Geophysical Research Letters®

Key Points:

- Our method to estimate effective radiative forcing based on common model diagnostics is consistent with Coupled Model Intercomparison Project 6 (CMIP6) fixed-SST estimates
- We present coherent estimates for abrupt CO_2, 1% CO_2, historical, and future scenario experiments
- Forcing estimates for the 21st century grow faster for CMIP6 than for CMIP5 models

21st Century Scenario Forcing Increases More for CMIP6 Than CMIP5 Models

Hege-Beate Fredriksen[1] ⬚, Christopher J. Smith[2,3] ⬚, Angshuman Modak[4,5] ⬚, and Maria Rugenstein[6] ⬚

[1]Department of Physics and Technology, UiT the Arctic University of Norway, Tromsø, Norway, [2]Priestley International Center for Climate, University of Leeds, Leeds, UK, [3]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, [4]Department of Meteorology (MISU), Stockholm University, Stockholm, Sweden, [5]Interdisciplinary Program in Climate Studies, Indian Institute of Technology Bombay, Mumbai, India, [6]Colorado State University, Fort Collins, CO, USA

Abstract Although the Coupled Model Intercomparison Project 6 (CMIP6) protocol provides an experiment to estimate effective radiative forcing (ERF), it is only quantified for few models. We present new estimates of ERF for models participating in CMIP6 by applying the method developed in Fredriksen et al. (2021, https://doi.org/10.1029/2020JD034145), and validate our approach with available fixed-SST forcing estimates. We estimate ERF for experiments with abrupt changes of CO_2, 1% increase of CO_2, historical forcings, and future scenarios, and demonstrate that CMIP6 ERF is lower than CMIP5 ERF at the end of the historical period, but grows faster than CMIP5 in the future scenarios, ending up at higher levels than CMIP5 at the end of the 21st century. The simulated radiative efficiency of CO_2 has not changed much, suggesting that the larger future increase in CO_2 concentrations in CMIP6 compared to CMIP5 is important for explaining the forcing difference.

Plain Language Summary To understand climate model responses, it is useful to separate between the drivers of climate change and their responses. We present new estimates of the drivers, called the effective radiative forcing, for the latest generation of climate models (Coupled Model Intercomparison Project 6 (CMIP6)). This estimates the energy imbalance at the top of the atmosphere and is a measure of human and natural influences on climate. Normally this requires additional climate model experiments to make these estimates, but since these have only been run for a few models, we are here aiming to make the best alternative estimates based on more widely available data, following the method in Fredriksen et al. (2021, https://doi.org/10.1029/2020JD034145). We show that our forcing estimates are growing faster during the 21st century for the new CMIP6 models than for the previous generation of models (CMIP5), and suggest this can be attributed to the higher CO_2 concentrations in future scenarios for CMIP6 compared to CMIP5.

1. Introduction

The [effective] radiative forcing (ERF) describes the energy imbalance in the Earth system that drives climate change. It is a common currency to compare the energetic impacts of different human and natural influences on the climate, and also used to develop scenarios characterizing possible futures, for example, in representative concentration pathway (RCP) and shared socioeconomic pathway (SSP) scenarios (Moss et al., 2010; O'Neill et al., 2016). However, ERF is impossible to observe and complex climate models such as the general circulation models (GCMs) developed as part of the Coupled Model Intercomparison Project Phase 6 (CMIP6) are often the best or only way to determine ERF. Accurately quantifying ERF allow us to attribute cause and effect in climate model behavior and better constrain climate sensitivity.

Unfortunately, only a small number of CMIP6 models—9 out of 51— provided estimates of ERF, for the historical period and one single scenario to 2100. These ERF estimates were derived from atmosphere-only runs of CMIP6 models using pre-industrial sea-surface temperatures and sea-ice distributions (P. M. Forster et al., 2016; Hansen et al., 2005), known as the fixed-SST method. The experiment is a Tier 2 simulation provided by the Radiative Forcing Model Intercomparison Project contribution to CMIP6 (Pincus et al., 2016). To obtain estimates of ERF from more models and scenarios, we can use estimates of the climate feedback parameter to relate outputs of modeled top-of-atmosphere energy imbalance and surface temperature to time-varying ERF (for a full description of this method, see P. M. Forster et al. (2013)). The climate feedback parameter is the change

in top-of-atmosphere energy imbalance per degree surface temperature change, here obtained from a Gregory regression of each model's abrupt-4xCO2 experiment (a mandatory experiment for all CMIP6 models). However, this method is biased, as it is now well-known that the climate feedback parameter is not constant in time (e.g., Andrews et al., 2022; Armour, 2017; Rugenstein et al., 2020; Senior & Mitchell, 2000; Winton et al., 2010). Comparing the two methods for historical ERF shows that the bias is worse in models that show significant non-stationarity in their climate feedback parameter (Smith & Forster, 2021).

