

ENVIRONMENTAL RESEARCH
LETTERS

LETTER





Assessing the robustness and implications of econometric estimates of climate sensitivity

OPEN ACCESS

RECEIVED
17 October 2024REVISED
10 January 2025ACCEPTED FOR PUBLICATION
20 January 2025PUBLISHED
4 February 2025

Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Trude Storelvmo^{1,2,*} , Menghan Yuan², Thomas Leirvik^{2,3} , Kari Alterskjær⁴, Peter C B Phillips^{5,6,7}  and Chris Smith^{8,9} ¹ University of Oslo, Oslo, Norway² Nord University, Bodø, Norway³ The Arctic University of Norway, Tromsø, Norway⁴ Center for International Climate and Environmental Research (Cicero), Oslo, Norway⁵ University of Auckland, Auckland, New Zealand⁶ Yale University, New Haven, CT, United States of America⁷ Singapore Management University, Singapore, Singapore⁸ University of Leeds, Leeds, United Kingdom⁹ International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

* Author to whom any correspondence should be addressed.

E-mail: trude.storelvmo@geo.uio.no**Keywords:** climate sensitivity, climate econometrics, remaining carbon budget, climate changeSupplementary material for this article is available [online](#)

Abstract

Earth's transient climate response (TCR) quantifies the global mean surface air temperature change due to a doubling of atmospheric CO₂ concentration after 70 years of a compounding 1% per year increase. TCR is highly correlated with near-term climate projections, and thus of relevance for climate policy, but remains poorly constrained in part due to uncertainties in the representation of key physical processes in Earth System Models (ESMs). Within state-of-the-art ESMs participating in the Coupled Model Intercomparison Project (CMIP6), the TCR range (1.1 °C–2.9 °C) is too wide to offer useful guidance to policymakers. Similarly, the sixth report of the Intergovernmental Panel on Climate Change, while not solely reliant on ESMs for its TCR assessment, produced a very likely range of 1.2 °C–2.4 °C. To complement earlier, ESM-based, estimates, we here present a new TCR estimate of 2.17 (1.72–2.77) °C (95% confidence interval), derived based on a statistical relationship between surface air temperature and observational proxies for its main drivers, i.e. changes in atmospheric greenhouse gases and aerosols. We show that, within uncertainty, this method correctly diagnoses TCR from 20 CMIP6 ESMs if the same input variables are taken from the ESMs that are available from observations. This increases confidence in the new observation-based central estimate and range, which is respectively higher and narrower than the mean and spread of the estimates from the entire ensemble of CMIP6. Many ESM-based estimates tend to produce TCRs lower than the observational range reported here. Our findings suggest that a misrepresentation of the aerosol cooling effect could be the cause of this discrepancy. Further, the revised TCR estimate suggests a downward revision of the remaining carbon budgets aligned with the overarching goal of the Paris agreement.

1. Introduction

The question of how sensitive Earth's climate is to atmospheric greenhouse gas perturbations has been long-standing in the climate research community and of mounting concern in society at large. Research over several decades has pursued several different

types of evidence to constrain standard metrics of climate sensitivity, like the Transient Climate Response (TCR). However, many estimates continue to, directly or indirectly, depend on Earth System Models (ESMs), which rely on simplified representations of a wide range of small-scale processes of relevance for forcing- and feedback mechanisms in the climate

system, resulting in a large spread in simulated climate sensitivity (Forster *et al* 2021). This uncertainty, in turn, translates into highly uncertain climate projections for a given future emission-scenario (Lee *et al* 2021, Tebaldi *et al* 2021), with 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, 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 very likely TCR range to be 1.2 °C–2.4 °C with a central estimate of 1.8 °C (Forster *et al* 2021). Multiple CMIP6 models produce TCR values well above the upper end of this range (Meehl *et al* 2020), raising questions about their plausibility.

This 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. This method makes use of a statistical relationship among surface air temperature, surface solar radiation (SSR), and greenhouse gas concentrations to estimate TCR empirically. An important innovation of the approach is that it uses an observational proxy, SSR, for the cooling effect of aerosols, in order to isolate the observed surface air temperature change attributable to atmospheric greenhouse gas changes, thus allowing for TCR inference. We present tests of the robustness of the statistical method by applying it to 'synthetic climate data' in the form of output from CMIP6 models, for which the true TCR value is known. Finally, we discuss implications of the observation-based TCR estimate for the remaining carbon budgets (RCBs) in line with the goals of the Paris agreement.

2. Data and methods

2.1. Data

Data used in this study come from both observations and ESMs. Observed surface air temperature data are available from the Climate Research Unit gridded Time Series (CRU TS V4) (Harris *et al* 2020). Observed SSR data are obtained from a spatially interpolated data set based on the Global Energy Balance Archive (GEBA, Wild *et al* 2017, Yuan *et al* 2021). GEBA provides the most extensive SSR observations available, with measurements from ~2500 stations worldwide. The quality of the data has been rigorously controlled and the random measurement error has been estimated to be 5% for monthly and 2% for annual mean values per station on average (Gilgen

et al 1998). In spite of GEBA's unparalleled spatial and temporal coverage, station data tend to cluster in developed regions, while only sparse data exist in rural regions. In order to obtain a balanced SSR data set, we apply the machine learning method of Yuan *et al* (2021) with minor modifications, to interpolate SSR values in locations with missing data using data from neighbouring locations and a range of climate variables. The CRU and SSR data sets provide complete gridded observations over land at 0.5° × 0.5° resolution. We aggregate these values into global time series for the purpose of the econometric analysis. The term 'SSR' hereinafter refers to land-based SSR unless specified otherwise.

Simulation counterparts to the observational data sets, 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 one realization is used here because 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 limits the study to the time period from 1964 to 2014.

SSR is in our analysis used as a proxy for aerosol forcing primarily, but also for other transient/minor forcings operating in the solar part of the spectrum (e.g. volcanic forcing and ozone forcing). Aerosols absorb and scatter sunlight and also affect the radiative properties of clouds (Forster *et al* 2021), and are viewed as the main cause of the so-called 'dimming' and 'brightening' periods observed in SSR decadal trends (see e.g. Ruckstuhl and Norris 2009, Kudo *et al* 2012, Wandji Nyamsi *et al* 2020, Wild *et al* 2021).

Table 1 reports annual trends of global average land temperature and SSR over two sub-periods representative of 'global dimming' (1964–1994) and 'global brightening' (1984–2014) for observations and the ESM average. As expected, observed temperature shows accelerated warming, with the trend of the later period about twice that of the earlier period. The ESMs show trends generally in line with observations, yet with slightly stronger warming in the later period. However, the ESMs generally show poor skill in capturing the SSR trends, with the average trend only about one quarter of the observed trend for the 1964–1994 period. This is a feature that has been discussed also in previous literature (see e.g. Moseid *et al* 2020).

Our source of global CO₂ equivalent concentrations is the estimated greenhouse gas longwave effective radiative forcing (LW ERF) from the Indicators of Global Climate Change 2023 (Forster *et al* 2024). We estimate LW greenhouse gas ERF to be the sum of

Table 1. Annual trends of global land average temperature and SSR for observation and the ensemble mean of 22 ESMs, obtained by regressing temperature or SSR on a linear time trend. Two sub-periods are chosen to account for the transition from global dimming to brightening and to ensure more than 30 years of duration to reduce the effect of internal variability. Individual ESMs trends are available in SI tables S1 and S2.

