Geophysical Research Letters

RESEARCH LETTER
10.1029/2021GL095778

Key Points:

- Estimates from historical energy budget constraints underestimate equilibrium climate sensitivity and transient climate response in models
- Atmosphere-only models forced by observed surface warming patterns produce lower estimates of ECS, in line with historical observations
- Discrepancies between modeled and observed historical surface warming patterns account for the differences in climate sensitivity estimates

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Dong,
dongy24@uw.edu

Citation:

Dong, Y., Armour, K. C., Proistosescu, C., Andrews, T., Battisti, D. S., Forster, P. M., et al. (2021). Biased estimates of equilibrium climate sensitivity and transient climate response derived from historical CMIP6 simulations. *Geophysical Research Letters*, *48*, e2021GL095778. https://doi.org/10.1029/2021GL095778

Received 25 AUG 2021
Accepted 3 DEC 2021

Biased Estimates of Equilibrium Climate Sensitivity and Transient Climate Response Derived From Historical CMIP6 Simulations

Yue Dong[1], Kyle C. Armour[1,2], Cristian Proistosescu[3], Timothy Andrews[4], David S. Battisti[1], Piers M. Forster[5], David Paynter[6], Christopher J. Smith[5,7], and Hideo Shiogama[8]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [2]School of Oceanography, University of Washington, Seattle, WA, USA, [3]Department of Atmospheric Sciences and Department of Geology, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA, [4]Met Office Hadley Centre, Exeter, UK, [5]School of Earth and Environment, University of Leeds, Leeds, UK, [6]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, NJ, USA, [7]International Institute for Applied Systems Analysis, Laxenburg, Austria, [8]National Institute for Environmental Studies, Tsukuba, Japan

Abstract This study assesses the effective climate sensitivity (EffCS) and transient climate response (TCR) derived from global energy budget constraints within historical simulations of eight CMIP6 global climate models (GCMs). These calculations are enabled by use of the Radiative Forcing Model Intercomparison Project (RFMIP) simulations, which permit accurate quantification of the radiative forcing. Long-term historical energy budget constraints generally underestimate EffCS from CO_2 quadrupling and TCR from CO_2 ramping, owing to changes in radiative feedbacks and changes in ocean heat uptake efficiency. Atmospheric GCMs forced by observed warming patterns produce lower values of EffCS that are more in line with those inferred from observed historical energy budget changes. The differences in the EffCS estimates from historical energy budget constraints of models and observations are traced to discrepancies between modeled and observed historical surface warming patterns.

Plain Language Summary Here we use climate models and observations to evaluate the extent to which future warming can be inferred from historical climate change. We assess the historical energy budget in 8 climate models by leveraging a recent community effort to quantify radiative forcing within their historical simulations. We find the historical energy budget tends to provide a biased-low constraint on future warming, represented by two metrics: effective climate sensitivity and transient climate response. Moreover, observations and simulations run with observed surface warming patterns produce even lower values of climate sensitivity. This difference in climate sensitivity estimates can be traced to discrepancies between modeled and observed surface warming patterns.

1. Introduction

Equilibrium climate sensitivity (ECS) and transient climate response (TCR) are two fundamental metrics for evaluating climate change projections. ECS represents the *equilibrium* global surface warming in response to a doubling of atmospheric CO_2 concentration relative to pre-industrial levels. Although idealized, ECS has been found to explain most of the spread in projected 21st century global temperature change under realistic emission scenarios (Grose et al., 2018; Sherwood et al., 2020). TCR represents the *transient* surface warming at the time of CO_2 doubling under an idealized 1% per year CO_2 increase. As a measure of transient response, TCR is better constrained and is also informative about the projected degree of global warming in the coming century.

In principle, ECS and TCR can be inferred from the global energy balance framework (Gregory et al., 2004):

$$\Delta N = \Delta F + \lambda \Delta T, \tag{1}$$

where ΔN is the global-mean top-of-atmosphere (TOA) radiation anomaly (approximately equal to ocean heat uptake), ΔF is the effective radiative forcing (ERF; Myhre et al., 2013), ΔT is the global-mean surface air temperature anomaly, and λ is the radiative feedback parameter (negative for a stable climate). By definition, ECS $= -F_{2x}/\lambda_{eq}$, where F_{2x} is the ERF from CO_2 doubling, and λ_{eq} is the radiative feedback when a new equilibrium is

reached ($\Delta N = 0$). ideally, ECS can be estimated from equilibrium states within fully-coupled atmosphere-ocean global climate models (AOGCM) forced by an abrupt CO_2 doubling (abrupt-2xCO2) or CO_2 quadrupling (abrupt-4xCO2), after sufficiently long integration (Rugenstein et al., 2020). In practice, ECS is often extrapolated from a linear regression of ΔN against ΔT for the first 150 years of abrupt-4xCO2 simulations (Gregory et al., 2004). this extrapolation generally underestimates the true ECS due to changes in radiative feedbacks as climate equilibrates (Dunne et al., 2020; Rugenstein et al., 2020), owing to time-evolving surface warming patterns (e.g., Andrews et al., 2015; Armour et al., 2013; Dong et al., 2020), and nonlinear state dependence of radiative feedbacks (e.g., Bloch-Johnson et al., 2015, 2021; Caballero & Huber, 2013). Therefore, we refer the ECS values estimated from these non-equilibrium states to as an *effective* climate sensitivity (EffCS; Andrews et al., 2015; Sherwood et al., 2020):

$$\mathrm{EffCS} = -\frac{F_{2x}}{\lambda_{\mathrm{eff}}}, \tag{2}$$

assuming the effective radiative feedback (λ_{eff}) at a transient state would remain constant to equilibrium. We use $\mathrm{EffCS}_{\mathrm{4xCO2}}$ to refer to ECS estimates from abrupt-4xCO2 simulations (through regressions of annual-mean ΔN against ΔT for the first 150 years; data from Zelinka et al., 2020), and use $\mathrm{EffCS}_{\mathrm{his}}$ to refer to ECS estimates from historical energy budget constraints (see more details in Section 3).

TCR is commonly calculated as the global-mean surface air temperature change averaged over a 20-year period centered on year 70 of the 1pctCO_2 simulations where CO_2 concentration is doubled (referred to as $\mathrm{TCR}_{\mathrm{1pct}}$ here). The values of TCR can also be estimated from historical energy budget constraints (see more details in Section 4), in which case we refer to it as $\mathrm{TCR}_{\mathrm{his}}$.

Estimates of $\mathrm{EffCS}_{\mathrm{his}}$ and $\mathrm{TCR}_{\mathrm{his}}$ from observed energy budget constraints (e.g., Lewis & Curry, 2015, 2018; Otto et al., 2013) have been found to be lower than values of ECS and TCR inferred from other lines of observational and proxy evidence (e.g., Forster et al., 2021; Sherwood et al., 2020). However, AOGCMs have also been found to produce values of $\mathrm{EffCS}_{\mathrm{his}}$ and $\mathrm{TCR}_{\mathrm{his}}$ that are lower than their corresponding values of $\mathrm{EffCS}_{\mathrm{4xCO2}}$ and $\mathrm{TCR}_{\mathrm{1pct}}$, at least within the few models tested (Winton et al., 2020 for GFDL-CM4; Andrews et al., 2019 for HadGEM3-GC3.1-LL; Dessler et al., 2018 for MPI-ESM1.1). The limited number of model studies reflects the fact that the time-varying historical ERF (ΔF in Equation 1) is not often diagnosed, precluding accurate calculation of radiative feedback and thus $\mathrm{EffCS}_{\mathrm{his}}$ and $\mathrm{TCR}_{\mathrm{his}}$. Some other studies have instead used abrupt-4xCO2 or 1pctCO_2 simulations as a surrogate for historical warming (Armour, 2017; Armour et al., 2013; Dong et al., 2020; Lewis & Curry, 2018; Proistosescu & Huybers, 2017), or used a rough estimate of historical ERF taken from IPCC AR5 (Myhre et al., 2013) for CMIP5 AOGCMs (Gregory et al., 2020; Marvel et al., 2018). These approaches generally find that $\mathrm{EffCS}_{\mathrm{4xCO2}}$ is larger than $\mathrm{EffCS}_{\mathrm{his}}$, but it is unclear how accurate their estimates are given that they do not use model-specific estimates of historical ERF.