We seek to improve ERF estimated from the abrupt-4xCO2 climate feedback by calculating ERF with a time-scale dependent feedback parameter. Fredriksen et al. (2021) showed that this approach well describes the surface temperature output of the historical and RCP scenarios for the majority of CMIP5 models. Here, we extend the analysis to CMIP6 models and scenarios, with the added confidence of comparing results with fixed-SST esti-mates in 10 cases, and also compare ERF in RCP (CMIP5) and SSP (CMIP6) scenarios with the same nominal year-2100 radiative forcing. The provided forcing will be useful for model evaluation and selection, including the discussion on "hot" models in CMIP6 (Hausfather et al., 2022).

2. Data

We study CMIP6 models that have published the four variables *tas* (near-surface air temperature), *rlut*, *rsut*, and *rsdt* (top of atmosphere longwave upwelling, shortwave upwelling, and shortwave downwelling radiation respec-tively) for both the piControl and the abrupt-4xCO2 experiment by March 2022. For these 51 models (listed in Tables S1 and S2 of the Supporting Information S1), we look at all members for the experiments abrupt-4xCO2, abrupt-2xCO2, abrupt-0p5xCO2, 1pctCO2, historical, hist-GHG, hist-aer, hist-nat, ssp119, ssp126, ssp245, ssp370, ssp585, piClim-4xCO2, piClim-control and piClim-histall. The number of members and a short descrip-tion of the experiments are given in Tables S1–S4 and Text S1 of the Supporting Information S1. For many models, the piClim-histall experiments used to compute historical fixed-SST forcing are extended with the SSP2-4.5 scenario to year 2100. In addition to the 9 publicly available piClim-histall experiments, we have included an experiment done with the model MPI-ESM1-2-LR not available yet through CMIP6. For each variable, we compute the annual anomalies relative to a linear trend over the corresponding period of the piControl run.

3. Method

The linear energy balance framework describes, to first order, the correspondence between forcing, feedbacks and global mean temperature:

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

where N is the top-of-the-atmosphere net radiative downward flux (in W/m²), λ is the climate feedback parame-ter (in W m⁻² K⁻¹), T is the surface air temperature change (in K) relative to an unperturbed steady state where $N = F = 0$ and F is the external radiative forcing (in W/m²), for instance due to a change in atmospheric composi-tion. λ is often determined from idealized experiments where the CO_2 concentration is abruptly quadrupled, using a regression of N against T (Gregory et al., 2004). Once λ is known, Equation 1 can be rearranged to determine $F(t)$ from any experiment where the evolution of $T(t)$ and $N(t)$ are known (P. M. Forster et al., 2013), here referred to as the $1-\lambda$ forcing.

We use the method described in detail in Fredriksen et al. (2021) (henceforth F21) to compute what we will refer to as the $3-\lambda$ forcing. We assume the global temperature responds linearly to the forcing, and can be decomposed as

$$T(t) = \sum_{n=1}^{K} T_n(t) = \sum_{n=1}^{K} c_n \exp(-t/\tau_n) * F(t) \tag{2}$$

where the $*$ denotes a convolution, K is the number of components, τ_n are time scales (in years) and c_n (in K m² W⁻¹) are amplitudes of the temperature responses. We assume that N can be decomposed similarly, where different λ's are associated with each component of $T(t)$:

$$N(t) = \sum_{n=1}^{K} N_n(t) = F(t) + \sum_{n=1}^{K} \lambda_n T_n(t) \tag{3}$$

For the abrupt-4xCO2 experiment we get $T_n(t) = a_n\left(1 - e^{-t/\tau_n}\right)$ and $N_n(t) = b_n e^{-t/\tau_n}$, where $a_n = c_n \tau_n F_{4\times CO_2}$ and $b_n = -a_n \lambda_n$. By fitting these expressions to T and N from the abrupt-4xCO2 experiment, we determine the c_n's and λ_n's for each model. We improve the method over F21 to use data points from all available ensemble members to better constrain the estimate. Additional members are averaged over when computing the parameters of the temperature response, and treated as extra data points when plotting T versus N to determine the λ_n's. We fit 1,000 different selections of time scales, and select the parameters providing the best least-squares fit to the temperature response. We use 150 years of data for estimation to treat all models equally. Many models have run the experiments for longer than that, and these extra years are included in the figures, allowing us to visually inspect how the fit performs at longer scales.