Temperature ($^{\circ}\text{C yr}^{-1}$)						
	Period	Slope	Slope std	<i>t</i> value	<i>P</i> val	<i>P</i> val.symbol ^a
Observation	1964–1994	0.017	0.002	8.384	<.001	***
	1984–2014	0.030	0.003	9.106	<.001	***
ESM mean	1964–1994	0.015	0.002	7.045	<.001	***
	1984–2014	0.038	0.002	16.153	<.001	***
SSR ($\text{Wm}^{-2} \text{ yr}^{-1}$)						
	Period	Slope	Slope std	<i>t</i> value	<i>P</i> val	<i>P</i> val.symbol ^a
Observation	1964–1994	−0.240	0.013	−18.314	<.001	***
	1984–2014	0.020	0.011	1.874	.071	.
ESM mean	1964–1994	−0.066	0.009	−6.940	<.001	***
	1984–2014	−0.009	0.010	−0.897	.377	

^a *** indicates $p \leq 0.001$, ** for $p \leq 0.01$, * for $p \leq 0.05$, . for $p \leq 0.1$, and no symbol if $p > 0.1$.

CO_2 , CH_4 , N_2O , all halogenated gases, stratospheric water vapour from CH_4 oxidation, and 50% of the ozone ERF. We use 50% of ozone ERF to estimate the longwave component, as ozone is also a significant shortwave (SW) absorber, and the SW and LW contributions to ERF are similar in ESMs (Skeie *et al* 2020). CO_2 is the dominant contributor to LW ERF, so modelling based on CO_2 equivalent concentrations is appropriate (Jenkins *et al* 2018). The time series from Forster *et al* (2024) is an update of the IPCC AR6 time series to include post-2019 data. The same CO_2 equivalent concentration data are applied for both observations and ESMs.

When empirically estimating TCR for the CMIP6 models, we use the reported TCR, regarded as the ‘true’ TCR for each model, 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. Confidence levels were found by bootstrapping the mean difference between the two 20 year segments with 10 000 realizations.

2.2. Econometric framework

TCR is in this study estimated using an empirical econometric framework. Econometric methodology is statistical in nature, but avoids some of the biases involved in standard statistical methods and can assist in validating inference, both of which are important in empirical work, whether it involves climate or economic data. Here, econometric methods are used to relate global average surface air temperature in year $t + 1$ (\bar{T}_{t+1}) to previous year’s temperature (\bar{T}_t),

global average SSR (\bar{R}_t), and the logarithm of CO_2 equivalent concentrations ($\text{CO}_{2e,t}$):

$$\bar{T}_{t+1} = \gamma_0 + \gamma_1 \bar{T}_t + \gamma_2 \bar{R}_t + \gamma_3 \ln(\text{CO}_{2e,t}) + u_{t+1}. \quad (1)$$

The equation error disturbance u_{t+1} embodies variability not captured by the explanatory regressors \bar{R}_t and $\text{CO}_{2e,t}$, and the γ_i are fixed parameters. Here, γ_1 can be viewed as a measure of the global temperature inertia, which is governed first and foremost by ocean heat uptake, while γ_2 and γ_3 quantify the transient sensitivity of the surface temperature to surface SW radiation and equivalent CO_2 , respectively, and thus are largely controlled by climate feedback. As analyzed and discussed in Phillips *et al* (2020), each of the main variables in (1) is nonstationary with stochastic trend characteristics analogous to those of random walks or random walks with drift. This nonstationarity can be removed by a simple transformation of equation (1) that leads to a new equation explaining the evolution of temperature differences over time, $\Delta \bar{T}_{t+1} = \bar{T}_{t+1} - \bar{T}_t$, in terms of variables that represent deviations from an equilibrium relationship between temperature, surface radiation and greenhouse gases. The reformulation of (1) is

$$\Delta \bar{T}_{t+1} = \gamma_0 - (1 - \gamma_1) \left[\bar{T}_t - \frac{\gamma_2}{1 - \gamma_1} \bar{R}_t - \frac{\gamma_3}{1 - \gamma_1} \ln(\text{CO}_{2e,t}) \right] + u_{t+1}, \quad (2)$$

which leads to the following *cointegrating* relationship among the main variables

$$\bar{T}_t = d_0 + d_1 \bar{R}_t + d_2 \ln(\text{CO}_{2e,t}) + v_t, \quad (3)$$

with $d_0 = \frac{\gamma_0}{1 - \gamma_1}$, $d_1 = \frac{\gamma_2}{1 - \gamma_1}$, $d_2 = \frac{\gamma_3}{1 - \gamma_1}$ and residual v_t . The linear relationship (3) is described as *cointegrating* because it captures the long term (multi-decadal)

relationship among the variables $\{\bar{T}_t, \bar{R}_t, \text{CO}_{2t}\}$. Methods for establishing the existence of this relationship and a fully modified least squares (FM-OLS) procedure for estimating the coefficients $\{d_0, d_1, d_2\}$ in (3) were given in Phillips *et al* (2020). FM-OLS estimation of the relationship (3) and its coefficients is important because it overcomes difficulties of both bias and inference that arise from direct estimation of equation (1) and focuses attention on the key long run parameters d_1 and d_2 . The modifications involved in the FM-OLS procedure take account of endogeneity (or joint determination) of the variables $\{\bar{T}_t, \bar{R}_t, \text{CO}_{2t}\}$ and the presence of trends and serial dependence in their generating mechanism, which assists in validating inference. Readers are referred to Phillips *et al* (2020) for detailed discussion of this approach.

Checking for cointegration using the augmented Dickey–Fuller test (Dickey and Fuller 1979, 1981, Phillips and Ouliaris 1990) (table S3 and figure S2), we indeed find that the three observed variables (temperature, $\text{CO}_{2e,t}$ and SSR) are cointegrated over the time period considered. Among the 22 ESMs in our dataset, BCC-ESM1 and SAM0-UNICON failed the cointegration test at the 1% significance level. Since cointegration is a crucial assumption for the validity of our empirical framework, as outline above, we exclude these two models from the TCR estimation in the subsequent analysis. It is also worth noting that for the remaining models and the observations, the same cointegration test fails if only two of the three variables are considered, independent of which two variables are selected. In other words, the temperature evolution cannot be statistically explained by CO_{2e} or SSR alone, only by the combination of the two.

Once the parameters of equation (3) have been estimated following the method of Phillips *et al* (2020), it follows from the same equation that TCR (i.e. the temperature change due to a doubling of CO_{2e}) can be expressed as

$$\text{TCR} = d_2 \times \ln(2). \quad (4)$$

Thereafter, one additional step is required to compute a globally representative TCR value. Since our observational data cover only land areas, equation (1) yields a TCR representative for land only. This value must therefore be converted to a global TCR value. Specifically,

$$\begin{aligned} \text{TCR}_G &= \text{TCR}_L \cdot \frac{A_L \cdot w_L + A_O \cdot w_O}{w_L} = \text{TCR}_L \\ &\cdot \left(A_L + \frac{A_O}{\text{WR}} \right) = \text{TCR}_L \cdot W_{\text{trans}} \end{aligned} \quad (5)$$

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. $\frac{1}{\text{WR}} = \frac{w_O}{w_L}$ stands for the inverse of the *land–ocean warming ratio* (WR), where w_O denotes the warming rate over ocean

and w_L over land. W_{trans} denotes the conversion factor for the central estimate.

