High-sensitivity Earth System Models Most Consistent with Observations

Menghan Yuan[a] Thomas Leirvik[b, c, d] Trude Storelvmo[b, e] Kari Alterskjær[b, f] Peter C.B. Phillips[g,h,i,j] Christopher J. Smith[k, l]

[a] *Nuffield College, University of Oxford, United Kingdom*

[b] *Nord University, Norway*

[c] *The Arctic University of Norway, Norway*

[d] *The Norwegian University of Science and Technology, Norway*

[e] *University of Oslo, Norway*

[f] *Center for International Climate and Environmental Research (Cicero), Norway*

[g] *University of Auckland, New Zealand*

[h] *Yale University, United States of America*

[i] *Singapore Management University, Singapore*

[j] *University of Southampton, United Kingdom*

[k] *University of Leeds, United Kingdom*

[l] *International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria*

Corresponding author: Menghan Yuan, menghan.yuan@nuffield.ox.ac.uk

ABSTRACT: Earth's transient climate response (TCR) quantifies the global mean surface air temperature change due to a doubling of atmospheric CO_2, at the time of doubling. TCR is highly correlated with near-term climate projections, and thus of utmost relevance for climate policy, but remains poorly constrained in part due to uncertainties of physical process simulations in Earth System Models (ESMs). Within state-of-the-art ESMs participating in the Coupled Model Intercomparison Project (CMIP6), the TCR range (1.1–2.9$°C$) is much too wide to offer useful guidance to policymakers on remaining carbon budgets aligned with the Paris agreement goals. To address this issue, we here present an observation-based TCR estimate of 2.3±0.4$°C$ (95% confidence interval). We show that this method correctly diagnoses TCR from 22 CMIP6 ESMs if the same variables are taken from the ESMs as are available from observations. This increases confidence in the new observation-based central estimate and range, which are respectively higher and narrower than the mean and spread of the estimates from the entire ensemble of CMIP6. Many ESMs tend to have TCRs lower than the observational range, for which our findings suggest that underestimation of the aerosol cooling effect could be a primary cause. This paper points to the need for ESMs to re-examine their aerosol cooling effect to achieve better correspondence with observational data. Further, the revised TCR estimate suggests a remaining carbon budget to 1.5$°C$ of around nine years of current CO_2 emissions.

SIGNIFICANCE STATEMENT: Understanding the relationship between temperature change and greenhouse gas emissions, also referred to as climate sensitivity, is essential to constrain global warming and its economic consequences. Current studies informing climate sensitivity rely heavily on climate model projections, which rely on highly uncertain parameterizations of a wide range of critical processes. However, observations could potentially provide a more reliable data source for climate evolution. We propose an observation-based framework for estimating climate sensitivity and validate it using the output of 22 Earth System Models. Using our framework, we provide an empirical climate sensitivity estimate simultaneously as producing a reduced uncertainty compared to the likely range suggested by the whole ESM ensemble in CMIP6 and the IPCC AR6 assessment. The observational estimate suggests a downward revision of the remaining carbon budget to $1.5°C$.

1. Introduction

The question of exactly how sensitive Earth's climate is to atmospheric greenhouse gas perturbations has been long-standing in the climate research community and is of mounting concern in society at large. Yet, arguably, we are no closer to the answer today than we were several decades ago (IPCC 2001; Forster et al. 2021). Assessments continue to depend on Earth System Models (ESMs), which rely on simplified representations of a wide range of small-scale physical processes of relevance for feedback mechanisms in the climate system, resulting in a large spread in simulated climate sensitivity. This uncertainty, in turn, translates into highly uncertain climate projections for a given future emission-scenario (Tebaldi et al. 2020), with obvious consequences for society's ability to determine necessary mitigation and adaptation action. TCR has been demonstrated to correlate well with near-term climate projections across a wide range of emission scenarios (see e.g., Grose et al. 2018; Huusko et al. 2021), and is therefore among the metrics of Earth's climate sensitivity most relevant for today's decision makers. The latest generation of ESMs in the CMIP6 ensemble produces a mean TCR of $2.0°C$ (Eyring et al. 2016), somewhat higher than the previous ESM generation (CMIP5 mean of $1.8°C$, Meehl et al. 2020). For context, the most recent report from the Intergovernmental Panel for Climate Change (IPCC AR6) assessed the likely TCR range to be 1.2 to $2.4°C$, based on multiple lines of evidence (Forster et al. 2021). Multiple CMIP6 models now produce TCR values well above the upper end of this range (Meehl et al. 2020), raising questions about the plausibility of some of the most sensitive ESMs.

The above serves as the backdrop for the research presented here, which takes advantage of a new observational approach proposed by Phillips et al. (2020) to determine TCR based on observations. This method makes use of an equilibrium relationship among surface air temperature, surface solar radiation, and greenhouse gas concentrations and estimates empirically the sensitivity of temperature to greenhouse gases. An important innovation of the approach is that it uses an observational proxy, surface solar radiation, for the cooling effect of aerosols, in order to isolate the observed surface air temperature change that can be attributed to atmospheric greenhouse gas changes, thus allowing for TCR inference. Other efforts to constrain TCR based on historical observations have generally relied on ESM output for aerosol cooling estimates (e.g., Otto et al. 2013) or have been based on the premise that aerosol cooling has remained nearly constant in recent decades (see e.g., Jiménez-de-la Cuesta and Mauritsen 2019; Tokarska et al. 2020; Nijsse et al. 2020). The latter is based on the fact that globally, emissions of aerosol particles and their precursors have been relatively stable since the mid-1970s. However, there is ample evidence that a near-constant global mean atmospheric aerosol burden does not directly translate to a constant global mean aerosol cooling, as the spatial distribution of aerosols is also of critical importance for the global mean aerosol effect on climate (Regayre et al. 2014; Shindell et al. 2015; Persad and Caldeira 2018). Indeed, the spatial distribution of atmospheric aerosols has changed considerably in recent decades (Hoesly et al. 2018), and the associated climate impacts are expected to be non-negligible (Marvel et al. 2016).

In this study, we merge observations of well-mixed atmospheric greenhouse gas concentrations with surface air temperature and surface radiation fluxes over land during the period 1964 –2014. Based on the constructed data set, we use statistical methods that are well established within the field of econometrics to indirectly determine TCR. Additionally, we shed light on implications of the observation-based TCR estimate on the remaining carbon budget in line with the Paris agreement warming goal.

The rest of the paper proceeds as follows: section 2 describes the data and methods used in the empirical estimation of TCR. Section 3 displays the main results. The findings are discussed in Section 4.

2. Data and Methods

a. Data

Data leading to the findings in this study come from both observations and ESM simulations. Observed surface air temperature data are available from the Climate Research Unit gridded Time Series (CRU TS V4) maintained by the University of East Anglia (Harris et al. 2020). Observed surface solar radiation (SSR) data are obtained from a spatially interpolated data set based on the Global Energy Balance Archive (Wild et al. 2017; Yuan et al. 2021). Both observational data sets provide complete gridded observations over land at 0.5° resolution.

Simulation counterparts, hereafter, termed 'synthetic observations', are obtained from historical simulations from 22 ESMs in CMIP6 (Eyring et al. 2016). The number of ESMs included was determined by the availability of model simulations and output variables required to calculate TCR at the time of the analysis. Some ESMs have several realizations, each started from slightly different initial conditions. Only the first realization ('r1') of each model is used here because we believe it is reasonably representative as ensemble members tend to converge and generate similar TCR estimates (supplementary information SI Figure S1). Reconciling the data availability of CMIP6 model simulations with that of observations we limit the study to the time period from 1964 to 2014.

In this study SSR is used as a proxy for aerosol forcing. Aerosols absorb and scatter sunlight and also affect the radiative properties of clouds (e.g., Forster et al. 2021), causing the dimming and brightening observed in SSR decadal trends, and are deemed as the major driver of long-term variations of SSR (see e.g., Wild et al. 2021; Kudo et al. 2012; Wandji Nyamsi et al. 2020; Ruckstuhl and Norris 2009). Quantitatively, a statistically significant positive correlation is found between SSR and aerosol forcing for the majority of ESMs (SI Table S1).

To obtain a global overview of temperature and SSR evolution, we aggregate grid cell values to global land averages, weighted by the cosine of latitude to account for the gridbox areas reducing with increasing latitude. Table 1 reports the summary statistics for annual changes in global average temperature and SSR for observations and ESM simulations over 1964–2014. The mean annual change in observed temperature is $0.025°C$, with a standard deviation of $0.248°C$. ESMs simulate comparable temperature trends. The mean of 22 ESMs shows an average annual change of 0.028°C, with a standard deviation of 0.221°C.

Over 1964–2014, observed SSR shows dimming trends, with a mean annual change of −0.11 Wm^{-2}, and a standard deviation of 0.588 Wm^{-2}. By contrast, the dimming trends are much weaker in the ESM simulations. The mean of the annual change of SSR in the ESMs is only about one fifth of the observed dimming trend (−0.023 vs. −0.11 Wm^{-2}). Even the model with the strongest dimming trends fails to fully replicate the magnitude of the observed dimming. The most negative simulated annual change of SSR is recorded in GISS-E2-1-G at −0.066 Wm^{-2}, only about 60% of the observed trends. Counterfactually, two models even report positive mean annual changes—HadGEM3-GC31-LL and UKESM1-0-LL at 0.026 and 0.002 Wm^{-2}, respectively (SI Table S3).

Table 1. Mean, standard deviation, minimum and maximum for the annual change, i.e., first difference, in global average temperature and SSR. Statistics are shown for observations and a summary of 22 ESMs over the period 1964–2014.

Temperature [°C/year]					
		Mean	St. dev	Min.	Max.
observation		0.025	0.248	−0.529	0.500
ESMs	Mean	0.028	0.221	−0.522	0.512
	Min.	0.013	0.146	−0.942	0.265
	Max.	0.046	0.317	−0.366	0.727
SSR [Wm^{-2}/year]					
Model		Mean	St. dev	Min.	Max.
observation		−0.110	0.588	−0.979	1.492
ESMs	Mean	−0.023	0.824	−2.124	1.894
	Min.	−0.066	0.504	−4.079	0.956
	Max.	0.026	1.317	−1.153	3.061

Refer to SI Tables S2 and S3 for the detailed statistics for each individual ESM.

Our source of global CO_2 equivalent concentrations is the National Oceanic and Atmospheric Administration (NOAA) Annual Greenhouse Gas Index (ACGI), which contains measures of the interannual variability of global forcing resulting from changes in greenhouse gases. CO_2 is known to be the largest contributor to the index, and all non-CO_2 greenhouse gas effects are converted into changes in global forcing and aggregated with that of CO_2. In other words, the AGGI is deemed as an instrument of equivalent CO_2 atmospheric concentrations.

We use the reported TCR, regarded as the 'true' TCR, as the reference for comparison with the empirically estimated TCR. The reported TCR is calculated as the change in global near surface temperature in a 20-year average around the time of CO_2 doubling (years 60-79 in simulations in which CO_2 was increased by 1% per year) as compared to the equivalent 20-year segment of each model's own pre-industrial control simulation. The equivalent time period was used to avoid influence from any drift due to remaining energy imbalance in the control. Confidence levels were found by bootstrapping the mean difference between the two 20-year segments with 10,000 realizations.

b. Econometric Framework

The transient climate response (TCR) in this study is estimated using an empirical econometric framework which relates global average surface air temperature in year $t+1$ ($\bar{T}_{t+1}$) to previous year's temperature ($\bar{T}_t$), global average surface solar radiation ($\bar{R}_t$), and the logarithm of CO_2 equivalent concentrations ($CO_{2,t}$). CO_2 is assumed uniformly distributed in the atmosphere, so no spatial averaging is needed in this case. The following time series representation, which is reduced from the original panel model established in Phillips et al. (2020), is used for the analysis in this paper

$$\bar{T}_{t+1} = \gamma_0 + \theta_1 \bar{T}_t + \theta_2 \bar{R}_t + \gamma_3 \ln(CO_{2,t}) + u_{t+1} \tag{1}$$

where u_{t+1} is the equation error disturbance at year $t+1$ that embodies variability not captured by the explanatory regressors. This global time series $\bar{T}_t$ and $\bar{R}_t$ are global averages aggregated by grid cells i and time periods t.

