# Observational Constraint on the Climate Sensitivity to Atmospheric $CO_2$ Concentrations Changes Derived from the 1971–2017 Global Energy Budget

JONATHAN CHENAL,[a] BENOÎT MEYSSIGNAC,[a] AURÉLIEN RIBES,[b] AND ROBIN GUILLAUME-CASTEL[a]

[a] *LEGOS, Université de Toulouse, CNES, CNRS, IRD, UPS, Toulouse, France*
[b] *Centre National de Recherches Météorologiques, Université de Toulouse, Météo France, CNRS, Toulouse, France*

(Manuscript received 28 July 2021, in final form 1 February 2022)

ABSTRACT: The estimate of the historical effective climate sensitivity (histeffCS) is revisited with updated historical observations of the global energy budget in order to derive an observational constraint on the effective sensitivity of climate to $CO_2$ ($CO_2$effCS). A regression method based on observations of the energy budget over 1971–2017 is used to estimate the histeffCS (4.34 [2.17; 22.83] K: median and 5%–95% range). Then, climate model simulations are used to evaluate the distance between the histeffCS and the $CO_2$effCS. The observational estimate of the histeffCS and the distance between the histeffCS and the $CO_2$effCS are combined to derive an observational constraint on $CO_2$effCS of 5.46 [2.40; 35.61] K. The main sources of uncertainty in the $CO_2$effCS estimate comes from the uncertainty in aerosol forcing and in the top of the atmosphere energy imbalance. Further uncertainty arises from the pattern effect correction estimated from climate models. There is confidence in the lower end of the 5%–95% range derived from our method because it relies only on reliable recent data and it makes full use of the observational record since 1971. This important result suggests that observations of the global energy budget since 1971 are poorly consistent with climate sensitivity to $CO_2$ below 2.4 K. Unfortunately, the upper end of the 5%–95% range derived from the regression method is above 30 K. This means that the observational constraint derived from observations of the global energy budget since 1971 is too weak (i.e., the uncertainty is too large) to provide any relevant information on the credibility of high $CO_2$effCS.

KEYWORDS: Climate change; Climate sensitivity; Radiative forcing

## 1. Introduction

The climate sensitivity characterizes the asymptotic response of the surface air temperature (SAT) of the climate system to a given increase in $CO_2$ concentrations. It is a parameter that depends on the definition of the climate system (i.e., which components it includes), the processes that are accounted for in the climate system, and the time scale that is considered (see, e.g., Heinze et al. 2019; Ghil and Lucarini 2020). Most recent studies adopt Charney et al.'s definition of the climate sensitivity that has been considered in the Charney report (National Research Council 1979) and in the Intergovernmental Panel on Climate Change (IPCC) reports. This definition excludes the carbon cycle response and the ice sheet response in the estimation of the asymptotic response of the SAT. By convention, in the Charney report, the climate sensitivity has been defined as the increase in SAT that is reached at equilibrium after an abrupt doubling of preindustrial atmospheric $CO_2$ concentrations. This climate sensitivity is often called the equilibrium climate sensitivity (ECS).

The ECS is difficult to estimate because climate change simulations need to be run for thousands of years to simulate the response of the deepest layers of the ocean and reach a new equilibrium (e.g., Rugenstein et al. 2020). In general, for practical reasons, climate simulations are not run so long and the climate sensitivity is actually derived from a few hundred years of climate change simulations while the climate system has not yet reached equilibrium. Such climate sensitivities that are estimated over periods of nonequilibrium, are called "effective climate sensitivities" (effCS). The common practice in the literature is to estimate an effective climate sensitivity from the first 150 years of climate change simulation under an abrupt quadrupling of $CO_2$ concentration (e.g., Zelinka et al. 2020). We call $CO_2$ effective climate sensitivity (CO2effCS) the climate sensitivity that is obtained by regressing the top-of-atmosphere (TOA) radiative imbalance against SAT and then linearly extrapolating to determine the temperature at which the radiative imbalance is zero (see Gregory et al. 2004 and following articles). This effective climate sensitivity does not account for the carbon cycle response and the ice sheet response like the ECS. In addition, it ignores the response of the ocean and sea ice beyond 150 years (unlike the ECS).

The recent LongrunMIP project (Rugenstein et al. 2019) has shown that $CO_2$effCS is actually a good proxy for ECS

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0565.s1.

Chenal's current affiliation: École Nationale des Ponts et Chaussées, Marne-la-Vallée, France.

*Corresponding author*: Jonathan Chenal, jonathan.chenal@legos.obs-mip.fr



DOI: 10.1175/JCLI-D-21-0565.1

because it is highly correlated with ECS (>0.9) although it is slightly biased low (by less than 20%). $CO_2$effCS is also, by itself, a key parameter of the climate response to changes in $CO_2$ concentrations because it is representative at centennial time scales of global warming under realistic scenarios of future climate change. Indeed, at centennial time scales, it correlates well with the global increase in SAT (e.g., Grose et al. 2018) making it a reliable index for future surface warming and for future changes in the many variables that are physically related to global SAT (such as global atmospheric water vapor, global precipitations, global ocean temperature, global sea level rise, etc.). In addition, because the pattern of many variables like surface temperature, ocean heat content or sea level, scale with global SAT, $CO_2$effCS is also a relevant index of future local changes (e.g., Santer et al. 1990; Mitchell 2003; Perrette et al. 2013; Bilbao et al. 2015; Grose et al. 2017).

For all of these reasons $CO_2$effCS has been the center of much attention in the climate community since the Charney report. However, despite important efforts deployed by the climate community, the estimates of $CO_2$effCS have had a persistently wide spread. Indeed, the 66% confidence level (CL) range for $CO_2$effCS has long been [1.5; 4.5] K (e.g., Knutti et al. 2017), which is very close to the range given at the time of the IPCC First Assessment Report (IPCC 1992).

Recently, significant progress has been achieved with large-eddy simulations and global climate model simulations on the understanding of physical processes that control the climate feedbacks and in particular the cloud radiative feedbacks (which remains the biggest driver of intermodel spread in effCS; Zelinka et al. 2020). An important effort on tropical marine low clouds over the past 10 years has produced new evidence that these clouds cause actually a positive feedback (see, e.g., Klein et al. 2017; Myers et al. 2021). The negative feedback due to transitions from ice to liquid in high-latitude clouds present in many climate models has also been investigated and is now considered as overestimated in absolute value (e.g., Frey and Kay 2018; Mülmenstädt et al. 2021). This progress has led to less negative estimates of the total climate feedback and suggests the lower bound of the 66% CL range in $CO_2$effCS should be increased to 2.3 K (Sherwood et al. 2020). This progress is encouraging to narrow the range in $CO_2$effCS. Combined with other lines of evidence from paleoclimatic data in particular, this recent progress has led to a revised range for CO2effCS in the 6th assessment report of the IPCC. The 66% confidence level range for the $CO_2$effCS from the IPCC is now [2.5; 4.0] K, which is slightly reduced relative to the IPCC AR5 estimate.

Our sense of progress, however, is not a substitute for verifying the response of the climate system to $CO_2$ concentrations change in the real world, in our time. Historical observations can provide crucial information on how the energy budget is actually changing in response to past and present greenhouse gases (GHG) concentrations. If historical observations are precise and long enough, they may lead in principle to accurate estimates of the historical effective climate sensitivity (histeffCS) and further provide an observational constraint on $CO_2$effCS. By comparing this observational constraint with $CO_2$effCS estimates derived from simulations, we should be able to evaluate the consistency of climate model simulations with the historical real climate response.

In the past, several studies analyzed the historical energy budget and derived a constraint on $CO_2$effCS from historical observations of the SAT and the Earth energy imbalance (EEI) at TOA (Lewis and Curry 2018; Skeie et al. 2018 are the most recent studies). They found ranges for $CO_2$effCS that are significantly smaller and tighter than the range derived from climate model simulations. In particular, they found an upper bound that does not exceed 3.1 K (95%CL), which disagrees with most phase 6 of the Climate Model Intercomparison Project (CMIP6; Eyring et al. 2016) simulations (Zelinka et al. 2020). Three major reasons have been put forward to explain this apparent disagreement between the observed historical energy budget and the climate sensitivity from climate model simulations. One reason is that past observational studies used an underestimated uncertainty in the aerosol effective radiative forcing (ERF; Bellouin et al. 2020) to derive the histeffCS (Sherwood et al. 2020). A second reason is that, in observational studies, the internal variability in SAT is not removed before deriving the histeffCS, leading to estimates that are biased high and artificially narrow (Gregory et al. 2020). A third reason is that past observational studies have not fully considered the dependence of the EEI to the geographic pattern in SAT (Sherwood et al. 2020; Gregory et al. 2020). This effect is called "the pattern effect." It arises from changes in the mix of radiative forcings, lag-dependent responses to forcings, or unforced variability and it leads to apparent time variations in the estimates of the histeffCS and to differences between histeffCS and $CO_2$effCS (see, e.g., Armour et al. 2013; Gregory and Andrews 2016; Andrews et al. 2018; Andrews and Webb 2018; Dong et al. 2019; Marvel et al. 2018; Paynter and Frölicher 2015; Winton et al. 2010; Zhou et al. 2017). Not accounting for these apparent time variations leads to biased estimates and artificially narrow ranges for the observational constraint on $CO_2$effCS.

Recently, important progress has been made on these issues. A new estimate of the aerosols ERF is available (Bellouin et al. 2020). In addition, the release of the data from the Radiative Forcing Model Intercomparison Project (RFMIP; Pincus et al. 2016) makes it now possible to evaluate the time-dependent radiative response of Earth in historical climate model simulations, and thus to evaluate the impact of the pattern effect on estimates of the histeffCS, so it is timely to revisit the historical energy budget from observations. We propose here to account for these different issues and see whether it changes the observational constraint on $CO_2$effCS.

In this paper we focus on the historical period during which climate has warmed significantly. This warming has been predominantly a response to anthropogenic GHG emissions (Pachauri et al. 2014). We use up-to-date observational estimates of the EEI changes, of the forcing changes and of the SAT changes to evaluate the changes experienced by the energy budget during this warming period and to derive an estimate of the histeffCS (section 3). We make a regression of the observed energy budget with a method similar to Gregory et al. (2020). We focus on the recent period 1971–2017, when the most reliable data are available. This time period is longer

than the time period in Gregory et al.'s (2020) study, who only considered data from ERBE and CERES over 1985–2011. In this approach, we evaluate the associated uncertainty accounting for all sources of errors in observations. We compare our estimates of the histeffCS with recent estimates from the literature and discuss the similarities and differences (section 3).