Land and ocean warming rates are calculated as the least squares estimate of the slope coefficient from regressing temperature on a linear time trend. Warming rates and ratios as well as the W_{trans} used for individual ESMs are reported in SI table S4. ESMs report WRs ranging from 1.29 to 2.02, with an average of 1.63 and standard deviation of 0.20. For the observations we adopt a WR of 1.62, which is generally supported by Gulev *et al* (2021) (their table 2.4). To account for uncertainty in the WR from the ESMs and observations, we add an uncertainty bound of ± 0.05 to $\frac{1}{\text{WR}}$. We multiply the lower bound of the confidence interval for TCR_L by W_{trans}^- and the upper bound by W_{trans}^+ given by

$$\begin{aligned} W_{\text{trans}}^- &= A_L + \frac{A_O}{\text{WR}} \cdot (1 - 0.05) \\ W_{\text{trans}}^+ &= A_L + \frac{A_O}{\text{WR}} \cdot (1 + 0.05). \end{aligned} \quad (6)$$

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.

Additionally, we imposed a $\pm 2\%$ uncertainty band to account for measurement errors in observed SSR (Gilgen *et al* 1998). We performed a bootstrap with 100 000 iterations. For each iteration, we generated a noise series that follows a normal distribution with a mean of zero and a standard deviation equal to 2% of the annual average SSR. The empirical confidence interval was then derived by taking the lower 95% CI boundary from the smallest TCR estimate and the upper 95% CI boundary from the largest TCR estimate.

In principle, since ESMs have global coverage, a direct global TCR estimate based on land and ocean data would be possible. However, in order to keep consistency in the estimation between observations and ESM simulations, we retain only the ESM output over land and convert the land estimate to the global estimate following the conversion procedure described above. A discussion of how the conversion impacts the global TCR estimate can be found in section 4.

2.3. RCB calculation

The RCB up to a particular temperature limit above pre-industrial ΔT_{lim} can be conceptualized as (Matthews *et al* 2021)

$$\text{RCB} = \frac{\Delta T_{\text{lim}}(1 - f_{\text{nc}}^*) - \Delta T_{\text{anth}}(1 - f_{\text{nc}})}{\text{TCRE}}, \quad (7)$$

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 TCRES is the TCR to cumulative emissions of CO_2 .

TCRE can be approximated as (Jones and Friedlingstein 2020)

$$\text{TCRE} = a_f \cdot \frac{\text{TCR}}{\Delta C_{2 \times \text{CO}_2}}, \quad (8)$$

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 and $\Delta C_{2 \times \text{CO}_2}$ 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) and a conversion of 1 ppm = 2.124 GtC (Friedlingstein *et al* 2020) gives $\Delta C_{2 \times \text{CO}_2} = 604$ GtC.

To generate probability distributions of the RCB to $\Delta T_{\text{lim}} = 1.5$ °C or 2 °C a 1 million member Monte Carlo ensemble was produced. TCR is sampled as gamma-distributed for reported TCR from CMIP6 models (see table SX), and as normally-distributed for the observational TCR using the mean of 2.17 °C and standard deviation 0.31 °C. For the estimate from IPCC AR6 we use a normal distribution with mean of 1.80 °C and standard deviation of 0.36 °C (Forster *et al* 2021). 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 RCB is computed by sampling the terms in equation (7) 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 – 0.33 , Rogelj *et al* 2018) and sampled using a kernel density estimate. The non- CO_2 forcing fraction at net-zero is $f_{\text{nc}}^* = 0.3081f_{\text{nc}} + 0.14 + \varepsilon$ where ε is sampled as a normal distribution (mean 0 and 5%–95% range of 0.05) that represents additional future socio-economic 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 2023 of 1.31 (1.1–1.7) °C (Forster *et al* 2024). 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 2024.

3. Results

3.1. Using SSR as a proxy for aerosol forcing

To show that SSR is a reasonable proxy for aerosol forcing (AER), we calculated correlation coefficients for 11 ESMs for which both SSR and AER data are available. AER data are from the Radiative Forcing Model Intercomparison Project (RFMIP) piclim-histaer simulations obtained from Smith *et al* (2021). An almost perfect linear correlation is found

Table 2. Correlation coefficients of surface solar radiation (SSR) and aerosol forcing (AER) over 1850–2014 (1964–2014 in parentheses). Five-year moving average was applied to SSR and AER to smooth random noise. Note that for E3SM-1-0, the correlation coefficient is calculated based on the period 1870–2014 due to data availability of AER. See SI figures S6–S16 for SSR and AER time series for individual ESMs.

Model	Corr coef.	Pval	Pval.symbol ^a
CanESM5	0.94 (0.04)	<.001	***
CNRM-CM6-1	0.93 (0.71)	<.001	***
E3SM-1-0	0.97 (0.60)	<.001	***
GFDL-ESM4	0.93 (0.47)	<.001	***
GISS-E2-1-G	0.97 (0.76)	<.001	***
HadGEM3-GC31-LL	0.97 (0.32)	<.001	***
IPSL-CM6A-LR	0.90 (0.76)	<.001	***
MIROC6	0.96 (0.79)	<.001	***
MRI-ESM2	0.97 (0.23)	<.001	***
NorESM2-LM	0.98 (0.59)	<.001	***
UKESM1-0-LL	0.95 (0.22)	<.001	***

^a Refer to table 1 for significance symbol representation.