The TCR can be estimated as a 'reduced form' parameter given by

$$TCR = \frac{\gamma_3}{1-\theta_1} \times \ln(2) \tag{2}$$

Estimates of the coefficients are obtained by fully modified least squares (FM-OLS, Phillips and Hansen 1990), using the econometric framework derived in Phillips et al. (2020), which allows for joint dependence and nonstationarity among variables as well as autocorrelation common in time series data and residuals[1].

Since our observational data cover only land areas, we need to follow a conversion procedure to convert the calculated TCR, which is valid for land only, to a global TCR value. Specifically,

$$TCR_G = TCR_L \cdot \frac{A_L \cdot w_L + A_O \cdot w_O}{w_L} = TCR_L \cdot \left(A_L + \frac{A_O}{WR} \right) = TCR_L \cdot W_{trans}, \tag{3}$$

where TCR_L and TCR_G denote land and global TCR, respectively. A_L and A_O are Earth's land and ocean area fractions which are set to 0.29 and 0.71. $\frac{1}{WR} = \frac{w_O}{w_L}$ stands for the inverse of the *land-ocean warming ratio*, where w_O denotes the warming rate over ocean and w_L over land. W_{trans} denotes the conversion factor for the central estimate. To obtain the confidence interval (CI) for TCR_G accounting for uncertainty in WR, we multiply the lower bound of the CI for TCR_L by W^-_{trans} and the upper bound by W^+_{trans} given by

$$\begin{aligned} W^-_{trans} &= A_L + \frac{A_O}{WR} \cdot (1 - 0.05) \\ W^+_{trans} &= A_L + \frac{A_O}{WR} \cdot (1 + 0.05) \end{aligned} \tag{4}$$

This adjustment leads to a slightly wider uncertainty range than the 95% CI of global TCR estimate based on the transformation factor W_{trans} alone.

Note that ESMs have global coverage, making a direct global TCR estimate without any conversion possible. However, in this way we will not be able to assess how the conversion, which is necessary for observational estimates, affects the final global TCR estimate. Therefore, in order to keep consistency in the estimation method for observations and ESM simulations, we first mask the ESM simulations to retain only the land part, and then convert the land estimate to the global estimate following the same conversion procedure as in the observational analysis. A discussion of how the conversion impacts the global TCR estimate can be found in section 4.

[1]Variables are nonstationary if the distribution changes over time and autocorrelation occurs if observations over successive time periods are correlated.

c. Remaining Carbon Budget Calculation

The remaining carbon budget (RCB) up to a particular temperature limit above pre-industrial ΔT_{lim}, such as 1.5°C, can be conceptualized as (Matthews et al. 2021)

$$RCB = \frac{\Delta T_{lim}(1 - f_{nc}^{*}) - \Delta T_{anth}(1 - f_{nc})}{TCRE}, \tag{5}$$

where ΔT_{anth} is the anthropogenic-attributed warming since pre-industrial, f_{nc} is the present-day fraction of anthropogenic effective radiative forcing from non-CO_2 sources, f_{nc}^{*} is the non-CO_2 forcing fraction at net-zero CO_2 emissions, and TCRE is the transient climate response to cumulative emissions of CO_2.

TCRE can be approximated as (Jones and Friedlingstein 2020)

$$TCRE = a_f \cdot \frac{TCR}{\Delta C_{2\times CO_2}}, \tag{6}$$

where a_f is the cumulative airborne fraction taken at the time of doubling of CO_2 in a 1% per year compound CO_2 increase (i.e., approximately after 70 years) and $\Delta C_{2\times CO_2}$ is the increase in atmospheric carbon mass for a doubling of pre-industrial CO_2. Using a pre-industrial CO_2 value of 284.32 ppm representative of 1850 conditions (Meinshausen et al. 2017) as used in CMIP6 and a conversion of 1 ppm = 2.124 GtC (Friedlingstein et al. 2020) gives $\Delta C_{2\times CO_2}$ = 604 GtC.

To generate distributions of the remaining carbon budget to ΔT_{lim} = 1.5°C a 1-million member Monte Carlo ensemble was produced. TCR is sampled as gamma distributed for reported TCR from CMIP6 models from the distribution in Figure 1, and as normally distributed for the observational TCR using the mean of 2.31 K and standard deviation 0.18 K. For the estimate from Sherwood et al. (2020) we use a normal distribution with mean of 1.85 K and standard deviation of 0.35 K to approximate the median and 66% range of 1.8 (1.5-2.2) K in Sherwood et al. (2020). In all cases, airborne fraction is sampled from a normal distribution using the results from 11 CMIP6 carbon-cycle models in Arora et al. (2020) with mean 0.532 and standard deviation 0.033.

From the derived TCRE distributions, the remaining carbon budget is computed by sampling the terms in Eqn.(5) from distributions in Matthews et al. (2021). f_{nc} is taken from mean 1990-2019 non-CO_2 forcing fractions from all 411 integrated assessment model (IAM) scenarios considered by the IPCC Special Report on 1.5°C (median 0.14, 5-95% range −0.11 to 0.33, Rogelj et al.

2018) and sampled using a kernel density estimate. The non-CO_2 forcing fraction at net-zero $f_{nc}^* = 0.3081 f_{nc} + 0.14 + \varepsilon$ where ε is sampled as a normal distribution (mean 0 and 5-95% range of 0.05) that represents additional future socioeconomic pathway uncertainty up to net-zero CO_2 emissions in IAM scenarios (Matthews et al. 2021). ΔT_{anth} is sampled as a skew-normal distribution fit to best-estimate and 5-95% uncertainty of anthropogenic warming from 1850-1900 to 2019 of 1.18 (1.05 to 1.41) °C (Matthews et al. 2021). RCB calculations are converted from units of GtC to $GtCO_2$ (multiplied by 3.664) and reported to the nearest 5 $GtCO_2$ from the beginning of 2020.

3. Results

a. New Observation–based TCR Estimate

Because our empirical estimation is observation-driven and independent from complex physical process simulations in ESMs, the method has the potential to serve as an important tool for evaluation of ESM-simulated TCR. In a first application of this method, TCR was estimated to be $2.0 \pm 0.8°C$ (Storelvmo et al. 2016), while in the present study updates to observational data sets and further development of the methodology (Phillips et al. 2020) produce a somewhat higher estimate and a considerably narrower uncertainty range of $2.3 \pm 0.4°C$, thus supporting some of the higher TCR estimates emerging from CMIP6. Compared to previous applications, a more extensive observational data set with complete land coverage is used, in contrast to the scattered station data used in Storelvmo et al. (2016) and Phillips et al. (2020).

Next, we present evidence that the observational method can in fact correctly diagnose TCR. This is done by comparing the standard TCR calculation from 22 CMIP6 models with the TCR values estimated when the same variables that are available from observations are also extracted from the 22 models and used in the same way in the observational analysis (the TCR values estimated from the statistical analysis will hereinafter be referred to as E-TCR).

b. Increasing Confidence in the New TCR Estimate

To determine whether any method can in fact correctly diagnose TCR, one could simply wait for a couple of decades, as the role of aerosol cooling is expected to diminish with time due to projected reductions in anthropogenic aerosol emissions (Gidden et al. 2019; Shindell and Smith 2019). The observed warming would therefore increasingly be attributable to greenhouse gas increases, and first and foremost CO_2 (Myhre et al. 2015). With time, it should therefore be possible to infer TCR from observations with a considerably reduced uncertainty range. However, important climate policy decisions cannot wait for the more constrained TCR estimates that would eventually emerge; for example, a halving of the uncertainty range for TCR has been estimated to have a net present value of about $9.7 trillion if accomplished by 2030 (Hope 2015).

Motivated by this urgency, we here test the new method on 'synthetic observations' from the aforementioned 22 ESMs, to confirm that it can correctly diagnose the 'true' TCR from each of the models (see section 2.1).

As evident from Figure 1, the TCR distribution based on the standard calculation and the E-TCR emerging from the synthetic observations extracted from the ESMs are indeed very similar, albeit the latter produces a slightly higher ensemble mean (E-TCR mean of $2.16°C$ vs. TCR mean of $2.05\ °C$).

As further evidence that the observational TCR estimate is reliable, there is also a statistically significant positive correlation between the estimated E-TCR values and the reported TCR values based on standard calculations for the CMIP6 models (r=0.61, Figure 2), with low-TCR models correctly being diagnosed as such, and vice versa. Nevertheless, we note a slight tendency for the method to overestimate TCR from low-sensitivity ESMs and underestimate high-sensitivity models' TCR.

We also note that there are a few ESMs with particularly high or low E-TCRs, which stand out from the cluster of models and the regression line of E-TCR on TCR. The magnitude of E-TCR is largely determined by the climate trends emerging from the ESM simulations. Higher E-TCR models tend to show stronger simulated trends of temperature and/or radiation, whereas lower E-TCR models are usually associated with weaker trends. For instance, CanESM5 (model 4) shows strong trends in both temperature and radiation and reports the highest E-TCR among all ESMs. Similarly, UKESM1-0-LL with the second highest E-TCR (model 22) shows strong trends in temperature yet modest trends in radiation, which suggests the predominant role of temperature over radiation. By contrast, low E-TCR models CAMS-CSM1-0 and SAM0-UNICON (models 3 and 21) simulate some of the weakest temperature and radiation trends of all ESMs. (SI Figure S3).

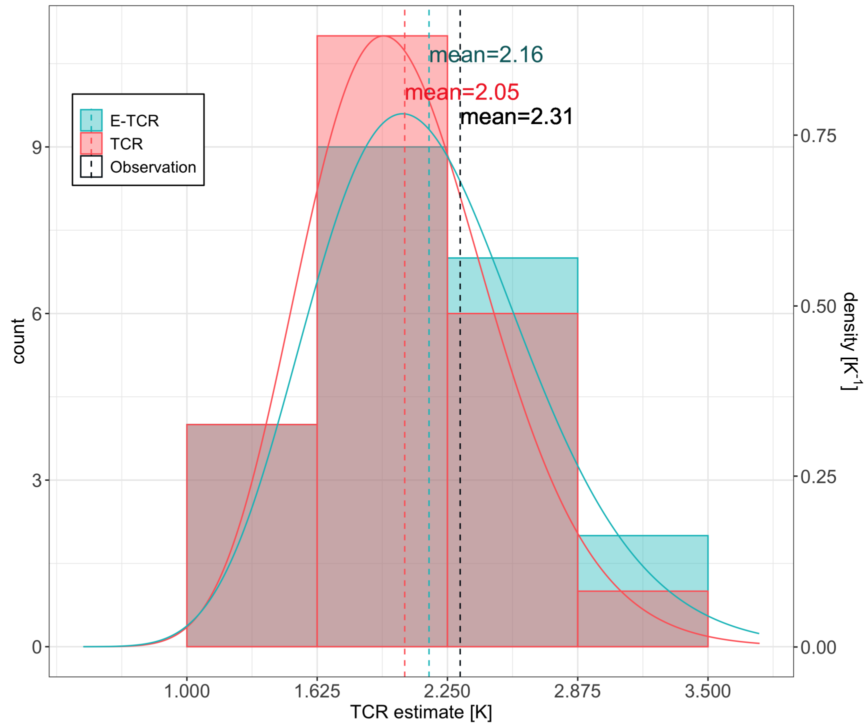

FIG. 1. Histograms of TCR from 22 CMIP6 models based on standard calculations (red bars) and estimated based on synthetic observations extracted from the ESMs (blue bars). Also shown are fitted gamma distributions for the standard TCR calculations (red curve) and the estimated values (blue curve). The dashed vertical lines show the mean for TCR (red), E–TCR (blue), respectively. The black line shows the TCR estimated from observations.