Because the current warming is predominantly a response to atmospheric $CO_2$ concentrations, the current changes of the energy budget are expected to reflect predominantly the response of the energy budget to $CO_2$ concentrations. As such the histeffCS should be close to the $CO_2$effCS. We use atmosphere–ocean general circulation models (AOGCM) simulations to evaluate the distance between the histeffCS and $CO_2$effCS accounting for all uncertainties (section 4). Then we use the observational estimate of the histeffCS to derive an observational constraint on $CO_2$effCS accounting for the estimated distance between both histeffCS and CO2effCS (section 5). We discuss our observational constraint on $CO_2$effCS in section 6 and conclude in section 7.

## 2. The global energy budget

The global conservation of energy in the climate system relates the ERF $F$ (Myhre et al. 2013) and the radiative response of the climate system $R$ with the changes $N$ in EEI at TOA:

$$N = F + R. \tag{1}$$

In this equation $F$, $R$, and $N$ are anomalies with respect to an unperturbed equilibrium in which $F = N = R = 0$; $F$ and $R$ are counted positive downward, and $N$ is counted positive when the incoming radiative flux at TOA is greater than the outgoing flux (i.e., $N$ is the heat flux into the climate system); $R$ is the sum of the radiative response of the climate system to the change in global SAT ($R_T$) and the radiative response that is unrelated to the change in global SAT ($R_T$). Following Budyko (1969) and the vast literature on the global energy budget that comes after [see, e.g., Knutti et al. (2017) and references therein] we assume that $R_T$ scales linearly with global changes in SAT and $R_T$ is a random variable with zero mean that is generated by the internal variability of the climate system. Now, if we call $T$ the global changes in SAT with respect to the unperturbed equilibrium and if we consider temporal means that are sufficiently long such that $R_T$ averages to zero, the global energy budget reads as follows:

$$N - F = R_T = \lambda T. \tag{2}$$

Note that here, $N$, $F$, $R_T$, and $T$ are now temporal averages, but we keep the same notation as in Eq. (1) for simplicity; $\lambda$ is the climate feedback parameter (W m$^{-2}$ K$^{-1}$). Because of the convention adopted in Eq. (2), $\lambda$ is a negative parameter.

In the perturbed equilibrium after a doubling of preindustrial atmospheric $CO_2$ concentrations, $N = 0$ (because equilibrium has been reached), $F = F_{2\times}$ (with $F_{2\times}$ being the forcing under 2-times-$CO_2$ concentrations in the atmosphere), and $T$ = ECS (by definition of ECS). Thus, the energy budget of the perturbed equilibrium described in Eq. (2) leads to the following formula for the ECS: ECS $= -F_{2\times}/\lambda$. Because ECS is inversely related to $\lambda$, the closer $\lambda$ gets to zero, the larger ECS becomes. Another consequence of this inverse relation is that an uncertainty with a normal distribution in $\lambda$ translates into an uncertainty in ECS that is skewed and shows a fat tail (see, e.g., Roe and Baker 2007). The tail becomes fatter as the distribution in $\lambda$ spreads and comes closer to zero.

In transient climate change, the climate system has not yet reached equilibrium and $N \neq 0$. In this case $\lambda$ verifies $\lambda = (N - F)/T$ and the quantity $-F_{2\times}/\lambda$ is the effCS. The formula for effCS is then effCS $= -F_{2\times}/\lambda = TF_{2\times}/(F - N)$. Following Andrews et al. (2012) and applying the method of Gregory et al. (2004), $CO_2$effCS corresponds to one-half of the effCS computed over the first 150 years of AOGCM simulations under an abrupt quadrupling of $CO_2$ concentration; histeffCS corresponds to the effCS in response to the historical forcing. It is estimated over historical periods (we mean by historical period here any period between 1850 and the present). If $\lambda$ is constant through the transient phase and the equilibrium phase of climate change then histeffCS = $CO_2$effCS = ECS. However, if $\lambda$ is not constant, the equality no longer holds and the inconstancy of $\lambda$ has to be taken into account to relate histeffCS with $CO_2$effCS and with ECS.

Because the historical record does not include any period with an unperturbed climate that is in equilibrium, the unperturbed equilibrium state is not known and the anomaly with respect to the equilibrium state cannot be estimated. To cope with this problem, Eq. (2) is generally applied to differences between two historical states: a base state and a present state [as in, e.g., Lewis and Curry (2018) and Sherwood et al. (2020)]. The difference enables to remove the reference to the unperturbed equilibrium state in Eq. (2). We call this method herein the "state-difference method." Other authors use an alternative method based on a differential form of Eq. (2) and estimate $\lambda$ by a regression of $N - F$ against $T$ (Gregory et al. 2020 is the most recent example). Hereinafter, we call this method the "regression method." Here we adapt the regression method for observations of the global energy budget since 1971. It allows for a full use of the available historical data and is a priori a better estimator of the slope than the state-difference method (Barnes and Barnes 2015). Note that both methods are affected by the internal variability and by the volcanic activity (e.g., Lewis and Curry 2018; Gregory and Forster 2008). Thus, these effects must be considered in the implementation of the regression method here (see section 3).

Recent studies based on AOGCM and atmosphere general circulation models (AGCMs) show that $\lambda$ is not constant for two reasons. First, $\lambda$ depends on the climate state, which means $\lambda$ depends on the magnitude of global mean $T$ or global mean $F$. This effect is small for small departures in $T$ or $F$, but it becomes significant under large forcing ($4 \times CO_2$ and larger) after 100 years, when $T$ reaches high values (Gregory et al. 2015; Bloch-Johnson et al. 2015, 2021; Rugenstein et al. 2019, 2020; Sherwood et al. 2020). Second, $R$ may vary because of changes in the pattern of sea surface temperature (SST) or, equivalently, $R$ may have a different value for the same change in global mean temperature (e.g., Armour et al. 2013; Andrews et al. 2015; Gregory et al. 2015; Knutti and

Rugenstein 2015). This effect is called the pattern effect. It arises either because of heterogeneous radiative forcing (e.g., Shindell 2014), lag-dependent responses to forcings (e.g., Held et al. 2010; Armour et al. 2013), or unforced variability (e.g., Proistosescu and Huybers 2017; Andrews and Webb 2018; Marvel et al. 2018). When $R$ is represented as $R = \lambda T$, the pattern effect means that $\lambda$ actually changes with the SST pattern (e.g., Stevens et al. 2016; Gregory and Andrews 2016).

As stated above a consequence of the inconstancy of $\lambda$ is that histeffCS $\neq$ $CO_2$effCS. Thus, estimates of histeffCS cannot be used as direct observational constraints on $CO_2$effCS. We use AOGCM simulations of the historical period and AOGCM simulations under an abrupt quadrupling of $CO_2$ concentration to evaluate the distance between histeffCS and $CO_2$effCS. Then we use this distance to derive an observational constraint on $CO_2$effCS from the estimate of histeffCS (see section 5).

## 3. Historical effective climate sensitivity

### a. Observations and associated uncertainties

In this section we estimate the histeffCS from the historical energy budget using observations of the historical anomalies in the ERF $F$, the TOA energy imbalance $N$, and the SAT $T$. Myhre et al. (2013) used radiative transfer models and estimates of the rapid atmospheric adjustments from climate models to estimate the total historical ERF over 1750–2011 from historical anthropogenic emissions and atmospheric composition. Dessler and Forster (2018) and then Sherwood et al. (2020) revised and extended this estimate to 2018. We use here the ERF updated version from Sherwood et al. (2020). In this version, as compared with Myhre et al. (2013), the forcing from ozone has been updated following Myhre et al. (2017) and the forcing from $CO_2$, $N_2O$, and $CH_4$ has been computed using concentrations and formulae to convert mixing ratios to forcing from Etminan et al. (2016). The aerosols ERFs have also been updated with the unconstrained estimate from Bellouin et al. (2020) (although Bellouin et al. 2020 considered their constrained estimate of aerosols ERFs more accurate, we cannot use it here because it is derived with an energy budget constraint and thus it would lead to circular reasoning).

For all ERFs except the aerosol ERF, the uncertainty in forcing is estimated using the radiative forcing uncertainties in 2011, from Myhre et al. (2013). In 2011, we assume the uncertainties are Gaussian distributed and we derive the Gaussian standard deviation from Myhre et al. (2013, their Table 8.6). For all other years, we also assume the uncertainty is Gaussian distributed with a standard deviation that is equal to the standard deviation of the 2011 distribution (see Fig. 1).

For the aerosol ERF, the uncertainty in forcing is estimated from Bellouin et al. (2020). We use a lognormal distribution fitted on Bellouin et al. (2020) to represent the distribution of the aerosol ERF in 2011. We normalize the lognormal distribution by its median and use 500 000 Monte Carlo draws of this normalized distribution to scale the time series of aerosol forcing. We get an ensemble of 500 000 time series, which represents the aerosol forcing and its uncertainty (see Fig. 1). Note that this approach means we have considered a time correlation of 1 in the aerosol ERF uncertainty. Assuming here a high level of time correlation is relevant because most of the uncertainty in aerosol ERF is coming from the poorly known interaction with clouds (which is rather unconstrained by process knowledge; Bellouin et al. 2020), and thus it is very likely a systematic uncertainty. However, assuming full time correlation is probably conservative and leads to a slightly overestimated uncertainty range.

For observed historical SAT changes, we use the updated global annual mean temperature anomaly dataset derived from blended SST and SAT from Cowtan and Way (2014), version 2. This dataset corrects for missing data in the observational network (in particular in the polar regions), which makes it less sensitive to coverage biases in particular in the beginning of the historical record (Fig. 1). Global mean SAT estimates from blended SST and SAT datasets are lower than estimates derived from SAT only (Richardson et al. 2016; Cowtan et al. 2015). We correct for this bias following Richardson et al. (2016) who found that climate models closely agreed on a scaling of ~1.09 from blended historical SST and SAT anomalies to SAT anomalies. The uncertainty in SAT changes is taken from Cowtan and Way (2014). It includes the temporal correlation in errors (Morice et al. 2012). Cowtan and Way's (2014) uncertainty in SAT changes represents only the instrumental uncertainty. It is small relative to the uncertainty in SAT due to internal variability so we neglect it in the rest of the study (we did run the regression method with the GISS SAT time series and could not find significant differences with the run with the Cowtan and Way time series). The uncertainty in $T$ generated by the internal variability is considered at a later stage in the energy budget (see section 3b).