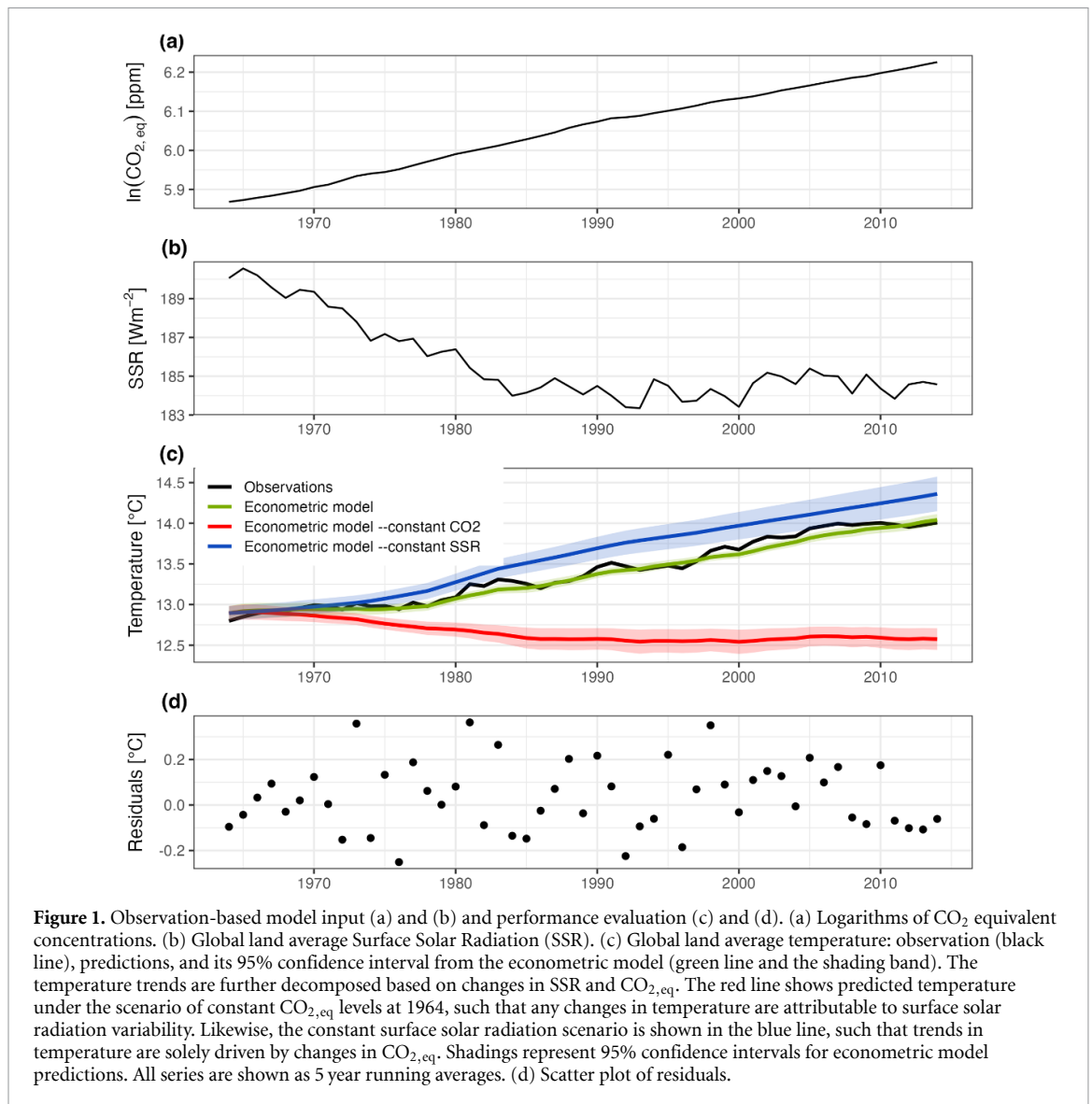
between SSR and AER for all ESMs when the entire length of the historical ESM simulations (1850–2014) is considered (all correlation coefficients > 0.9 in table 2). This is remarkably high considering that SSR captures all influence on downwelling SW radiation at the surface, including effects from volcanic eruptions and ozone. When the shorter period of 1964–2014 is considered, the influence of large episodic volcanic eruptions (e.g. El Chicon of 1984 and Mt Pinatubo in 1991), ozone variations and interannual cloud variability naturally weakens the correlation. Nevertheless, the correlation between SSR and AER remains positive for all ESMs and is statistically significant at the 95% confidence level for 8 out of 11 models with an average correlation coefficient of 0.53.

3.2. New observation-based TCR estimate

In a first application of the method, TCR was estimated to be 2.0 ± 0.8 °C (Storelvmo *et al* 2016), while in the present study updates to the observational data sets and further development of the methodology (see Phillips *et al* 2020) produce a somewhat higher estimate and a narrower uncertainty range of 2.17 (1.72–2.77) °C (95% confidence interval).

Figure 1 shows trends in the observational data sets and indicators confirming that the econometric model performs as intended. CO_2 equivalent concentrations show log-linear trends; SSR shows significant dimming until the early 1990s and stability thereafter. As expected, temperature predictions by the econometric model are well in line with long-term trends in the observations, with less year-to-year volatility. This confirms that the parameter estimation procedure is sound. The residuals fluctuate around zero, showing fair performance of the econometric model.

Next, we present evidence that the observational method can correctly diagnose TCR. This is done by comparing the standard TCR calculation from 20



CMIP6 models with the TCR values estimated when the same variables that are available from observations are extracted from the models and used in the observational analysis (hereinafter referred to as E-TCR). Numerical TCR and E-TCR values for individual models are presented in table S6.

3.3. Increasing confidence in the new TCR estimate

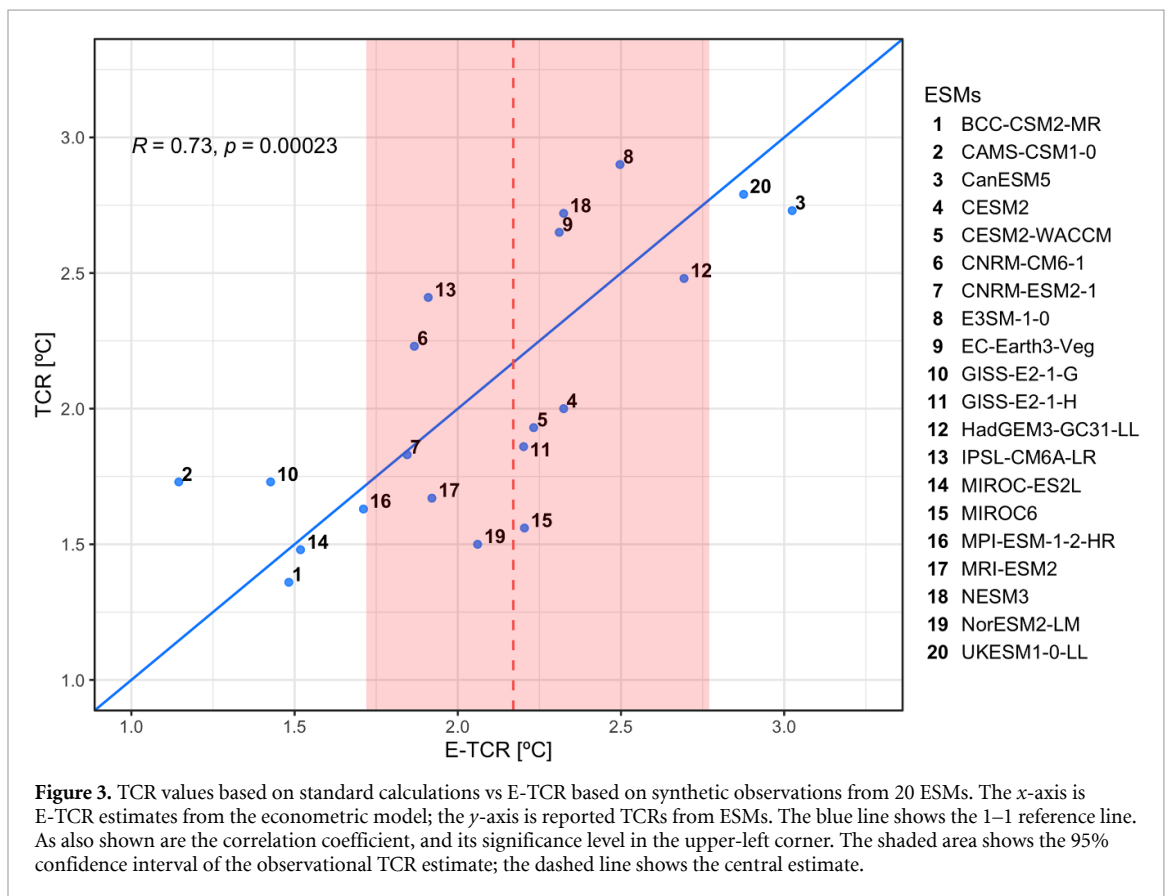
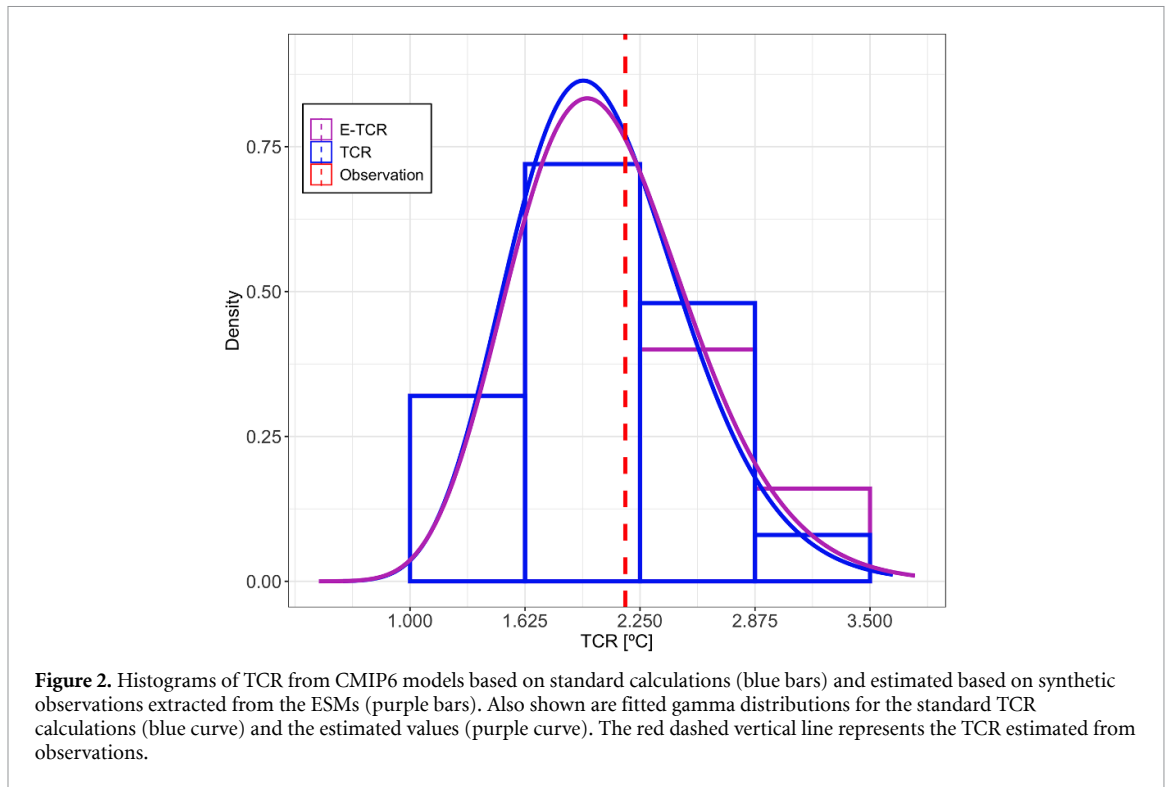
As evident from figure 2, the TCR distribution based on the standard calculation and the E-TCR emerging from the synthetic observations extracted from the ESMs are indeed very similar. Accordingly, the ensemble means for the two distributions are also near-identical, with an E-TCR mean of 2.08 °C vs TCR mean of 2.06 °C. However, this excellent agreement to some extent hides somewhat larger discrepancies for individual models, for which E-TCR sometimes underestimates and sometimes overestimates the true TCR.