Finally, Figure 3 shows that among the 22 ESMs considered, 20 have reported TCR values that lie within the empirically estimated 95% confidence interval, while the remaining two (NorESM2-LM and GISS-E2-1-H) have reported TCR values lying marginally outside the confidence intervals.

In other words, Figure 2 shows that our observation-based method has skill. The figure also shows that the method cannot always perfectly diagnose the exact value of the true TCR, but Figure 3 importantly shows that the true TCR is always within or at the margin of the estimated E-TCR range. Based on this evidence we can have high confidence in the ability of the empirical TCR estimation method to correctly diagnose the TCR of the real climate system, which is thus very likely to lie in the estimated observation-based 95% confidence interval of 1.9 to 2.7, centered on $2.3°C$. Notably, only about half the CMIP6 models analyzed here produce TCRs that lie within this range. Out of the ones that do not, ten underestimate TCR relative to the observation-based

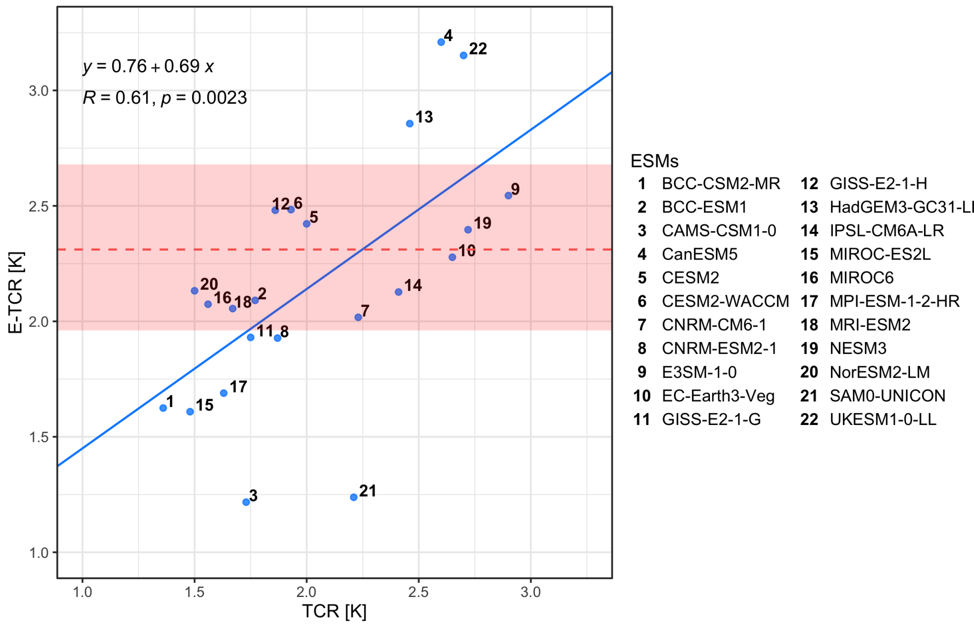

FIG. 2. TCR values based on standard calculations vs. E-TCR based on synthetic observations from 22 CMIP6 models. The blue line shows a regression line of E–TCR on TCR. As also shown are the regression equation, correlation coefficient, and its significant level in the upper-left corner. The shading area shows the 95% confidence interval of the observational TCR estimate; the dashed line shows the central estimate.

range, while only one overestimates it. In other words, the higher CMIP6 ensemble mean TCR relative to previous ESM generations is strongly supported by the findings presented here. This stands in contrast to recent studies that have attempted to use the rate of warming in recent decades to constrain TCR, arriving at best estimates of TCR as low as 1.6°C (see e.g., Tokarska et al. 2020). However, these studies rely heavily on the accuracy of the assumption of a near-constant aerosol cooling in recent decades, as simulated by CMIP6 models, which is not supported by the present observational framework. Our observational estimate relies on observations only and stands independent from ESMs widely applied in other studies.

Using SSR as a proxy for aerosol forcing, we note that ESMs tend to markedly underestimate aerosol cooling compared to observations, whereas temperature simulations reproduce historical warming reasonably well (Figure 4). The underestimation of aerosol cooling contributes to the divergence between E-TCR of the ESMs and the observational TCR. In our empirical method, we disentangle temperature change attributable to greenhouse gas warming and aerosol cooling effect. We find that greenhouse gases have driven up global land temperature by 1.5°C over 1964–2014,

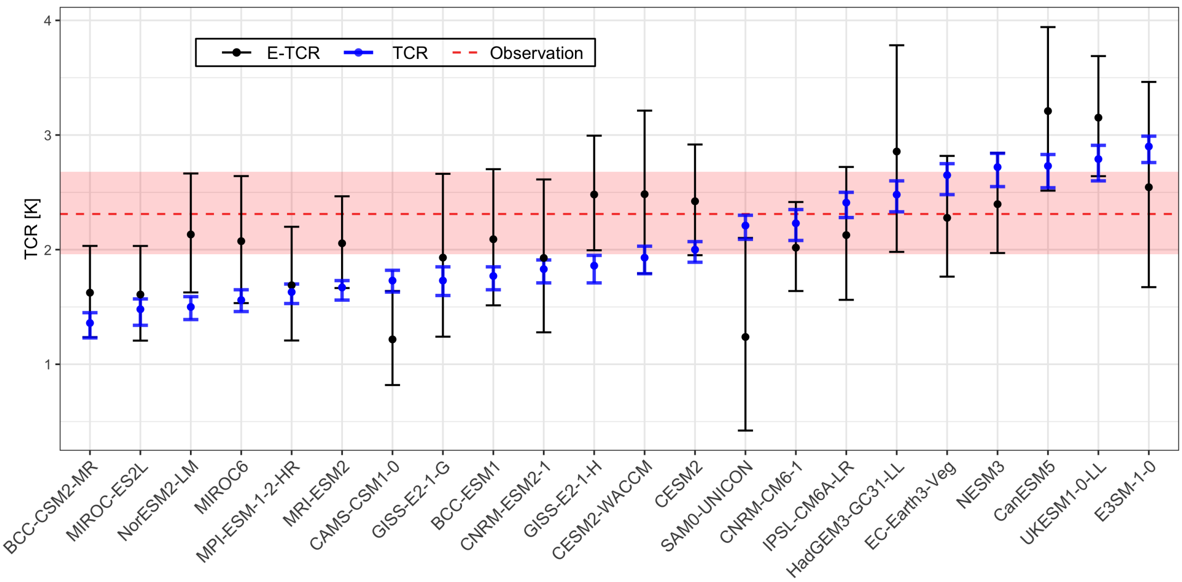

FIG. 3. TCR values based on standard calculations for 22 CMIP6 models (blue points and bars showing central values and 95% confidence intervals, respectively) and the corresponding E–TCR values (black points and bars) using the exact same data and method as were used to produce the observational estimate. The horizontal dashed red line shows the central observational estimate, while the pale red shaded band shows the observational 95% confidence interval.

about 0.4°C of which has been offset by aerosol cooling, resulting in an overall observed warming over land of 1.1°C (SI Figure S4). Our study reinforces the findings in Storelvmo et al. (2016), which concluded that aerosol loading has masked a substantial fraction of continental warming over the past half-century. The average of temperature simulations in ESMs shows a comparable warming of 1.2°C but a different decomposition, of which greenhouse warming has driven the temperature up by 1.4°C and aerosols have cooled it down by 0.2°C (SI Figure S5). A similar temperature decomposition for CMIP6 models is also reported in Tokarska et al. (2020). We note that the aerosol cooling effect is considerably weaker in ESMs than in observations (0.2 vs. 0.4°C), which consequently requires a lower sensitivity of temperature to CO_2 in order to simulate a realistic net historical warming. Specifically, ESMs simulate weak trends in SSR and by proxy in aerosol forcing, thus less CO_2 warming would be needed to counterbalance the cooling effect, leaving more CO_2 variation to contribute to the warming, which implies a smaller sensitivity of temperature to CO_2, i.e., a smaller TCR.

Notably, observed SSR shows more than three times stronger trends than the average trend of ESMs, reporting annual trends of −0.24 vs. −0.07 Wm^{-2}/year, respectively, over 1964 –1994 (see SI Table S4 for individual ESM trends). The reporting period is chosen over the time during which the differences between observed and simulated dimming are particularly large (Figure 4 (a)), and meanwhile covering more than 30 years of duration in order to reduce the effect of internal variability. In contrast to the discrepancy in SSR trends, temperature shows a fairly good agreement between observations and ESM simulations—observed temperature generally fluctuates within the 66% uncertainty band of ESMs for most of the time (Figure 4 (b)). Recalling that given fixed trends in temperature and CO_2 equivalent concentrations, weak trends in SSR would result in a smaller TCR, we expect that a natural remedy for the divergence of ESMs from observations is to strengthen their SSR trends. To demonstrate this point, we estimate E-TCR based on a counterfactual scenario in which the empirical framework uses observed SSR and ESM simulated temperature. The results conform to our expectation that the underestimation of E-TCR relative to the observational TCR would be mitigated significantly by the reinforced SSR trends (SI Figure S6).

In addition to biasing the E-TCR values, the weak SSR trends in ESMs also lead to larger uncertainty in the estimation of E-TCR compared to that of observation-based TCR. In our empirical framework, temperature is a function of SSR and CO_2. For many of the ESMs, there is little trend in SSR, so that CO_2 carries a greater burden in explaining the trend and variation in temperature. By contrast, the observational data display a strong trend with high variability in SSR. Thus, in the observational regressions, SSR has a strong signal that helps to explain the variation in temperature much more so than in the ESMs. Overall, the result is less uncertainty associated with the impact on temperature from CO_2 which manifests in the narrower confidence interval from the observational data.

c. Implications for Climate Projections and Remaining Carbon Budgets

The implications of these findings are wide-reaching. Using statistical methods suited to the nonstationary and jointly dependent properties of the data we have shown that the CMIP6 models with higher TCR are generally more consistent with observations. The results further demonstrate that the approach used to estimate TCR from observations (E-TCR) is capable of diagnosing the true TCR when applied to synthetic observations from 22 CMIP6 ESMs. This capability reinforces the method used here to produce an observational best TCR estimate of $2.3°C$. This estimate is substantially higher than the assessed best TCR estimate from IPCC AR6 of $1.8°C$. The AR6 assessment was based on three semi-independent lines of evidence, namely process understanding, the instrumental record, and so-called emergent constraints. These three lines of evidence in isolation yielded best estimates for TCR of 2.0, 1.9 and $1.7°C$, respectively. While the former two estimates fall within our observational 95% confidence interval, the latter (based on emergent constraints) does not, and neither does the overall best TCR estimate from AR6.

The divergence of emergent constraint estimates from our observational analysis has several causes. Most importantly, the methodologies are entirely different. Emergent constraint studies usually screen and subset ESMs that are most consistent with observed temperature trends over a specified period and report TCR for the filtered sample (see e.g., Tokarska et al. 2020). One noteworthy issue is that they assume the fact that ESMs correctly reproduce observed temperature indicates the models' capability of capturing the underlying atmospheric mechanism determining temperature changes, while evidence shows otherwise. Even though ESMs unanimously underestimate SSR trends (see Figure 4 (a)), which are a main driver of temperature changes, they are still able to reproduce historical temperature trends reasonably well. In other words, ESMs are susceptible to the risk that they capture the correct temperature trends for the wrong reason, and the emergent constraint literature may overlook this possibility. Many an over-warming model is readily discarded by emergent constraints, whereas in the current study we stress that such models can in fact generate a TCR that is more consistent with observations when other observables in addition to surface air temperature are considered. The rationale is that their over warming trends compensate for the bias from the underestimation of SSR trends, such that they end up with a TCR more consistent with observations. Secondly, emergent constraint studies often apply a shorter time period than the time frame used in the current study, which may lead to year-to-year variability (noise) dominating over long-term trends.

