced slope and thus should not be used to estimate ECS. The temporal stability of the forced (−dN/dT) slope makes it possible for using

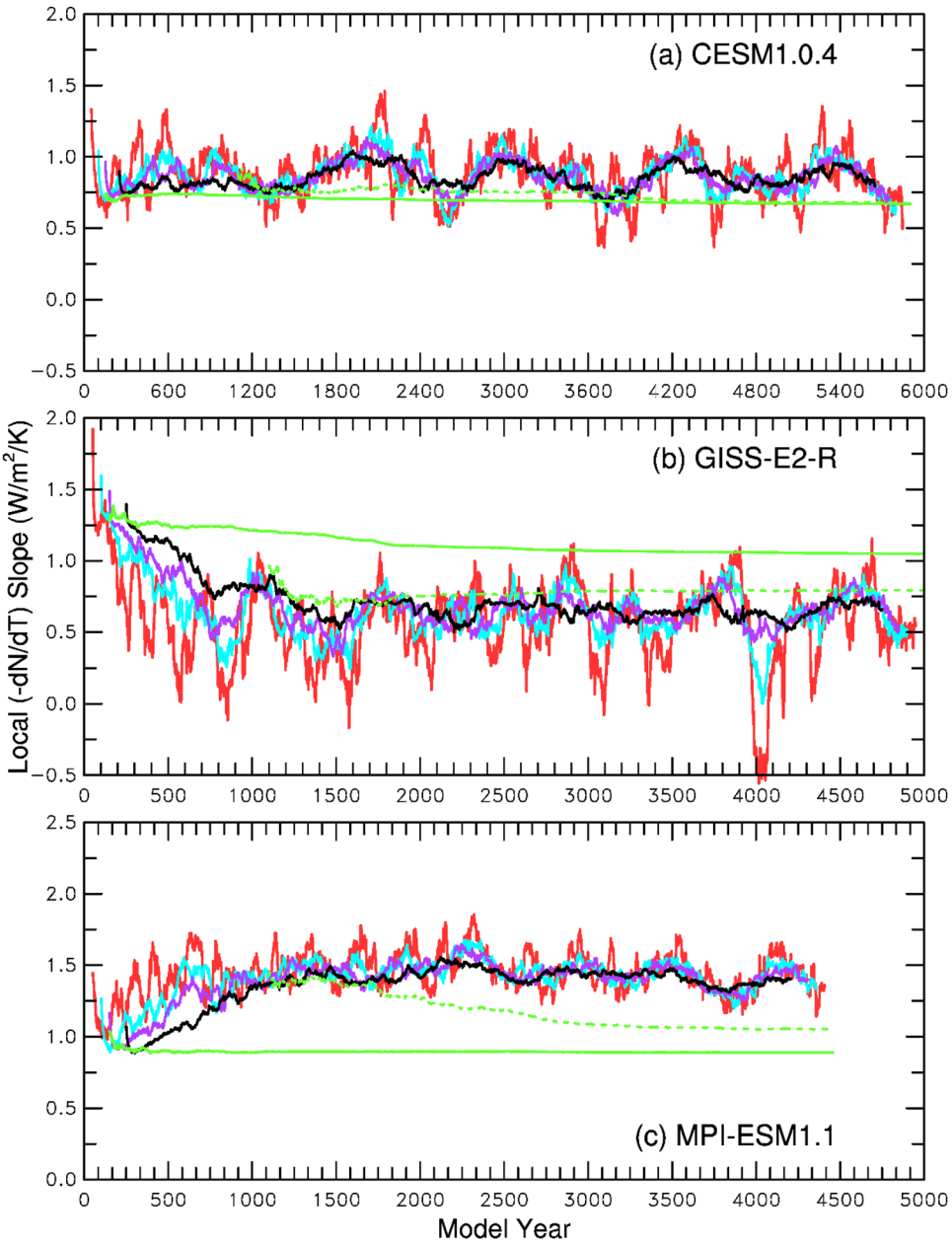

Fig. 10 Time series of the local (−dN/dT) slope estimated using the local 101 (red line), 201 (blue line), 301 (magenta line), and 501 (black line) years of data centered around the plotted year from the abrupt $4\times CO_2$ experiment using **a** CESM1.0.4, **b** GISS-E2-R, and **c** MPI-ESM1.1. The green solid line is the (−dN/dT) slope estimated using the data from year 41 to the plotted year on the x axis, which is the slope used in Methods 2–3. The green dashed line is the (−dN/dT) slope estimated using data from year 1001 to the plotted year

our methods 2–4 to estimate ECS with a small error within about 0.1 K using relatively short simulations of a few hundred years.

Previous studies (e.g., Andrews et al. 2015; Gregory et al. 2015; Rose and Rayborn 2016; Rugenstein et al. 2019b) suggest that the feedback parameter λ may change over time, often based on simulations less than 2000 years. These early findings are not inconsistent with our results (Fig. 10). However, we show that there exist two different types of the (−dN/dT) slopes, one due to forced changes that dominate during the early part (up to year 1000) of the simulation and one due to internal variations that dominate after year ~ 1000. The forced slope from year 41 up to 1000 does not change greatly, which allows one to estimate ECS using relatively short simulations (200–1000 years).

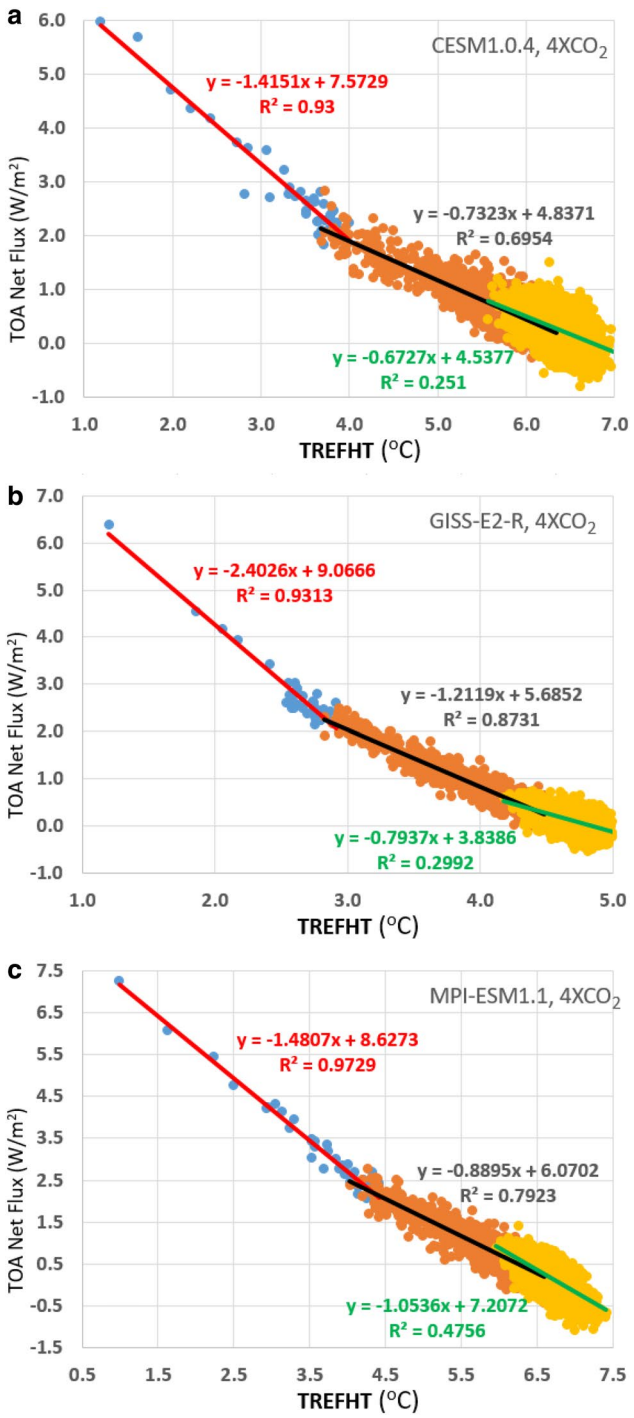

Fig. 11 Scatter plot of the anomalies (relative to the control-run climatology) of the global-mean and annual-mean surface air temperature (TREFHT, x axis) and top-of-atmosphere (TOA) net radiation flux (y axis, positive downward) from year 1 to 40 (blue dots, red line = linear fit), year 41 to 1000 (red dots, some are behind the orange dots, black line = linear fit), and year 1001 to the end of the simulation (orange dots, green line = linear fit) from the abrupt $4\times CO_2$ simulation using **a** CESM1.0.4, **b** GISS-E2-R, and **c** MPI-ESM1.1. The regression equation with the squared correlation (in the same color as the fit line) is also shown