Nevertheless, there is also a statistically significant positive correlation between the estimated E-TCR

values and the TCR values based on standard calculations for the ESMs ($r = 0.73$, figure 3), with low-TCR models correctly being diagnosed as such, and vice versa. While we note relatively large deviations from the 1–1 line for some models, the E-TCR confidence interval (figure 4) encompasses or hugs this line for all but one model.

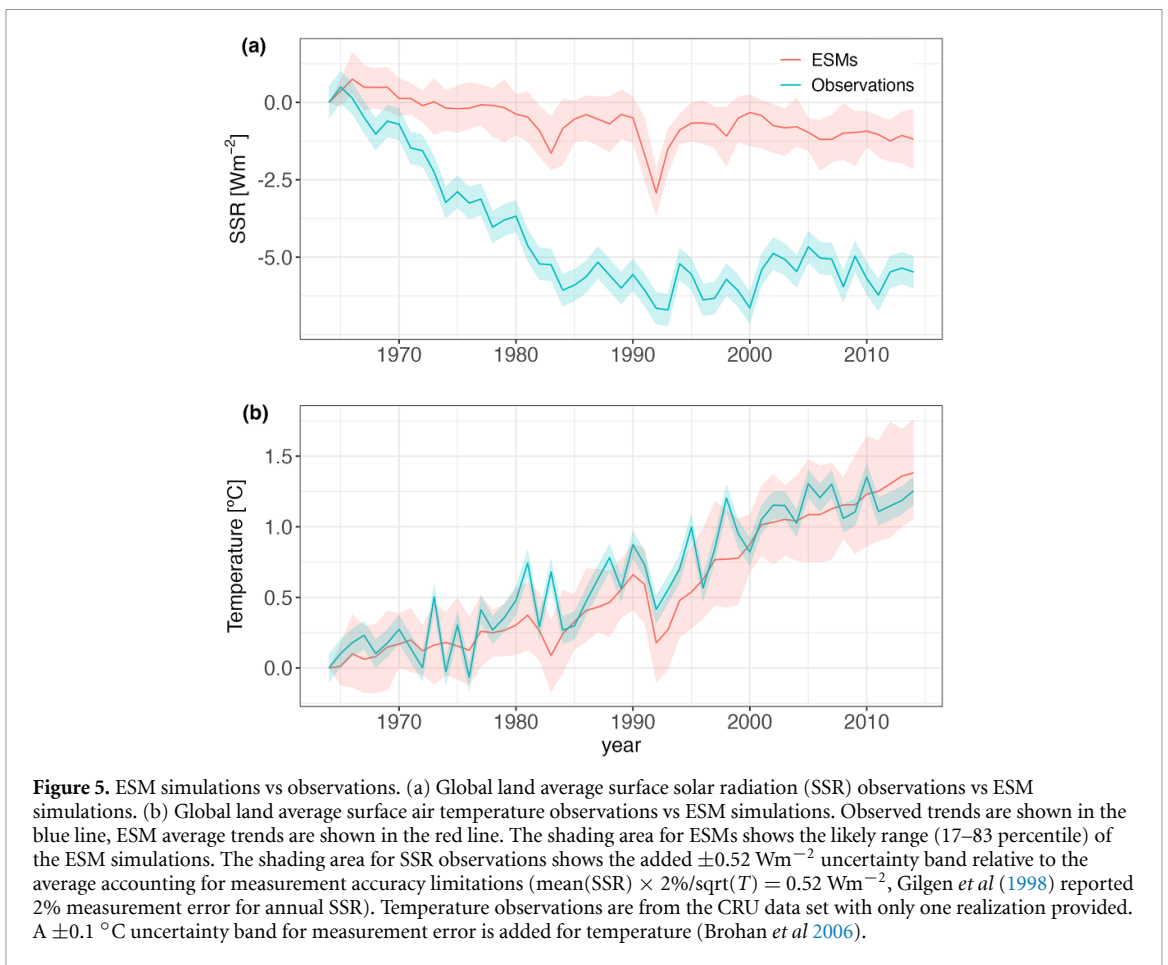
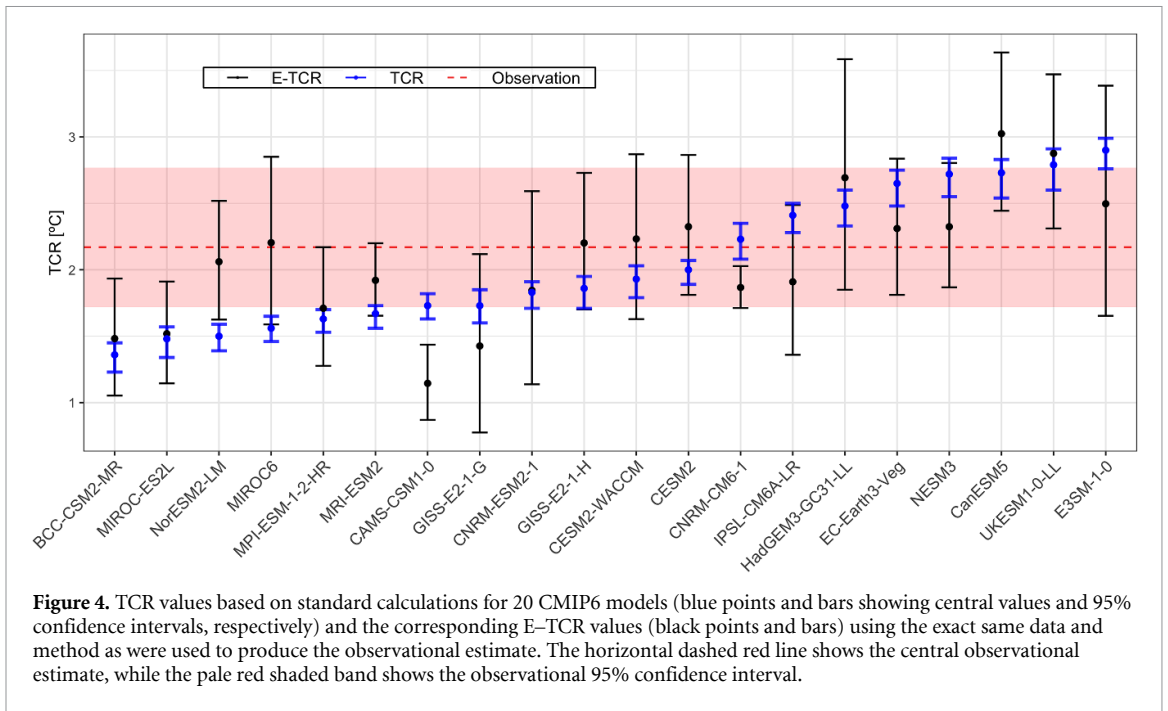
Higher E-TCR models (e.g. CanESM5) tend to show stronger simulated trends of temperature and/or SSR, whereas lower E-TCR models (e.g. CAMS-CSM 1-0) are usually associated with weaker trends. Other things being equal, i.e. fixing CO₂ and radiation variations, a stronger temperature trend means a larger response of temperature given a unit change of CO₂, that is, a larger TCR, and vice versa. Aerosol cooling effects counteract CO₂ warming, so a stronger trend in SSR requires a larger TCR in order for the same amount of CO₂ change to achieve the same rise in temperature.