This work is thus motivated by two key questions: (a) how robust is the finding that values of $\mathrm{EffCS}_{\mathrm{4xCO2}}$ and $\mathrm{TCR}_{\mathrm{1pct}}$ are higher than values of $\mathrm{EffCS}_{\mathrm{his}}$ and $\mathrm{TCR}_{\mathrm{his}}$ estimated using historical energy budget constraints? (b) How do the estimates of $\mathrm{EffCS}_{\mathrm{his}}$ and $\mathrm{TCR}_{\mathrm{his}}$ from models compare to those from observations? The answers to these questions have major implications for how the historical record informs future climate projections. Here, we employ simulations of the Radiative Forcing Model Intercomparison Project (RFMIP; Pincus et al., 2016), which provide the time series of historical ERF for eight CMIP6 AOGCMs (Section 2). With ERF in hand, we assess $\mathrm{EffCS}_{\mathrm{his}}$ and $\mathrm{TCR}_{\mathrm{his}}$ values within historical simulations, and compare to the corresponding values of $\mathrm{EffCS}_{\mathrm{4xCO2}}$ and $\mathrm{TCR}_{\mathrm{1pct}}$ within models and the estimates from observations.

2. Data

2.1. Historical Effective Radiative Forcing From RFMIP Simulations

The ERF includes rapid adjustments from the atmosphere in response to changes in CO_2 or other forcing agents (Myhre et al., 2013). It can be quantified from the TOA radiation changes within atmosphere-only GCM (AGCM) simulations wherein forcing agents are changed while SST and sea-ice concentration (SIC) fields are fixed at pre-industrial values (Forster et al., 2016). Here we make use of the fixed-SST simulations of RFMIP that are currently available for eight CMIP6 models (CanESM5, CNRM-CM6-1, GFDL-CM4, GISS-E2-1-G, HadGEM3-GC31-LL, IPSL-CM6A-LR, MIROC6, and NorESM2-LM). The timeseries of historical ERF is calculated as the

difference of net TOA radiative flux between a 30-year control run (piClim-control), where all forcing agents are fixed to pre-industrial levels, and a forcing run (piClim-histall), where time-varying atmospheric concentrations of all historical forcing agents are imposed. ERF from a single group of forcing agents (e.g., greenhouse gases, anthropogenic aerosols, natural forcings including volcanoes and solar variability) can also be estimated using single-forcing runs of RFMIP (piClim-histghg, piClim-histaer, and piClim-histnat, respectively). We also estimate ERF of CO_2 doubling, F_{2x}, from RFMIP piClim-4xCO2 simulations, where CO_2 is abruptly quadrupled and held constant for 30 years while SST and SIC fields are fixed. F_{2x} is computed from the TOA radiation changes of the 30years-average (scaled by 1/2 to account for CO_2 doubling). For all RFMIP simulations, the ensemble mean is used when more than one member of the simulation exist.

Note that the TOA radiation flux changes derived from the fixed-SST simulations includes the effect of temperature changes over land and sea ice, which should be considered as part of the radiative response rather than ERF. We remove this portion of radiative effects by subtracting off the global-mean surface air temperature change scaled by each model's radiative feedback parameter from its abrupt-4xCO2 simulation–the method proposed in Hansen et al. (2005). Recent studies find advantages in several new correction methods, such as fixing both SST and land-surface temperatures in AGCM (Andrews et al., 2021), or using surface temperature radiative kernels (Smith et al., 2020). We choose to apply the Hansen et al. (2005) method here because it is a widely used method and readily improves ERF estimates using the available output of the RFMIP simulations. All the historical ERFs are calculated as global and annual means, spanning the period 1850–2014 (the same interval of fully-coupled historical simulations).

2.2. Historical Simulations of AOGCMs and AGCMs

Within historical simulations of AOGCMs, we compute global mean N and T from the mean of all available ensemble members (Table S2 in Supporting Information S1), in attempt to reduce noises from internal variability. The annual mean changes of N and T relative to pre-industrial levels are calculated by subtracting a linear fit of the annual global-mean piControl values to remove model drift. Note that ΔN, ΔF, and ΔT in the energy budget framework (Equation 1) can also be defined as differences between two specific historical states. We will elaborate the periods over which we compute the historical energy balance in the following two sections. In order to examine the contributions of individual forcing agents to historical climate change, we also employ single-forcing historical simulations (hist-GHG, hist-aer, and hist-nat), described by the Detection and Attribution Model Intercomparison Project (DAMIP; Gillett et al., 2016), where only one type of forcing agent is changed while all other forcing agents are fixed at preindustrial levels.

Results from the coupled AOGCMs are compared to two sets of AGCM simulations performed as part of the Atmospheric Model Intercomparison Project (AMIP) protocol. One is called "amip," where AGCMs are forced by time-evolving observed SST and SIC fields and by time-varying historical forcing agents. While amip simulations are available for all eight CMIP6 models assessed here, they are performed only over 1979–2015. The other one is called "amip-piForcing," described by the Cloud Feedback Model Intercomparison Project (CFMIP; Webb et al., 2017), where AGCMs are forced by the same observed SST and SIC fields over 1870–2014, except all radiative forcing agents are fixed at pre-industrial levels. The use of AMIP simulations to study feedbacks is based on the expectation that a given SST pattern will result in a unique radiative response whether that pattern is prescribed or internally produced within a model (Haugstad et al., 2017). In Section 3, we show that EffCS estimates from amip and amip-piForcing runs are indeed generally consistent with each other. A caveat is that only six out of eight models used have amip-piForcing experiments available. For both sets of AGCM simulations we only use the first realization of each model, given that most of the variability in TOA radiative fluxes comes about through variations in SSTs, which are the same in these simulations.