The planetary heat uptake essentially occurs in the ocean (~93%; Church et al. 2011; Levitus et al. 2012; Meyssignac et al. 2019; von Schuckmann et al. 2020). Thus, the ocean heat uptake (OHU) places a strong constraint on the planetary heat uptake. Here, following Melet and Meyssignac (2015), we estimate $N$ from observations of OHU divided by the fraction of energy entering the ocean (0.93). The observations of OHU are derived from the in situ record of the ocean subsurface temperatures using the International Thermodynamic Equation Of Seawater—2010 (TEOS-10; see Melet and Meyssignac 2015). For the regression method, we need time series of $N$ as long as possible to make full use of available information. The longest global and continuous ocean temperature datasets are provided by four research groups. They are gridded estimates of the ocean temperature over 1955–2017 either with a statistical interpolation of in situ data (Good et al. 2013; Ishii et al. 2017; Levitus et al. 2012) or with a combination of statistical interpolation of in situ data and climate model information (Cheng et al. 2017). Although the four products are based on the same database of in situ data, they show differences because they use different in-filling strategy for data gaps, different corrections for instrumental biases, and different climatologies. To account for these differences, we use an ensemble of five datasets. This ensemble comprises the NOAA dataset (Levitus et al. 2012) and the Meteorological Research Institute (MRI)-JMA dataset (Ishii et al. 2017), plus two versions of the

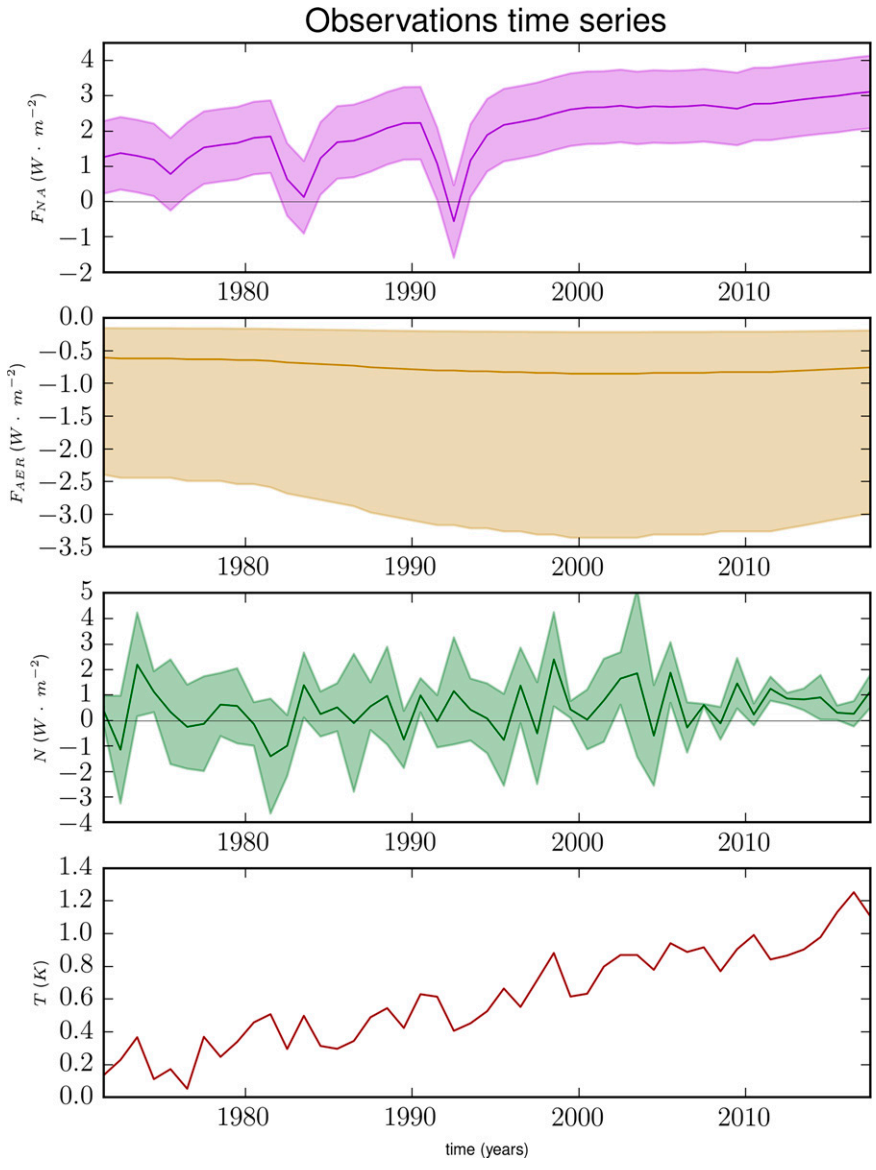


FIG. 1. Time series of annual-mean nonaerosols radiative forcing $F_{NA}$, aerosols radiative forcing $F_{AER}$, TOA radiative imbalance $N$, and global mean surface temperature $T$ (with respect to their respective time average over 1869–82 when available) that are used for the regression method. The shaded envelopes indicate the 5%–95% confidence intervals.

Met Office enhanced ocean data assimilation and climate prediction (ENACT)/ENSEMBLES, version 4 (EN4), dataset (Good et al. 2013) and the Institute of Atmospheric Physics (IAP) dataset [each version using a different correction of the instrumental biases based on Gouretski and Koltermann (2007) and Levitus et al. (2009)]. We take as best estimate of OHU the average of the ensemble because averaging allows to reduce the effect of particular errors of individual datasets and emphasizes the common variability among datasets (Fig. 1). The uncertainty in OHU is characterized with the spread of the ensemble. Measurements of ocean subsurface temperature are particularly sparse before 1971. We use time series of OHU beginning in 1971 when quasi-global coverage of ocean subsurface temperature becomes available. It starts to be quasi-global only in 1971 and it provides rare measurements below 2000-m depth (Abraham et al. 2013). To cope with these issues, we use the OHU time series only from 1971 to 2017 and we add an extra-deep OHU below 2000-m depth of 0.0 ± 0.04 W $m^2$ between 1971 and 1990 and then 0.07 ± 0.04 W $m^2$ from 1991 to 2017 [following von Schuckmann et al. (2020)].

### b. *Estimate of the histeffCS with the regression method*

In the regression method, we estimate $\lambda$ as the slope of the linear regression of $N - F$ over $T$ ($N$, $F$, and $T$ are from the observational datasets described in section 3a). The regression method is applied over the period 1971–2017 because it is the

longest period covered by the three datasets of $F$, $N$, and $T$ ($N$ being the limiting dataset). This 47-yr-long period is marked by two major volcanic eruptions (El Chichón in 1982 and Pinatubo in 1991), however. To mitigate the effect of volcanic eruptions on the estimate of the $CO_2$effCS, we remove from all time series the 5 years following each volcanic eruption, following Church et al. (2005). Then, we build 500 000 time series of $N$ with 500 000 Monte Carlo draws of the distribution of $N$ at each epoch (this approach means that we assume no correlation in time for the uncertainty in $N$). The same is done for the nonaerosol ERF. We also use the 500 000 aerosols radiative forcing time series described in section 3. Then we use 500 000 times an ordinary least squares scheme to regress $N - F_{NA} - F_{AER}$ over $T$. The output is an ensemble of 500 000 $\lambda$ that represents the distribution in the climate feedback parameter (Fig. S1 in the online supplemental material shows the distribution of $N - F_{NA} - F_{AER}$ against $T$, and supplemental Fig. S2 shows an example of regression for one draw of $N - F_{NA} - F_{AER}$ and $T$). Here, we do not account for the internal variability in $N$ and $T$ that can blur the estimate of the histeffCS. Both these sources of uncertainty are dealt with in the next paragraph.

Gregory et al. (2020) showed that estimates of the historical $\lambda$ made by least squares regressions (either ordinary or generalized least squares regressions) are negatively (low) biased and lead to artificially narrow ranges in historical $\lambda$ because they do not account for noise in $T$ and $N$ (here the noise in $N$ is due to instrumental noise and internal variability). Given our definition of the energy budget in Eq. (1), the noise in $T$ here corresponds to any signal $T'$ in $T$ that does not produce proportionate variability $\lambda T'$ in $R$ (as a reminder, $\lambda T'$ here is the forced radiative response of climate to the historical forcing). This noise $T'$ can be due to instrumental noise in $T$ or it can be due to internal variability in $T$ that produces a response in $R$ that either is not proportional with $T$ or is proportional with $T$ with a different constant than $\lambda$ (like in the case of the pattern effect). Another possible source for $T'$ is some persistence in internal variability of $R$ that generates a response in $T$ with thermal inertia [see appendix C and D in Gregory et al. (2020) and also Proistosescu et al. (2018) (their second case)]. The instrumental noise in $T$ identified in Morice et al. (2012) is significantly smaller than the noise $T'$ generated by the internal variability (Dessler et al. 2018) so we neglect it (as indicated above). The appendix goes into the computation of the bias and uncertainty in $\lambda$ from these effects in more detail.

Following Gregory et al. (2020), we evaluate the bias and the range in $\lambda$ that is generated by the internal variability in $T$ and $N$ with AOGCMs historical simulations. $N$ is estimated from the TOA energy budget of the historical simulations and $T$ from the SAT of the historical simulations. The effective radiative forcing $F$ is estimated with the piClim-histall simulations from the RFMIP project (Pincus et al. 2016). The piClim-histall simulations are run with historical forcing agents, keeping the surface temperature at its preindustrial climatology such that the radiative response of the Earth $R$ is zero and the TOA energy imbalance $N$ in these simulations equals the historical forcing $F$. Thus, we use the TOA energy imbalance $N$ of the piClim-histall simulations to estimate the forcing $F$ of historical simulations. At the time we worked on this study, RFMIP piClim-histall full simulations were available for only five models, namely, IPSL-CM6A-LR (Boucher et al. 2018), CanESM5 (Swart et al. 2019), MIROC6 (Tatebe et al. 2019), GFDL CM4 (Winton et al. 2020), and GISS-E2-1-G (Kelley et al. 2020). These five models provide in total 103 realization of the historical simulations. For each realization of the historical simulation, we regress the differential form of the energy budget on sliding windows whose length is equal to the duration of the observation dataset (i.e., 47 years). Since historical simulations start in 1850 and end in 2014, they include 117 different overlapping 47-yr periods). To sample as much as possible the effect of internal variability we run the regression method for each of these periods (see Fig. S3 in the online supplemental material). We do the same for the ensemble mean of the historical simulations of each model. We evaluate the bias and the range in $\lambda$ induced by the internal variability by calculating for each model the difference between the time averaged $\lambda$ of each historical realization and the time averaged $\lambda$ of the ensemble mean. We find that the internal variability generates a bias in $\lambda$ of $-0.005$ W m$^{-2}$ K$^{-1}$ and a range of $\pm 0.136$ W m$^{-2}$ K$^{-1}$ ($1\sigma$). It corresponds to a bias of $-0.5\%$ and a range of $\pm 14\%$ around the best estimate of the historical $\lambda$. The range is consistent with the finding from Gregory et al. (2020) (for a regression over a period of 30 years) but the bias is smaller. This is likely because there are few realizations per model, making our estimate of the bias less accurate than in Gregory et al. (2020). Note that the range is also close to the estimate from Dessler et al. (2018) who find that internal climate variability alone results in an uncertainty in $\lambda$ of $\pm 0.14$ W m$^{-2}$ K$^{-1}$ in the historical simulation ensemble of the MPI-ESM model.