The higher observational TCR in turn implies a substantial downward revision of how much additional burning of coal, gas and oil is allowable without considerable risk of exceeding $1.5°C$

of warming relative to pre-industrial times, as most previous calculations have assumed a TCR that is well below the observation-based estimate presented here (see e.g., Millar et al. 2017).

Using the distribution of observation-based TCR of 2.3±0.4°C, convoluted with other uncertainties in the remaining carbon budget (Matthews et al. 2021), leads to a remaining carbon budget to 1.5°C of 360 (245-470) $GtCO_2$ (median and 33-67% range) from 2020, or around nine years of current CO_2 emissions (Friedlingstein et al. 2020). Reported CMIP6 TCR values provide a remaining carbon budget of 405 (275-535) $GtCO_2$ from 2020, hence the revised TCR results in a median reduction in the remaining carbon budget of approximately one year of allowable CO_2 emissions. This reduction can be compared with a recent assessment of TCR from other lines of evidence (Sherwood et al. 2020) that results in a remaining carbon budget of 450 (305-590) $GtCO_2$, or approximately two more years' allowable CO_2 emissions for a 50% chance of remaining below 1.5°C compared to the observational TCR estimate. The narrower distribution and higher central value of observational TCR compared to other estimates also reduce the uncertainty in the remaining carbon budget (Figure 5), and one effect of this is to reduce the probability that larger values of cumulative emissions are consistent with a 1.5°C carbon budget. These estimates can be compared to the process-based estimate of 440 (230-670) $GtCO_2$ using the TCRE distribution in Matthews et al. (2021). The remaining carbon budget estimates presented from the TCR assessments here have less spread than the range presented in Matthews et al. (2021), which is likely a consequence of the relatively small spread in the airborne fraction distribution.

4. Discussion

Using the econometric framework in Phillips et al. (2020), this study provides an update on the observation-based TCR estimate over an extended time period from 1964 to 2014. Our empirical estimation reveals a higher observational TCR with narrowed uncertainty of 2.3±0.4°C (95% confidence interval). Compared with ESM reported TCRs in CMIP6, half of the ESMs report TCR falling within the observational range. Among the other ESMs with TCR falling outside the range, we notice a prominent tendency toward underestimation, which could be attributable to their too weak simulated trends and variability of surface solar radiation and by proxy aerosol cooling—less CO_2 needed to counteract aerosol cooling and more CO_2 left for explaining the warming effect, and thereby a smaller sensitivity of temperature to CO_2. We therefore suggest that it is imperative for ESMs to adjust for their underestimation of surface solar radiation trends and variability in order to better reproduce observations and provide more reliable guidance in climate projections and climate policy decisions.

The observational approach has several caveats to bear in mind. First, it might not sufficiently take account of internal variability due to the limited temporal coverage of observational data. Unlike ESMs usually with climate simulations covering hundreds of years, surface solar radiation observations are not available until recent decades. One of the issues of a short history is that in the short term climate might diverge temporarily from the long-term equilibrium, and these deviations might result in the TCR estimate varying based on the choice of time period. To examine how the TCR estimate responds to alternative time periods, we estimate TCR based on an extended time period for an additional five years; the results show a very similar estimate with reduced uncertainty (2.29±0.3$°C$ vs. 2.31±0.4$°C$ for periods ending in 2019 and 2014, respectively, see SI Figure S7). The central TCR estimate is fairly stable as we extend the estimation period, which proves its applicability over different times. Second, the observational analysis is limited to land areas and needs to convert to global TCR using a conversion procedure based on the land-ocean warming ratio. However, there is evidence indicating stronger aerosol cooling over ocean than land (see e.g., Christensen et al. 2016), which might indicate a more complicated relationship between land and global TCR. We therefore evaluate the impacts of the conversion on global E-TCR for ESMs by comparing a direct estimate based on global data with a converted estimate based on land data in conjunction with the conversion. The results show comparable estimates using the two approaches and indicate that the conversion does not make a significant difference on the final estimate (SI Figure S8).

Furthermore, the econometric approach simplifies atmospheric representations and makes use of the long-run equilibrium among three climatic variables—temperature, radiation, and CO_2 equivalent concentrations. More climatic variables could be integrated to explain temporary deviations from the equilibrium, such as effects of Interdecadal Pacific Oscillation (see e.g., Fyfe et al. 2016; Su et al. 2017; Hu and Fedorov 2017). Lastly, observational data are prone to being affected by observational bias. However, such bias should not be a major concern here as it would be greatly mitigated by the spatial aggregation of the data.

ESMs are the key to climate projections and the foundation of climate change adaptation and mitigation. They importantly illuminate TCR from a geophysical understanding of climate system dynamics. Moreover, a vast variety of ESMs together with their respective ensemble members allow for wide-ranging scenarios of future climate, which is of essential importance to prepare for various social and economic consequences. However, not all ESMs are equally consistent with observations. Our paper presents an important and different perspective, based on a novel econometric approach that is importantly independent of global climate models, and therefore well-suited for their evaluation.

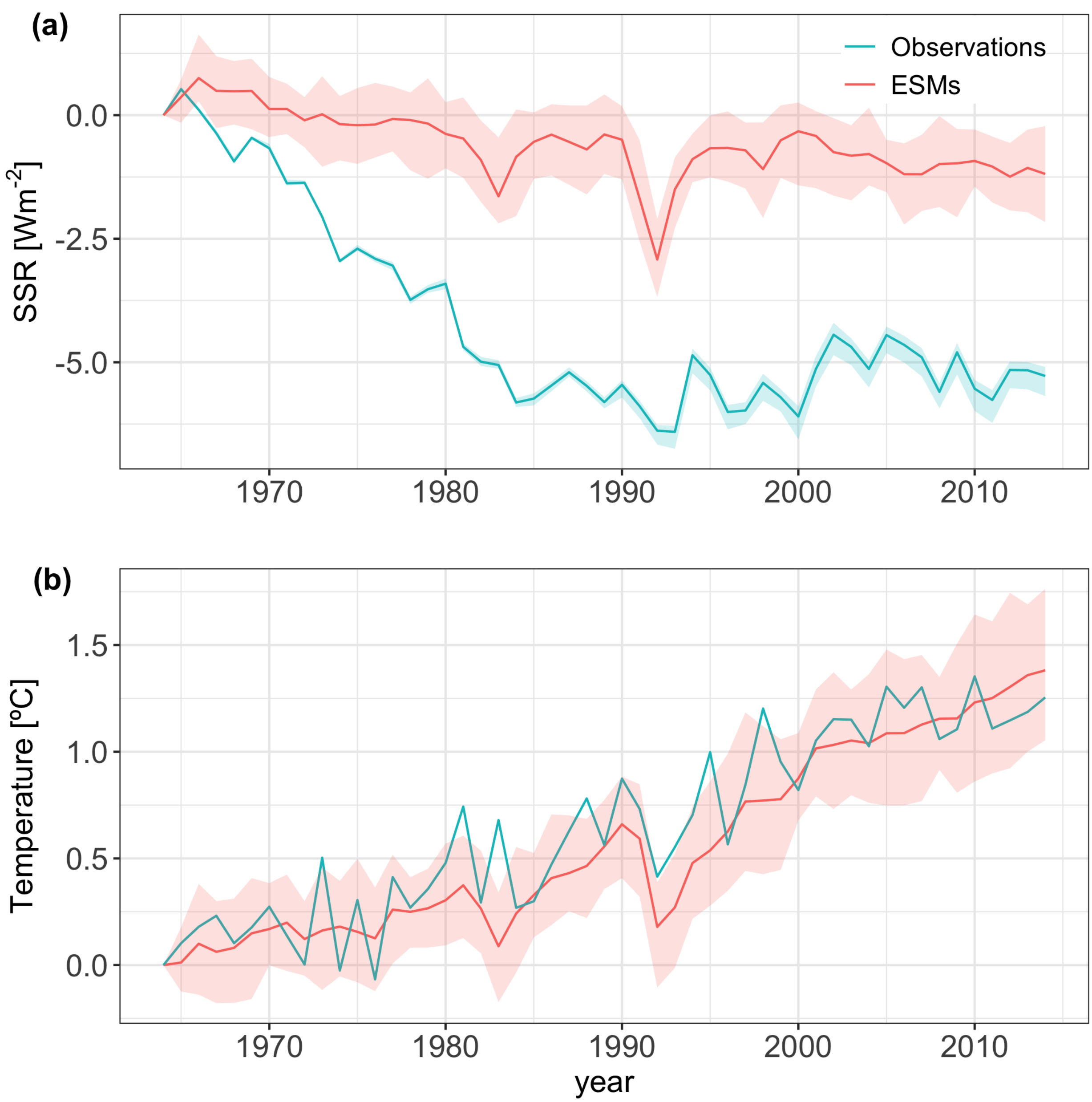

FIG. 4. ESM simulations vs. observations. (a) Global land average surface solar radiation (SSR) observations vs. ESM simulations. (b) Global land average surface air temperature observations vs. ESM simulations. Observed trends are shown in the blue line, ESM average trends are shown in the red line. The shading area for ESMs shows the likely range (17 to 83% percentile) of the ESM simulations. The shading area for SSR observations shows the added ± 5% uncertainty band relative to the average accounting for measurement accuracy limitations (Wild et al. 2017). Temperature observations are from the CRU data set with only one realization provided.

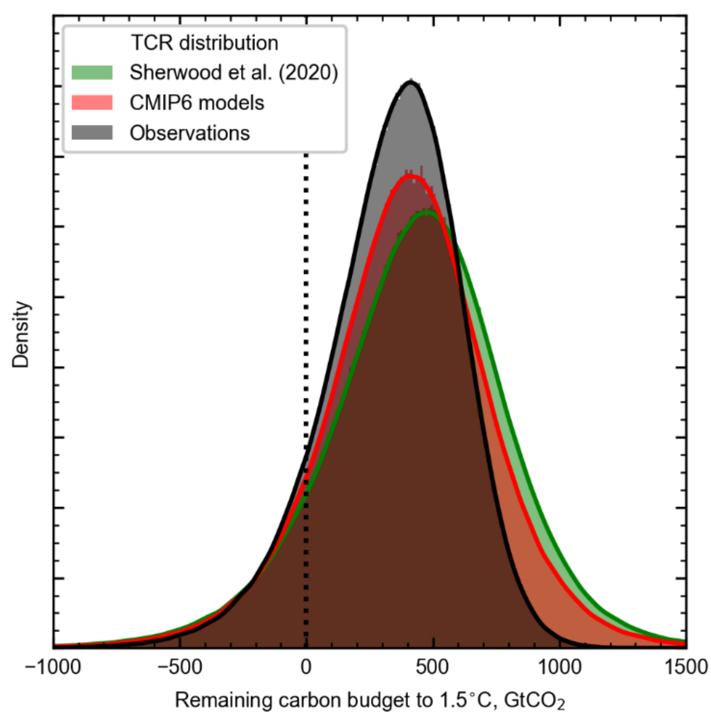

Fig. 5. Remaining carbon budget to 1.5°C using the distribution of TCR from observations (black), reported TCR values from CMIP6 models (red), and the distribution of TCR from the assessment of Sherwood et al. (2020).

Acknowledgments. This research was funded by Norwegian Research Council (grant No. 281071), under the project of "Climate Change Modelling and Prediction of Economic Impact". T.S. also acknowledges funding from the European Union's Horizon 2020 research and innovation programme project FORCeS through grant agreement No. 821205. P.C.B.P. acknowledges support from the NSF under Grant No. SES 18-50860 and the Kelly Fund at the University of Auckland. C.S. was supported by a NERC/IIASA Collaborative Research Fellowship (NE/T009381/1). We acknowledge the climate modeling groups in CMIP6 for producing and making available their model output.