3. Historical Energy Budget Constraints on Radiative Feedbacks and EffCS

In the energy budget framework, $\mathrm{EffCS}_{\mathrm{his}}$ can be written as:

$$\mathrm{EffCS}_{\mathrm{his}} = -\frac{F_{2x}}{\lambda_{\mathrm{his}}}, \tag{3}$$

where the historical effective radiative feedback parameter (λ_{his}) is given by:

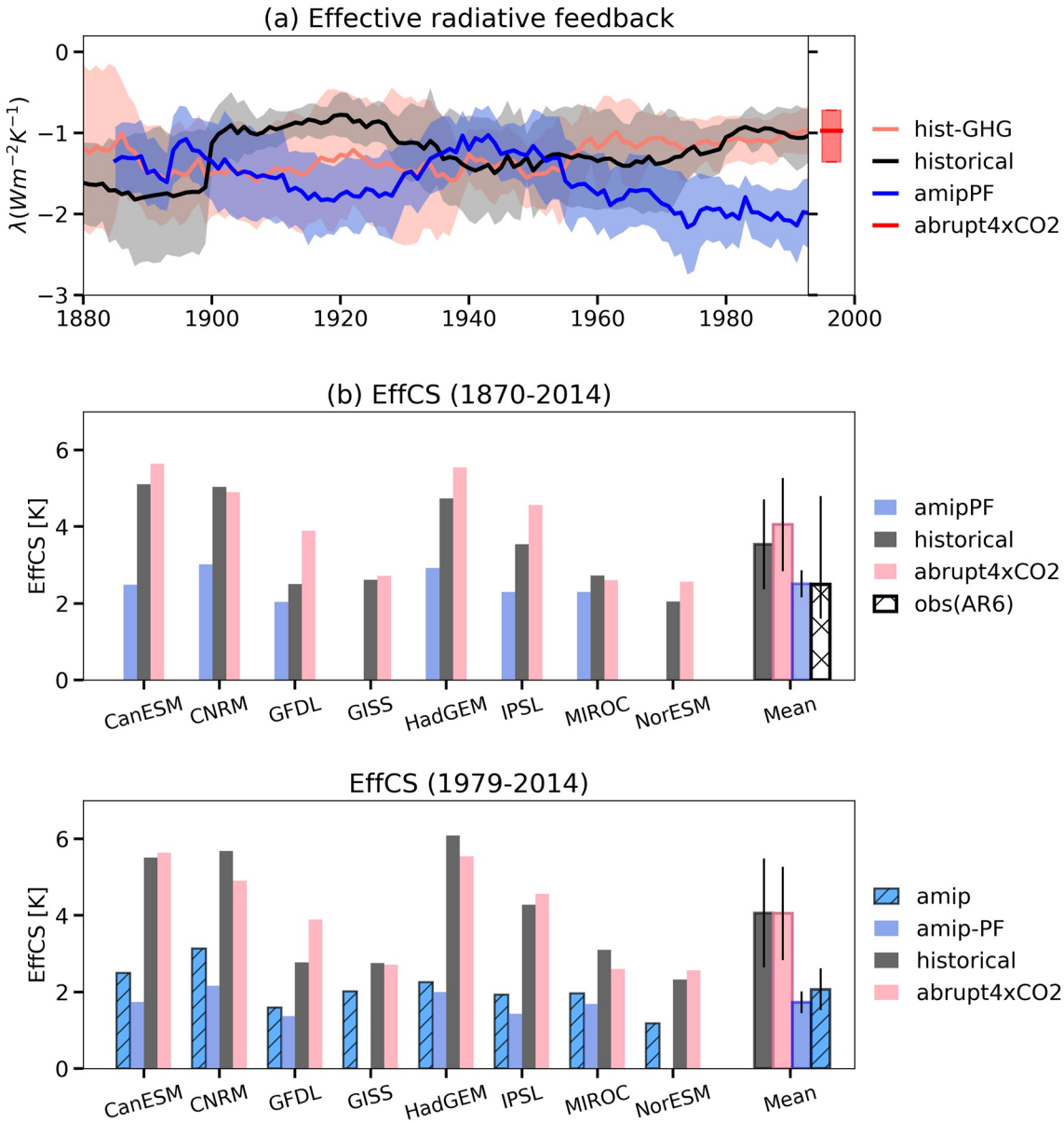

Figure 1. (a) Time series of the estimated λ_{his}. Thick lines denote multi-model means, shadings denote one standard deviation across models. The box plot on the right denotes the interquartile range (box) and the mean value (red line) of λ_{4xCO2} across 8 models. (b and c) EffCS estimated from the energy budget of (b) full historical record (1870–2014) and (c) recent decades (1979–2014). The outlined colored bars on the right in (b), (c) denote the multi-model mean values of EffCS from corresponding simulations, with error bars indicating one standard deviation across models. The white hatched bar in (b) denotes the median $EffCS_{his}$ value of 2.5 K reported in AR6 (Forster et al., 2021) based on observed energy budget changes, with the error bars denoting 5%–95% range of 1.6–4.8 (k) Models listed (from the left to right) are: CanESM5, CNRM-CM6-1, GFDL-CM4, GISS-E2-1-G, HadGEM3-GC31-LL, IPSL-CM6A-LR, MIROC6, and NorESM2-LM.

$$\lambda_{his} = \frac{\Delta N - \Delta F}{\Delta T}. \tag{4}$$

For the historical and amip simulations, we calculate λ_{his} and $EffCS_{his}$ using ΔN and ΔT taken from those simulations combined with ΔF from each model's RFMIP simulation. For the amip-piForcing simulations, which have constant forcing, we calculate λ_{his} and $EffCS_{his}$ only using ΔN and ΔT taken from those simulations (i.e., $\Delta F = 0$).

We first show historical variations in λ_{his}, calculated by a linear regression form of Equation 4 in a sliding 30-year window. We find remarkable differences in decadal-scale radiative feedbacks between historical simulations (black line in Figure 1a) and amip-piForcing simulations (blue line in Figure 1a). While natural variability may

have played a dominant role in the first half of the twentieth century, where net ERF was relatively small (Figure S1 in Supporting Information S1), the discrepancy between AOGCMs and AGCMs persists throughout the full historical period toward early 21st century. Notably, λ_{his} in the amip-piForcing simulations of AGCMs trends toward more-negative values since 1970s to present, consistent with earlier studies using CMIP5 models (Andrews et al., 2018; Dong et al., 2019; Gregory et al., 2020; Silvers et al., 2018); whereas in the historical simulations of AOGCMs, λ_{his} trends slightly toward more-positive values, and at the end of the century becomes comparable to the values of λ_{4xCO2} from abrupt-4xCO2 runs. During the second half of the century, λ_{his} values from AOGCMs track those in hist-GHG simulations (red line), suggesting the simulated feedbacks are primarily driven by GHG forcing, which has dominated global net ERF over this period (Figure S1 in Supporting Information S1).

We next assess $EffCS_{his}$ from energy budget constraints within the historical simulations of the AOGCMs and the AGCMs. To compute the energy budget in Equations 3 and 4, the time interval over which anomalies (Δ) are calculated needs to be carefully chosen to avoid short-term variability and effects of volcanic eruptions (Forster, 2016; Lewis & Curry, 2015). Previous studies have often used two methods: (a) taking finite differences between a base period and a final period (Lewis & Curry, 2015, 2018; Sherwood et al., 2020; Winton et al., 2020); or (b) using regression over the full period of interest (Andrews et al., 2019; Gregory et al., 2020). Since we are comparing EffCS between AOGCMs and AGCMs (including amip simulations which are only available from 1979 onwards), we choose to use the regression method here. That is, the λ_{his} used to compute $EffCS_{his}$ (Equation 3) is calculated via ordinary least squares regression of Equation 4, over two periods of interest: the full historical period 1870–2014 (Figure 1b), and the recent decades of the Satellite Era 1979–2014 (Figure 1c).