3.4 Comparison of the ECS estimates from $2\times CO_2$ and $4\times CO_2$ experiments

Figure 12 compares the ECS values estimated using the ΔT and ΔN data for $L = 100$–500 years from CESM1 $2\times CO_2$ and $4\times CO_2$ experiments using methods 1–4. Using the $2\times CO_2$ experiment, all the methods underestimate the equilibrium T response significantly when $L < 240$ years, whereas it requires 180 or more years for the $4\times CO_2$ experiment and methods 2–4 to produce relatively stable ECS estimates. If the simulation length is shorter than about 240 years, all the methods would produce considerably higher ECS values using the $4\times CO_2$ experiment than those using the $2\times CO_2$ experiment. This difference becomes smaller for $L \geq 250$ years (especially for methods 1–2), but its sign reverses; that is, the ECS estimates from the $2\times CO_2$ experiment exceeds those based on the $4\times CO_2$ experiment. Here, we used a conversion ratio of 0.4769 (the ratio of the initial TOA forcing from $2\times CO_2$ and $4\times CO_2$ based on Byrne and Goldblatt 2014) to convert the equilibrium ΔT into ECS for the $4\times CO_2$ case. Using a conversion ratio of 0.5 (as in Gregory et al. 2004) would make the difference even larger for $L < 240$ years, while the ECS from the $4\times CO_2$ would be considerably larger than that from the $2\times CO_2$ for $L > 240$ years (not shown). Our tests showed that using a conversion ratio of 0.485 yielded closer matches for $L > 240$ years for the ECS estimates from the two types of experiments.

Given the lower signal to noise ratio in the $2\times CO_2$ experiment than in the $4\times CO_2$ experiment, we would expect that a longer simulation is needed for estimating ECS, and our results from the CESM1 experiments indicate that the extra length of simulation is about 60 years (240 vs. 180 years). Thus, using the $4\times CO_2$ experiment could potentially save up to about 1/4 of the simulation time needed for a $2\times CO_2$ experiment. However, in this case choosing a proper conversion ratio may induce additional uncertainties for the estimated ECS.

3.5 Results from CMIP5 and CMIP6 models

We were able to download the necessary model data from only 20 CMIP5 models listed in Table 1 and 19 CMIP6 models listed in Table 2, and repeated the above analyses for each of these models. Figures 13 and 14 shows three examples from the CMIP5 and CMIP6 models, respectively, of the fitting using the two-layer model solutions (Eqs. 1, 2) to the model output of ΔT(t) and ΔN(t) data. Similar to Fig. 1 for the CESM1 and Fig. 8 for the three millennium simulations, the two-layer model solutions with a non-unit efficacy of deep ocean heat uptake from Geoffroy et al. (2013b) provide

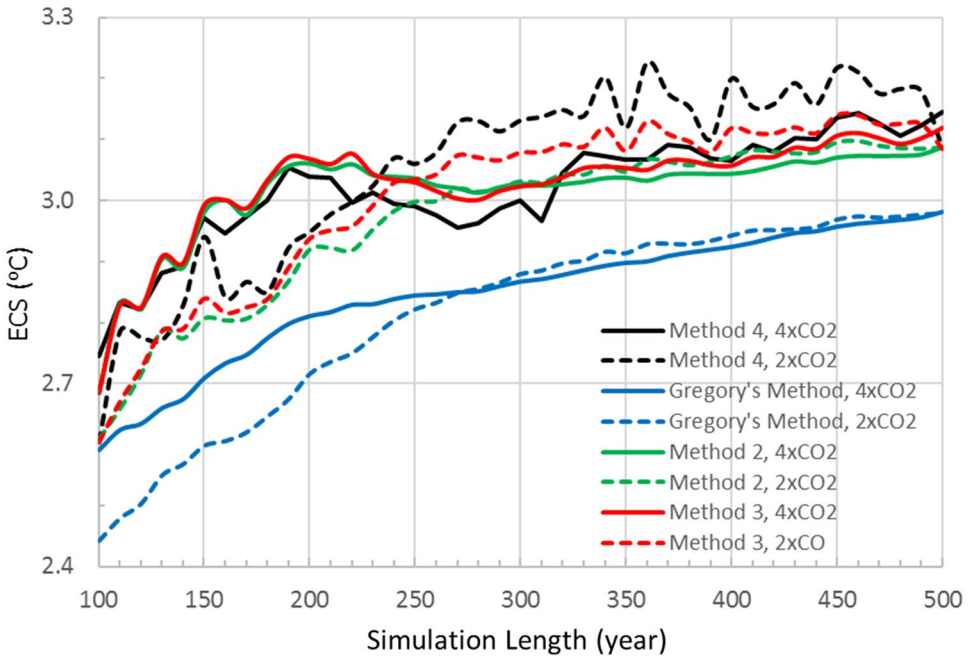

Fig. 12 Comparison of the ECS estimates using the $2\times CO_2$ (dashed) and $4\times CO_2$ (solid) CESM1 experiments and Method 1 (Gregory's, blue lines), 2 (green lines), 3 (red lines) and 4 (black lines) as a function of the simulation length. The equilibrium temperature response from the $4\times CO_2$ experiment was multiplied by 0.4769 (the ratio of the TOA forcing between $2\times CO_2$ and $4\times CO_2$ according to Byrne and Goldblatt 2014) to derive ECS for the $4\times CO_2$ case

a good fit to the model data for all the CMIP5 and CMIP6 models analyzed here, while the one time-scale solution (red line in Fig. 13) from Gregory et al. (2015) does not fit the $\Delta T(t)$ data well. The fast response time (τ_f) from the fitting (see Eq. 1) ranges from 1.5 to 6.4 years with a mean of 3.7 years among the 20 CMIP5 models, while the slow response time (τ_s) is around 200–400 years for most of the models (Table 1; similar τ_f and a slightly wider range for τ_s for the 19 CMIP6 models, Table 2). As implied by the multi-millennium simulations analyzed above, the ~ 150 years of simulation from these models may be sufficient for estimating ECS using our methods 2–4, although longer simulations would help reduce the error. Ideally, such $4\times CO2$ simulations should be at least 180 years in order to avoid the large underestimate by all the methods examined here, as discussed above (cf. Figs. 6, 7).

Similar to Figs. 5 and 6, Fig. 15 shows that method 4 reproduces well the *pre-defined* (−dN/dT) slope and ECS (through fitted Eqs. 1, 2, see Sects. 2.3, 2.4) in the synthetic data generated by combining the fitting functions and the randomly sampled noise (through 10-year block sampling of the residual series of the fitting), while method 1 (i.e., the method of Gregory et al. 2004) overestimates the (−dN/dT) slope and thus underestimates ECS (by 0.42 K on average) for most of the 20 CMIP5 models with a simulation length $L = 150$ or 140 years (Table 1). Methods 2 and 3 perform better than method 1, but still have a tendency to underestimate ECS (by ~ 0.18 K on average for both of them, compared with a mean underestimate of ~ 0.10 K by method 4 (for $L = 140$–150 years) (Fig. 15b). Similar results are seen for the 19 CMIP6 models (Fig. 15c, d).