Finally, figure 4 shows that while the method cannot always perfectly diagnose the true TCR, it is (with



the exception of the CAMS-CSM1-0 model) within or at the edge of the estimated E-TCR range. This strengthens confidence in the ability of the empirical TCR estimation to correctly diagnose the TCR of the real climate system, which is therefore very likely to lie

in the estimated observation-based 95% confidence interval. Notably, only four ESMs have a TCR that breaches the observational-based range, and all of them undershoot, indicating a tendency for ESMs to underestimate TCR compared to observational data.



The difference between the TCR of ESMs and that based on the observational method can likely be attributed to differences in the aerosol cooling effect; on the premise that SSR is a valid proxy for aerosol forcing, ESMs appear to underestimate

aerosol cooling compared to observations, whereas historical warming is reproduced reasonably well (figure 5). Our method also allows us to disentangle temperature change attributable to greenhouse gas warming and aerosol cooling effects. We find that

greenhouse gases have increased global land temperature by about ~ 1.5 °C over 1964–2014, about 0.4 °C of which has been offset by aerosol cooling, resulting in an overall observed warming of 1.1 °C over land areas (figure 1(c)). This reinforces previous findings (Storelvmo *et al* 2016) concluding that aerosols have masked a substantial fraction of continental warming in the latter part of the 20th century and the early 2000s. The ESM average shows a comparable warming of 1.2 °C but a different decomposition, in which greenhouse warming has increased temperature by 1.4 °C and aerosols have reduced it by only 0.2 °C (not shown). A similar temperature decomposition for CMIP6 models is also reported in Tokarska *et al* (2020). Since the aerosol cooling effect is considerably weaker in the ESMs than suggested by the observation-based method used here, ESMs require a lower TCR in order to simulate a realistic net historical warming. 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 difference in E-TCR relative to the observational TCR would be mitigated by stronger SSR trends (figure S3).

3.4. Implications for climate projections and RCBs

The implications of these findings are wide-reaching. Firstly, we 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 20 CMIP6 ESMs. This increases confidence in the method, which produces an observational best TCR estimate of 2.17 (1.72–2.77) °C (95% confidence interval). 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. Emergent constraint studies of TCR 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). The methods thus indirectly assume that ESMs that correctly reproduce observed temperature over a certain period simulate a more accurate response to CO₂. This only holds if the contributions from all other forcing agents are accurately simulated, some of which are poorly understood and constrained (Forster *et al* 2021). Some ESMs may therefore be capturing the correct temperature trend for the wrong reason, for example by compensating too weak aerosol forcing with a low climate sensitivity, as is strongly

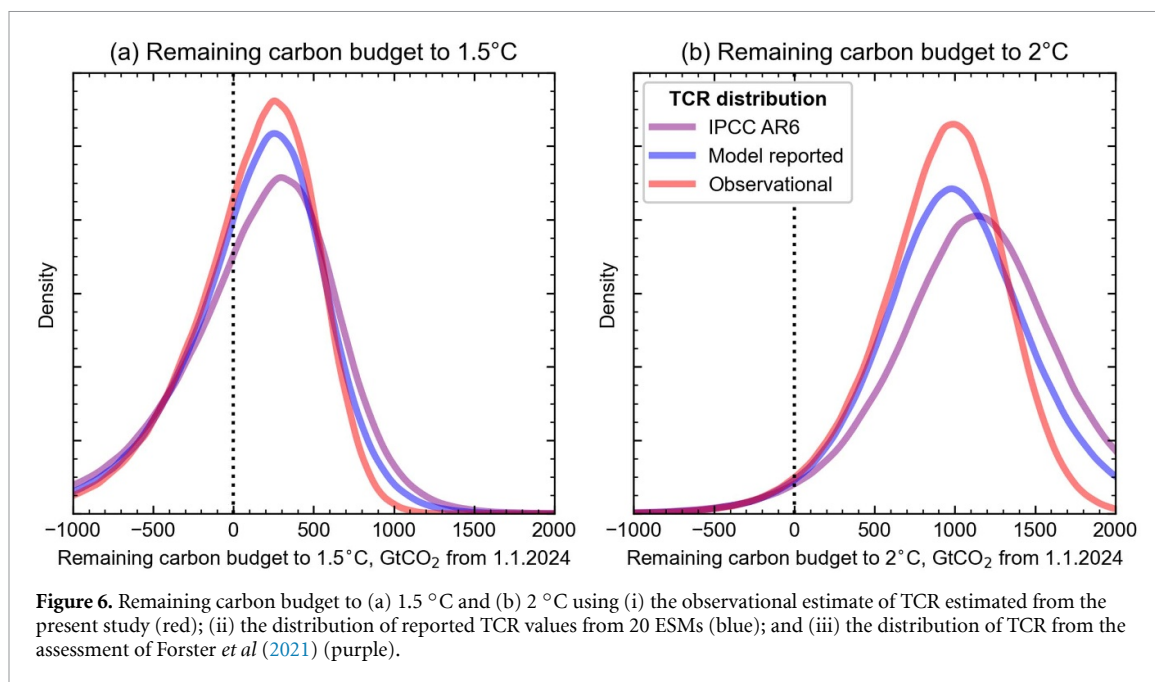
indicated by our analysis. ESMs with too strong warming trends would be discarded by emergent constraints, whereas we argue that such models may in fact generate a TCR that is more consistent with observations. While this reasoning goes against some previous literature (see e.g. Golaz *et al* 2019, Meehl *et al* 2020, Wang *et al* 2021), it is strengthened by the recent accelerated warming trend (see e.g. Hansen *et al* 2023, Hodnebrog *et al* 2024), although contributions from unforced climate variability must also be taken into account here (see e.g. Raghuraman *et al* 2024).