Data availability statement. The data sets generated during and/or analysed during the current study are available from the corresponding author on request.

References

Arora, V. K., and Coauthors, 2020: Carbon-concentration and carbon-climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, **17 (16)**, 4173–4222, https://doi.org/10.5194/bg-17-4173-2020.

Christensen, M. W., Y. C. Chen, and G. L. Stephens, 2016: Aerosol indirect effect dictated by liquid clouds. *Journal of Geophysical Research: Atmospheres*, **121 (24)**, 14,636–14,650, https://doi.org/10.1002/2016JD025245, URL https://onlinelibrary.wiley.com/doi/full/10.1002/2016JD025245https://onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025245https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2016JD025245.

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. *Geoscientific Model Development*, **9 (5)**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, URL https://gmd.copernicus.org/articles/9/1937/2016/.

Forster, P., and Coauthors, 2021: The earth's energy budget, climate feedbacks, and climate sensitivity. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, O. Yeleçki,

R. Yu, and B. Zhou, Eds., Cambridge University Press, 923–1054, https://doi.org/10.1017/9781009157896.009.

Friedlingstein, P., and Coauthors, 2020: Global Carbon Budget 2020. *Earth System Science Data*, **12 (4)**, 3269–3340, https://doi.org/10.5194/essd-12-3269-2020.

Fyfe, J. C., and Coauthors, 2016: Making sense of the early-2000s warming slowdown. *Nature Climate Change 2016 6:3*, **6 (3)**, 224–228, https://doi.org/10.1038/nclimate2938, URL https://www.nature.com/articles/nclimate2938.

Gidden, M. J., and Coauthors, 2019: Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development*, **12 (4)**, 1443–1475, https://doi.org/10.5194/gmd-12-1443-2019, URL https://gmd.copernicus.org/articles/12/1443/2019/.

Grose, M. R., J. Gregory, R. Colman, and T. Andrews, 2018: What Climate Sensitivity Index Is Most Useful for Projections? *Geophysical Research Letters*, https://doi.org/10.1002/2017GL075742.

Harris, I., T. J. Osborn, P. Jones, and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, **7 (1)**, https://doi.org/10.1038/s41597-020-0453-3, URL https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9.

Hoesly, R. M., and Coauthors, 2018: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, **11 (1)**, 369–408, https://doi.org/10.5194/gmd-11-369-2018, URL https://gmd.copernicus.org/articles/11/369/2018/.

Hope, C., 2015: The $10 trillion value of better information about the transient climate response. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **373 (2054)**, 20140 429, https://doi.org/10.1098/rsta.2014.0429, URL https://royalsocietypublishing.org/doi/10.1098/rsta.2014.0429.

Hu, S., and A. V. Fedorov, 2017: The extreme El Niño of 2015–2016 and the end of global warming hiatus. *Geophysical Research Letters*, **44 (8)**, 3816–3824, https://doi.org/10.1002/2017GL072908, URL https://onlinelibrary.wiley.com/doi/full/10.1002/2017GL072908https://onlinelibrary.wiley.com/doi/abs/10.1002/2017GL072908https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GL072908.

Huusko, L. L., F. A. Bender, A. M. Ekman, and T. Storelvmo, 2021: Climate sensitivity indices and their relation with projected temperature change in CMIP6 models. *Environmental Research Letters*, **16 (6)**, 064 095, https://doi.org/10.1088/1748-9326/AC0748, URL https://iopscience.iop.org/article/10.1088/1748-9326/ac0748https://iopscience.iop.org/article/10.1088/1748-9326/ac0748/meta.

IPCC, 2001: Third Assessment Report, Climate Change 2001: The Scientific Basis. *Climate Change 2001: The Scientific Basis*.

Jiménez-de-la Cuesta, D., and T. Mauritsen, 2019: Emergent constraints on Earth's transient and equilibrium response to doubled CO2 from post-1970s global warming. *Nature Geoscience*, **12 (11)**, 902–905, https://doi.org/10.1038/s41561-019-0463-y, URL https://doi.org/10.1038/s41561-019-0463-y.

Jones, C. D., and P. Friedlingstein, 2020: Quantifying process-level uncertainty contributions to TCRE and carbon budgets for meeting Paris Agreement climate targets. *Environmental Research Letters*, **15 (7)**, 074 019, https://doi.org/10.1088/1748-9326/ab858a, URL https://doi.org/10.1088/1748-9326/ab858a.

Kudo, R., A. Uchiyama, O. Ijima, N. Ohkawara, and S. Ohta, 2012: Aerosol impact on the brightening in Japan. *Journal of Geophysical Research: Atmospheres*, **117 (D7)**, 7208, https://doi.org/10.1029/2011JD017158, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2011JD017158https://onlinelibrary.wiley.com/doi/abs/10.1029/2011JD017158https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011JD017158.

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

Matthews, H., and Coauthors, 2021: An integrated approach to quantifying uncertainties in the remaining carbon budget. *Communications Earth & Environment*, **2 (1)**, 1–11, https://doi.org/10.1038/s43247-020-00064-9, URL https://doi.org/10.1038/s43247-020-00064-9.

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

Meinshausen, M., and Coauthors, 2017: Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, **10 (5)**, 2057–2116, https://doi.org/10.5194/gmd-10-2057-2017.

Millar, R. J., and Coauthors, 2017: Emission budgets and pathways consistent with limiting warming to 1.5 °C. *Nature Geoscience*, **10 (10)**, 741–747, https://doi.org/10.1038/NGEO3031, URL www.nature.com/naturegeoscience.

Myhre, G., O. Boucher, F. M. Bréon, P. Forster, and D. Shindell, 2015: Declining uncertainty in transient climate response as CO2 forcing dominates future climate change. *Nature Geoscience*, **8 (3)**, 181–185, https://doi.org/10.1038/ngeo2371, URL www.nature.com/naturegeoscience.

Nijsse, F. J., P. M. Cox, and M. S. Williamson, 2020: Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, **11 (3)**, 737–750, https://doi.org/10.5194/ESD-11-737-2020.

Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. Nature Publishing Group, URL www.nature.com/naturegeoscience, 415–416 pp., https://doi.org/10.1038/ngeo1836.

Persad, G. G., and K. Caldeira, 2018: Divergent global-scale temperature effects from identical aerosols emitted in different regions. *Nature Communications*, **9 (1)**, 1–9, https://doi.org/10.1038/s41467-018-05838-6, URL www.nature.com/naturecommunications.

Phillips, P. C., T. Leirvik, and T. Storelvmo, 2020: Econometric estimates of Earth's transient climate sensitivity. *Journal of Econometrics*, **214 (1)**, 6–32, https://doi.org/10.1016/j.jeconom.2019.05.002.

Phillips, P. C. B., and B. E. Hansen, 1990: Statistical Inference in Instrumental Variables Regression with I(1) Processes. *The Review of Economic Studies*, **57 (1)**, 99, https://doi.org/10.2307/2297545, URL https://academic.oup.com/restud/article-lookup/doi/10.2307/2297545.

Regayre, L. A., and Coauthors, 2014: Uncertainty in the magnitude of aerosol-cloud radiative forcing over recent decades. *Geophysical Research Letters*, **41 (24)**, 9040–9049, https://doi.org/10.1002/2014GL062029, URL http://doi.wiley.com/10.1002/2014GL062029.

Rogelj, J., and Coauthors, 2018: Scenarios towards limiting global mean temperature increase below 1.5 °C. *Nature Climate Change*, **8 (4)**, 325–332, https://doi.org/10.1038/s41558-018-0091-3, URL https://www.nature.com/articles/s41558-018-0091-3.

Ruckstuhl, C., and J. R. Norris, 2009: How do aerosol histories affect solar "dimming" and "brightening" over Europe?: IPCC-AR4 models versus observations. *Journal of Geophysical Research: Atmospheres*, **114 (D10)**, https://doi.org/10.1029/2008JD011066, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2008JD011066https://onlinelibrary.wiley.com/doi/abs/10.1029/2008JD011066https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008JD011066.

Sherwood, S. C., and Coauthors, 2020: An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence. *Reviews of Geophysics*, **58 (4)**, e2019RG000 678, https://doi.org/10.1029/2019RG000678, URL https://onlinelibrary.wiley.com/doi/10.1029/2019RG000678.

Shindell, D., and C. J. Smith, 2019: Climate and air-quality benefits of a realistic phase-out of fossil fuels. *Nature*, **573 (7774)**, 408–411, https://doi.org/10.1038/s41586-019-1554-z, URL https://doi.org/10.1038/s41586-019-1554-z.

Shindell, D. T., G. Faluvegi, L. Rotstayn, and G. Milly, 2015: Spatial patterns of radiative forcing and surface temperature response. *Journal of Geophysical Research: Atmospheres*, **120 (11)**, 5385–5403, https://doi.org/10.1002/2014JD022752, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/2014JD022752.

Storelvmo, T., T. Leirvik, U. Lohmann, P. C. B. Phillips, and M. Wild, 2016: Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity. *Nature Geoscience*, **9 (4)**, 286–289, https://doi.org/10.1038/ngeo2670, URL http://www.nature.com/articles/ngeo2670.

Su, J., R. Zhang, and H. Wang, 2017: Consecutive record-breaking high temperatures marked the handover from hiatus to accelerated warming. *Scientific Reports 2017 7:1*, **7 (1)**, 1–9, https://doi.org/10.1038/srep43735, URL https://www.nature.com/articles/srep43735.

Tebaldi, C., and Coauthors, 2020: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. *Earth System Dynamics Discussions*, 1–50, https://doi.org/10.5194/esd-2020-68.

Tokarska, K. B., M. B. Stolpe, S. Sippel, E. M. Fischer, C. J. Smith, F. Lehner, and R. Knutti, 2020: Past warming trend constrains future warming in CMIP6 models. *Science Advances*, **6 (12)**, https://doi.org/10.1126/sciadv.aaz9549.

Wandji Nyamsi, W., A. Lipponen, A. Sanchez-Lorenzo, M. Wild, and A. Arola, 2020: A hybrid method for reconstructing the historical evolution of aerosol optical depth from sunshine duration measurements. *Atmospheric Measurement Techniques*, **13 (6)**, 3061–3079, https://doi.org/10.5194/AMT-13-3061-2020.

Wild, M., A. Ohmura, C. Schär, G. Müller, D. Folini, M. Schwarz, M. Zyta Hakuba, and A. Sanchez-Lorenzo, 2017: The Global Energy Balance Archive (GEBA) version 2017: A database for worldwide measured surface energy fluxes. *Earth System Science Data*, **9 (2)**, 601–613, https://doi.org/10.5194/essd-9-601-2017, URL http://www.geba.ethz.ch.supplementarydataareavailableathttps//doi.org/10.1594/PANGAEA.873078.

Wild, M., S. Wacker, S. Yang, and A. Sanchez-Lorenzo, 2021: Evidence for Clear-Sky Dimming and Brightening in Central Europe. *Geophysical Research Letters*, **48 (6)**, e2020GL092 216, https://doi.org/10.1029/2020GL092216, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2020GL092216https://onlinelibrary.wiley.com/doi/abs/10.1029/2020GL092216https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL092216.

Yuan, M., T. Leirvik, and M. Wild, 2021: Global trends in downward surface solar radiation from spatial interpolated ground observations during 1961-2019. *Journal of Climate*,

-1, 1–56, https://doi.org/10.1175/JCLI-D-21-0165.1, URL https://journals.ametsoc.org/view/journals/clim/aop/JCLI-D-21-0165.1/JCLI-D-21-0165.1.xml.