Table 1 shows our estimate of historical effective $\lambda$ from the regression method with associated uncertainty based on observations of the energy budget over 1971–2017. The uncertainty comprises all sources of uncertainty described above, which include the measurement uncertainty in $N$ and in $F$ and the bias and the uncertainty that are both due to the internal variability (it includes the effect of internal variability on $N$ and $T$). The internal variability bias and the associated uncertainty in historical $\lambda$ is added to the measurement uncertainty by applying a Monte Carlo simulation. On the overall, the regression method indicates a historical $\lambda$ of $-0.92$ [$-1.82$; $-0.18$] W m$^{-2}$ K$^{-1}$ for the period 1971–2017 (the best estimate indicates the median and the range is the 5%–95% confidence level range). This corresponds to an histeffCS of 4.34 [2.17; 22.69] K for the period 1971–2017 (in the translation from historical $\lambda$ to histeffCS we use the mean estimate of $F_{2\times} = 3.99 \pm 0.27$ W m$^{-2}$ K$^{-1}$ from Smith et al. (2020). Note that in the translation, $\lambda$ values that are above zero are truncated because they are physically inconsistent (due to the substantial noise in each fit, some fraction of the $\lambda$ estimates are greater than zero, here 9%). A consequence of the truncation of these positive $\lambda$ is that the distribution in $\lambda$ is slightly changed leading to small changes ($<0.18$ W m$^{-2}$ K$^{-1}$) in the median and the limits of the 5%–95% range. This means that our method is intrinsically uncertain at the level of $\pm 0.18$ W m$^{-2}$ K$^{-1}$.

In the histeffCS uncertainty, the observational uncertainty in $N$ and the uncertainty in $F$ dominates over the uncertainty

TABLE 1. Estimate of the historical effective $\lambda$ (first row) and the histeffCS (second row) from the regression method with the classical energy budget. The historical effective $\lambda$ and the histeffCS characterize the climate response to the historical forcing (see the text). The numbers indicate the median, and the ranges indicate the 5%–95% CL. In the second column, all sources of uncertainty are considered: the instrumental uncertainty in $N$ and in $F$, the bias due to internal variability, the uncertainty due to internal variability, and the uncertainty in $F_{2\times}$. In the column named "no uncertainty in $X$" the range and the best estimate of historical effective $\lambda$ and histeffCS have been computed with a value for $X$ set at the median of $X$ and with a range in $X$ set at 0. Numbers have been computed with the full 64 bits of precision and then rounded at the second digit. For this reason, the same rounded value in $\lambda$ can correspond to a slightly different rounded value in histeffCS.

| Variable | All uncertainties included | No uncertainty in $F_{2\times}$ | No uncertainty in $N$ | No uncertainty in nonaerosols $F$ | No uncertainty in aerosols $F$ | No uncertainty in $F$ and $N$ | No realization bias | No uncertainty from internal variability |
|---|---|---|---|---|---|---|---|---|
| Historical effective $\lambda$ | −0.92 [−1.82; −0.18] | — | −0.89 [−1.52; −0.24] | −0.90 [−1.65; −0.21] | −1.16 [−2.05; −0.33] | −1.15 [−1.37; −0.93] | −0.92 [−1.81; −0.17] | −0.92 [−1.79; −0.18] |
| histeffCS | 4.34 [2.17; 22.83] | 4.34 [2.19; 22.83] | 4.50 [2.59; 16.34] | 4.45 [2.39; 19.50] | 3.44 [1.92; 12.16] | 3.47 [2.81; 4.41] | 4.36 [2.17; 22.95] | 4.36 [2.20; 22.37] |

due to internal variability. If there was no observational uncertainty (i.e., no uncertainty in estimates of $N$ and $F$) the 5%–95% range of the histeffCS would be reduced by 86% and most of this reduction (79%) would occur in the upper limit of the range (see Table 1). In contrast if there was no internal variability the range would be reduced by less than 3%. It means that the uncertainty due to internal variability is small and plays only a marginal role in the range of the histeffCS estimate. As for the bias that is generated by the internal variability in histeffCS, it also plays a minor role. It slightly shifts the range of historical $\lambda$ by a few percent toward less negative values, which results in a very small upward shift of the histeffCS upper range.

Among observational uncertainties, the uncertainty from the aerosol forcing dominates (see Fig. 2). If there was no uncertainty in the aerosol forcing, then the range of histeffCS would be reduced by 50% and this reduction occurs almost entirely on the upper limit of the range (see Table 1). This result confirms earlier studies that show that aerosols ERFs are the main source of uncertainty in the estimate of the historical $\lambda$ and the histeffCS. It also shows that the uncertainty in aerosols is the main cause for the fat tail in high values of the histeffCS (note that removing the uncertainty in the aerosol ERF also shifts the range in historical $\lambda$ toward more negative values by a few percent leading to smaller histeffCS estimates. This is because of the skewness of the distribution in $F_{\mathrm{AER}}$).

In comparison, the role of the uncertainty in $N$ or in nonaerosol forcing on the uncertainty in histeffCS is smaller; but it is sizable. If there was no uncertainty in $N$ then the range of histeffCS would be reduced by 33% and this reduction would occur essentially on the upper limit of the range (31% decrease of the upper limit and 2% increase of the lower limit, see Table 1). The same is true for the nonaerosol $F$. If there was no uncertainty in nonaerosol $F$ then the range of histeffCS would be reduced by 17% and this reduction would occur essentially on the upper limit of the range (16% decrease of the upper limit and 1% increase of the lower limit, see Table 1). These results show that the uncertainty in $N$ and nonaerosol $F$ are far from negligible in the total uncertainty in histeffCS and that they play a sizable role in the fat tail in high values of the histeffCS.

Lewis and Curry (2018) estimated the histeffCS with a state difference method between the periods of 1869–82 and 2007–16. They found an histeffCS of 1.5 [1.05; 2.45] K (median and 5%–95% range). This is significantly smaller than our estimate. And their fat tail in high histeffCS values is much thinner than the tail here. This is essentially because they used an estimate of the aerosol forcing based on the AR5 report (Boucher et al. 2013) that is significantly more certain and not skewed relative to the aerosol forcing from Bellouin et al. (2020); see also Sherwood et al. (2020).

Sherwood et al. (2020) estimated the histeffCS with a state difference method over similar base period and final period as Lewis and Curry (2018). They used the aerosol forcing from Bellouin et al. (2020) and found an histeffCS of 3.11 [1.86; 14.41] K [see Sherwood et al. 2020, their green curve on Fig. 11b and their Table 5 Eq. (19)]. This is close and consistent with our estimate but slightly smaller on the lower end of the 5%–95% range. Sherwood et al. (2020) estimate is also about 8 K lower than our estimate on the high end of the 5%–95%

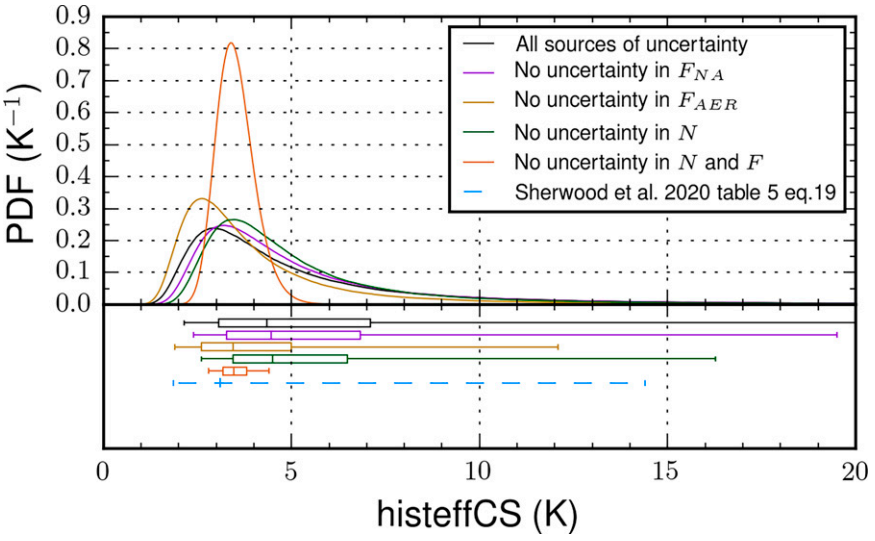


FIG. 2. Probability density functions (above the zero line) and the associated whisker plots (below the zero line) of the estimate of the histeffCS derived from the regression method with all sources of uncertainty (black), without the uncertainty in $F_{NA}$ (violet), without the uncertainty in $F_{AER}$ (brown), without the uncertainty in $N$ (green), and without the uncertainty in both $N$ and $F$. In the whisker plots, the boxes indicate the 25th–75th percentiles, the whiskers indicate the 5th and the 95th percentiles, and the vertical lines indicate the medians. The blue dashed line and vertical line respectively indicate the 5%–95% range and the median of the Sherwood et al. (2020) estimate [see their Table 5 Eq. (19)].

range. The difference on the high end is not significant because it corresponds to a difference in $\lambda$ of + 0.04 W m$^{-2}$ K$^{-1}$ (this is below the intrinsic uncertainty of ±0.18 W m$^{-2}$ K$^{-1}$ of our method; see section 3). This is different for the lower end of the 5%–95% range. The small difference of 0.31 K is significant because it corresponds to a difference in $\lambda$ of + 0.31 W m$^{-2}$ K$^{-1}$ (this is above the intrinsic uncertainty of ±0.18 W m$^{-2}$ K$^{-1}$ of our method; see section 3). Most of the difference in low end $\lambda$ can be actually explained by a potential bias in the base-state EEI estimate in Sherwood et al. (2020). See the discussion section.