We use $K = 3$ time scales except for the responses to abrupt-4xCO2. There $K = 4$, but we assume the slowest response to be so slow, that it can be approximated as a constant heat flux $N_4 = b_4$ going into the deeper oceans without affecting the surface temperature during the first 150 years after quadrupling.

In Equation 2, we note that we can move an arbitrary factor between the c_n and $F(t)$ without changing the temperature response, so different definitions of the forcing can in fact be used in a linear/impulse response model. Here we strive to make a forcing definition that does not involve adjustments from surface temperature responses. When the parameters c_n and λ_n have been estimated from the abrupt-4xCO2 experiment, we define a separation of forcing and response, which is used to compute $F(t)$ for other experiments by rearranging Equation 3. Since this equation needs to know the components of $T(t)$, we need to iterate until convergence between (a) performing the convolutions in Equation 2 to find the components and (b) computing the forcing which is needed for the convolutions. We assume an efficacy of 1 for all forcing agents (P. Forster et al., 2021; Hansen et al., 2005; Smith & Forster, 2021).

The method relies on a linear response model for predicting the temperature components, so a criterion for making good forcing estimates is that the linear model actually predicts the temperature well. In several figures we therefore include the difference between the temperature predicted by the linear model and the temperature output of the GCM. A difference close to zero is considered a necessary, but not sufficient criterion that we have a good forcing estimate. More importantly, we are interested in estimates that are consistent with fixed-SST forcing corrected for land temperature responses (Andrews et al., 2021). This definition of the forcing has efficacy factors closest to 1 (Richardson et al., 2019). Whenever available, our forcing estimates are compared to fixed-SST estimates of the forcing. Several methods exist for adjusting these forcing estimates to isolate the forcing at zero temperature response, and in Figure 1 we include the tropospherically corrected ERF (ERF_trop) estimates from Smith et al. (2020), where the fixed-SST forcing is corrected for land warming using radiative kernels.

As a thought experiment of why we think several timescales for λ are necessary for correct forcing estimates, we consider the result of using $1-\lambda$ methods for estimating time-varying forcing for abrupt-4xCO2, and test how close this is to a constant. Assuming we have a typical model behavior where feedbacks become less negative with time and we regress the first 150 years, the time-varying forcing $F(t) = N(t) - \lambda T(t)$ will have higher values in the beginning where the values of $N(t)$ are above the straight line. Similarly, if making a regression for the first 20 years, then the later time period will get stronger forcing estimates. So if these forcing estimation methods cannot reproduce the constant $4xCO_2$ forcing, we would expect them to give biased time-varying forcings also for other experiments.