Finally, the higher observational TCR implies a substantial downward revision of how much additional burning of coal, gas and oil is allowable without considerable risk of exceeding 2 °C of warming relative to pre-industrial times (the 1.5 °C budget is less affected, given the limited headroom to remain under this threshold), 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 2.17 (1.72–2.77) °C (95% CI), convoluted with other uncertainties (Matthews *et al* 2021), leads to a RCB to 1.5 °C of 170 (0–325) GtCO₂ (median and 33%–67% range) from 2024, or around four years of current CO₂ emissions (Friedlingstein *et al* 2023). This can be compared with the AR6 assessment of TCR that results in a RCB of 205 (0–395) GtCO₂, or five years' emissions for a 50% chance of remaining below 1.5 °C. This central estimate corresponds well to the 1.5 °C budget in Forster *et al* (2024) of 200 GtCO₂, though with a larger spread (150–300 GtCO₂ in Forster *et al* 2024). For a 2 °C budget, the observation-based TCR gives a median and 33%–67% range of 940 (765–1105) GtCO₂ (about 23 years of current CO₂ emissions), and the AR6 distribution gives 1140 (920–1355) GtCO₂ (or 28 years of current CO₂ emissions), again broadly consistent with Forster *et al* (2024). Therefore, the median estimate for crossing 2 °C at present CO₂ emissions rates is brought forward by about five years.

4. Discussion and conclusion

As discussed in section 2.1, the observational record used to estimate TCR is relatively short (1964–2014), which increases uncertainty due to internal variability and weakens confidence in SSR as a valid proxy for aerosol forcing. These concerns could result in the TCR estimate being strongly dependent on the choice of period. To further examine this, we estimate TCR based on an extended time period that includes an additional five years (1964–2019). The results show a very similar estimate with reduced uncertainty (centered on 2.20 °C and 1.90 °C–2.56 °C for the 95% confidence interval, see figure S4). The central TCR estimate is in other words not sensitive to additional



years of observations while the confidence interval is slightly reduced with increasing data record length, as expected. Another potential caveat of this study is that the observational analysis is limited to land areas and requires conversion to global TCR using the land–ocean WR. 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 that varies over time. 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. The two approaches yield comparable E-TCR estimates, indicating that the conversion does not significantly influence the final estimate (figure S5). However, more generally, the geographical distribution of aerosol forcing may affect its global temperature influence, and this is not taken into account here. While the demonstration that our method nevertheless captures the actual TCR of the CMIP6 models relatively well indicates that using a global proxy for aerosol cooling is reasonable, it is possible that an even better agreement between TCR and E-TCR would have been achieved if a measure of the geographical distribution of aerosol forcing had been accounted for. This is beyond the scope of the present study, but could be incorporated in future work.

Finally, the present TCR estimate implicitly assumes that patterns of warming will be broadly like the historical at the time of (hypothetical) CO₂ doubling. Extensive literature has shown that if this pattern of warming changes, so too will the corresponding radiative response, and thus TCR (see e.g. Andrews *et al* 2022, Armour *et al* 2024, Forster *et al* 2024). Omission of such effects is a weakness of the present study. A pattern effect could in principle be

built into our framework by adding a term representing it in equation (1), but its inclusion would be non-trivial and goes beyond the scope of the present study.

We emphasize here the value of strictly observational estimates of TCR, but also acknowledge that ESMs remain a key tool in climate projections. However, we stress that not all ESMs are equally consistent with observations, or may be consistent with observations for the wrong reason. We therefore argue that temperature evolution alone is an insufficient metric for ESM assessment.

Data availability statement

Original observational temperature data were downloaded from the Climate Research Unit gridded Time Series (CRU TS V4, Harris *et al* 2020). Observational SSR data were obtained from Yuan *et al* (2021). ESM simulations are available via the Earth System Grid Federation (<https://esgf-ui.ceda.ac.uk/cog/search/cmip6-ceda/>). Consolidated datasets can be found at <https://doi.org/10.17632/gmdsws36sy.1> (Yuan 2024).

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.14734091> (Smith 2025).

Acknowledgments

This research was funded by Norwegian Research Council (Grant No. 281071), under the project of ‘Climate Change Modelling and Prediction of Economic Impact’. TS also acknowledges funding from the European Union’s Horizon 2020 research and innovation programme project FORCeS through Grant Agreement No. 821205. PCBP acknowledges

support from the NSF under Grant No. SES 18-50860 and the Kelly Fund at the University of Auckland. CS 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.

Author Contributions

TL and TS designed the project. PCBP proposed the econometric framework. MY performed data integration and technical analysis relevant to TCR estimation. KA provided the reported TCR and calculated its confidence intervals. CS performed analysis relating to the carbon budget. TS, MY and CS wrote the paper with contributions from all co-authors.

Conflict of interest

The authors declare no competing financial interests.

Code availability

The code used to generate the results in the current study are available at <https://doi.org/10.17632/gmdsws36sy.1>. Code to reproduce figure 6 is available from <https://doi.org/10.5281/zenodo.7948070>.