When applied to the $\Delta T(t)$ and $\Delta N(t)$ data from the CMIP5 $4\times CO_2$ simulations, method 1 produces ECS estimates that are about 25% lower than our new estimates using method 4 for models with an ECS above about 3.0 K, while the two estimates are similar for models with an ECS around 2.5 K (see the regression line in Fig. 16a). When averaged over the 20 CMIP5 models, the ECS from method 1 is about 10% lower than that from method 4 (Table 1). Thus, the ECS values reported previously by Andrews et al. (2012), Forster et al. (2013) and Flato et al. (2013) are likely to be underestimated by about 10% on average and as much as 25% for models with medium–high ECS (above 3 K). These biases are consistent with the findings of Paynter et al. (2018), who found that the ECS values reported by Flato et al. (2013) based on Gregory et al. (2004)'s method for two GFDL models are about 0.8–0.9 K lower than those estimated from multi-millennium simulations. On the other hand, our ECS estimates appear to be comparable to those implied by the equilibrium warming under $4\times CO_2$ estimated by Geoffroy et al. (2013b) for the limited number of models with estimates available in both our analyses and their study (Table 1). This is expected because both our method 4 and Geoffroy et al. (2013b) rely on the fitting to the 2-layer model solutions to extrapolate the T response as N approaches zero, although the actual implementation of our method (including the fitting procedure and estimating the realized warming, see Sects. 2.2, 2.3) differs from Geoffroy et al. (2013b), who used a number of steps (described in section 2d of Geoffroy et al. 2013b and section of Geoffroy et al. 2013a) to estimate the various parameters in their 2-layer model and the equilibrium T response using the data

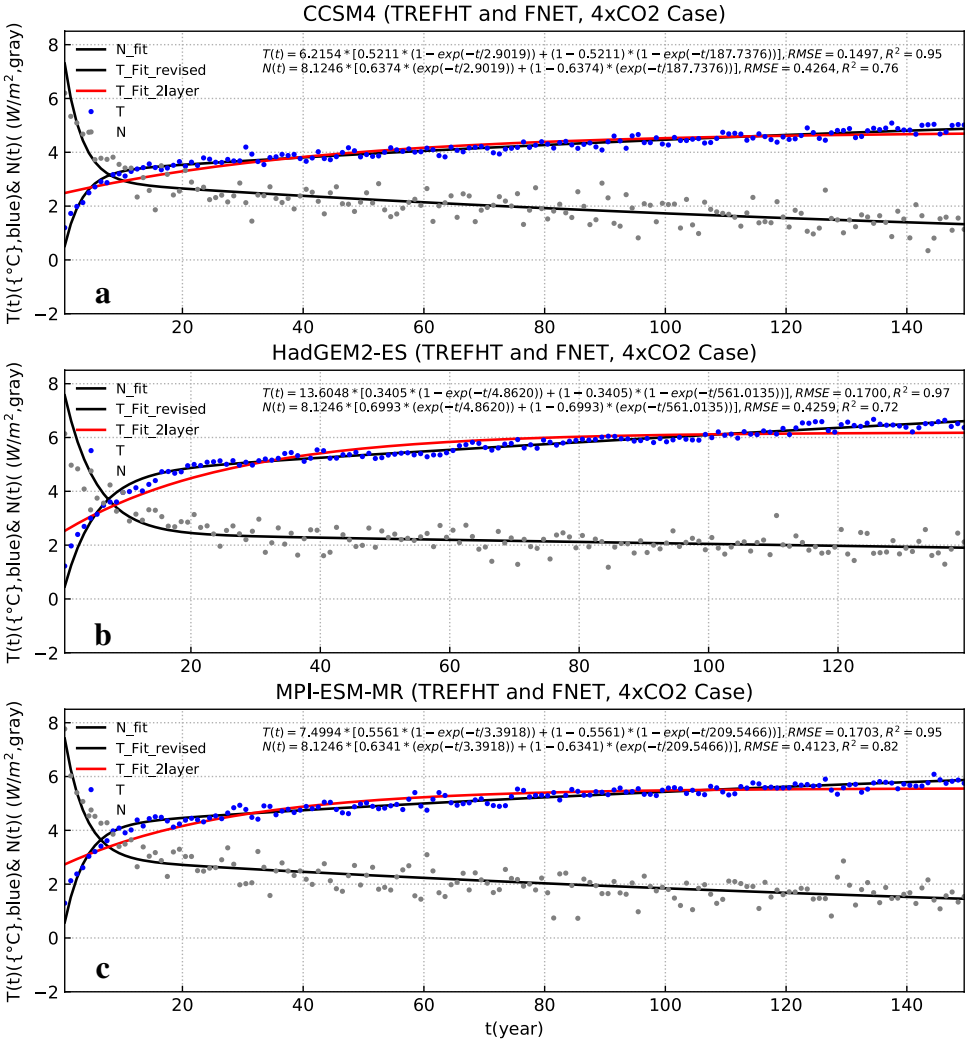

Fig. 13 Time series of anomalies (relative to the control climate) of global-mean and annual-mean surface air temperature (T(t) or TREFHT, blue dots) and TOA net radiation flux (N(t) or FNET, gray dots, positive downward) from the $4\times CO_2$ experiment from three select CMIP5 models. The black line is the best fit using the two-layer model solution of Geoffroy et al. (2013b), with the fitting results shown on the figure. The red line is a fit of the temperature data following the one-time-scale solution of the two-layer model of Gregory et al. (2015)

from CMIP5 $4\times CO_2$ simulations. In our method, we first estimated the α parameter in Eq. (1) using our Method 4 and specified the F_o parameter in Eq. (2) based on Byrne and Goldblatt (2014), and then used the least squares fitting (Eq. 4) to estimate the best combination of the other parameters in Eqs. (1, 2), which were then used to update the estimate of parameter α for the next iteration. Also, the slope from the fitted model is only used to estimate the unrealized warming, whereas Geoffroy et al. (2013b) used the fitted model to estimate the entire ECS.

The ECS results for CMIP6 models (Table 2 and Fig. 16b) show a mean underestimate bias of ~ 0.27 K for the Gregory method compared with our Method 4, and as much as 1.0 K for models with large ECS (e.g., E3SM-1-0). When a conversion ratio of 0.5 is used, our method 1 (i.e., the Gregroy method) yielded ECS values for both CMIP5 and CMIP6 models that are very similar to those of Zelinka et al. (20,202), who used the Gregory method. Our new ECS estimates for the 19 CMIP6 models show a mean of 3.43 K with a range of 1.79–5.93 K, very similar to the mean of 3.45 K and slightly higher than the range of 1.75–5.36 K for the 20 CMIP5 models (Tables 1, 2). Thus, while the ECS range from CMIP6 is higher than that from CMIP5, their multi-model ensemble means are very close.

Tables 1, 2 and Fig. 16 show that the difference of the ECS estimates from the Gregory method and our method 4 is smaller for models with lower ECS than for models with higher ECS for both CMIP5 and CMIP6 models. We are unsure about the cause behind this, but Fig. 9 shows that for GISS-E2-R, which has a low ECS compared with the other two models, methods 2–4 all show noticeable underestimates compared with ECS_{best} for $L < 1000$ years, leading to a smaller difference of the ECS estimates between method 1 (i.e., the Gregory method) and method 4 than for the other two models. Since the ECS from our method 4 is very close to the equilibrium warming represented by the fitted two layer model, the fact that the ECS from method 4 has a relatively large bias in Fig. 9b for short L suggests that the fitting for short L has a relatively large error in representing the long-term evolution of the dT for this model,