4. Results

For the 18 models where fixed-SST forcing is available for abrupt-4xCO$_2$, we find a generally good correspondence between our forcing estimates and ERF_trop (see Figure 1 and estimated parameters in Tables S6 and S7 of the Supporting Information S1). An exception is CNRM-ESM2-1, which used a different method than other models to prescribe CO$_2$ (see Text S2 in Supporting Information S1). Since land temperatures have responded in the fixed-SST estimates, they are comparable to our curve after a few months of response. The light blue curves provide insight into the uncertainties associated with our estimates, and we note that their spread varies substantially between models. We expect eventual over- or under-estimations of the $4\times CO_2$ forcings here to follow the transient forcing estimates presented in Figures 2–4. Uncertainties in the forcing estimates are often larger if the model's surface temperature responds quickly, as there will be fewer points close to the y-axis to constrain the intercept. Similarly, if a model is still far from equilibrium after 150 years we can expect a larger spread in the
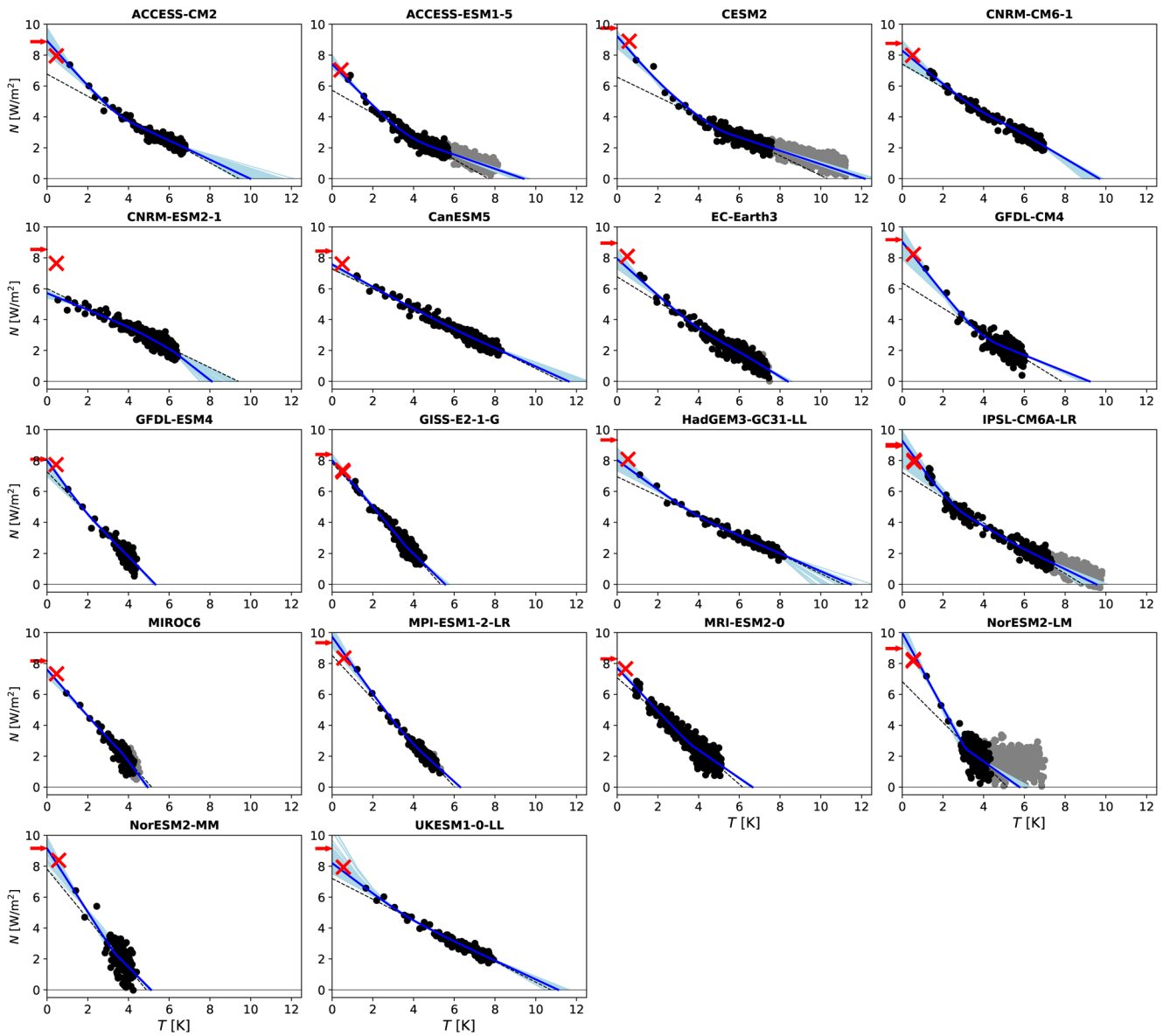
Figure 1. The top-of-the-atmosphere net radiation anomalies (N) versus the temperature anomalies (T) for the 18 models where we know the fixed-SST forcing (plotted as red crosses) for abrupt-4xCO2 simulations. The black dots are annual mean values (gray beyond year 150), and all available ensemble members are included. The black dashed fit is the standard 150-year linear regression still used in the Sixth Assessment Report. The light blue curves are fits done to the first 150 years of the response with the $3-\lambda$ method for 1,000 different random choices of time scales, and the dark blue curves show the best (least-squares) fit for the temperature response. The red arrows show the ERF_trop forcing estimates from Smith et al. (2020), which corrects for the land surface temperature response in the fixed-SST estimates.

estimated climate sensitivity (as derived from the intercept with the x-axis). Internal variability in T and N also plays a role in determining the uncertainty of the fit (e.g., Gregory et al., 2020).