## 4. Distance between the histeffCS and $CO_2$effCS

Recent model studies showed that the climate parameter $\lambda$ is not constant (see section 2). A consequence of the inconstancy of $\lambda$ is that histeffCS is different from $CO_2$effCS. Thus, the observational constraint derived on histeffCS from the historical climate record cannot be used as direct observational constraint on $CO_2$effCS. To derive an observational constraint on $CO_2$effCS, we need first to evaluate the distance between histeffCS and $CO_2$effCS (and the associated uncertainty).

In AOGCMs simulations of the climate response to abrupt $CO_2$ forcing, $\lambda$ is not constant and tend to become less negative with time, as climate approaches the equilibrium (see, e.g., Armour et al. 2013; Andrews et al. 2015; Gregory et al. 2015; Rugenstein et al. 2016; Bloch-Johnson et al. 2015; Armour 2017; and many others). As we have seen in section 2, this temporal variation of $\lambda$ occurs because $\lambda$ depends on the amplitude of global warming (e.g., Bloch-Johnson et al. 2015) and because it depends on the SST pattern (the so-called pattern effect; e.g., Armour et al. 2013). A consequence of this behavior is that $\lambda$ in abrupt-4×CO2 simulations will, in general, be less negative than the apparent $\lambda$ in historical simulations. To evaluate this difference between both $\lambda$s, we use CMIP6 historical simulations and abrupt-4×CO2 simulations to compute the difference between the simulated abrupt-4×CO2 $\lambda$ and the historical $\lambda$. We use the estimates of $\lambda$ over the first 150 years of abrupt-4×CO2 simulations derived by Zelinka et al. (2020). The historical $\lambda$ is estimated for each model as well, by regressing $N$–$F$ over $T$ on sliding windows on the ensemble mean of the historical realizations (the sliding windows being of the same duration as the observed dataset). Then, this $\lambda(t)$ time series is averaged across time. We find a bias in $\lambda$ between the abrupt-4×CO2 $\lambda$ and the historical $\lambda$ of +0.30 ± 0.30 W m$^{-2}$ K$^{-1}$ ($1\sigma$) for the period covered by the regression method. The range associated to this estimate is a range across models (see Fig. S4 in the online supplemental material).

Different models lead to different biases because they show different patterns in SST and thus different distance in $\lambda$ between historical simulations and the abrupt-4×CO2 simulation. So, the range can be seen as a simple estimate of the uncertainty in the bias estimate that is due to the representation of the pattern effect in AOGCM simulations. Our estimate of the bias and the uncertainty associated to the representation of the pattern effect is consistent with a previous study from Andrews et al. (2019), where the authors used the historical simulations of the HadGEM3 model and found a bias of +0.2 ± 0.4 W m$^{-2}$ K$^{-1}$ ($1\sigma$). The biases and uncertainties in $\lambda$ induced by the representation of the pattern

TABLE 2. Estimate of the $CO_2$eff$\lambda$ and the $CO_2$effCS from the regression method with the classical energy budget. The numbers indicate the median, and the ranges indicate the 5%–95% CL. In the second column, all sources of uncertainty are considered: the uncertainty in histeffCS (see Table 1 for more details on the estimate of histeffCS), the bias between histeffCS and $CO_2$effCS, the uncertainty in the bias due to divergence in the representation of the pattern effect among AOGCMs, and the uncertainty in the bias due to the internal variability. In the column named "no uncertainty in $X$" the range and the best estimate of $CO_2$eff$\lambda$ and $CO_2$effCS have been computed with a value for $X$ set at the median of $X$ and with a range in $X$ set at 0. Numbers have been computed with the full 64 bits of precision and then rounded at the second digit. For this reason, the same rounded value in $\lambda$ can correspond to a slightly different rounded value in effCS.

| Variable | All uncertainties included | No uncertainty in historical $\lambda$ | No bias between $CO_2$effCS and histeffCS | No uncertainty in the bias due to the misrepresentation of the pattern effect in AOGCMs |
|---|---|---|---|---|
| $CO_2$eff$\lambda$ | −0.73 [−1.63; −0.11] | −0.58 [−1.10; −0.13] | −0.94 [−1.93; −0.17] | −0.69 [−1.52; −0.11] |
| $CO_2$effCS | 5.46 [2.40; 35.61] | 6.92 [3.58; 31.53] | 4.23 [2.03; 23.79] | 5.75 [2.59; 36.01] |

effect is added to the measurement uncertainty with a Monte Carlo simulation.

## 5. Observational constraint on $CO_2$effCS

Table 2 shows the resulting estimate of the $CO_2$effCS (see Fig. 3). Table 2 also shows the effect of the bias between the historical and the abrupt-4×CO2 $\lambda$ and the effect of its uncertainty on the estimate of the $CO_2$effCS. The $CO_2$effCS estimate is greater and more uncertain than the histeffCS estimate. This is because of the systematic positive bias in $\lambda$ between abrupt-4×CO2 simulations and historical simulations. This positive bias also induces an increased uncertainty and a fatter tail in $CO_2$effCS relative to histeffCS (see Table 2) because of the inverse relation between $\lambda$ and effCS. Indeed as $\lambda$ is inversely related to effCS, the same range in $\lambda$ systematically leads to a larger range (and fatter tail) in effCS when it is less negative and shifted closer to 0.

We find that the dominant source of uncertainty in $CO_2$effCS comes from the uncertainty in the estimate of histeffCS (see Table 2). The second largest source of uncertainty comes from the bias in $\lambda$ that is generated by the pattern effect. As explained above this bias inflates systematically the uncertainty in $CO_2$effCS because it shifts $\lambda$ closer to 0 (see Table 2). The third largest source of uncertainty is the uncertainty in the bias associated to the representation of the pattern effect. Its effect is small

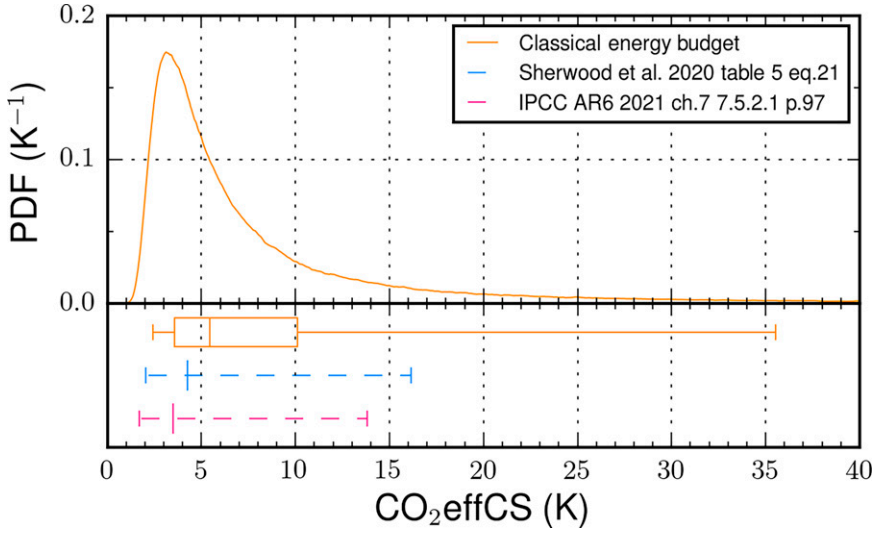


FIG. 3. Probability density function (above the zero line) and associated whisker plot (below the zero line) of the estimate of the $CO_2$effCS derived from the regression method with a classical energy budget and AOGCM simulations to estimate the distance between histeffCS and $CO_2$effCS (yellow). In the whisker plot, the boxes indicate the 25th–75th percentiles, the whiskers indicate the 5th and the 95th percentiles, and the vertical lines in the box indicate the medians. The blue dashed line and vertical line respectively indicate the 5%–95% range and the median of the Sherwood et al. (2020) estimate [see their Table 5 Eq. (21)], and the pink dashed line and vertical line respectively indicate the 5%–95% range and the median of IPPC AR6 (Forster et al. 2021) (see their chapter 7, section 7.5.2.1).

relative to the effect of the uncertainty in histeffCS and the effect of the bias induced by the pattern effect (see Table 2).

With the regression method applied on the global energy budget since 1971, we find an estimate of $CO_2$effCS of 5.46 [2.40; 35.61] K (5%–95% CL). This observational constraint means that the historical energy budget imposes that the $CO_2$effCS is above 2.4 K and below 35.61 K (5%–95% CL). This is similar to the result of Sherwood et al. (2020) who find a range for $CO_2$effCS of [2.04; 16.13] K [see Sherwood et al. 2020, their Table 5 and Eq. (21)]. This similarity of the results despite the different approaches gives confidence in both Sherwood et al.'s (2020) results and our results [Sherwood et al. (2020) uses a state different method and a Bayesian approach, while we use a regression method and a frequentist approach].

Beside the general similarity, there are some discrepancies between Sherwood et al.'s (2020) estimate of the $CO_2$effCS and our estimate. Our lower end of the 5%–95% range is higher by 0.36 K while our upper end is higher by 19.5 K. The 19.5 K difference in upper ends is not significant because it corresponds to a difference in $CO_2$eff$\lambda$ of only + 0.13 W m$^{-2}$ K$^{-1}$ (here such a small difference in $CO_2$eff$\lambda$ of + 0.13 W m$^{-2}$ K$^{-1}$ turns into a large difference in $CO_2$effCS of 19.5 K just because of the inverse relation between $\lambda$ and effCS and because the upper end $CO_2$eff$\lambda$ is very close to 0). In contrast, the 0.36 K difference in the lower end is significant because it corresponds to a difference in $CO_2$eff$\lambda$ of + 0.30 W m$^{-2}$ K$^{-1}$. That is unexpected because both methods (the regression method and the state difference method) should lead to consistent estimates of $CO_2$eff$\lambda$. This difference in $CO_2$effCS lower end estimates can be actually attributed to the difference in histeffCS lower end estimates identified earlier in section 3. We suspect that the original cause for this difference is mostly a potential bias in the base-state EEI estimate in Sherwood et al. (2020). See the discussion section below.