3.1. $EffCS_{his}$ From Long-Term Historical Energy Budget (1870–2014)

The values of $EffCS_{his}$ inferred from long-term energy budget in historical simulations are generally lower than $EffCS_{4xCO2}$ from abrupt-4xCO2 simulations (Figure 1b). As noted above, the difference between $EffCS_{his}$ and $EffCS_{4xCO2}$ has been documented in a few models. For GFDL-CM4, Winton et al. (2020) found an $EffCS_{his}$ of 1.8 K and an $EffCS_{4xCO2}$ of 4 K ($EffCS_{4xCO2}$ = 5K if using yrs 51–300 of the model's extended abrupt-4xCO2 simulation). For HadGEM3-GC3.1-LL, Andrews et al. (2019) found an effective F_{2x} of 3.49 Wm^{-2} and a historical feedback of 0.86 $Wm^{-2}K^{-1}$ (average of four ensembles), implying an $EffCS_{his}$ of 4.1 K, in contrast to the model's $EffCS_{4xCO2}$ of 5.5 K. Here, we show that, within six out of eight CMIP6 AOGCMs assessed, historical energy budget constraints underestimate $EffCS_{4xCO2}$ from CO_2 quadrupling. Values of $EffCS_{his}$ range from 4.6% above to 35.6% below values of $EffCS_{4xCO2}$, with an average of 12.6% below across these 8 models (Table S1 in Supporting Information S1). Averaging over all these eight AOGCMs that are currently available, $EffCS_{his}$ is 3.54 K ($\pm$1.17 K; one standard deviation across models, unless noted elsewhere) and $EffCS_{4xCO2}$ is 4.05 K ($\pm$1.46 K), corresponding to an averaged λ_{his} of -1.16 $Wm^{-2}K^{-1}$ ($\pm$0.37 $Wm^{-2}K^{-1}$) and λ_{4xCO2} (from the regression of 150 years abrupt-4xCO2 simulations) of -0.97 $Wm^{-2}K^{-1}$ ($\pm$0.34 $Wm^{-2}K^{-1}$), respectively. Lower values of $EffCS_{his}$ are found in AGCM amip-piForcing experiments over the same historical period, with a mean $EffCS_{his}$ value of 2.51 K ($\pm$0.35 K) across six available models, which is lower than the mean $EffCS_{4xCO2}$ value of 4.52 K ($\pm$1.04 K) across the same 6 models by 44%. Using the Winton et al. (2020) method, that is, taking ΔN, ΔT, and ΔF as differences between 1869–1882 and 1995–2014, yields nearly the same result: the mean value of $EffCS_{his}$ is 3.42 K ($\pm$1.24 K) from historical simulations across all 8 AOGCMs and 2.54 K ($\pm$0.4 K) from amip-piForcing simulations across six available AGCMs.

The EffCS and radiative feedback differences between historical energy budget constraints and CO_2 quadrupling in models arise primarily from differences between historical and near-equilibrium warming patterns (Figure 2). Under CO_2 quadrupling, AOGCMs generally project an equilibrium warming pattern featuring polar amplification and weakened tropical Pacific west-east SST gradient (Figure 2c; Andrews et al., 2015; Andrews & Webb, 2018; Ceppi & Gregory, 2017; Dong et al., 2020); whereas the SST trend patterns in ensemble-mean historical simulations are more spatially uniform (Figure 2a). It has been argued that the projected enhancement of warming in the tropical eastern Pacific relative to the tropical western Pacific in models tends to weaken the lower tropospheric stability, thereby weakening the negative low-cloud feedback and lapse-rate feedback, producing a higher EffCS (Andrews & Webb, 2018; Ceppi & Gregory, 2017; Dong et al., 2019; Zhou et al., 2016). In contrast, the relatively uniform tropical warming patterns simulated in historical simulations would maintain negative cloud feedback and therefore lower EffCS. The fact that $EffCS_{his}$ estimates from amip-piForcing simulations are even lower can be traced to their SST patterns prescribed from observations, featuring slightly enhanced warming

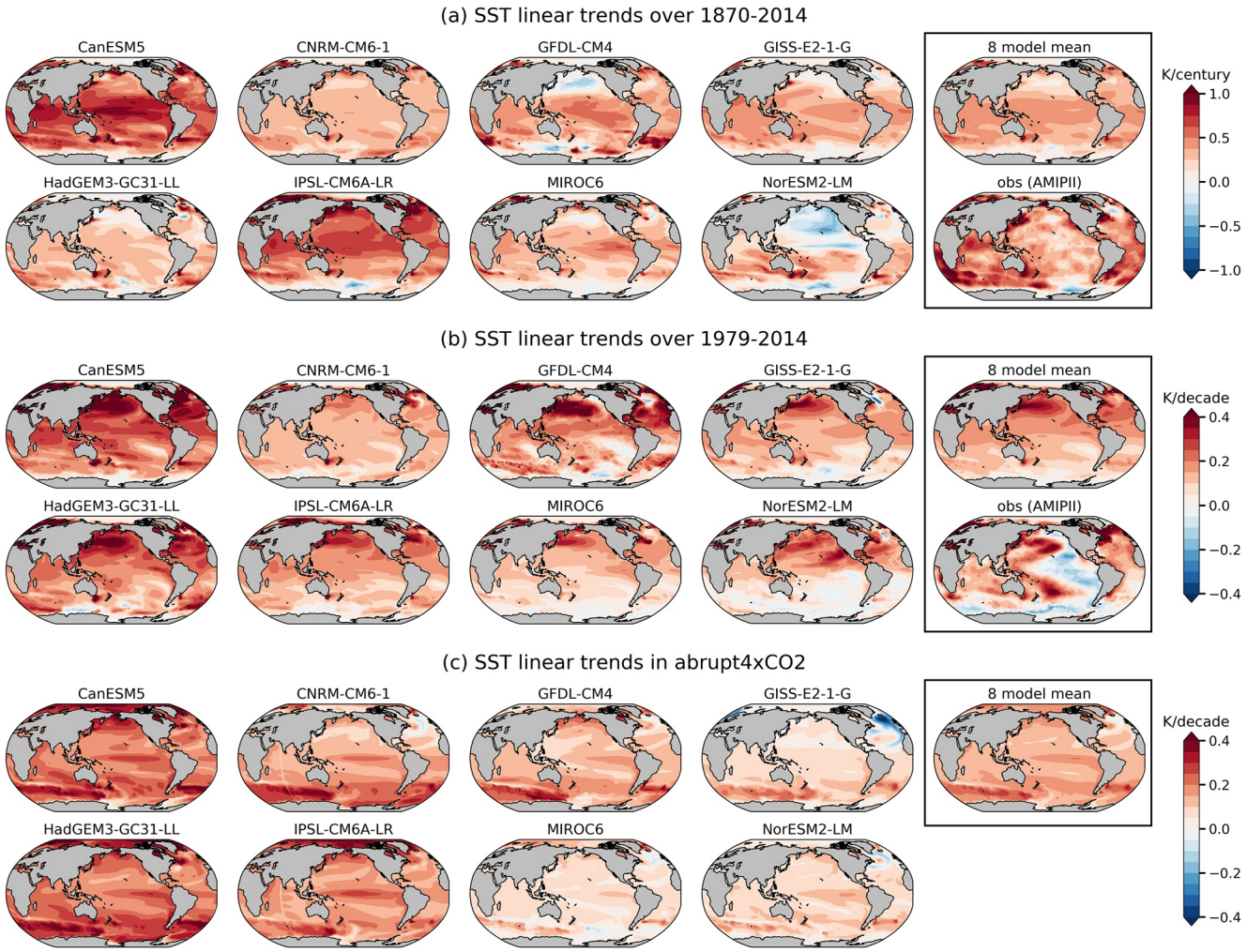

Figure 2. Historical and equilibrium SST trend patterns. Annual-mean SST linear trends over (a) 1870–2014, (b) 1979–2014, and (c) 150 years of abrupt-4xCO2 simulations. The observed SST trend patterns in (a), (b) are calculated using AMIPII dataset (Hurrell et al., 2008). Note that the color scales in (a) and (b and c) are different.

in the Indo-Pacific Ocean and delayed warming in both the eastern Pacific Ocean and the Southern Ocean (e.g., Dong et al., 2020, 2019; Silvers et al., 2018; Zhou et al., 2016).