ORCID iDs

Trude Storelvmo  <https://orcid.org/0000-0002-0068-2430>

Thomas Leirvik  <https://orcid.org/0000-0002-8174-601X>

Peter C B Phillips  <https://orcid.org/0000-0003-2341-0451>

Chris Smith  <https://orcid.org/0000-0003-0599-4633>

References

- Andrews T *et al* 2022 On the effect of historical SST patterns on radiative feedback *J. Geophys. Res. Atmos.* **127** e2022JD036675
- Armour K C *et al* 2024 Sea-surface temperature pattern effects have slowed global warming and biased warming-based constraints on climate sensitivity *Proc. Natl Acad. Sci.* **121** e2312093121
- Arora V K *et al* 2020 Carbon-concentration and carbon-climate feedbacks in CMIP6 models and their comparison to CMIP5 models *Biogeosciences* **17** 4173–222
- Brohan P, Kennedy J J, Harris I, Tett S F and Jones P D 2006 Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850 *J. Geophys. Res. Atmos.* **111** 12106
- Christensen M W, Chen Y C and Stephens G L 2016 Aerosol indirect effect dictated by liquid clouds *J. Geophys. Res. Atmos.* **121** 14636–50
- Dickey D A and Fuller W A 1979 Distribution of the estimators for autoregressive time series with a unit root *J. Am. Stat. Assoc.* **74** 427
- Dickey D and Fuller W A 1981 Likelihood ratio statistics for autoregressive time series with a unit root *Econometrica* **49** 1057–72
- Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58
- Forster P M *et al* 2024 Indicators of global climate change 2023: annual update of key indicators of the state of the climate system and human influence *Earth Syst. Sci. Data* **16** 2625–58
- Forster P *et al* 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* ed V Masson-Delmotte *et al* (Cambridge University Press) ch 7, pp 923–1054
- Friedlingstein P *et al* 2020 Global carbon budget 2020 *Earth Syst. Sci. Data* **12** 3269–340
- Friedlingstein P *et al* 2023 Global carbon budget 2023 *Earth Syst. Sci. Data* **15** 5301–69
- Gilgen H, Wild M and Ohmura A 1998 Means and trends of shortwave irradiance at the surface estimated from global energy balance archive data *J. Clim.* **11** 2042–61
- Golaz J-C *et al* 2019 The DOE E3SM coupled model version 1: overview and evaluation at standard resolution *J. Adv. Model. Earth Syst.* **11** 2089–129
- Grose M R, Gregory J, Colman R and Andrews T 2018 What climate sensitivity index is most useful for projections? *Geophys. Res. Lett.* **45** 1559–66
- Gulev S *et al* 2021 Changing state of the climate system *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* ed V Masson-Delmotte *et al* (Cambridge University Press) ch 2, pp 287–422
- Hansen J E *et al* 2023 Global warming in the pipeline *Oxf. Open Clim. Change* **3** kgad008
- Harris I, Osborn T J, Jones P and Lister D 2020 Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset *Sci. Data* **7** 109
- Hodnebrog Ø *et al* 2024 Recent reductions in aerosol emissions have increased Earth's energy imbalance *Commun. Earth Environ.* **5** 166
- Huusko L L, Bender F A-M, Ekman A M L and Storelvmo T 2021 Climate sensitivity indices and their relation with projected temperature change in CMIP6 models *Environ. Res. Lett.* **16** 064095
- Jenkins S, Millar R J, Leach N and Allen M R 2018 Framing climate goals in terms of cumulative CO₂-forcing-equivalent emissions *Geophys. Res. Lett.* **45** 2795–804
- Jones C D and Friedlingstein P 2020 Quantifying process-level uncertainty contributions to TCRe and carbon budgets for meeting Paris Agreement climate targets *Environ. Res. Lett.* **15** 074019
- Kudo R, Uchiyama A, Ijima O, Ohkawara N and Ohta S 2012 Aerosol impact on the brightening in Japan *J. Geophys. Res. Atmos.* **117** 7208
- Lee J-Y *et al* 2021 Changing state of the climate system *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* ed V Masson-Delmotte *et al* (Cambridge University Press) ch 4, pp 553–672
- Matthews H, Tokarska K B, Rogelj J, Smith C J, MacDougall A H, Hausteim K, Mengis N, Sippel S, Forster P M and Knutti R 2021 An integrated approach to quantifying uncertainties in the remaining carbon budget *Commun. Earth Environ.* **2** 1–11
- Meehl G A, Senior C A, Eyring V, Flato G, Lamarque J E, Stouffer R J, Taylor K E and Schlund M 2020 Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth System Models *Sci. Adv.* **6** eaba1981

- Meinshausen M *et al* 2017 Historical greenhouse gas concentrations for climate modelling (CMIP6) *Geosci. Model Dev.* **10** 2057–116
- Millar R J, Fuglestedt J S, Friedlingstein P, Rogelj J, Grubb M J, Matthews H D, Skeie R B, Forster P M, Frame D J and Allen M R 2017 Emission budgets and pathways consistent with limiting warming to 1.5 °C *Nat. Geosci.* **10** 741–7
- Moseid K O *et al* 2020 Bias in CMIP6 models as compared to observed regional dimming and brightening *Atmos. Chem. Phys.* **20** 16023–40
- Phillips P C B, Leirvik T and Storelvmo T 2020 Econometric estimates of Earth's transient climate sensitivity *J. Econom.* **214** 6–32
- Phillips P C B and Ouliaris S 1990 Asymptotic properties of residual based tests for cointegration *Econometrica* **58** 165–93
- Raghuraman S P, Soden B, Clement A, Vecchi G, Menemenlis S and Yang W 2024 The 2023 global warming spike was driven by the El Niño–Southern Oscillation *Atmos. Chem. Phys.* **24** 11275–83
- Rogelj J *et al* 2018 Scenarios towards limiting global mean temperature increase below 1.5 °C *Nat. Clim. Change* **8** 325–32
- Ruckstuhl C and Norris J R 2009 How do aerosol histories affect solar “dimming” and “brightening” over Europe?: IPCC-AR4 models versus observations *J. Geophys. Res. Atmos.* **114** D00D04
- Skeie R B *et al* 2020 Historical total ozone radiative forcing derived from CMIP6 simulations *npj Clim. Atmos. Sci.* **3** 32
- Smith C J *et al* 2021 Energy budget constraints on the time history of aerosol forcing and climate sensitivity *J. Geophys. Res. Atmos.* **126** e2020JD033622
- Smith C 2025 Calculation of remaining carbon budget to 1.5°C and 2°C in Storelvmo *et al* (2025) (v2.0) *Zenodo*
- Storelvmo T, Leirvik T, Lohmann U, Phillips P C B and Wild M 2016 Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity *Nat. Geosci.* **9** 286–9
- Tebaldi C *et al* 2021 Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6 *Earth Syst. Dyn. Discuss.* **12** 1–50
- Tokarska K B, Stolpe M B, Sippel S, Fischer E M, Smith C J, Lehner F and Knutti R 2020 Past warming trend constrains future warming in CMIP6 models *Sci. Adv.* **6** eaaz9549
- Wandji Nyamsi W, Lipponen A, Sanchez-Lorenzo A, Wild M and Arola A 2020 A hybrid method for reconstructing the historical evolution of aerosol optical depth from sunshine duration measurements *Atmos. Meas. Tech.* **13** 3061–79
- Wang C, Soden B J, Yang W and Vecchi G A 2021 Compensation between cloud feedback and aerosol-cloud interaction in CMIP6 models *Geophys. Res. Lett.* **48** e2020GL091024
- Wild M, Ohmura A, Schär C, Müller G, Folini D, Schwarz M, Zytka M and Sanchez-Lorenzo A 2017 The Global Energy Balance Archive (GEBA) version 2017: a database for worldwide measured surface energy fluxes *Earth Syst. Sci. Data* **9** 601–13
- Wild M, Wacker S, Yang S and Sanchez-Lorenzo A 2021 Evidence for clear-sky dimming and brightening in central Europe *Geophys. Res. Lett.* **48** e2020GL092216
- Yuan M 2024 Supplementary Data: assessing the robustness and implications of econometric estimates of climate sensitivity *Mendeley Data*, V1 [10.17632/gmdswws36sy.1](https://doi.org/10.17632/gmdswws36sy.1)
- Yuan M, Leirvik T and Wild M 2021 Global trends in downward surface solar radiation from spatial interpolated ground observations during 1961–2019 *J. Clim.* **34** 1–56