## 6. Discussion

In this study we have estimated the historical effective sensitivity from the latest historical observations and derived an observational constraint on the effective sensitivity of climate to $CO_2$ concentrations. We used AOGCMs simulations to evaluate the distance between the histeffCS and the $CO_2$effCS. Our approach accounts for all sources of uncertainty including the uncertainty in aerosols forcing (from Bellouin et al. 2020) and it quantifies the role of each source of uncertainty in the total uncertainty of the estimated $CO_2$effCS.

We used a regression method to estimate the $CO_2$effCS in order to make full use of the available data since 1971. The regression method yields a wide 5%–95% range of $CO_2$effCS of [2.40; 35.61] K. This is broadly consistent with Sherwood et al.'s (2020) (and Forster et al. 2022) estimate. However, we find a significant difference of +0.36 K in the lower end of the $CO_2$effCS 5%–95% range. Indeed, the regression method shows that the historical energy budget imposes that the $CO_2$effCS is actually above 2.4 K at the 95%CL, which is 0.36 K above the low end estimate from Sherwood et al. (2020).

We find that this difference in low end estimates of $CO_2$effCS is actually due to the difference in low end estimates of histeffCS (cf. section 3). This difference can be due to differences in the $N$ and $T$ datasets used. We tested the state difference method, as in Sherwood et al. (2020), but with present day estimates of $T$ and $N$ derived from our own datasets in $N$ and $T$. We found an histeffCS of 2.73 [1.66; 11.52] K (see Table S1 in the online supplemental material), which is consistent with Sherwood et al. (2020) estimate [when they use the Cowtan and Way (2014) blended dataset; see their Table 5, column 8, line 2]. Thus, it cannot be the different datasets in $T$ and $N$ that explain the difference between the regression method and the state different method from Sherwood et al. (2020).

The difference in low end could be explained by a bias in $N$ during the base period chosen for the state difference method. Indeed, while there is confidence in the global SAT record at the end of the nineteenth century, there are reasons to doubt Sherwood et al.'s (2020) estimate of EEI over the base period 1861–80. Sherwood et al. (2020) chose an EEI of +0.2 W m$^{-2}$ over the base period 1861–80 from AOGCM simulations. This corresponds to an OHU positive and thus it corresponds to an ocean that is already warming in response to GHG emissions. However, there is evidence that, in 1861–80, the OHU was probably negative because the climate was still responding to the Little Ice Age (Paasche and Bakke 2010). The data from the 1872–76 mission of Her Majesty's Ship (HMS) *Challenger* indicates a basinwide cooling of ~0.1 K between the 1870s and the 2000s in middepth in the Pacific Ocean. This spatial pattern of temperature change is consistent with a long-term response to the little ice age (Gebbie and Huybers 2011, 2012). In Gebbie and Huybers (2019, their Fig. 1) the global mixed-layer temperature anomaly in 1870 is around −0.15 K reflecting the response to the Little Ice Age while the subsurface anomaly is warmer and close to zero, reflecting the residual warmth of the Medieval Warm Period. This difference of temperature suggests an import of heat in the mixed layer in the 1870s and thus an OHU that is negative (and close to 0) rather than positive. With an ocean heat uptake efficiency of about 1.2 W m$^{-2}$ K$^{-1}$ [mean of Gregory (2000) estimate and Dufresne and Bony (2008)], it would correspond to an OHU of about −0.2 W m$^{-2}$, which is consistent with the finding of Bagnell and DeVries (2021). We tested this hypothesis in a sensitivity study of the state difference method.

We changed $N$ in the base period to $-0.2 \pm 0.2$ W m$^{-2}$ and run the state difference method with all other parameters as in Sherwood et al. (2020). We found that changing $N$ to −0.2 W m$^{-2}$ in the base period leads to an histeffCS of 3.60 [1.98; 17.76] K (see Table S1 in the supplemental material), which explains most of the difference in the low end histeffCS between the regression method and Sherwood et al.'s (2020) state difference method. A small difference of 0.18 K remains that drives a difference in $CO_2$effCS of about 0.15 K. This small residual could be due to a bias in the $F$ record between the beginning of the record (before 1900) and the end of the record (after 1971) or due to the changing pattern effect that may have led to an historical $\lambda$ that was slightly less negative

(closer to 0) over the second half of the twentieth century than over the entire twentieth century.

Another possible reason is the errors in the estimate of the bias induced by the pattern effect. Indeed, our estimate of the historical pattern effect, as well as Sherwood et al.'s (2020) estimate, both rely on AOGCM simulations of the historical period. There are reasons to doubt the capacity of AOGCM simulations to accurately capture the relative patterns of historical temperature change. Recent studies (e.g., Andrews et al. 2018; Marvel et al. 2018) show that historical simulations of AOGCMs generally produce patterns of warming that resemble that of their abrupt-4×CO2 simulations rather than that of observed warming (Seager et al. 2019). Here, potentially, AOGCMs do not perfectly capture the different pattern effect over the second half of the twentieth century in comparison with the pattern effect over the base period and the recent period of the state difference method. This would lead to an erroneous estimate of the distance between histeffCS and CO2effCS and could explain part of the discrepancy in the lower range of the estimates of CO2effCS between the regression method and the state difference method. In particular, Andrews et al. (2018) argues that comparing abrupt-4×CO2 simulations with AMIP-piForcing simulations yields more accurate estimates of the pattern effect. It would be interesting to test this option and see if it explains the residual discrepancy in lower end estimates of the $CO_2$effCS.

When we compare our estimate of the $CO_2$effCS with Lewis and Curry's (2018) estimate, we find a significantly large difference. Sherwood et al. (2020) showed that an important cause for the difference between Lewis and Curry (2018) estimate and their own estimate is the use of a different aerosol forcing. Since we use here the recent estimate of the aerosol forcing from Bellouin et al. (2020) as in Sherwood et al. (2020) we suspect that the aerosol forcing could explain a significant part of the difference between the regression estimate of the $CO_2$effCS and Lewis and Curry (2018) estimate. And as Lewis and Curry (2018) rely on the same assumption as Sherwood et al. (2020) concerning the base-state EEI, we suspect that the bias in the base-state EEI could explain the potential residual after correcting the aerosol forcing. To test this hypothesis, we run a state difference method with a present day estimate of $T$ and $N$ derived from our datasets, with the AR5 aerosol forcing and a base-state EEI set to $+0.2$ W m$^{-2}$ [close to Lewis and Curry's (2018) assumptions]. With this approach we find an histeffCS of 1.75 [1.06; 3.01] K that is very close to Lewis and Curry's (2018) estimate. It confirms that the difference in aerosol forcing and a potential bias in the 1861–80 EEI estimate explain the difference in $CO_2$effCS between Lewis and Curry (2018) and the regression method.

Our analysis of the uncertainty budget of the regression method shows that, in estimates of the $CO_2$effCS, the major source of uncertainty (86% of the total uncertainty) arises from the uncertainty in the observations (primarily the uncertainty in the aerosol forcing but also the uncertainty in $N$ and in the non-aerosol forcing). The uncertainty in observations is also the major responsible for the fat tail in high values of $CO_2$effCS, confirming earlier studies (it should also be noted that removing dispersion around aerosols radiative forcing median gives a lower estimate of histeffCS [3.44 [1.92; 12.16] K)]. The second source of uncertainty arises from the bias between the $CO_2$effCS and the histeffCS that is induced by the pattern effect. This bias adds on the historical $\lambda$ and shifts it closer to 0. This shift inflates the uncertainty when $\lambda$ is further translated into $CO_2$effCS because of the inverse relationship between $\lambda$ and effCS.

On the overall we find that the upper limit of the 5%–95% range derived from the regression method (and from the state difference method) is too high to provide any valuable observational constraint on the real-world $CO_2$effCS. However, the lower limit provides interesting observational information. But it differs substantially when it is derived from the regression method or from the state difference method [as in Sherwood et al. (2020) and Lewis and Curry (2018)]. We have seen that most of these differences can be attributed to differences in the aerosol forcing that is used and to a possible bias in the estimate of the EEI over 1861–80 in the state difference method. There is some evidence of such a bias, but they are based on very few measurements from the HMS *Challenger*. So, which solution should be trusted—the regression estimate or the state difference estimate? It is difficult to tell and probably impossible to determine for now. More research is needed on this aspect (in particular on the late nineteenth-century EEI estimate).

In any case, we argue there is confidence in the results from the regression method for two reasons. The first reason is that observations are the main source of uncertainty in the estimate of histeffCS and we have more confidence on the recent observations (as the one used in the regression method) than on old observations (especially on observations of a reference period as old as 1861–80). This is particularly true for observations of $N$ and of $F_{\mathrm{AER}}$. The second reason is that the regression method makes full use of the observation data over a period as long as 47 years. Such long periods are expected to be less randomly affected by the pattern effect than short periods of 15 years and thus better capture the long term change in the pattern effect.

## 7. Conclusions

In this work, we propose a robust approach to estimate an observational constraint on the real-world $CO_2$effCS, based on regression method that uses the pattern effect deduced from AOGCMs simulations. This approach is different from other recent attempts that are based on the state difference method between a recent period and a base period at the end of the nineteenth century (Sherwood et al. 2020). The regression approach is based on the most recent and robust observations, and it uses the best representation of the pattern effect that is currently available. It shows a best estimate for the climate sensitivity to $CO_2$ concentrations of 5.46 K and a 5%–95% range of [2.40; 35.61] K. This is consistent with the recent estimates from Sherwood et al. (2020) although the constraint is tighter on the lower end of the 5%–95% range. Indeed, the lower end of the 5%–95% range derived from the regression method is 0.4 K higher. We suspect most of the difference comes from a bias in the late nineteenth-century EEI estimate in the state difference method of Sherwood et al. (2020). There is confidence in the lower end of the 5%–95% range derived from the regression method as it relies only on reliable recent data, and it makes full use of the observational record since 1971. It makes

us conclude that a climate sensitivity to $CO_2$ concentrations below 2.40 K is very unlikely (probability $< 0.05$) as it is not consistent with the observed 1971–2017 global energy budget. Unfortunately, on the high end of the 5%–95% range, the observational constraint from the regression method is not tight enough to provide any relevant information on the credibility of high climate sensitivity to $CO_2$ concentrations.

In the short term, a way forward to improve the observational constraint on $CO_2$effCS from the observed energy budget since 1971 is to tame down the uncertainty contribution from internal variability and from the short term forcing (such as volcanic eruptions and variations in the solar cycle) by separating a priori the long term forced signal in the ocean heat content (OHC) data and the SAT data, before applying the regression method. Because 2D observations of OHC and SAT are available since 1971, we could use signal-to-noise-maximizing pattern filtering methods that have proven to be efficient at identifying the forced response in individual ensemble members (and thus also in observations, e.g., Wills et al. 2021).