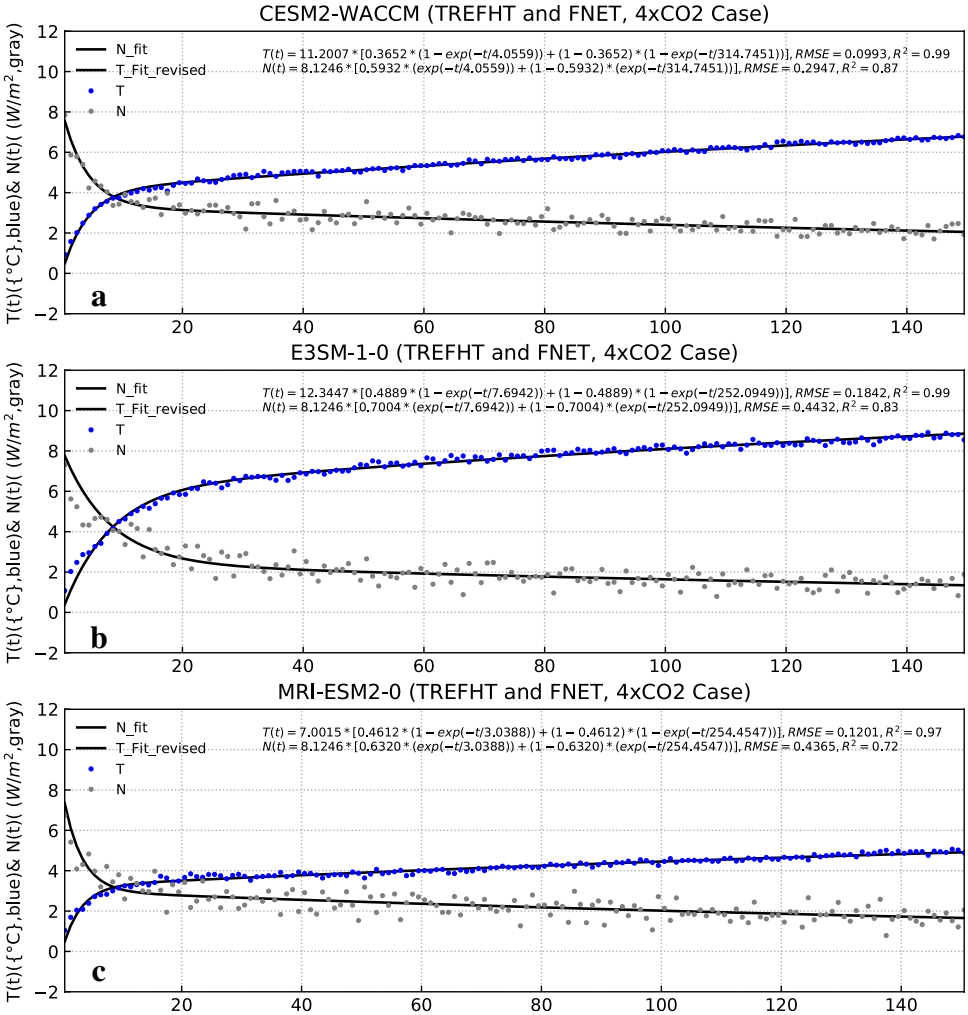

Fig. 14 Same as Fig. 13 but for three select models from CMIP6 without the red line

which seems to be confirmed by the long-dashed green line in Fig. 8b. Thus, it is possible that due to the particular dT response curve during the first several hundred years from models like GISS-E2-R that has a low ECS, the fitted two-layer model has a relatively large underestimate bias in representing the ECS. In other words, the difference in the dT response curves for short L between models with low and high ECS might be behind the different bias in the ECS from method 1 relative to method 4 among the CMIP models. This difference is also reflected by the different deviations between the forced long-term (solid green in Fig. 10) and unforced local (red, blue and black lines in Fig. 10) (−dN/dT) slopes between GISS-E2-R and the other two models in Fig. 10. The underlying physical processes leading to such differences require further investigation.

Given the comparable performance of method 2 shown in Fig. 9, one may want to apply it to the CMIP models due to its simplicity. Tables 1, 2 and Fig. 16c, d show that indeed the ECS estimates from method 2 are very close to those from method 4, with a mean ECS of 3.36 K from method 2 vs. 3.45 K from method 4 for the CMIP5 models and a mean ECS of 3.34 K from method 2 vs. 3.44 K from method 4 for the CMIP6 models.

Figure 17 shows that the transient climate response (TCR) at the time of CO_2 doubling in a 1% per year increase run is correlated with ECS among the CMIP5 and CMIP6 models, as shown previously (Flato et al. 2013). However, this correlation becomes weaker for our new ECS estimates (Fig. 17b, d). Furthermore, because of our increased ECS estimates, on average the TCR accounts for only about 53% of the ECS from our method 4, compared with 58% of the ECS based on Gregory et al. (2004)'s method for the CMIP5 models (Table 1; 57% vs. 62% for CMIP6 models, Table 2). Thus, the differences in the ECS estimates also substantially alter this realized warming fraction at the time of CO_2 doubling in a transient run. Our analyses did not reveal a strong relationship between the TCR/ECS ratio and the long-term response time (τ_s) listed in Tables 1, 2.

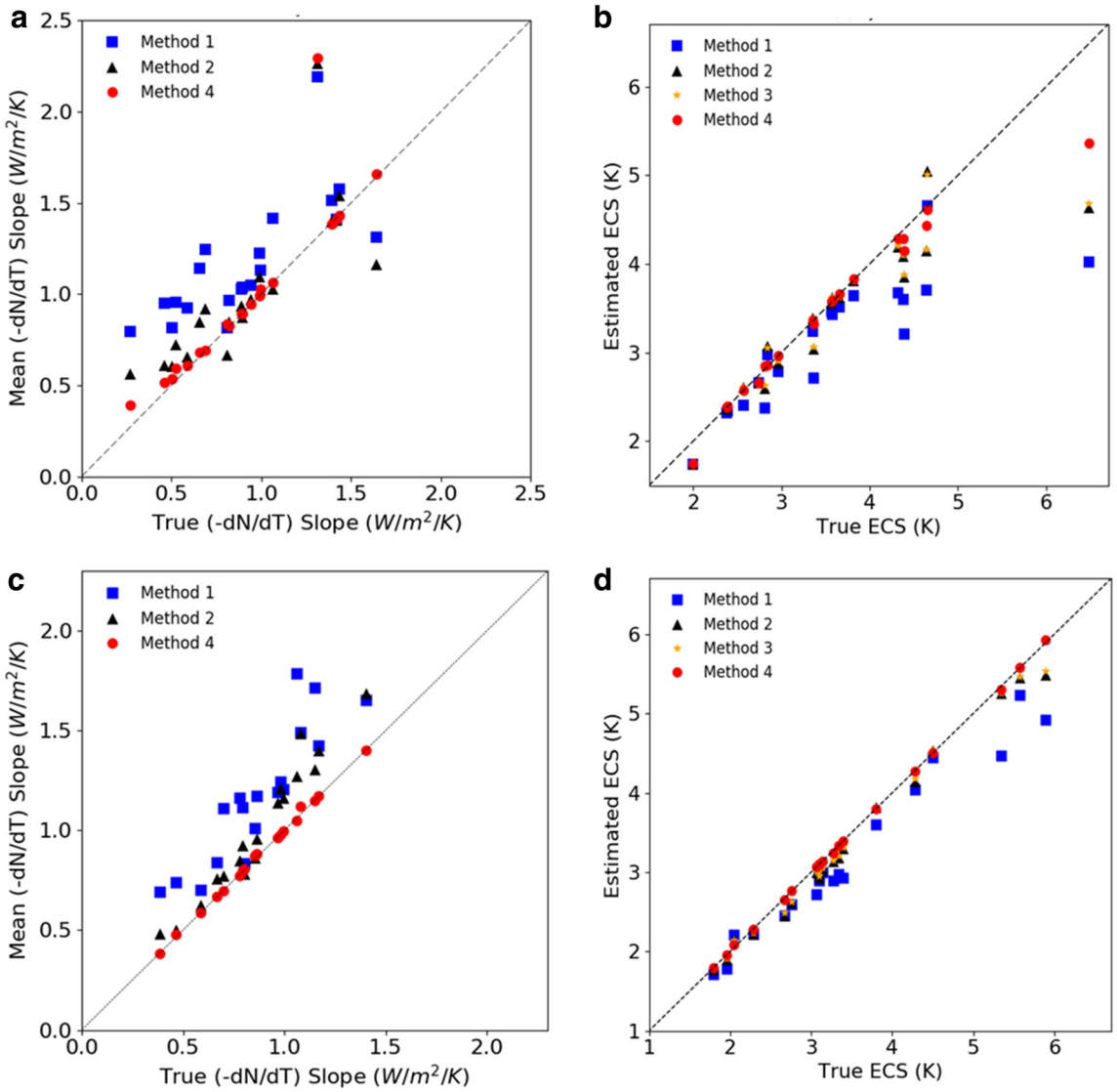

Fig. 15 **a**, **c** Scatter plot of the pre-defined true (−dN/dT) slope (embedded in the synthetic data) (as the x axis) vs. the mean (−dN/dT) slope estimated by methods 1, 2, and 4 from the noisy data (for length of 150 years, y axis) from **a** 20 CMIP5 and **c** 19 CMIP6 models, with each different symbol representing one method and each data point for one CMIP model. **b** Same as **a** but for the pre-defined true ECS (x axis) vs. the ECS estimated by the four different methods (y axis). **d** Same as **b** but for 19 CMIP6 models

4 Summary and discussions

In this study, we first examined the nonlinearity in the global-mean T response to TOA forcing N and discussed its implications for estimating ECS using the slope between ΔT(t) and ΔN(t). We then designed a benchmarking framework using the ΔT(t) and ΔN(t) data from a 1000-year abrupt $4\times CO_2$ simulation by the CESM1, a fully coupled model, to quantify the performance of four methods for estimating ECS using data with varying lengths of simulation (*L*). These methods were further evaluated using three multi-millennium (~5000 year) $4\times CO_2$ simulations from three different coupled models. These analyses were repeated using the abrupt $4\times CO_2$ experiment with 140–160 year integrations from 20 CMIP5 and 19 CMIP6 models. The ECS estimates from the $4\times CO_2$ experiment by the CESM1 were also compared with those based an abrupt $2\times CO_2$ experiment by the same model to verify whether a simple re-scaling of the equilibrium T response from the $4\times CO_2$ experiment would yield an ECS estimate similar to that from the $2\times CO_2$ experiment, as ECS is conventionally defined as the equilibrium warming after an abrupt CO2 doubling. The main results are summarized below.