The historical pattern effect that leads to lower values of $EffCS_{his}$ may partially result from various non-CO_2 forcing agents that have operated in the historical period (e.g., Forster, 2016; Marvel et al., 2016). Gregory et al. (2020) suggest that volcanic forcing may bias estimate of EffCS from CO_2 quadrupling by causing different surface warming patterns in CMIP5 models. Winton et al. (2020) find that a large portion of the $EffCS_{his}$ underestimate in GFDL-CM4 is attributable to its large efficacy of aerosol forcing. To test this possibility within other CMIP6 models, we make use of the DAMIP non-GHG forcing simulations, namely, hist-aer and hist-nat (Figure S2 in Supporting Information S1). Within all but one model, natural forcing alone produces even lower values of $EffCS_{his}$ than those from historical simulations (i.e., a larger historical pattern effect). In comparison, when forced by anthropogenic aerosol forcing alone, four models show a larger historical pattern effect while three models show a reduced pattern effect. These results suggest that non-GHG forcing may largely account for the historical pattern effect, though the impact of aerosol forcing is less robust across models.

3.2. $EffCS_{his}$ From Recent Energy Budget (1979–2014)

Having quantified the long-term historical energy budget constraint on EffCS, we next focus on the most recent decades 1979–2014 (Figure 1c), where observations of global SSTs have been improved by satellite products and better in-situ sampling. This is also the period where GHG forcing has increased dramatically while aerosol forcing trends are relatively small (Figure S1 in Supporting Information S1). With stronger ERF having operated over this period, nearly all coupled AOGCMs produce higher values of $EffCS_{his}$, with a multi-model mean $EffCS_{his}$ of 4.06 K (corresponding to a mean radiative feedback of -1.02 $Wm^{-2}K^{-1}$), comparable to the mean $EffCS_{4xCO2}$ of 4.05 K.

Does this imply that the historical pattern effect is weak in recent decades? In fact, the $EffCS_{his}$ values over this period from all eight AOGCMs are substantially higher than the values from their AGCM counterparts driven by observed warming patterns within amip and amip-piForcing simulations (Figure 1c). Averaging over all these AGCMs, the mean $EffCS_{his}$ and the corresponding λ_{his} from amip simulations is 2.07 K ($\pm$0.57 K) and -1.92 $Wm^{-2}K^{-1}$ ($\pm$0.58 $Wm^{-2}K^{-1}$), respectively. Note that $EffCS_{his}$ estimates from amip runs are slightly higher than those from amip-piForcing runs, potentially due to minor biases in the estimates of ERF from RFMIP fixed-SST simulations (Andrews et al., 2021), but the difference between these two AGCM simulations is much smaller than the difference between AOGCM historical simulations and AGCM simulations.

The $EffCS_{his}$ difference between AOGCMs and their counterpart AGCMs can be traced to the difference between modeled and observed SST patterns over recent decades. The ensemble-mean SST trend pattern in historical simulations of AOGCMs fails to capture many key features in observations (Figure 2b), including the pronounced cooling trends over the eastern Pacific and Southern Ocean. The enhanced tropical Pacific zonal SST gradient has been linked to the observed increase in low clouds over the stratocumulus deck, which contributes to a more-negative radiative feedback and lower EffCS (Ceppi & Gregory, 2017; Dong et al., 2019; Fueglistaler, 2019; Zhou et al., 2016). The Southern Ocean SSTs have also been found to have large impacts on low-cloud feedbacks over the Southern Ocean and therefore EffCS (Dong et al., 2020; Gjermundsen et al., 2021). Moreover, potential teleconnections from the Southern Ocean to the tropical Pacific may also contribute to the observed changes in the tropical cloud feedback and lower EffCS over this period (e.g., Hwang et al., 2017; Kang et al., 2020).

A few studies have argued that the observed tropical Pacific SST pattern may be driven by aerosol forcing (Takahashi & Watanabe, 2016) or volcanic forcings (Gregory et al., 2020). Using the DAMIP simulations, we found that the SST trend patterns driven by anthropogenic aerosol forcing and natural forcing are indeed more spatially heterogeneous, with some models showing weak cooling in the tropical eastern Pacific (Figure S3 in Supporting Information S1). However, the cooling trends produced in these non-GHG simulations are much weaker than that observed, and are generally overwhelmed by the warming trends produced by GHG forcing (Figure S3a in Supporting Information S1). It is also possible that the observed warming pattern in part results from natural variability. For example, Watanabe et al. (2021) found that the observed equatorial Pacific west-east SST gradient over a longer period (1951–2010) lies within the range of large ensembles of model simulations. However, such regional analyses may be insufficient to explain the observed SST trend pattern beyond the equatorial Pacific, and their results may be sensitive to the time interval selected. We have examined $EffCS_{his}$ and the equatorial Pacific zonal SST gradient for all individual members of historical simulations. We define the zonal SST gradient following Watanabe et al. (2021): the difference between the eastern Pacific (180–80W, 5S–5N) and the western Pacific (110E−180, 5S–5N). But we calculate SST linear trends over 1979–2014 instead of 1951–2010. Over these recent decades, nearly all of the 201 ensemble members fail to capture the low $EffCS_{his}$ values from the corresponding amip simulations and the observed zonal SST gradient (Figure S4 in Supporting Information S1), suggesting a significant discrepancy in the pattern effect between AOGCMs and observations.

Identifying the causes of the recent observed SST trend pattern is beyond the scope of this study. Our results on the historical energy budget constraints suggest that $EffCS_{his}$ estimates from historical simulations generally underestimate $EffCS_{4xCO2}$ due to the pattern effect. However, the historical pattern effect is relatively small over recent decades in AOGCMs, owing to the bias of their historical warming patterns.

4. Historical Energy Budget Constraints on TCR

In the energy budget framework, TCR can be inferred from sufficiently long-term historical record where ΔT increases approximately proportional to ΔF:

$$\mathrm{TCR}_{\mathrm{his}} = \Delta T \frac{F_{2\mathrm{x}}}{\Delta F}. \tag{5}$$

Under the global energy framework (Equation 1), $\mathrm{TCR}_{\mathrm{his}}$ is governed by both historical radiative feedback (λ_{his}) and ocean heat uptake (OHU) efficiency (κ_{his}), with the relationship between these approximated as (Gregory et al., 2015; Gregory & Forster, 2008; Gregory & Mitchell, 1997; Raper et al., 2002):

$$\mathrm{TCR}_{his} = \frac{F_{2\mathrm{x}}}{\kappa_{\mathrm{his}} - \lambda_{\mathrm{his}}}, \tag{6}$$

where κ_{his} is defined as:

$$\kappa_{\mathrm{his}} = \frac{\Delta N}{\Delta T}. \tag{7}$$

Here, we calculate $\mathrm{TCR}_{\mathrm{his}}$ from historical simulations using Equation 5, where anomalies (Δ) are averaged over 1995–2014 relative to 1869–1882. This period is chosen to cover a sufficiently long time of historical record, and also to be largely consistent with several recent studies (Lewis & Curry, 2018; Winton et al., 2020). As noted above, Winton et al. (2020) found a $\mathrm{TCR}_{\mathrm{his}}$ of 1.27 K for GFDL-CM4, lower than the model's $\mathrm{TCR}_{\mathrm{1pct}}$ of 2.05 K. Here we find that seven out eight AOGCMs assessed are consistent with GFDL-CM4—the historical energy budget constraint underestimates TCR values from 1pctCO_2 simulations. Values of $\mathrm{TCR}_{\mathrm{his}}$ range from 9.7% above to 32.2% below values of $\mathrm{TCR}_{\mathrm{1pct}}$, with an average of 13% below across all eight models. The mean $\mathrm{TCR}_{\mathrm{his}}$ value across 8 AOGCMs is 1.81 K ($\pm$0.51 K), lower than the mean $\mathrm{TCR}_{\mathrm{1pct}}$ value of 2.08 K ($\pm$0.43 K).