*Acknowledgments.* We are very grateful to the editor Isaac Held and three anonymous reviewers for constructive comments on the paper. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output, particularly those participating to the RFMIP project. We thank the scientific groups that produce the observation datasets for making available continuous, up-to-date, quality-controlled products. We also thank the International Argo Program and the national Argo programs for collecting the Argo data and making them freely available. The Argo Program is part of the Global Ocean Observing System. This work has been possible because of the funding from the H2020 COCLICO project. Author Chenal is grateful to the French Ministry of Ecological Transition for his funding.

*Data availability statement.* We used CMIP6 data available on the IPSL internal infrastructure or, alternatively, from the ESGF web node (https://esgf-node.ipsl.upmc.fr/projects/cmip6-ipsl/). OHC products come from internal estimates. The ERF time series comes from Sherwood et al. (2020). The temperature time series was obtained online (https://www-users.york.ac.uk/~kdc3/papers/coverage2013/series.html).

## APPENDIX

### Calculating Bias and Uncertainty in λ

In this appendix, we explain how the bias and the uncertainty in $\lambda$ from internal variability (in histeffCS) and from the pattern effect (in CO2effCS) are computed.

#### a. Notations

Let $z$ be the $z$th climate model and let $j$ be the $j$th realization of the historical experiment of model $z$. We note that $\langle X\rangle_j$ is the average across realizations of the variable $X$ and $\langle X\rangle_z$ is the average across models. We note that $\langle j\rangle$ is the ensemble mean of the ensemble of all historical realizations of a given model. (Note that $\langle\lambda\rangle_j \neq \lambda_{\langle j\rangle}$ because $\lambda$ is a nonlinear variable with respect to $T$; see also Fig S1 in the online supplemental material.) In general, in this study, variables $X$ are computed over sliding windows of 47 years, and therefore they are functions of time $X(t)$. For a time-varying variable $X(t)$, we note that $\overline{X}$ is the average over sliding widows.

#### b. Bias and the uncertainty in λ from internal variability

##### 1) Individual model

For a given model $z$, the bias in $\lambda$ due to internal variability is

$$\mathrm{Biv}_{\lambda_z} = \left\langle \overline{\lambda_{j,z}} \right\rangle_j - \overline{\lambda_{\langle j\rangle,z}}$$

and the uncertainty due to the internal variability is

$$\Xi\mathrm{iv}_{\lambda_z} = \sqrt{\left\langle \left( \overline{\lambda_{j,z}} - \left\langle \overline{\lambda_{j,z}} \right\rangle_j \right)^2 \right\rangle_j}.$$

##### 2) Ensemble of several models

Given $Z$ models in total, the bias in $\lambda$ due to internal variability (Table 1, column 8) is then computed as

$$\mathrm{Biv}_{\lambda} = \left\langle \mathrm{Biv}_{\lambda_z} \right\rangle_z,$$

and the uncertainty in $\lambda$ due to internal variability (Table 1, column 9) is computed as

$$\Xi\mathrm{iv}_{\lambda} = \sqrt{\frac{1}{Z}\sum_{1}^{Z} \Xi_{\lambda_z}^2}.$$

#### c. Bias and the uncertainty in λ from the pattern effect

##### 1) Individual model

Given a model $z$, the pattern effect bias in $\lambda$ is

$$\mathrm{Bp}_{\lambda_z} = \overline{\lambda_{\langle j\rangle,z}} - \lambda_{4\times \mathrm{CO}_{2z}},$$

where $\lambda_{4\times \mathrm{CO}_{2z}}$ is derived from the regression of $N$ against $T$ on the abrupt-4×CO2 experiment of model $z$, following the method of Gregory et al. (2004) (see main text).

##### 2) Ensemble of several models

Given $Z$ models in total, the pattern effect bias $\mathrm{Bp}_{\lambda}$ (Table 2, column 4) and the pattern effect uncertainty $\Xi\mathrm{p}_{\lambda}$ (Table 2, column 5) are respectively

$$\mathrm{Bp}_{\lambda} = \left\langle \mathrm{Bp}_{\lambda_z} \right\rangle_z \text{and}$$

$$\Xi\mathrm{p}_{\lambda} = \sqrt{\left\langle \left( \mathrm{Bp}_{\lambda_z} - \mathrm{Bp}_{\lambda} \right)^2 \right\rangle_z}.$$

## REFERENCES

Abraham, J. P., and Coauthors, 2013: A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Rev. Geophys.*, **51**, 450–483, https://doi.org/10.1002/rog.20022.

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

——, J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, **39**, L09712, https://doi.org/10.1029/2012GL051607.

——, ——, and ——, 2015: The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Climate*, **28**, 1630–1648, https://doi.org/10.1175/JCLI-D-14-00545.1.

——, ——, D. Paynter, L. G. Silvers, C. Zhou, and T. Mauritsen, 2018: Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophys. Res. Lett.*, **45**, 8490–8499, https://doi.org/10.1029/2018GL078887.

——, and Coauthors, 2019: Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3.1 and UKESM1. *J. Adv. Model. Earth Syst.*, **11**, 4377–4394, https://doi.org/10.1029/2019MS001866.

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

——, C. M. Bitz, and G. H. Roe, 2013: Time-varying climate sensitivity from regional feedbacks. *J. Climate*, **26**, 4518–4534, https://doi.org/10.1175/JCLI-D-12-00544.1.

Bagnell, A., and T. DeVries, 2021: 20th century cooling of the deep ocean contributed to delayed acceleration of Earth's energy imbalance. *Nat. Commun.*, **12**, 4604, https://doi.org/10.1038/s41467-021-24472-3.

Barnes, E. A., and R. J. Barnes, 2015: Estimating linear trends: Simple linear regression versus epoch differences. *J. Climate*, **28**, 9969–9976, https://doi.org/10.1175/JCLI-D-15-0032.1.

Bellouin, N., and Coauthors, 2020: Bounding global aerosol radiative forcing of climate change. *Rev. Geophys.*, **58**, e2019RG000660, https://doi.org/10.1029/2019RG000660.

Bilbao, R. A., J. M. Gregory, and N. Bouttes, 2015: Analysis of the regional pattern of sea level change due to ocean dynamics and density change for 1993–2099 in observations and CMIP5 AOGCMs. *Climate Dyn.*, **45**, 2647–2666, https://doi.org/10.1007/s00382-015-2499-z.

Bloch-Johnson, J., R. T. Pierrehumbert, and D. S. Abbot, 2015: Feedback temperature dependence determines the risk of high warming. *Geophys. Res. Lett.*, **42**, 4973–4980, https://doi.org/10.1002/2015GL064240.

——, M. Rugenstein, M. B. Stolpe, T. Rohrschneider, Y. Zheng, and J. M. Gregory, 2021: Climate sensitivity increases under higher $CO_2$ levels due to feedback temperature dependence. *Geophys. Res. Lett.*, **48**, e2020GL089074, https://doi.org/10.1029/2020GL089074.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate change 2013: The physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 571–657.

——, and Coauthors, 2018: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP. Earth System Grid Federation, accessed 15 September 2019, https://doi.org/10.22033/ESGF/CMIP6.1534.

Budyko, M. I., 1969: The effect of solar radiation variations on the climate of the Earth. *Tellus*, **21**, 611–619, https://doi.org/10.3402/tellusa.v21i5.10109.

Cheng, L., K. E. Trenberth, J. Fasullo, T. Boyer, J. Abraham, and J. Zhu, 2017: Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.*, **3**, e1601545, https://doi.org/10.1126/sciadv.1601545.

Church, J. A., N. J. White, and J. M. Arblaster, 2005: Significant decadal-scale impact of volcanic eruptions on sea level and ocean heat content. *Nature*, **438**, 74–77, https://doi.org/10.1038/nature04237.

——, and Coauthors, 2011: Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophys. Res. Lett.*, **38**, L18601, https://doi.org/10.1029/2011GL048794.

Cowtan, K., and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, https://doi.org/10.1002/qj.2297.

——, and Coauthors, 2015: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.*, **42**, 6526–6534, https://doi.org/10.1002/2015GL064888.

Dessler, A. E., and P. M. Forster, 2018: An estimate of equilibrium climate sensitivity from interannual variability. *J. Geophys. Res. Atmos.*, **123**, 8634–8645, https://doi.org/10.1029/2018JD028481.

——, T. Mauritsen, and B. Stevens, 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.

Dong, Y., C. Proistosescu, K. C. Armour, and D. S. Battisti, 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. *J. Climate*, **32**, 5471–5491, https://doi.org/10.1175/JCLI-D-18-0843.1.

Dufresne, J.-L., and S. Bony, 2008: An assessment of the primary sources of spread of global warming estimates from coupled atmosphere–ocean models. *J. Climate*, **21**, 5135–5144, https://doi.org/10.1175/2008JCLI2239.1.

Etminan, M., G. Myhre, E. Highwood, and K. Shine, 2016: Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.*, **43**, 12 614–12 623, https://doi.org/10.1002/2016GL071930.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 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.

Forster, P., and Coauthors, 2022: The Earth's energy budget, climate feedbacks, and climate sensitivity. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_07.pdf, in press.

Frey, W. R., and J. E. Kay, 2018: The influence of extratropical cloud phase and amount feedbacks on climate sensitivity. *Climate Dyn.*, **50**, 3097–3116, https://doi.org/10.1007/s00382-017-3796-5.

Gebbie, G., and P. Huybers, 2011: How is the ocean filled? *Geophys. Res. Lett.*, **38**, L06604, https://doi.org/10.1029/2011GL046769.

——, and ——, 2012: The mean age of ocean waters inferred from radiocarbon observations: Sensitivity to surface sources and accounting for mixing histories. *J. Phys. Oceanogr.*, **42**, 291–305, https://doi.org/10.1175/JPO-D-11-043.1.

——, and ——, 2019: The Little Ice Age and 20th-century deep Pacific cooling. *Science*, **363**, 70–74, https://doi.org/10.1126/science.aar8413.

Ghil, M., and V. Lucarini, 2020: The physics of climate variability and climate change. *Rev. Mod. Phys.*, **92**, 035002, https://doi.org/10.1103/RevModPhys.92.035002.

Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans*, **118**, 6704–6716, https://doi.org/10.1002/2013JC009067.