The first 40 years or so show a steeper (−dN/dT) slope than the later years in the $4\times CO_2$ (and $2\times CO_2$) experiment; thus one should exclude these first 40 years in estimating

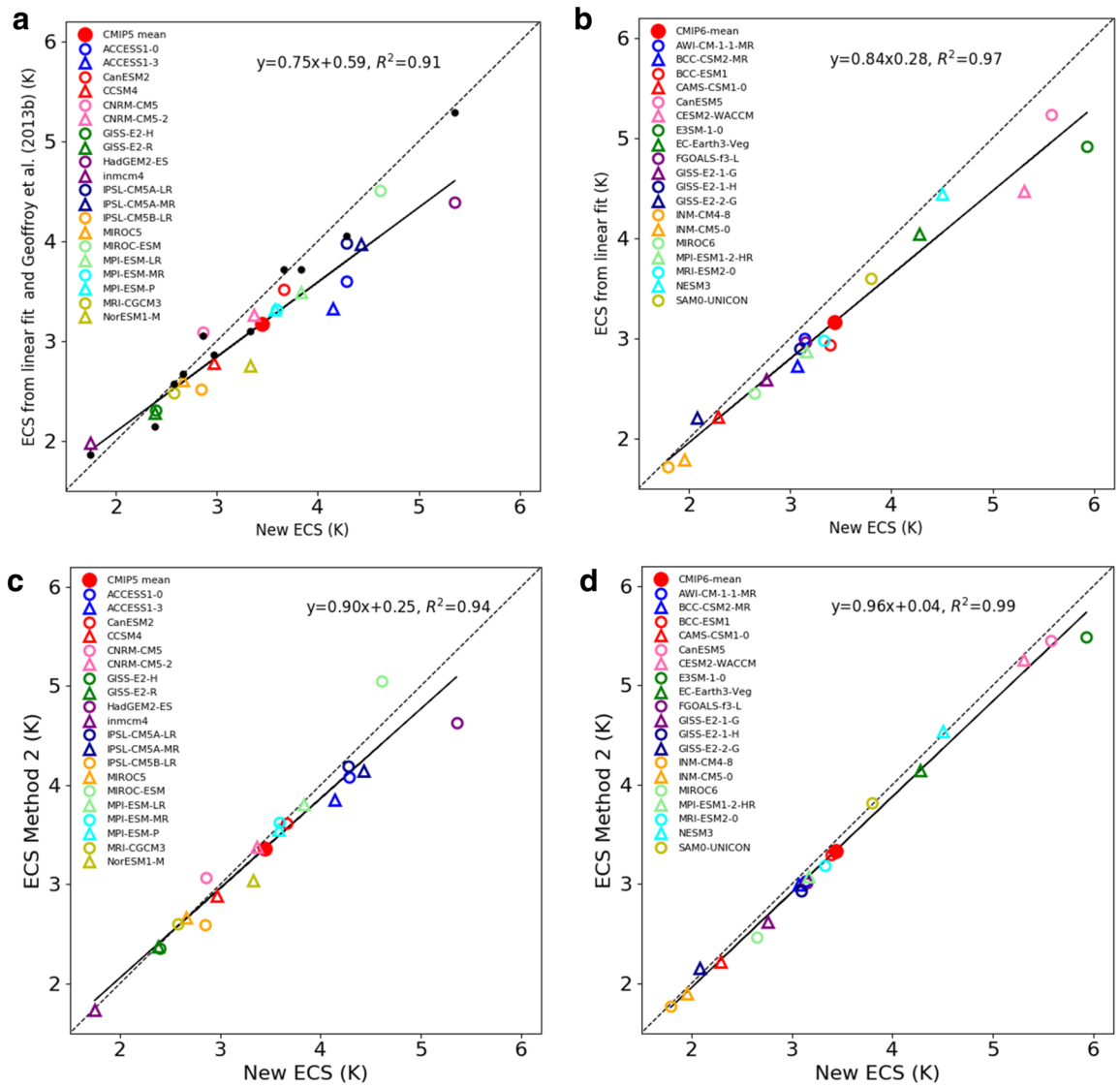

Fig. 16 Scatter plots of the estimated equilibrium climate sensitivity (ECS) based on Gregory et al. (2004)'s linear fitting method (y axis, colored symbols) and our new method 4 (x axis) among the **a** 20 CMIP5 and **b** 19 CMIP6 models estimated using their 150-year abrupt $4\times CO_2$ simulations. Also shown in **a** (black dots, on y axis) are the ECS estimates based on Table 1 of Geoffroy et al. (2013b). **c**, **d** Similar to **a**, **b** except that the y axis is the ECS estimated by our method 2 for the **c** 20 CMIP5 and **d** 19 CMIP6 models. The dashed line represents the 1:1 ratio and the solid line is the regression line of the colored symbols, with the regression equation shown

the (−dN/dT) slope for extrapolating the equilibrium T response as N approaches zero. Further, we show that the estimated (−dN/dT) slope is a better choice than the (−dT/dN) slope for estimating the long-term relationship (needed for estimating ECS) between ΔT(t) and ΔN(t) due to the different signal-to-noise levels in ΔT(t) and ΔN(t) series. Internal variability such as that associated with Arctic sea–ice fluctuations may cause global-mean T to vary on decadal to centennial time scales, but these short-term variations appear to have little influence on ECS, although they can complicate the estimation of the long-term (−dN/dT) slope using simulations with limited length. In particular, the three multi-millennium simulations do not show large changes in the externally-forced feedback parameter (λ = −dN/dT) that is largely reflected in the early part (from years ~ 40 – 1000) of the simulation. This provides a physical basis for using a constant λ to estimate ECS. Furthermore, using local ΔT(t) and ΔN(t) or their values after year 1000 would yield a (−dN/dT) slope that mainly reflects the relationship resulting from internal variations, and such a slope can vary greatly on 10–10^2 year time scales and may differ from the externally-forced slope; thus one should not use the internally-generated slope to estimate ECS.