Supplementary Information for High-sensitivity Earth System Models Most Consistent with Observations

Menghan Yuan*[a], Thomas Leirvik[b,c,d], Trude Storelvmo[b,e], Kari Alterskjær[b,f], Peter C.B. Phillips[g,h,i,j], and Christopher J. Smith[k,l]

[a] Nuffield College, University of Oxford, United Kingdom
[b] Nord University, Norway
[c] The Arctic University of Norway, Norway
[d] The Norwegian University of Science and Technology, Norway
[e] University of Oslo, Norway
[f] Center for International Climate and Environmental Research (Cicero), Norway
[g] University of Auckland, New Zealand
[h] Yale University, United States of America
[i] Singapore Management University, Singapore
[j] University of Southampton, United Kingdom
[k] University of Leeds, United Kingdom
[l] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

*Corresponding author: menghan.yuan@nuffield.ox.ac.uk

LIST OF TABLES

Table S1 Correlation coefficients of surface solar radiation (SSR) and aerosol forcing (AER)

Model	**Corr coef**	**Pval**	**Pval.symbol**[a]
CanESM5	0.038	0.793	
CNRM-CM6-1	0.714	0.000	***
GFDL-ESM4	0.467	0.001	***
GISS-E2-1-G	0.763	0.000	***
HadGEM3-GC31-LL	0.319	0.023	*
IPSL-CM6A-LR	0.755	0.000	***
MIROC6	0.790	0.000	***
NorESM2-LM	0.588	0.000	***
UKESM1-0-LL	0.215	0.129	

[a] Significance symbol representation: *** indicates $p < 0.001$, ** for $p < 0.01$, * for $p \leq 0.05$, . for $p \leq 0.1$, and no symbol if $p > 0.1$.
[b] Aerosol forcing data source: Smith et al. (2021).

Table S2 Mean, standard deviation, minimum and maximum for the annual change in global average temperature. Unit: $^{\circ}C$ per year.

Model	**Mean**	**St. dev**	**Min.**	**Max.**
observation	0.025	0.248	-0.529	0.500
BCC-CSM2-MR	0.028	0.240	-0.547	0.489
BCC-ESM1	0.022	0.146	-0.366	0.265
CAMS-CSM1-0	0.013	0.211	-0.423	0.582
CanESM5	0.046	0.214	-0.452	0.507
CESM2	0.026	0.220	-0.588	0.433
CESM2-WACCM	0.024	0.215	-0.467	0.574
CNRM-CM6-1	0.021	0.245	-0.375	0.603
CNRM-ESM2-1	0.024	0.216	-0.652	0.467
E3SM-1-0	0.041	0.183	-0.379	0.546
EC-Earth3-Veg	0.037	0.231	-0.589	0.480
GISS-E2-1-G	0.028	0.304	-0.466	0.720
GISS-E2-1-H	0.032	0.235	-0.703	0.442
HadGEM3-GC31-LL	0.036	0.201	-0.419	0.459
IPSL-CM6A-LR	0.028	0.233	-0.494	0.653
MIROC-ES2L	0.028	0.317	-0.942	0.579
MIROC6	0.030	0.293	-0.624	0.727
MPI-ESM-1-2-HR	0.018	0.203	-0.589	0.384
MRI-ESM2	0.024	0.175	-0.462	0.597
NESM3	0.021	0.228	-0.567	0.568
NorESM2-LM	0.034	0.188	-0.462	0.373
SAM0-UNICON	0.024	0.190	-0.454	0.430
UKESM1-0-LL	0.040	0.184	-0.454	0.393
Summary of ESMs				
Mean	0.028	0.221	-0.522	0.512
Min.	0.013	0.146	-0.942	0.265
Max.	0.046	0.317	-0.366	0.727

Table S3 Mean, standard deviation, minimum and maximum for the annual change in global average surface solar radiation. Unit: Wm^{-2} per year.

Model	Mean	St. dev	Min.	Max.
observation	-0.110	0.588	-0.979	1.492
BCC-CSM2-MR	-0.011	0.906	-1.867	2.258
BCC-ESM1	-0.002	0.610	-1.596	1.243
CAMS-CSM1-0	0.000	1.119	-4.079	2.923
CanESM5	-0.062	0.959	-2.233	2.008
CESM2	-0.018	0.972	-2.012	1.856
CESM2-WACCM	-0.045	1.040	-3.348	2.365
CNRM-CM6-1	-0.021	0.643	-1.916	1.446
CNRM-ESM2-1	-0.018	0.553	-1.153	1.354
E3SM-1-0	-0.023	0.684	-1.953	1.606
EC-Earth3-Veg	-0.042	0.946	-2.073	2.123
GISS-E2-1-G	-0.066	1.095	-3.331	2.398
GISS-E2-1-H	-0.041	1.317	-2.576	3.061
HadGEM3-GC31-LL	0.026	0.678	-2.044	1.426
IPSL-CM6A-LR	-0.024	0.780	-1.942	1.596
MIROC-ES2L	-0.006	0.684	-2.364	1.765
MIROC6	-0.016	0.711	-1.615	1.398
MPI-ESM-1-2-HR	-0.013	0.768	-1.770	1.771
MRI-ESM2	-0.050	0.727	-1.798	1.596
NESM3	-0.024	0.877	-2.046	2.787
NorESM2-LM	-0.021	0.794	-1.334	2.241
SAM0-UNICON	-0.040	0.504	-1.283	0.956
UKESM1-0-LL	0.002	0.760	-2.400	1.489
Summary of ESMs				
Mean	-0.023	0.824	-2.124	1.894
Min.	-0.066	0.504	-4.079	0.956
Max.	0.026	1.317	-1.153	3.061

Table S4 Annual radiation trends over the global dimming period 1964–1994.

Model	Slope[a]	Slope std	t value	Pval	Pval.symbol[b]
observation	-0.240	0.013	-18.314	0.000	***
BCC-CSM2-MR	-0.059	0.019	-3.076	0.005	**
BCC-ESM1	-0.043	0.011	-3.900	0.001	***
CAMS-CSM1-0	-0.055	0.019	-2.928	0.007	**
CanESM5	-0.098	0.015	-6.326	0.000	***
CESM2	-0.064	0.015	-4.356	0.000	***
CESM2-WACCM	-0.061	0.016	-3.912	0.001	***
CNRM-CM6-1	-0.064	0.012	-5.493	0.000	***
CNRM-ESM2-1	-0.057	0.011	-5.303	0.000	***
E3SM-1-0	-0.061	0.013	-4.521	0.000	***
EC-Earth3-Veg	-0.087	0.013	-6.741	0.000	***
GISS-E2-1-G	-0.104	0.019	-5.392	0.000	***
GISS-E2-1-H	-0.101	0.020	-5.041	0.000	***
HadGEM3-GC31-LL	-0.059	0.012	-4.908	0.000	***
IPSL-CM6A-LR	-0.066	0.012	-5.425	0.000	***
MIROC-ES2L	-0.059	0.013	-4.639	0.000	***
MIROC6	-0.053	0.011	-5.011	0.000	***
MPI-ESM-1-2-HR	-0.072	0.013	-5.711	0.000	***
MRI-ESM2	-0.078	0.011	-6.937	0.000	***
NESM3	-0.081	0.018	-4.420	0.000	***
NorESM2-LM	-0.050	0.012	-4.109	0.000	***
SAM0-UNICON	-0.036	0.008	-4.377	0.000	***
UKESM1-0-LL	-0.044	0.013	-3.269	0.003	**

[a] Slope unit: Wm^{-2} per year. The slope is the slope coefficient obtained from regressing SSR on a linear time trend.
[b] Significance symbol representation: *** indicates $p < 0.001$, ** for $p < 0.01$, * for $p \leq 0.05$, . for $p \leq 0.1$, and no symbol if $p > 0.1$.

Table S5 Annual temperature trends over 1984 –2014.

Model	Slope[a]	Slope std	$t$value	Pval	Pval.symbol[b]
observation	0.030	0.003	9.106	0.000	***
BCC-CSM2-MR	0.038	0.005	7.869	0.000	***
BCC-ESM1	0.032	0.003	10.726	0.000	***
CAMS-CSM1-0	0.018	0.004	4.518	0.000	***
CanESM5	0.049	0.004	12.330	0.000	***
CESM2	0.039	0.004	8.984	0.000	***
CESM2-WACCM	0.048	0.004	12.470	0.000	***
CNRM-CM6-1	0.026	0.003	7.861	0.000	***
CNRM-ESM2-1	0.031	0.003	11.152	0.000	***
E3SM-1-0	0.052	0.004	12.744	0.000	***
EC-Earth3-Veg	0.039	0.003	11.725	0.000	***
GISS-E2-1-G	0.032	0.005	7.052	0.000	***
GISS-E2-1-H	0.032	0.004	8.215	0.000	***
HadGEM3-GC31-LL	0.057	0.004	13.211	0.000	***
IPSL-CM6A-LR	0.039	0.004	8.731	0.000	***
MIROC-ES2L	0.026	0.006	4.503	0.000	***
MIROC6	0.034	0.006	6.067	0.000	***
MPI-ESM-1-2-HR	0.030	0.004	7.582	0.000	***
MRI-ESM2	0.034	0.003	10.486	0.000	***
NESM3	0.051	0.004	14.156	0.000	***
NorESM2-LM	0.041	0.004	9.762	0.000	***
SAM0-UNICON	0.038	0.004	10.195	0.000	***
UKESM1-0-LL	0.053	0.003	15.248	0.000	***

[a] Slope unit: $^{\circ}C$ per year.
[b] Significance symbol representation: refer to Table S4.

Table S6 ESM warming ratio and conversion factor. Warming over the globe ($\boldsymbol{w_G}$) is calculated using complete ESM data; warming over land ($\boldsymbol{w_L}$) is obtained by masking global ESM to retain only land areas; and warming over ocean ($\boldsymbol{w_O}$) can be obtained using the formula in the footnote[a]. $\boldsymbol{WR}$ is the land-ocean warming ratio; $\boldsymbol{W_{tran}}$ is the conversion factor transforming the land TCR to the global TCR.

Model	$\boldsymbol{w_G}$ $[°C/dec]$	$\boldsymbol{w_L}$ $[°C/dec]$	$\boldsymbol{w_O}$[a] $[°C/dec]$	$\boldsymbol{WR}$[b]	$\boldsymbol{W_{tran}}$[c]
BCC-CSM2-MR	0.14	0.21	0.12	1.83	0.68
BCC-ESM1	0.16	0.20	0.15	1.29	0.84
CAMS-CSM1-0	0.11	0.13	0.10	1.31	0.83
CanESM5	0.25	0.34	0.22	1.53	0.76
CESM2	0.21	0.29	0.17	1.69	0.71
CESM2-WACCM	0.20	0.28	0.17	1.65	0.72
CNRM-CM6-1	0.19	0.26	0.16	1.68	0.71
CNRM-ESM2-1	0.17	0.24	0.14	1.71	0.70
E3SM-1-0	0.20	0.29	0.16	1.85	0.67
EC-Earth3-Veg	0.23	0.32	0.19	1.68	0.71
GISS-E2-1-G	0.17	0.21	0.15	1.41	0.79
GISS-E2-1-H	0.22	0.27	0.20	1.36	0.81
HadGEM3-GC31-LL	0.23	0.30	0.20	1.46	0.78
IPSL-CM6A-LR	0.17	0.25	0.13	1.87	0.67
MIROC-ES2L	0.13	0.19	0.11	1.80	0.68
MIROC6	0.13	0.21	0.10	2.02	0.64
MPI-ESM-1-2-HR	0.14	0.17	0.13	1.35	0.82
MRI-ESM2	0.16	0.23	0.13	1.77	0.69
NESM3	0.17	0.24	0.14	1.67	0.71
NorESM2-LM	0.18	0.26	0.15	1.78	0.69
SAM0-UNICON	0.16	0.23	0.14	1.68	0.71
UKESM1-0-LL	0.26	0.33	0.23	1.47	0.77
ESM Mean	**0.18**	**0.25**	**0.15**	**1.63**	**0.73**
ESM St. Dev.	**0.04**	**0.05**	**0.04**	**0.20**	**0.06**

[a] $w_O = (w_G - w_l \cdot A_L)/A_O$

[b] (4)=(2)/(3)

[c] (5)=$A_L + A_O$/(4).