We also estimate the forcing for the 33 models without fixed-SST forcing estimates, and for the 12 models with abrupt-2xCO2 and the 9 models with abrupt-0p5xCO2 experiments (Figures S1–S4 in Supporting Information S1). Our curved fit through the points appears to be a generally better fit than straight lines, so we expect to find reasonable forcing estimates (i.e., relative to fixed-SST forcing estimates) also for these experiments. $4\times CO_2$ forcing is on average 2.11 times stronger than the $2\times CO_2$ forcing, and the absolute value of the $0.5\times CO_2$ forcing is a little weaker than the $2\times CO_2$ forcing for most models (Table S8 in Supporting Information S1), consistent with a radiative forcing depending superlogarithmically on the CO_2 concentration (Etminan et al., 2016). However, the smaller signal-to-noise ratio makes these lower forcing estimates more uncertain.
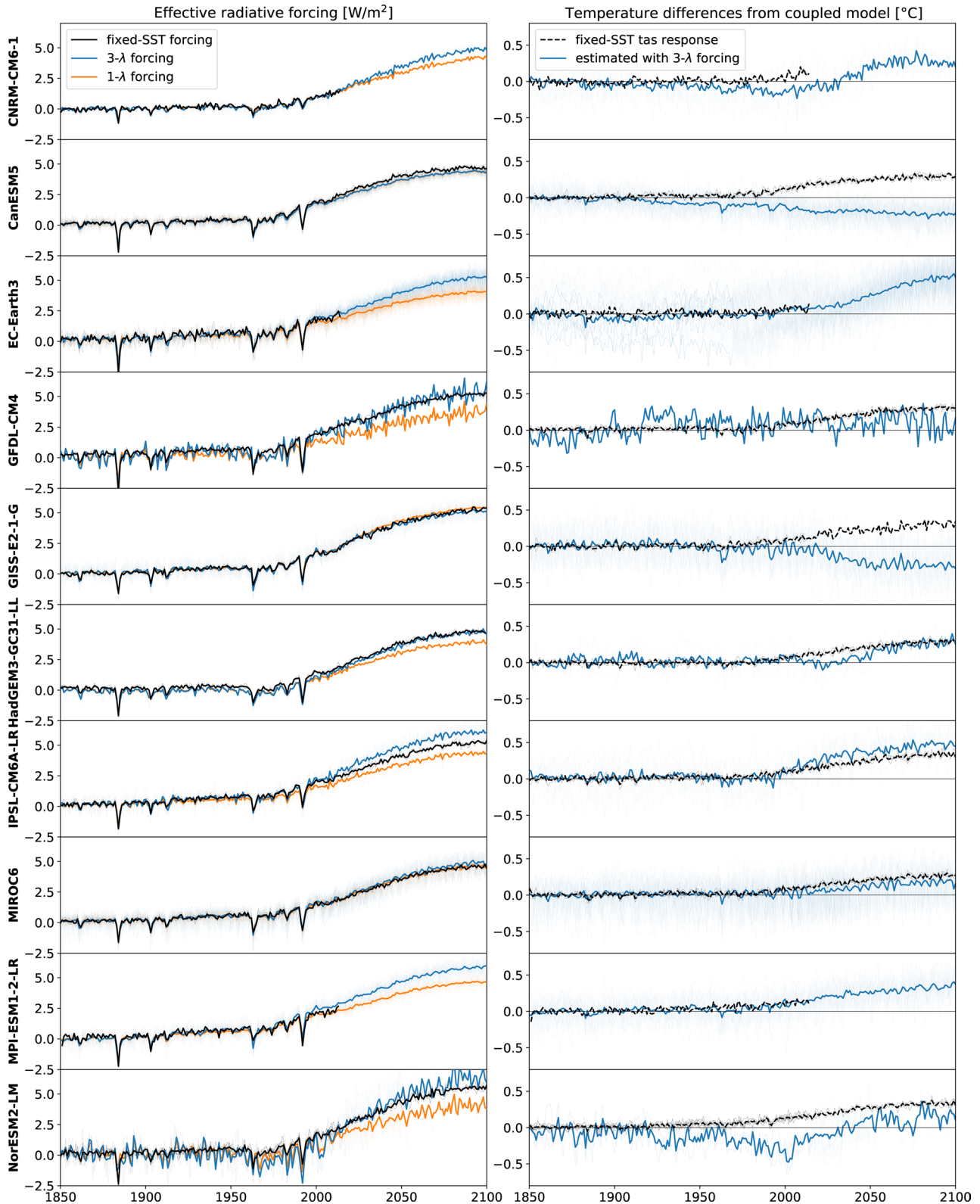
Figure 2. Left: 1−λ (orange), 3−λ (blue) and fixed-SST (black) forcing estimates, for the models where transient fixed-SST estimates are available (10 models). Thin lines are estimates from single members, and thicker lines are ensemble means. Right: surface air temperature differences between the output of the coupled model and the estimated response to the 3−λ forcing (blue) and the change in surface air temperature from the fixed-SST runs (black).