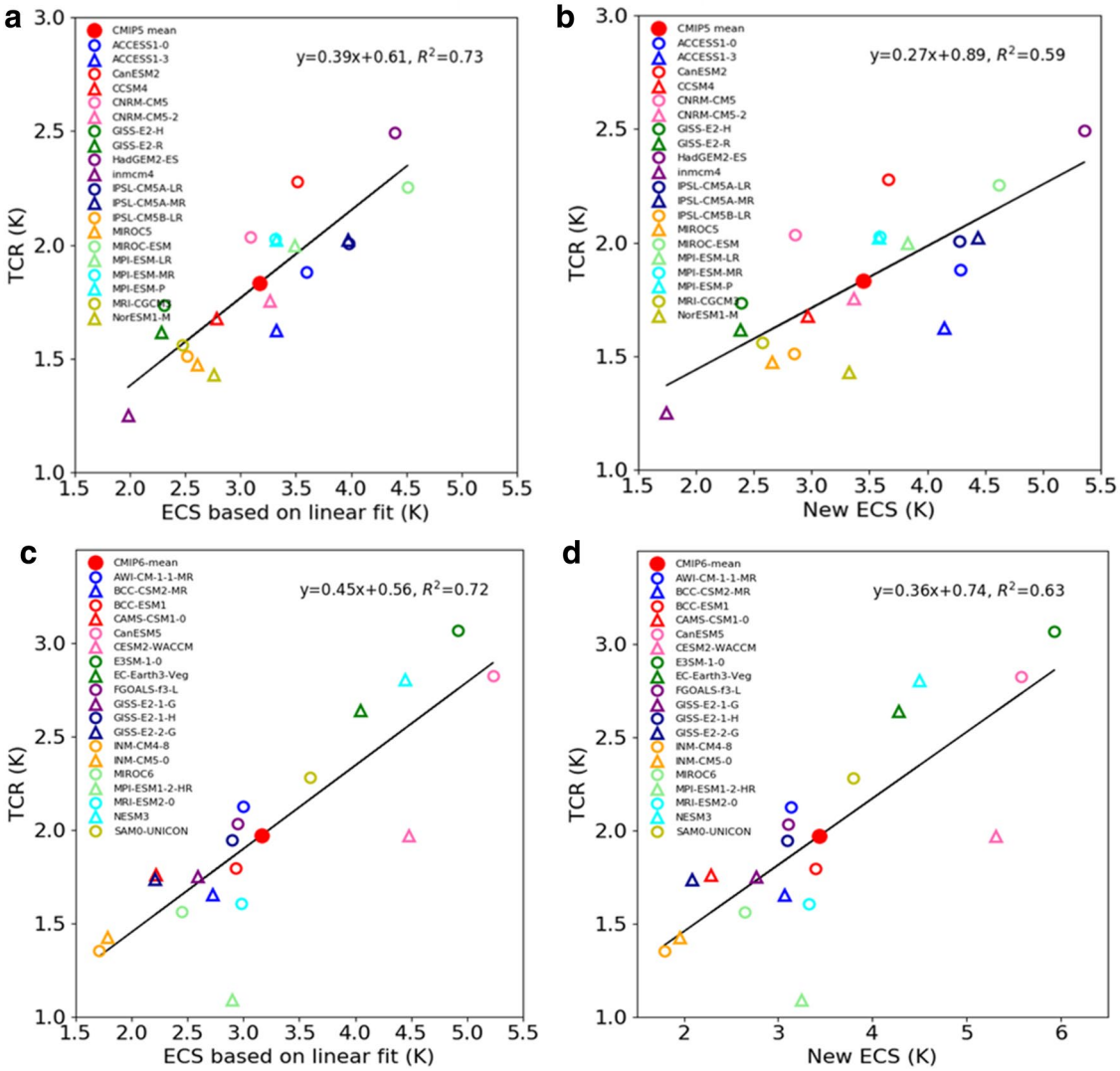

Fig. 17 Scatter plots of the transient climate response (TCR, y axis) vs. equilibrium climate sensitivity (ECS, x axis) estimated based on **a** the linear fitting of Gregory et al. (2004) and **b** our new method 4 among the 20 CMIP5 models. **c** Same as **a** but for the 19 CMIP6 models. **d** Same as **b** but for the 19 CMIP6 models

The change in the (−dN/dT) slope during the earliest part of the simulation has been noticed previously (e.g., Andrews et al. 2015; Gregory et al. 2015; Rose and Rayborn 2016; Armour 2017), but still the first 40 years of data were used in estimating the ECS of the CMIP5 and CMIP6 models (Andrews et al. 2012; Forster et al. 2013; Zelinka et al. 2020). However, the previous studies did not explicitly discuss the difference between the externally-forced and internally-generated (−dN/dT) slopes. Because of this changing slope, the ECS estimates based on the (−dN/dT) slope over the whole simulation period proposed by Gregory et al. (2004) and used by Andrews et al. (2012), Forster et al. (2013) and Flato et al. (2013) for CMIP5 models, and recently by Zelinka et al. (2020) for CMIP6 models would underestimate the true ECS, likely by about 10% on average and by as much as 25% for models with medium–high ECS (above 3 K), consistent with findings from Geoffroy et al. (2013b), Paynter et al. (2018) and Rugenstein et al. (2019b). Simply excluding the first 40 years in the analysis would reduce the underestimation by more than half, and using the slope from the best fit to the 2-layer model solutions for $\Delta T(t)$ and $\Delta N(t)$ would further reduce the underestimation to be < 0.03 K for $L \geq 180$ years based on our tests with synthetic data.

Our analyses of the three multi-millennium simulations confirm our test results using synthetic data, and suggest that ECS can be reliably estimated by our methods 2–4 with an error within about 0.1 K even using relatively short simulations of 150–500 years. The results also show that the two-layer model (Eqs. 1, 2) fitted to either the first 1000 years of

data or the entire simulation captures well the warming at the end of the long simulation. This suggests that the two-layer model, which has a constant (−dN/dT) slope after the first 40–50 years, can be used to depict the long-term evolution of the ΔT(t) and ΔN(t) series from an abrupt fully coupled $4\times CO_2$ simulation.

Our analyses of the CESM1 experiments with $2\times CO_2$ and $4\times CO_2$ showed that one needs at least 240 (180) years of simulation when abrupt $2\times CO_2$ ($4\times CO_2$) experiments are used to estimate ECS; otherwise, the underestimation will be large for all the methods. Furthermore, the ECS estimates using the $4\times CO_2$ experiment and a conversion ratio of 0.4769 (as implied by the TOA forcing for $2\times CO_2$ and $4\times CO_2$ given by Byrne and Goldblatt 2014) yielded slightly lower values than those based on the $2\times CO_2$ experiment for $L>240$ years. Using a conversion ratio of 0.5 would lead to substantially higher ECS than that using the $2\times CO_2$ experiment. For simulations longer than 240 years, using a conversion ratio of 0.485 yielded better agreements between the ECS estimates from the two types of experiments. Thus, we recommend to multiply the estimate of the equilibrium T response to $4\times CO_2$ by a factor of 0.485 to derive ECS. Our best estimate of the ECS for the CESM1 (with CAM4, ~2° grid) is around 3.17 K, close to the slab-ocean based estimate of 3.2 K by Gettelman et al. (2012) and that implied by the long simulation shown in Figs. 8a and 9a.

Thus, we recommend following steps to estimate the ECS of a fully coupled climate model. Fortran codes are available upon request from the lead author for carrying out Steps 2–3:

1. make an abrupt $4\times CO_2$ (or $2\times CO_2$) simulation with at least 180 (240) years of integration, with a 500 year integration is preferred;
2. fit the global-mean and annual-mean surface air temperature ΔT(t) and TOA net radiation ΔN(t) anomaly (relative to a pre-industrial control run climatology) time series from this experiment to the two-layer model solutions (Eqs. 1, 2) by minimizing the cost function (Eq. 4) and with pre-specified parameter F_o (= 8.1246 W/m^2 for $4\times CO_2$ and 3.8749 W/m^2 for $2\times CO_2$ according to Byrne and Goldblatt 2014) and α, which is the equilibrium T response and is estimated iteratively (after each fitting to the ΔT(t) and ΔN(t) data) as described in the 2nd paragraph of Sect. 2.2;
3. Estimate ECS either as the final $\alpha\times\theta$, where θ is 1.0 if the $2\times CO_2$ experiment is used or 0.485 if the $4\times CO_2$ experiment is used, or as $ECS=\theta\times(\Delta T_{mean}+\Delta N_{mean}/b)$, where ΔT_{mean} and ΔN_{mean} are the ΔT and ΔN, respectively, averaged over the last 50 years of the simulation (referred to as the realized warming and the remaining forcing), and b is the (−dN/dT) slope estimated from the fitted T(t) and N(t) functions. The two estimates should be very similar. Our benchmark tests using synthetic data indicated that this ECS estimate may still underestimate the true ECS by up to 0.03 K for using $4\times CO_2$ simulations exceeding 180 years.

The ECS estimates using our new method for 20 CMIP5 models yielded a range from 1.75 to 5.36 K with a mean of 3.45 K (Table 1), which is about 10% higher than the mean of Flato et al. (2013) and may be as much as 25% higher for models with relatively high ECS. The ensemble-mean ECS estimate for 19 CMIP6 models is similar to that for the CMIP5 models, but with a higher range. Furthermore, because the use of a relative short simulation of 140 or 150 years, these estimates contain a mean bias of about − 0.10 K. Accounting for this underestimation and using 0.485 (instead of 0.4769 as in Tables 1, 2) for the $4\times CO_2$ to $2\times CO_2$ conversion, the revised mean ECS should be around 3.61 K with a range of about 1.78–5.45 K for the 20 CMIP5 models and around 3.60 K with a range of 1.85–6.25 K for the 19 CMIP6 models analyzed here (note that the range is not affected noticeably by the mean bias, see Fig. 16a, b). Thus, the ECS reported previously by the IPCC AR5 is likely to be underestimated by about 0.4 K for the multi-model ensemble mean, especially for the models with an ECS above 3 K. Also, while the transient climate response (TCR) is still correlated with our new ECS estimates among the CMIP5 models, the correlation becomes weaker with the TCR accounting for a smaller fraction of the new ECS estimates.