As shown in Equation 6, the difference between $\mathrm{TCR}_{\mathrm{his}}$ and $\mathrm{TCR}_{\mathrm{1pct}}$ could arise from changes in radiative feedbacks and/or changes in OHU efficiency over time (Gregory et al., 2015). To separate these two factors, we estimate λ and κ from historical and 1pctCO_2 simulations, following Equation 4 and Equation 7, respectively. For historical estimates, ΔN, ΔT and ΔF are taken as finite differences between 1995–2014 and 1869–1882. For 1pctCO_2 estimates, ΔN and ΔT are from the 20-year period centered on year 70 of the simulation when CO_2 is doubled; ΔF at the time of CO_2 doubling is approximated by $F_{2\mathrm{x}}$, with a caveat that the true $F_{2\mathrm{x}}$ in 1pctCO_2 simulations was found slightly non-logarithmic (Gregory et al., 2015, 2020).

In all models, κ_{his} is larger than κ_{1pct}, which could contribute to the lower values of $\mathrm{TCR}_{\mathrm{his}}$ relative to $\mathrm{TCR}_{\mathrm{1pct}}$ (Figure3b). The difference between κ_{1pct} and κ_{his} could be associated with ocean stratification response on different time scales, or could arise from changes in Atlantic meridional overturning circulation or Southern Ocean meridional overturning circulation driven by historical non-CO_2 forcings. On the other hand, the difference between λ_{his} and λ_{1pct} varies by models (Figure 3c). Two models show λ_{his} more negative than λ_{1pct}, along with their large κ_{his}, suggesting that the lower values of $\mathrm{TCR}_{\mathrm{his}}$ in these models are owing to changes in both radiative feedbacks and OHU efficiency. The rest of the models show λ_{his} either very close to or slightly less negative than λ_{1pct}, suggesting a dominant role of changes in κ.

In summary, we find an overall underestimate of TCR of about 0.2 K using historical energy budget constraints within AOGCMs, owing to the combination of more-negative radiative feedback and/or larger OHU efficiency during the historical period. The differences in λ and κ between historical and 1pctCO_2 are largely ameliorated when using hist-GHG simulations (Figure S5 in Supporting Information S1), suggesting that the underestimate of $\mathrm{TCR}_{\mathrm{his}}$ is mostly driven by historical non-GHG forcings. These results suggest that as time evolves and CO_2 forcing increases, the weakening of both λ and κ could lead to higher values of TCR than those inferred from historical energy budget constraints.

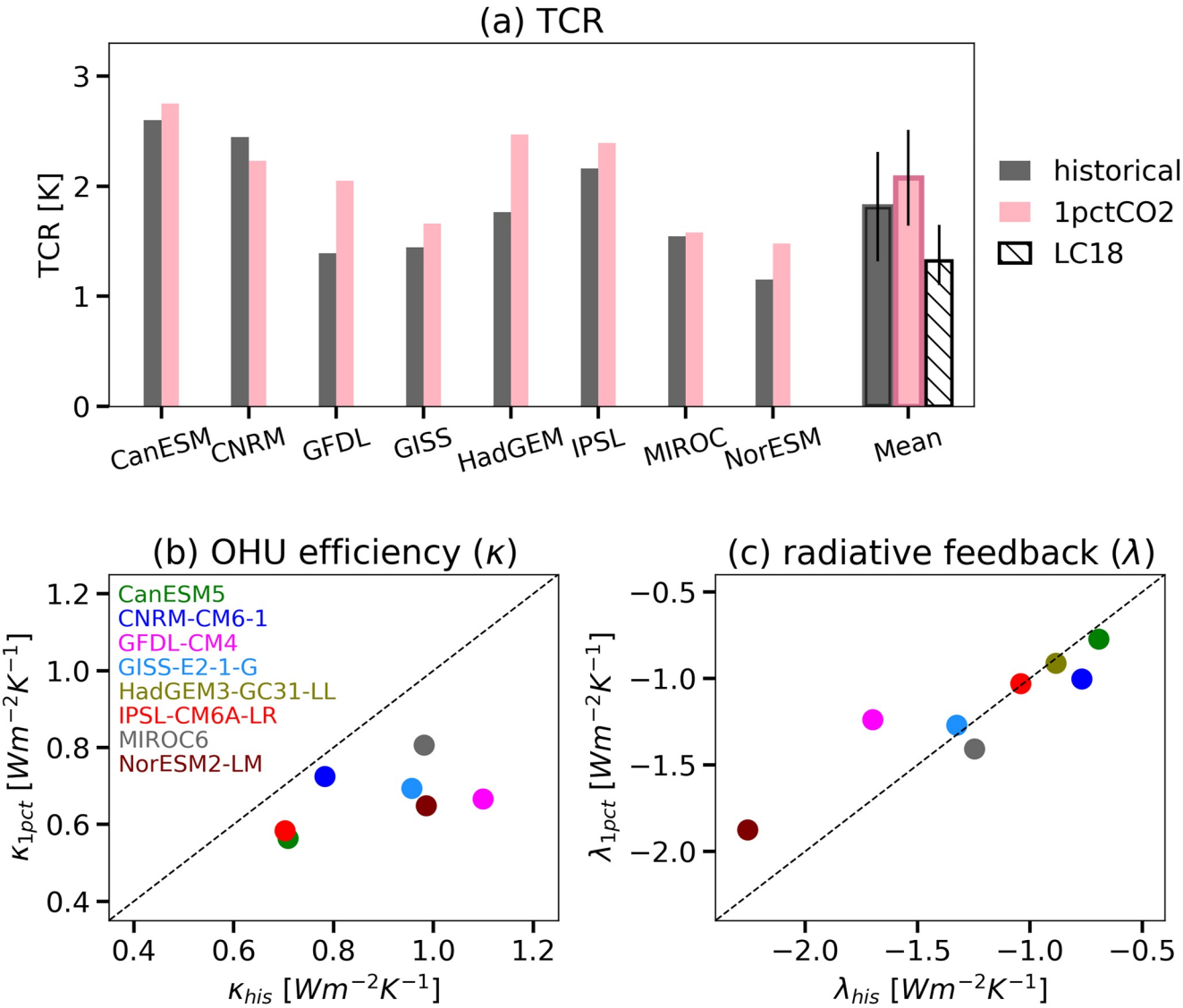

Figure 3. (a) TCR estimates from fully-coupled historical simulations (TCR_{his}; black bars) and 1pctCO_2 simulations (TCR_{1pct}; red bars). The white hatched bar denotes the best estimate of TCR_{his} of 1.32 K inferred by Lewis and Curry (2018) based on estimates of observed energy budget changes, with a 17%–83% range of 1.1–1.65 K. (b) Ocean heat uptake efficiency and (c) radiative feedback from historical and 1pctCO_2 simulations of all eight AOGCMs.