Table S7 E–TCR, reported TCR and their respective 95% confidence interval. Unit: $^{\circ}C$.

Model	E–TCR	Estimated 95% CI	Reported TCR	Reported 95% CI
Observation	2.31	(1.96, 2.68)	-	-
BCC-CSM2-MR	1.62	(1.24, 2.03)	1.36	(1.23, 1.45)
BCC-ESM1	2.09	(1.51, 2.70)	1.77	(1.65, 1.85)
CAMS-CSM1-0	1.22	(0.82, 1.64)	1.73	(1.63, 1.82)
CanESM5	3.21	(2.51, 3.94)	2.73	(2.54, 2.83)
CESM2	2.42	(1.95, 2.92)	2.00	(1.89, 2.07)
CESM2-WACCM	2.48	(1.79, 3.21)	1.93	(1.79, 2.03)
CNRM-CM6-1	2.02	(1.64, 2.42)	2.23	(2.08, 2.35)
CNRM-ESM2-1	1.93	(1.28, 2.61)	1.83	(1.71, 1.91)
E3SM-1-0	2.54	(1.67, 3.46)	2.90	(2.76, 2.99)
EC-Earth3-Veg	2.28	(1.76, 2.82)	2.65	(2.48, 2.75)
GISS-E2-1-G	1.93	(1.24, 2.66)	1.73	(1.60, 1.85)
GISS-E2-1-H	2.48	(1.99, 2.99)	1.86	(1.71, 1.95)
HadGEM3-GC31-LL	2.86	(1.98, 3.78)	2.48	(2.33, 2.60)
IPSL-CM6A-LR	2.13	(1.56, 2.72)	2.41	(2.28, 2.50)
MIROC-ES2L	1.61	(1.21, 2.03)	1.48	(1.34, 1.57)
MIROC6	2.07	(1.53, 2.64)	1.56	(1.46, 1.65)
MPI-ESM-1-2-HR	1.69	(1.21, 2.20)	1.63	(1.53, 1.70)
MRI-ESM2	2.06	(1.66, 2.47)	1.67	(1.56, 1.73)
NESM3	2.40	(1.97, 2.84)	2.72	(2.55, 2.84)
NorESM2-LM	2.13	(1.63, 2.66)	1.50	(1.39, 1.59)
SAM0-UNICON	1.24	(0.42, 2.10)	2.21	(2.09, 2.30)
UKESM1-0-LL	3.15	(2.64, 3.69)	2.79	(2.60, 2.91)

LIST OF FIGURES

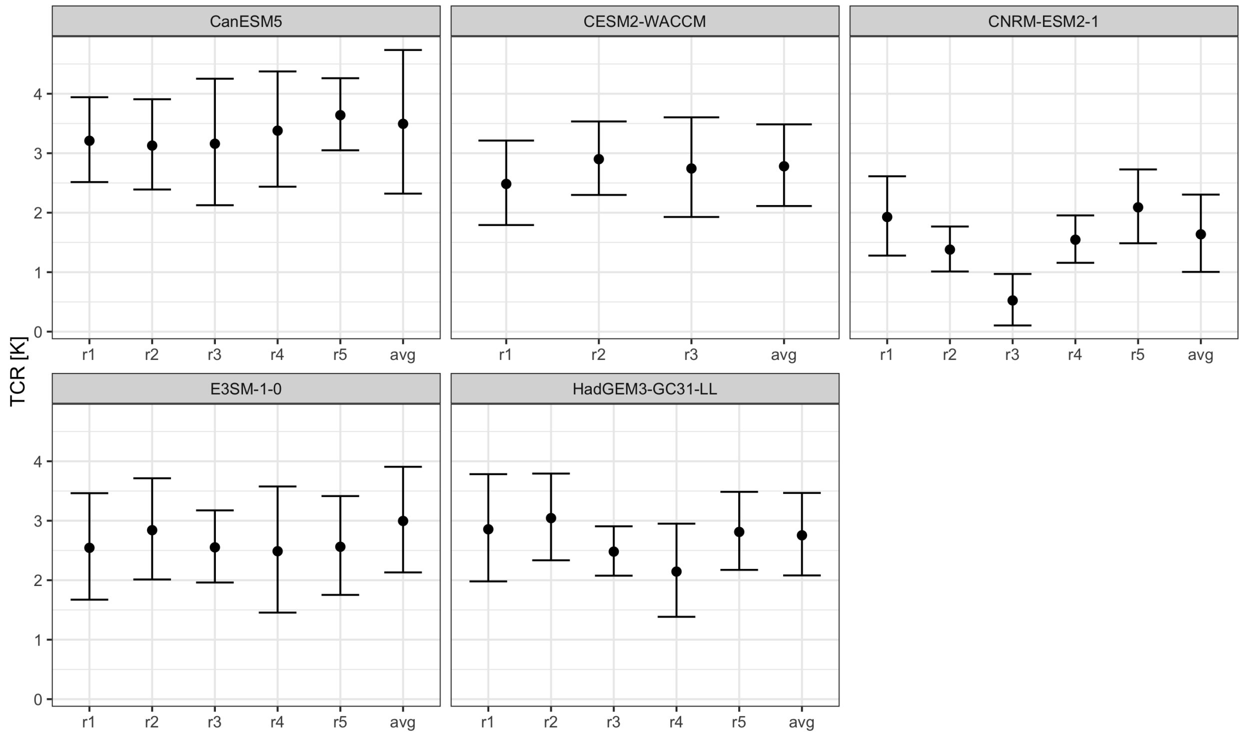

Figure S1 TCR estimates based on ensemble members. For each climate model, different scenarios are presented: E–TCR using separate ensemble members (r1-r5), and the average of ensemble members (avg). Note that there are only three realizations available for CESM2-WACCM while the other models have five. The 'r3' realization of CNRM-ESM2-1 has a particularly low estimate because of its weak temperature trends away from others (see Figure S2).

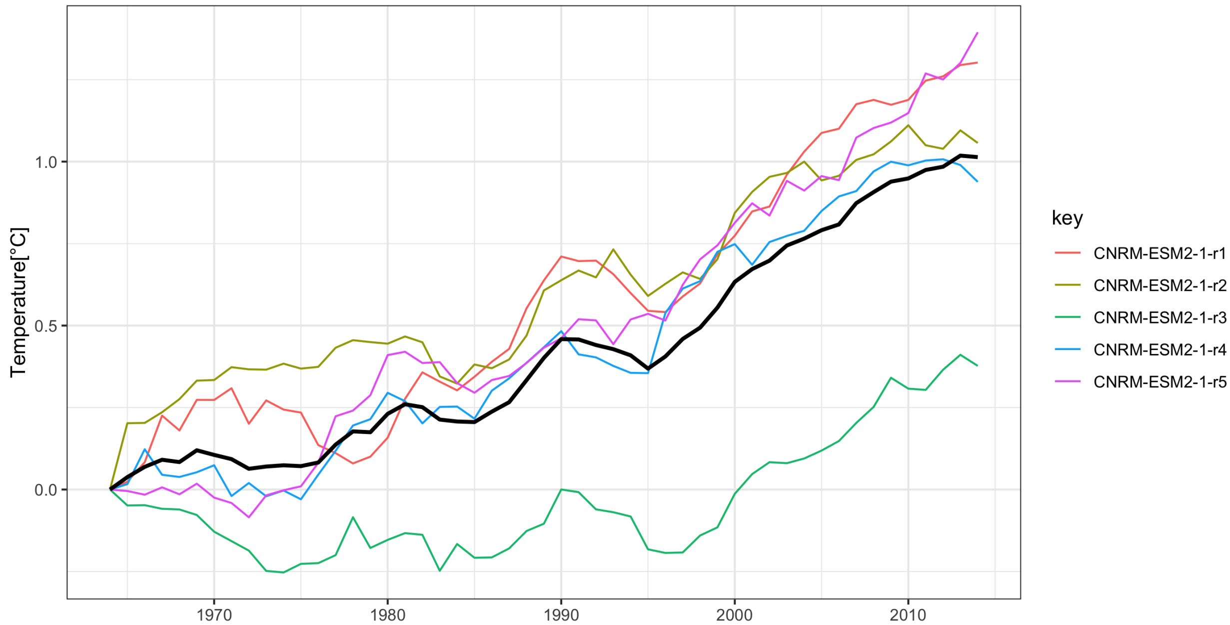

Figure S2 Temperature trends from ensemble members of CNRM-ESM2-1. The colorful lines represent individual ensemble members; the black line is the average of all ensemble members. It is noteworthy that the 'r3' realization has a particularly weak trend laying well below other ensemble members and the average trend.

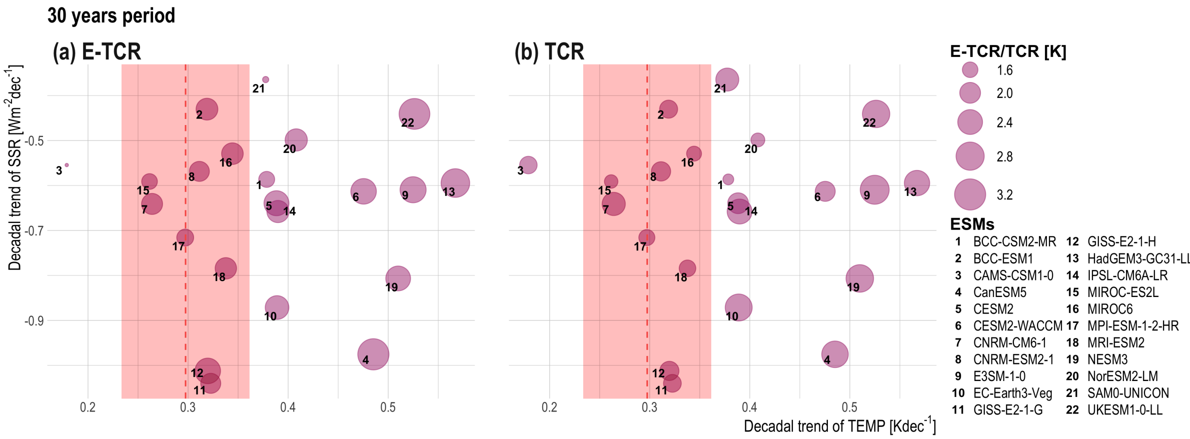

Figure S3 Relationship between E–TCR/TCR and climate trends in ESMs. X-axes show temperature trends over 1984–2014 (intensified warming period) and y-axes show surface solar radiation trends over 1964–1994 (dimming period). The corresponding periods are chosen over the time during which temperature and SSR show prominent trends while ensuring at least 30 years of duration to reduce the effect of internal variability. Refer to Table S4 and Table S5 for the trends in detail. The point size indicates the values of E–TCR (panel a) and reported TCR (panel b). The vertical dashed lines show the central estimate of the decadal trends of observed temperature; the shading areas show the 95% confidence interval. Given the large difference between observed and ESM simulated SSR trends we did not add observational constraint band for SSR, otherwise it will add a horizontal band way below the range of ESM SSR trends and distort the height of the figure.