Li et al. (2013) showed that the (−dN/dT) slope becomes extremely large during the quasi-equilibrium (year 1200–4600) and equilibrium (year 4600–6080) periods when ΔT is very small while ΔN continues to fluctuate presumably due to internal variability. Clearly, such slopes are irrelevant for estimating ECS because they result from internal variability rather than the CO_2 forcing, and because the ECS estimates based on recent climate variability are shown to be a poor proxy of ECS (Dessler et al. 2018; Marvel et al. 2018b). This raises the issue of the externally-forced vs. internally-generated (−dN/dT) slopes and how to separate them for estimating ECS as the internally-generated slope produces poor estimates of ECS. We briefly discussed this issue using Figs. 10 and 11 in Sect. 3.3, but further investigation is needed. Local (−dN/dT) slopes over periods of 10^1–10^2 years tend to vary greatly (Fig. 3a), presumably due to different internal processes involved over different time periods; this further suggests that they should not be used to estimate ECS. On the other hand, using all the data of the transient period (years 41–1000 in our case, year 141–1200 in Li et al. 2013) produces a stable slope (for simulations longer than about 180 years) mainly reflecting the forced ΔN and ΔT changes, and this estimate is close to the long-term slope implied by the fitted two-layer model solutions (Fig. 3a) and by the multi-millennium simulations (Fig. 10).

This suggests that one should use all the data from about year 40 to about year 1000 to estimate the (−dN/dT) slope for estimating ECS. According to Li et al. (2013), the simulation may enter a quasi-equilibrium period after year 1200, when the (−dN/dT) slope becomes extremely large and is not related to the CO_2 forcing (however, the impact of the later years on the (−dN/dT) slope is small if all the data after year 40 are used due to the small changes in later years, see Figs. 10 and 11). Using the (−dN/dT) slope from the transient period of Li et al. (2013) would yield an overestimation of ECS, in contrast to other studies and our estimate which suggest an underestimation by the Gregory method. This difference is likely due to the use of the 1%/year ramping-up simulation during the first 140 years to reach the $4\times CO_2$ level that was kept constant thereafter in Li et al. (2013), in contrast to an instantaneous CO_2 quadrupling in standard $4\times CO_2$ experiments commonly used to estimate ECS.

Previous studies using multi-millennium simulations by fully coupled models with coarse resolutions have shown that the slab ocean run may either slightly overestimate (Li et al. 2013) or underestimate (Danabasoglu and Gent 2009) the ECS of the fully coupled model, but this bias seems to be quite small. Our estimated ECS for the CESM1 is also very close to that based on the slab ocean simulations (Gettelman et al. 2012). Unfortunately, the CMIP5 and CMIP6 data archives do not contain slab ocean runs for the $2\times CO_2$ or $4\times CO_2$ experiments, so we cannot verify this for other models. But given these results and the relatively low costs, using slab ocean runs to estimate ECS may still be a valid option, although properly specifying the horizontal heat fluxes in the slab ocean requires some effort.

Acknowledgements We thank Bo Dong for helping set up some of the CESM1 simulations and making the control run used in this study, and Dr. Cao Li of MPI and Dr. David Paytner of GFDL for providing us their model data. We acknowledge the CMIP5 and CMIP6 modeling groups and NCAR CESM project, the Program for Climate Model Diagnosis and Intercomparison and the WCRP's Working Group on Coupled Modelling for their roles in making available the WCRP CMIP multi-model datasets. A. Dai acknowledges the funding support from the U.S. National Science Foundation (Grant No. AGS–1353740 and OISE-1743738), the U.S. Department of Energy's Office of Science (Award No. DE–SC0012602), and the U.S. National Oceanic and Atmospheric Administration (Award nos. NA15OAR4310086 and NA18OAR4310425). B. Rose was supported by NSF (Grant no. AGS-1455071).

Appendix: a note on the estimates of the slopes in noisy data

For two correlated noisy time series x_i and y_i, we can use least-squares fitting to estimate the slopes in the following equations

$$y_i = a_y + b_y x_i + \varepsilon_{yi} \tag{5}$$

$$x_i = a_x + b_x y_i + \varepsilon_{xi} \tag{6}$$

as

$$b_y = r_{xy}\frac{\sigma_y}{\sigma_x}, \quad \text{and} \quad b_x = r_{xy}\frac{\sigma_x}{\sigma_y} \tag{7}$$

where $r_{xy} = r(x,y) = \frac{cov(x,y)}{\sigma_x\sigma_y}$ is the correlation coefficient between x_i and y_i, σ_x and σ_y are the standard deviation of x_i and y_i, respectively, and ε_{yi} and ε_{xi} are the residuals from the fitting and are considered as noise here. Thus, $b_y < b_x$ if $\sigma_y < \sigma_x$, and $b_y \neq 1/b_x$ if $r_{xy} \neq 1$.

For an exact relationship: y = a + b x, we have $r_{xy} = 1$, $\sigma_y = b\,\sigma_x$, so that $b_y = b$, $b_x = 1/b$. Adding weakly correlated noise (with zero mean) to x and y to form two new variables: $X = x + \varepsilon_x$, and $Y = y + \varepsilon_y$ with $r(\varepsilon x, \varepsilon y) \approx 0$. Then, we have $\sigma_X^2 = \sigma_x^2 + \sigma_{\varepsilon x}^2$ and $\sigma_Y^2 = \sigma_y^2 + \sigma_{\varepsilon y}^2$, and

$$\begin{aligned} r_{XY} &= \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{cov(x,y) + cov\left(\varepsilon_x, \varepsilon_y\right)}{\sigma_X\sigma_Y} \approx \frac{cov(x,y)}{\sigma_X\sigma_Y} \\ &= \frac{cov(x,y)}{\sigma_x\sigma_y}\frac{\sigma_x\sigma_y}{\sigma_X\sigma_Y} = r_{xy}\frac{\sigma_x\sigma_y}{\sigma_X\sigma_Y} = \frac{\sigma_x\sigma_y}{\sigma_X\sigma_Y} \end{aligned} \tag{8}$$

Following (7), the slope between X (as the predictor) and Y (as the predictand) is

$$b_Y = r_{XY}\frac{\sigma_Y}{\sigma_X} \approx \frac{\sigma_x\sigma_y}{\sigma_X\sigma_Y}\frac{\sigma_Y}{\sigma_X} = \frac{\sigma_x\sigma_y}{\sigma_X^2} = \frac{b\sigma_x^2}{\sigma_x^2 + \sigma_{\varepsilon x}^2} = \frac{b}{1 + \sigma_{\varepsilon x}^2/\sigma_x^2}. \tag{9}$$

Thus, $|b_Y| < |b|$, and the difference between the estimated slope b_Y and the true slope b increases with the squared noise-to-signal ratio ($\sigma_{\varepsilon x}^2/\sigma_x^2$). Therefore, one should avoid using the variable with large noise (e.g., N(t)) as the predictor in estimating the slope between two data series. For X = T(t) (i.e., the global-mean temperature change series) and Y = = N(t) (i.e., the TOA net radiation change series), the estimated slope (−dN/dT) using least squares fitting should underestimate the true slope between the forced T and N changes (the signals) due to the existence of the noise induced by internal variability (Dai and Bloecker 2019). Since ECS = F/(−dN/dT), this underestimation should lead to an overestimation of ECS in Gregory et al. (2004)'s method. However, as stated in the main text of this paper, the use of the data from the first 40 years or so greatly increase the magnitude of the slope (−dN/dT), whose effect dominates over the effect of noise and leads an underestimation of ECS by the Gregory's method.