5. Discussions and Conclusions

In the previous two sections, we have compared estimates of EffCS and TCR between different simulations of coupled and atmosphere-only GCMs. How do the model results compare to values of $EffCS_{his}$ and TCR_{his} from observations, and what implications do the results have for our interpretation of the observed energy budget constraints?

In Figures 1 and 3 we show that the reported values of $EffCS_{his}$ and TCR_{his} from observations are lower than the values of $EffCS_{4xCO2}$ and TCR_{1pct} from CMIP6 models. For an observation-based estimate of $EffCS_{his}$, we use values reported in IPCC AR6 (Forster et al., 2021): a median value of 2.5 K and 5%–95% range of 1.6–4.8 K based on observed energy budget changes from 1850–1900 to 2006–2019 (Figure 1b). For TCR, we use values reported by Lewis and Curry (2018): a median value of 1.32 K and a 17%–83% range of 1.1–1.65 K based on observed energy budget changes over 1869–1882 to 1995–2016 (Figure 3). Values of $EffCS_{his}$ from AGCM simulations forced by observed SST patterns are well in line with observation-based values of $EffCS_{his}$, despite the fact that the values of $EffCS_{his}$ and $EffCS_{4xCO2}$ from their counterpart AOGCMs are both higher. The difference

between estimates of $EffCS_{4xCO2}$ from abrupt-4xCO2 simulations and estimates of $EffCS_{his}$ from AGCM simulations with observed surface warming is thus owing to changes in SST patterns with time. It implies that if nature evolves toward equilibrium in the way that AOGCMs project, we should expect higher values of EffCS and TCR (i.e., evolving toward $EffCS_{4xCO2}$ and TCR_{1pct}) in the future than those inferred from observed energy budget constraints.

Our findings are broadly consistent with earlier studies focusing on two individual CMIP6 models (Andrews et al., 2019; Winton et al., 2020): historical energy budget constraints generally (within 6 out of 8 AOGCMs) underestimate the values of EffCS from CO_2 quadrupling and TCR from CO_2 ramping. The underestimate of $EffCS_{his}$ is owing to differences in radiative feedbacks induced by the pattern effect; the underestimate of TCR_{his} is owing to a combination of differences in both radiative feedbacks and OHU efficiency. Accounting the pattern effect and assuming the observed SST pattern will evolve toward the projected equilibrium warming pattern, the observed historical energy budget may provide a biased-low constraint on EffCS and TCR.

That said, the projections by GCMs are confronted by not only uncertainties associated with atmospheric physics, for example, cloud feedback response to a given SST pattern, but also an open question: how reliable are model projections of future SST patterns? AOGCMs generally fail to reproduce the observed historical SST pattern, which led to an inconsistency between EffCS estimates from coupled historical runs and those from amip runs and observations. If the observed SST trend pattern is caused by natural variability, which will reverse sign in the coming decades according to AOGCM projections (Watanabe et al., 2021), then the higher values of EffCS and TCR found within AOGCMs may be more informative about near-future climate change under continued CO_2 forcing. If the recently observed SST trend pattern is a result of model biases in the response to GHG forcing (e.g., Coats & Karnauskas, 2017; Seager et al., 2019), the lower values of $EffCS_{his}$ and TCR_{his} from observations may persist over the coming decades, in which Case 21st century warming may be lower than that projected even by GCMs with realistic ECS values. This work suggests that both understanding the causes of the recent observed warming pattern and making accurate projections of future warming patterns are important for constraining transient and near-equilibrium climate change.

Data Availability Statement

The CMIP6 simulations used in this study and their DOIs are listed in Table S2 in Supporting Information S1. The processed data are available at http://hdl.handle.net/1773/48133.

Acknowledgments
We thank Michael Winton, Nadir Jeevanjee and two anonymous reviewers for constructive comments. Y Dong, K. C. Armour were supported by National Science Foundation Grant AGS-1752796 and the National Oceanic and Atmospheric Administration MAPP Program (Award NA20OAR4310391). K. C. Armour was supported by an Alfred P. Sloan Research Fellowship (grant number FG-2020-13,568). C. Proistosescu was supported by National Science Foundation grant OCE-2002385. D. S. Battisti was supported by the Tamaki Foundation. T Andrews was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra, and the European Union's Horizon 2020 research and innovation program under grant agreement No. 820829 (CONSTRAIN project). Funding for P. M. Forster was provided by the European Union's Horizon 2020 Research and Innovation Programme under grant no. 820829 (CONSTRAIN). C. J. Smith was supported by a NERC/IIASA Collaborative Research Fellowship (NE/T009381/1). H. Shiogama was supported by the Integrated Research Program for Advancing Climate Models (JPMXD0717935457) and Grants-in-Aid for Scientific Research (21H01161) of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., et al. (2019). Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3.1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4377–4394. https://doi.org/10.1029/2019ms001866

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., et al. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, *45*(16), 8490–8499. https://doi.org/10.1029/2018gl078887

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *Journal of Climate*, *28*(4), 1630–1648. https://doi.org/10.1175/jcli-d-14-00545.1

Andrews, T., Smith, C. J., Myhre, G., Forster, P. M., Chadwick, R., & Ackerley, D. (2021). Effective radiative forcing in a GCM with fixed surface temperatures. *Journal of Geophysical Research: Atmospheres*, *126*(4), e2020JD033880. https://doi.org/10.1029/2020jd033880

Andrews, T., & Webb, M. J. (2018). The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical Pacific warming. *Journal of Climate*, *31*(2), 641–654. https://doi.org/10.1175/jcli-d-17-0087.1

Armour, K. C. (2017). Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, *7*(5), 331–335. https://doi.org/10.1038/nclimate3278

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

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, *42*(12), 4973–4980. https://doi.org/10.1002/2015gl064240

Bloch-Johnson, J., Rugenstein, M., Stolpe, M. B., Rohrschneider, T., Zheng, Y., Gregory, J. M. (2021). Climate sensitivity increases under higher CO_2 levels due to feedback temperature dependence. *Geophysical Research Letters*, *48*(4), e2020GL089074. https://doi.org/10.1029/2020GL089074

Caballero, R., & Huber, M. (2013). State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proceedings of the National Academy of Sciences*, *110*(35), 14162–14167. https://doi.org/10.1073/pnas.1303365110

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

Coats, S., & Karnauskas, K. B. (2017). Are simulated and observed twentieth century tropical Pacific sea surface temperature trends significant relative to internal variability? *Geophysical Research Letters*, *44*(19), 9928–9937. https://doi.org/10.1002/2017gl074622

Dessler, A. E., Mauritsen, T., & Stevens, B. (2018). The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity. *Atmospheric Chemistry and Physics*, *18*(7), 5147–5155. https://doi.org/10.5194/acp-18-5147-2018

Dong, Y., Armour, K. C., Zelinka, M. D., Proistosescu, C., Battisti, D. S., Zhou, C., & Andrews, T. (2020). Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. *Journal of Climate*, *33*(18), 7755–7775. https://doi.org/10.1175/jcli-d-19-1011.1

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

Dunne, J. P., Winton, M., Bacmeister, J., Danabasoglu, G., Gettelman, A., Golaz, J.-C., et al. (2020). Comparison of equilibrium climate sensitivity estimates from Slab ocean, 150-year, and longer simulations. *Geophysical Research Letters*, *47*(16), e2020GL088852. https://doi.org/10.1029/2020gl088852

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

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., et al. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *Journal of Geophysical Research: Atmospheres*, *121*(20), 12–460. https://doi.org/10.1002/2016jd025320

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J., Frame, D., et al. (2021). The Earth's energy budget, climate feedbacks, and climate sensitivity. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. PÃ©an, S. Berger, et al. (Eds.), *Climate change 2021: The Physical Science Basis*. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on climate change*. Cambridge University Press.