Gouretski, V., and K. P. Koltermann, 2007: How much is the ocean really warming? *Geophys. Res. Lett.*, **34**, L01610, https://doi.org/10.1029/2006GL027834.

Gregory, J. M., 2000: Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dyn.*, **16**, 501–515, https://doi.org/10.1007/s003820000059.

——, and P. M. Forster, 2008: Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res.*, **113**, D23105, https://doi.org/10.1029/2008JD010405.

——, and T. Andrews, 2016: Variation in climate sensitivity and feedback parameters during the historical period. *Geophys. Res. Lett.*, **43**, 3911–3920, https://doi.org/10.1002/2016GL068406.

——, and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, https://doi.org/10.1029/2003GL018747.

——, T. Andrews, and P. Good, 2015: The inconstancy of the transient climate response parameter under increasing $CO_2$. *Philos. Trans. Roy. Soc.*, **A373**, 20140417, https://doi.org/10.1098/rsta.2014.0417.

——, ——, P. Ceppi, T. Mauritsen, and M. Webb, 2020: How accurately can the climate sensitivity to $CO_2$ be estimated from historical climate change? *Climate Dyn.*, **54**, 129–157, https://doi.org/10.1007/s00382-019-04991-y.

Grose, M. R., R. Colman, J. Bhend, and A. F. Moise, 2017: Limits to global and Australian temperature change this century based on expert judgment of climate sensitivity. *Climate Dyn.*, **48**, 3325–3339, https://doi.org/10.1007/s00382-016-3269-2.

——, J. Gregory, R. Colman, and T. Andrews, 2018: What climate sensitivity index is most useful for projections? *Geophys. Res. Lett.*, **45**, 1559–1566, https://doi.org/10.1002/2017GL075742.

Heinze, C., and Coauthors, 2019: Climate feedbacks in the Earth system and prospects for their evaluation. *Earth Syst. Dyn.*, **10**, 379–452, https://doi.org/10.5194/esd-10-379-2019.

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

IPCC, 1992: *Climate Change: The IPCC 1990 and 1992 Assessments.* WMO, 165 pp., https://www.ipcc.ch/site/assets/uploads/2018/05/ipcc_90_92_assessments_far_full_report.pdf.

Ishii, M., Y. Fukuda, S. Hirahara, S. Yasui, T. Suzuki, and K. Sato, 2017: Accuracy of global upper ocean heat content estimation expected from present observational data sets. *SOLA*, **13**, 163–167, https://doi.org/10.2151/sola.2017-030.

Kelley, M., and Coauthors, 2020: GISS-E2.1: Configurations and climatology. *J. Adv. Mod. Earth Syst.*, **12**, e2019MS002025, https://doi.org/10.1029/2019MS002025.

Klein, S. A., A. Hall, J. R. Norris, and R. Pincus, 2017: Low-cloud feedbacks from cloud-controlling factors: A review. *Shallow Clouds, Water Vapor, Circulation, and Climate Sensitivity*, R. Pincus et al., Eds., Springer, 135–157.

Knutti, R., and M. A. Rugenstein, 2015: Feedbacks, climate sensitivity and the limits of linear models. *Philos. Trans. Roy. Soc.*, **A373**, 20150146, https://doi.org/10.1098/rsta.2015.0146.

——, ——, and G. C. Hegerl, 2017: Beyond equilibrium climate sensitivity. *Nat. Geosci.*, **10**, 727–736, https://doi.org/10.1038/ngeo3017.

Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov, 2009: Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.*, **36**, L07608, https://doi.org/10.1029/2008GL037155.

——, and Coauthors, 2012: World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.*, **39**, L10603, https://doi.org/10.1029/2012GL051106.

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

Marvel, K., R. Pincus, G. A. Schmidt, and R. L. Miller, 2018: Internal variability and disequilibrium confound estimates of climate sensitivity from observations. *Geophys. Res. Lett.*, **45**, 1595–1601, https://doi.org/10.1002/2017GL076468.

Melet, A., and B. Meyssignac, 2015: Explaining the spread in global mean thermosteric sea level rise in CMIP5 climate models. *J. Climate*, **28**, 9918–9940, https://doi.org/10.1175/JCLI-D-15-0200.1.

Meyssignac, B., and Coauthors, 2019: Measuring global ocean heat content to estimate the Earth energy imbalance. *Front. Mar. Sci.*, **6**, 432, https://doi.org/10.3389/fmars.2019.00432.

Mitchell, T. D., 2003: Pattern scaling: An examination of the accuracy of the technique for describing future climates. *Climatic Change*, **60**, 217–242, https://doi.org/10.1023/A:1026035305597.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, https://doi.org/10.1029/2011JD017187.

Mülmenstädt, J., and Coauthors, 2021: An underestimated negative cloud feedback from cloud lifetime changes. *Nat. Climate Change*, **11**, 508–513, https://doi.org/10.1038/s41558-021-01038-1.

Myhre, G., and Coauthors, 2013: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 659–740, https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter08_FINAL.pdf.

——, and Coauthors, 2017: Multi-model simulations of aerosol and ozone radiative forcing due to anthropogenic emission changes during the period 1990–2015. *Atmos. Chem. Phys.*, **17**, 2709–2720, https://doi.org/10.5194/acp-17-2709-2017.

Myers, T. A., R. C. Scott, M. D. Zelinka, S. A. Klein, J. R. Norris, and P. M. Caldwell, 2021: Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity. *Nat. Climate Change*, **11**, 501–507, https://doi.org/10.1038/s41558-021-01039-0.

National Research Council, 1979: *Carbon Dioxide and Climate: A Scientific Assessment.* National Academies Press, 34 pp., https://doi.org/10.17226/12181.

Paasche, Ø., and J. Bakke, 2010: Defining the little ice age. *Climate Past Discuss.*, **6**, 2159–2175, https://doi.org/10.5194/cpd-6-2159-2010.

Pachauri, R. K., and Coauthors, 2014: *Climate Change 2014: Synthesis Report.* Cambridge University Press, 151 pp., https://www.ipcc.ch/site/assets/uploads/2018/02/SYR_AR5_FINAL_full.pdf.

Paynter, D., and T. L. Frölicher, 2015: Sensitivity of radiative forcing, ocean heat uptake, and climate feedback to changes in anthropogenic greenhouse gases and aerosols. *J. Geophys. Res. Atmos.*, **120**, 9837–9854, https://doi.org/10.1002/2015JD023364.

Perrette, M., F. Landerer, R. Riva, K. Frieler, and M. Meinshausen, 2013: A scaling approach to project regional sea level rise and its uncertainties. *Earth Syst. Dyn.*, **4**, 11–29, https://doi.org/10.5194/esd-4-11-2013.

Pincus, R., P. M. Forster, and B. Stevens, 2016: The Radiative Forcing Model Intercomparison Project (RFMIP): Experimental protocol for CMIP6. *Geosci. Model Dev.*, **9**, 3447–3460, https://doi.org/10.5194/gmd-9-3447-2016.

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

——, A. Donohoe, K. C. Armour, G. H. Roe, M. F. Stuecker, and C. M. Bitz, 2018: Radiative feedbacks from stochastic variability in surface temperature and radiative imbalance. *Geophys. Res. Lett.*, **45**, 5082–5094, https://doi.org/10.1029/2018GL077678.

Richardson, M., K. Cowtan, E. Hawkins, and M. B. Stolpe, 2016: Reconciled climate response estimates from climate models and the energy budget of earth. *Nat. Climate Change*, **6**, 931–935, https://doi.org/10.1038/nclimate3066.

Roe, G. H., and M. B. Baker, 2007: Why is climate sensitivity so unpredictable? *Science*, **318**, 629–632, https://doi.org/10.1126/science.1144735.

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

——, and Coauthors, 2019: LongRunMIP: Motivation and design for a large collection of millennial-length AOGCM simulations. *Bull. Amer. Meteor. Soc.*, **100**, 2551–2570, https://doi.org/10.1175/BAMS-D-19-0068.1.

——, and Coauthors, 2020: Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophys. Res. Lett.*, **47**, e2019GL083898, https://doi.org/10.1029/2019GL083898.

Santer, B. D., T. M. Wigley, M. E. Schlesinger, and J. F. Mitchell, 1990: Developing climate scenarios from equilibrium GCM results. Max-Planck-Institut-für-Meteorologie Rep. 47, 29 pp.

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

Sherwood, S., and Coauthors, 2020: An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.*, **58**, e2019RG000678, https://doi.org/10.1029/2019RG000678.

Shindell, D. T., 2014: Inhomogeneous forcing and transient climate sensitivity. *Nat. Climate Change*, **4**, 274–277, https://doi.org/10.1038/nclimate2136.

Skeie, R. B., T. Berntsen, M. Aldrin, M. Holden, and G. Myhre, 2018: Climate sensitivity estimates—Sensitivity to radiative forcing time series and observational data. *Earth Syst. Dyn.*, **9**, 879–894, https://doi.org/10.5194/esd-9-879-2018.

Smith, C. J., and Coauthors, 2020: Effective radiative forcing and adjustments in cmip6 models. *Atmos. Chem. Phys.*, **20**, 9591–9618, https://doi.org/10.5194/acp-20-9591-2020.

Stevens, B., S. C. Sherwood, S. Bony, and M. J. Webb, 2016: Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earth's Future*, **4**, 512–522, https://doi.org/10.1002/2016EF000376.

Swart, N. C., and Coauthors, 2019: The Canadian Earth System Model version 5 (CanESM5. 0.3). *Geosci. Model Dev.*, **12**, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019.

Tatebe, H., and Coauthors, 2019: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12**, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019.

von Schuckmann, K., and Coauthors, 2020: Heat stored in the Earth system: Where does the energy go? *Earth Syst. Sci. Data*, **12**, 2013–2041, https://doi.org/10.5194/essd-12-2013-2020.

Wills, R. C., K. C. Armour, D. S. Battisti, C. Proistosescu, and L. A. Parsons, 2021: Slow modes of global temperature variability and their impact on climate sensitivity estimates. *J. Climate*, **34**, 8717–8738, https://doi.org/10.1175/JCLI-D-20-1013.1.

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

——, and Coauthors, 2020: Climate Sensitivity of GFDL's CM4.0. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001838, https://doi.org/10.1029/2019MS001838.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.

Zhou, C., M. D. Zelinka, and S. A. Klein, 2017: Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach. *J. Adv. Model. Earth Syst.*, **9**, 2174–2189, https://doi.org/10.1002/2017MS001096.