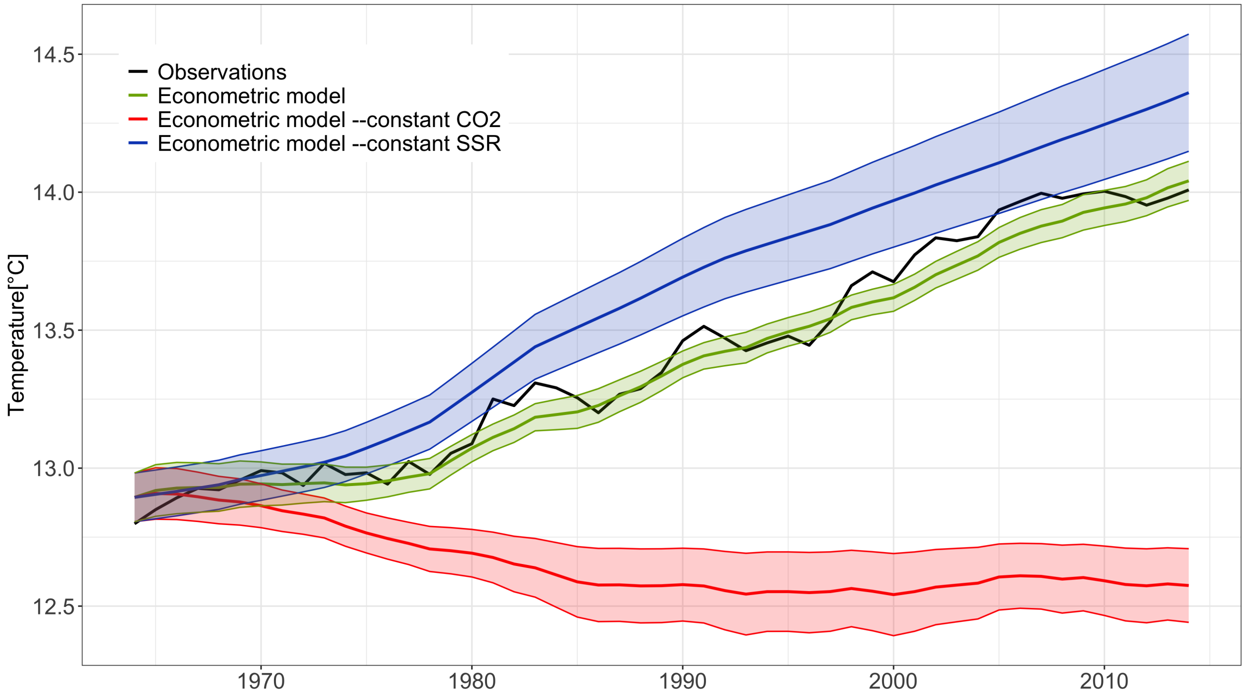

Figure S4 Observed temperature decomposition. Observation is shown in the green line; econometric model prediction is shown in the black line. Also shown is predicted temperature under the scenario of constant CO_2 levels at 1964 (red line), such that any changes in temperature are attributable to surface solar radiation variability. Likewise, the constant surface solar radiation scenario is shown in the blue line, such that trends in temperature are determined by changes in CO_2. Shadings represent 95% confidence intervals for econometric model predictions. All series are shown as 5-year running means.

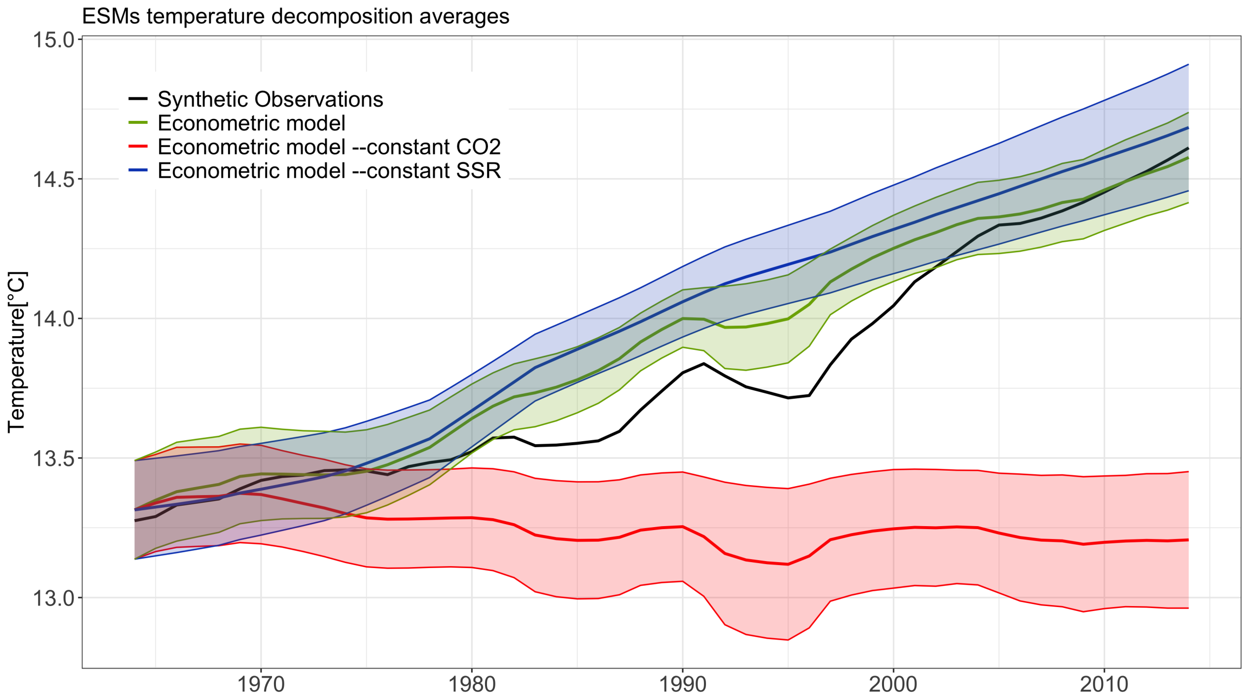

Figure S5 Average temperature decomposition for ESMs. This figure shows the average of temperature decomposition for 22 ESMs. Refer to Figure S4 for legend definitions.

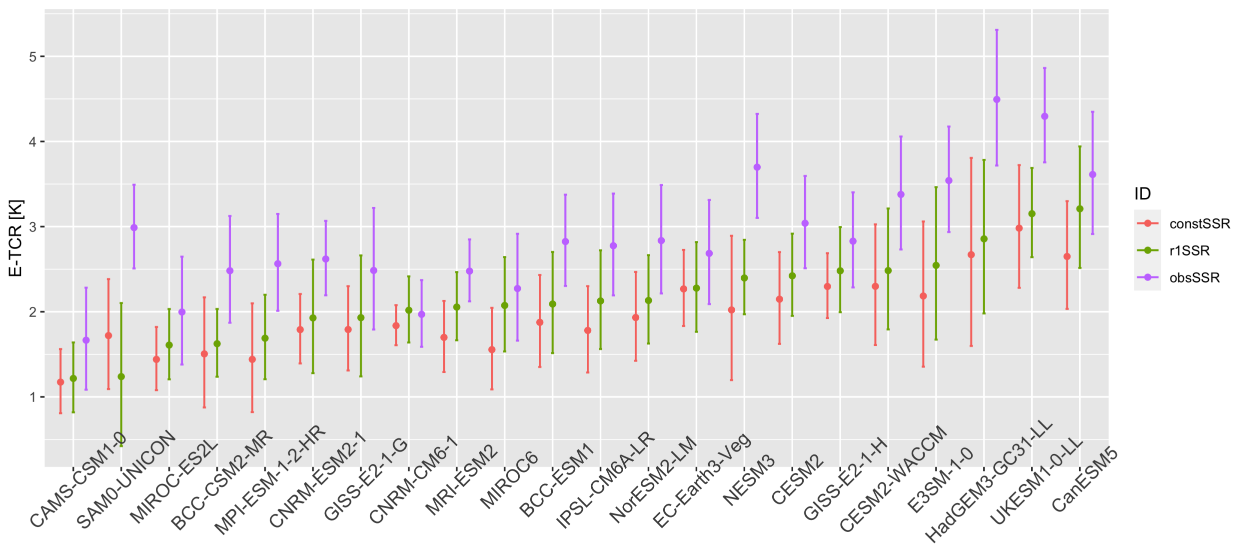

Figure S6 E–TCR under alternative scenarios for radiation data. The green series ('r1SSR') use ESM simulated radiation and provide a baseline for the changes of E–TCR under the other two alternative scenarios. The coral series ('constSSR') shows the E–TCRs estimated under constant radiation. The purple series ('obsSSR') shows the E–TCRs estimated by replacing ESM simulated radiation with observed radiation.

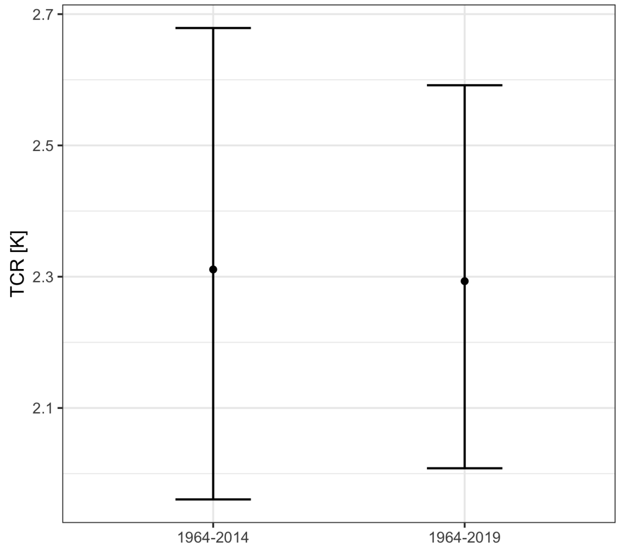

Figure S7 Observational TCR over the original (1964-2014) and the extended time period (1964-2019). The dots show the central TCR estimates; the error bars show the 95% confidence interval.

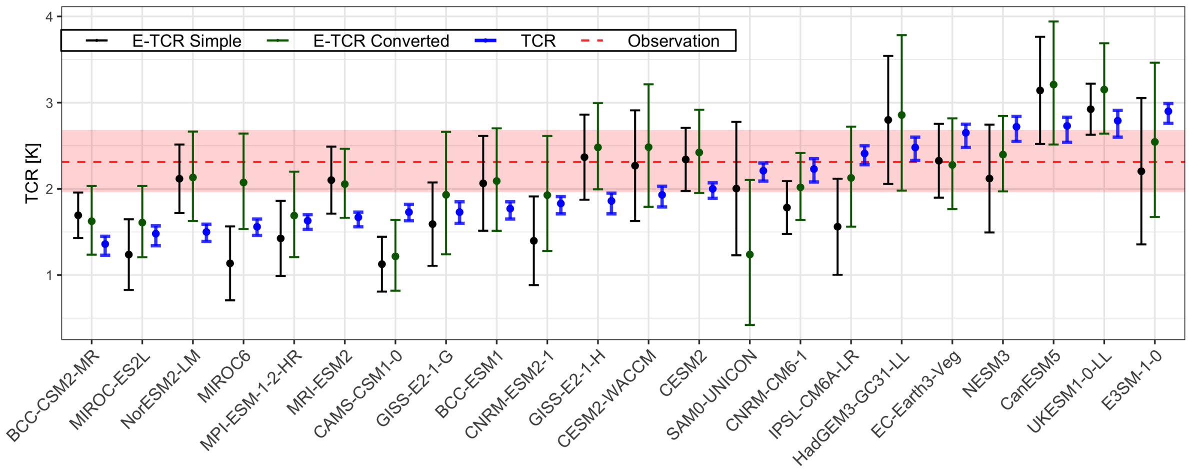

Figure S8 TCR estimates using global (black series) and land datasets (green series), as well as reported TCR (blue series) from ESMs. Error bars show the 95% confidence intervals.The horizontal dashed red line shows the central observational estimate, while the pale red shaded band shows the observational 95% confidence interval.

References

C. J. Smith, G. R. Harris, M. D. Palmer, N. Bellouin, W. Collins, G. Myhre, M. Schulz, J. C. Golaz, M. Ringer, T. Storelvmo, and P. M. Forster. Energy Budget Constraints on the Time History of Aerosol Forcing and Climate Sensitivity. *Journal of Geophysical Research: Atmospheres*, 126(13): e2020JD033622, jul 2021. ISSN 2169-8996. doi: 10.1029/2020JD033622.