References

Andrews T, Gregory JM, Webb MJ, Taylor KE (2012) Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere–ocean climate models. Geophys Res Lett 39:L09712

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

Armour KC (2017) Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. Nat Clim Change 7:331–335. https://doi.org/10.1038/NCLIMATE3278

Armour KC, Bitz CM, Roe GH (2013) Time-varying climate sensitivity from regional feedbacks. J Clim 26:4518–4534

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

Ceppi P, Gregory JM (2019) A refined model for the Earth's global energy balance. Clim Dyn 53:4781–4797. https://doi.org/10.1007/s00382-019-04825-x

Charney JG, Arakawa A, Baker DJ, Bolin B, Dickinson RE, Goody RM, Leith CE, Stommel HM, Wunsch CI (1979) Carbon dioxide and climate: a scientific assessment. National Academy of Sciences, Washington, DC

Cubasch U, Meehl GA, Boer GJ, Stouffer RJ, Dix M, Noda A, Senior CA, Raper S, Yap KS (2001) Projections of future climate change. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (eds) Climate change 2001: The scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom New York, NY, USA, pp 881

Dai A, Bloecker CE (2019) Impacts of internal variability on temperature and precipitation trends in large ensemble simulations by two climate models. Clim Dyn 52:289–306. https://doi.org/10.1007/s00382-018-4132-4

Dai A, Luo D, Song M, Liu J (2019) Arctic amplification is caused by sea–ice loss under increasing CO_2. Nat Commun 10:121. https://doi.org/10.1038/s41467-018-07954-9

Danabasoglu G, Gent PR (2009) Equilibrium climate sensitivity: is it accurate to use a slab ocean model? J Clim 22(9):2494–2499. https://doi.org/10.1175/2008JCLI2596.1

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

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9:1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Feldl N, Roe GH (2013) The nonlinear and nonlocal nature of climate feedbacks. J Clim 26:8289–8304

Flato G, Marotzke J, Abiodun B, Braconnot P, Chou SC, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M (2013) Evaluation of Climate Models. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (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. Cambridge University Press, Cambridge

Forster PM, Andrews T, Good P, Gregory JM, Jackson LS, Zelinka M (2013) Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. J Geophys Res Atmos 118:1139–1150

Garuba OA, Lu J, Liu F, Singh HA (2018) The active role of the ocean in the temporal evolution of climate sensitivity. Geophys Res Lett 45:306–315. https://doi.org/10.1002/2017GL075633

Geoffroy O, Saint-Martin D, Olivié DJL, Voldoire A, Bellon G, Tytéca S (2013a) Transient climate response in a two-layer energy-balance model. Part I: analytical solution and parameter calibration using CMIP5 AOGCM experiments. J Clim 26:1841–1857. https://doi.org/10.1175/JCLI-D-12-00195.1

Geoffroy O, Saint-Martin D, Bellon G, Voldoire A, Olivié DJL, Tytéca S (2013b) 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. J Clim 26:1859–1876

Gettelman A, Kay JE, Shell KM (2012) The evolution of climate sensitivity and climate feedbacks in the community atmosphere model. J Clim 25(5):1453–1469. https://doi.org/10.1175/JCLI-D-11-00197.1

Gregory JM, Ingram WJ, Palmer MA, Jones GS, Stott PA, Thorpe RB, Lowe JA, Johns TC, Williams KD (2004) A new method for diagnosing radiative forcing and climate sensitivity. Geophys Res Lett 31:L03205. https://doi.org/10.1029/2003GL018747

Gregory JM, Andrews T, Good P (2015) The inconstancy of the transient climate response parameter under increasing CO_2. Philos Trans R Soc A 373:20140417

Grose MR, Gregory J, Colman R, Andrews T (2018) What climate sensitivity index is most useful for projections? Geophys Res Lett 45(3):1559–1566

Hansen J, Lacis A, Rind D, Russell G, Stone P, Fung I, Ruedy R, Lerner J (1984) Climate sensitivity: analysis of feedback mechanisms. Clim Process Clim Sensit (AGU Geophysical Monograph Series 29) 5(29):130–163

Hansen J, Sato M, Ruedy R (1997) Radiative forcing and climate response. J Geophys Res 102:6831. https://doi.org/10.1029/96JD03436

Hansen J et al (2005) Efficacy of climate forcings. J Geophys Res 110:D18104. https://doi.org/10.1029/2005JD005776

Haugstad AD, Armour KC, Battisti DS, Rose BEJ (2017) Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. Geophys Res Lett. https://doi.org/10.1002/2017GL074372

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

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

Hurrell JW et al (2013) The Community Earth System Model: a framework for collaborative research. Bull Am Meteorol Soc 94:1339–1360

Jonko AK, Shell KM, Sanderson BM, Danabasoglu G (2013) Climate feedbacks in CCSM3 under changing CO_2 forcing. Part II: variation of climate feedbacks and sensitivity with forcing. J Clim 26:2784–2795. https://doi.org/10.1175/JCLI-D-12-00479.1

Kiehl JT, Shields CA, Hack JJ, Collins WD (2006) The climate sensitivity of the Community Climate System Model version 3 (CCSM3). J Clim 19:2584–2596

Knutti R, Rugenstein MAA (2015) Feedbacks, climate sensitivity and the limits of linear models. Philos Trans R Soc A 373:20150146

Knutti R, Rugenstein MAA, Hegerl GC (2017) Beyond equilibrium climate sensitivity. Nat Geosci 10(10):727–736

Li C, Storch J-S, Marotzke J (2013) Deep-ocean heat uptake and equilibrium climate response. Clim Dyn 41:1071–1086. https://doi.org/10.1007/s00382-012-1350-z

Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. Math Program 45:503–528

Marvel K, Schmidt GA, Miller RL, Nazarenko LS (2018a) Implications for climate sensitivity from the response to individual forcings. Nat Clim Change 6(4):386–389. https://doi.org/10.1038/nclimate2888

Marvel K, Pincus R, Schmidt GA, Miller RL (2018b) Internal variability and disequilibrium confound estimates of climate sensitivity from observations. Geophys Res Lett 45:1595–1601. https://doi.org/10.1002/2017GL076468

Meraner K, Mauritsen T, Voigt A (2013) Robust increase in equilibrium climate sensitivity under global warming. Geophys Res Lett 40(22):5944–5948. https://doi.org/10.1002/2013GL058118

PALEOSENS Project Members (2012) Making sense of palaeoclimate sensitivity. Nature 491:683–691

Paynter D, Frölicher TL, Horowitz LW, Silvers LG (2018) Equilibrium climate sensitivity obtained from multi-millennial runs of two GFDL climate models. J Geophys Res Atmos. https://doi.org/10.1002/2017JD027885

Proistosescu C, Huybers PJ (2017) Slow climate mode reconciles historical and model-based estimates of climate sensitivity. Sci Adv 3:e1602821. https://doi.org/10.1126/sciadv.1602821

Proistosescu C, Donohoe A, Armour KC, Roe GH, Stuecker MF, Bitz CM (2018) Radiative feedbacks from stochastic variability in surface temperature and radiative imbalance. Geophys Res Lett 45:5082–5094. https://doi.org/10.1029/2018GL077678

Randall DA, Wood RA, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer RJ, Sumi A, Taylor KE (2007) Cilmate models and their evaluation. In: olomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom, New York, NY, USA

Rohrschneider T, Stevens B, Mauritsen T (2019) On simple representations of the climate response to external radiative forcing. Clim Dyn. https://doi.org/10.1007/s00382-019-04686-4

Rose BEJ, Rayborn L (2016) The effects of ocean heat uptake on transient climate sensitivity. Curr Clim Change Rep 2:190–201

Rose BEJ, Armour KC, Battisti DS, Feldl N, Koll DDB (2014) The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. Geophys Res Lett 41:1071–1078

Rugenstein MAA, Caldeira K, Knutti R (2016) Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. Geophys Res Lett 43:9877–9885

Rugenstein MAA et al (2019a) Equilibrium climate sensitivity estimated by equilibrating climate models. Geophys Res Lett 47:e2019GL083898. https://doi.org/10.1029/2019GL083898

Rugenstein MAA et al (2019b) LongRunMIP: motivation and design for a large collection of millennial-length AOGCM simulations. Bull Am Meteorol Soc 100:2551–2570. https://doi.org/10.1175/BAMS-D-19-0068.1

Screen JA, Simmonds I (2010) The central role of diminishing sea–ice in recent Arctic temperature amplification. Nature 464:1334–1337

Senior CA, Mitchell JFB (2000) The time-dependence of climate sensitivity. Geophys Res Lett 27(17):2685–2688. https://doi.org/10.1029/2000GL011373

Stevens B, Sherwood SC, Bony S, Webb MJ (2016) Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. Earth's Future 4:512–522

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Winton M, Takahashi K, Held IM (2010) Importance of ocean heat uptake efficacy to transient climate change. J Clim 23:2333–2344

Yoshimori M et al (2016) A review of progress towards understanding the transient global mean surface temperature response to radiative perturbation. Prog Earth Planet Sci 3:21

Zelinka MD, Myers TA, McCoy DT, Po-Chedley S, Caldwell PM, Ceppi P, Klein SA, Taylor KE (2020) Causes of higher climate sensitivity in CMIP6 models. Geophys. Res Lett 47:e2019GL085782. https://doi.org/10.1029/2019GL085782