Fueglistaler, S. (2019). Observational evidence for two modes of coupling between sea surface temperatures, tropospheric temperature profile, and shortwave cloud radiative effect in the tropics. *Geophysical Research Letters*, *46*(16), 9890–9898. https://doi.org/10.1029/2019gl083990

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., et al. (2016). The detection and attribution model intercomparison project (DAMIP v1.0) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3685–3697. https://doi.org/10.5194/gmd-9-3685-2016

Gjermundsen, A., Nummelin, A., Olivié, D., Bentsen, M., Seland, Ø., & Schulz, M. (2021). Shutdown of southern ocean convection controls long-term greenhouse gas-induced warming. *Nature Geoscience*, *14*, 724–731. https://doi.org/10.1038/s41561-021-00825-x

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., & Webb, M. (2020). How accurately can the climate sensitivity to CO_2 be estimated from historical climate change? *Climate Dynamics*, *54*(1), 129–157. https://doi.org/10.1007/s00382-019-04991-y

Gregory, J. M., Andrews, T., & Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO_2. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *373*(2054), 20140417. https://doi.org/10.1098/rsta.2014.0417

Gregory, J. M., & Forster, P. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research*, *113*(D23). https://doi.org/10.1029/2008jd010405

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3). https://doi.org/10.1029/2003gl018747

Gregory, J. M., & Mitchell, J. F. (1997). The climate response to CO_2 of the Hadley Centre coupled AOGCM with and without flux adjustment. *Geophysical Research Letters*, *24*(15), 1943–1946. https://doi.org/10.1029/97gl01930

Grose, M. R., Gregory, J., Colman, R., & Andrews, T. (2018). What climate sensitivity index is most useful for projections? *Geophysical Research Letters*, *45*(3), 1559–1566. https://doi.org/10.1002/2017gl075742

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G., et al. (2005). Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, *110*(D18). https://doi.org/10.1029/2005jd005776

Haugstad, A., Armour, K., Battisti, D., & Rose, B. (2017). Relative roles of surface temperature and climate forcing patterns in the inconstancy of radiative feedbacks. *Geophysical Research Letters*, *44*(14), 7455–7463. https://doi.org/10.1002/2017gl074372

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., & Rosinski, J. (2008). A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *Journal of Climate*, *21*(19), 5145–5153. https://doi.org/10.1175/2008jcli2292.1

Hwang, Y.-T., Xie, S.-P., Deser, C., & Kang, S. M. (2017). Connecting tropical climate change with southern ocean heat uptake. *Geophysical Research Letters*, *44*(18), 9449–9457. https://doi.org/10.1002/2017gl074972

Kang, S. M., Xie, S.-P., Shin, Y., Kim, H., Hwang, Y.-T., Stuecker, M. F., et al. (2020). Walker circulation response to extratropical radiative forcing. *Science advances*, *6*(47), eabd3021. https://doi.org/10.1126/sciadv.abd3021

Lewis, N., & Curry, J. A. (2015). The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dynamics*, *45*(3), 1009–1023. https://doi.org/10.1007/s00382-014-2342-y

Lewis, N., & Curry, J. A. (2018). The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *Journal of Climate*, *31*(15), 6051–6071. https://doi.org/10.1175/jcli-d-17-0667.1

Marvel, K., Pincus, R., Schmidt, G. A., & Miller, R. L. (2018). Internal variability and disequilibrium confound estimates of climate sensitivity from observations. *Geophysical Research Letters*, *45*(3), 1595–1601. https://doi.org/10.1002/2017gl076468

Marvel, K., Schmidt, G. A., Miller, R. L., & Nazarenko, L. S. (2016). Implications for climate sensitivity from the response to individual forcings. *Nature Climate Change*, *6*(4), 386–389. https://doi.org/10.1038/nclimate2888

Myhre, G., Shindell, D., Bréon, F., Collins, W., Fuglestvedt, J., Huang, J., et al. (2013). Anthropogenic and natural radiative forcing. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Doschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis*. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

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

Pincus, R., Forster, P. M., & Stevens, B. (2016). The radiative forcing model Intercomparison project (RFMIP): Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(9), 3447–3460. https://doi.org/10.5194/gmd-9-3447-2016

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*(7), e1602821. https://doi.org/10.1126/sciadv.1602821

Raper, S. C., Gregory, J. M., & Stouffer, R. J. (2002). The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *Journal of Climate*, *15*(1), 124–130. https://doi.org/10.1175/1520-0442(2002)015<0124:TROCSA>2.0.CO;2

Rugenstein, M., Bloch-Johnson, J., Gregory, J. M., Andrews, T., Mauritsen, T., Li, C., et al. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, *47*(4), e2019GL083898. https://doi.org/10.1029/2019gl083898

Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., & Zhang, H. (2019). Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Climate Change*, *9*(7), 517–522. https://doi.org/10.1038/s41558-019-0505-x

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K., Forster, P., Hargreaves, J., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678. https://doi.org/10.1029/2019rg000678

Silvers, L. G., Paynter, D., & Zhao, M. (2018). The diversity of cloud responses to twentieth century sea surface temperatures. *Geophysical Research Letters*, *45*(1), 391–400. https://doi.org/10.1002/2017gl075583

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., et al. (2020). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chemistry and Physics*, *20*(16), 9591–9618. https://doi.org/10.5194/acp-20-9591-2020

Takahashi, C., & Watanabe, M. (2016). Pacific trade winds accelerated by aerosol forcing over the past two decades. *Nature Climate Change*, *6*(8), 768–772. https://doi.org/10.1038/nclimate2996

Watanabe, M., Dufresne, J.-L., Kosaka, Y., Mauritsen, T., & Tatebe, H. (2021). Enhanced warming constrained by past trends in equatorial Pacific sea surface temperature gradient. *Nature Climate Change*, *11*(1), 33–37. https://doi.org/10.1038/s41558-020-00933-3

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., et al. (2017). The cloud feedback model intercomparison project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, *10*(1), 359–384. https://doi.org/10.5194/gmd-10-359-2017

Winton, M., Adcroft, A., Dunne, J., Held, I., Shevliakova, E., Zhao, M., et al. (2020). Climate sensitivity of GFDL's CM4.0. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001838. https://doi.org/10.1029/2019ms001838

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782. https://doi.org/10.1029/2019GL085782

Zhou, C., Zelinka, M. D., & Klein, S. A. (2016). Impact of decadal cloud variations on the Earth's energy budget. *Nature Geoscience*, *9*(12), 871–874. https://doi.org/10.1038/ngeo2828

References From the Supporting Information

Hirahara, S., Ishii, M., & Fukuda, Y. (2014). Centennial-scale sea surface temperature analysis and its uncertainty. *Journal of Climate*, *27*(1), 57–75. https://doi.org/10.1175/jcli-d-12-00837.1

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. https://doi.org/10.1175/jcli-d-16-0836.1

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14). https://doi.org/10.1029/2002jd002670