The historical and SSP2-4.5 $3-\lambda$ forcing is consistent with the fixed-SST forcing for most models and always better than or as good as the $1-\lambda$ forcing estimates (Figure 2). Hence we expect the $3-\lambda$ ERF to be a good approximation also for the many models and experiments that lack fixed-SST forcing. The black curves in the right column show that land temperatures have not responded much in these fixed-SST experiments compared to abrupt-4xCO2 experiments, so these forcing estimates probably do not need to be corrected for land responses to the same degree, but we can expect a small negative bias when the temperature response grows. For models with little curvatures in Figure 1 (e.g., GISS-E2-1-G and CanESM5), the $1-\lambda$ forcing is as expected very similar to the $3-\lambda$ forcing. For IPSL-CM6A-LR, the fixed-SST forcing falls in the middle of the $1-\lambda$ and $3-\lambda$ forcing, suggesting that both the $3-\lambda$ and $1-\lambda$ forcings are slightly biased in different directions.

In addition to the comparison with fixed-SST forcing, the ability to predict the GCM temperature also serves as a measure of how good the forcing estimate is, therefore we have included in the right column of Figure 2 the difference between the temperature predicted from our $3-\lambda$ forcing and linear response model and the output of the complex model. For positive differences, the forcing is overestimated, and vice versa. Temperature differences are typically within a $\pm0.5°C$ interval, suggesting that our combination of forcing and linear response can generally well describe global mean temperatures.

Figure 3 shows the multi-model mean $3-\lambda$ forcing from all available models for seven different experiments (left column), and the corresponding global mean temperature difference between the linear responses and the GCMs (right column). The large ensembles of temperature differences show that temperature responses are on average slightly overestimated for the 1pctCO2 and future scenario experiments, and hence the forcings in the left column are probably slightly overestimated too. We hypothesize that this could be due to state-dependencies in the feedback parameters. Differentiating how much of a feedback change is due to state-dependence (inconsistent with linear response (Bloch-Johnson et al., 2021)) or due to pattern effect/time-scale dependence (can be consistent with linear response) will be important in future work. We include also the CMIP5 estimates from F21 in this figure for comparison, and note that for the 1pctCO2 experiment (top row) the average forcing estimates are remarkably similar for CMIP5 and CMIP6.

For the last decades of the historical experiment, we find the CMIP6 forcing to be lower than the CMIP5 forcing, but maybe slightly underestimated, as evidenced by the comparison between our estimated temperatures and those from the GCMs. For the future comparable scenarios however (SSP1-2.6, SSP2-4.5, and SSP5-8.5 for CMIP6; RCP2.6, RCP4.5, and RCP8.5 for CMIP5), the CMIP6 forcing grows more than the CMIP5 forcing, and ends up at higher values than CMIP5 at the end of the 21st century with no clear difference in the bias in temperatures compared to CMIP5 models. This suggests that CMIP6 ERF is higher than equivalent nominal scenarios in CMIP5. The multi-model mean difference in year 2100 is 0.18, 0.46, and 0.55 W/m^2 for RCP2.6/SSP1-2.6, RCP4.5/SSP2-4.5, RCP8.5/SSP5-8.5, respectively. The respective mean temperature differences in year 2100 are 0.20/0.12, 0.17/0.16, and 0.09/0.12 K.

A closer look at the historical period (Figure 4) shows that our $3-\lambda$ total forcing for around 1995 onwards is a little stronger than the $1-\lambda$ forcing, as used in Smith and Forster (2021). Studying the components separately, we find that the greenhouse gas forcing becomes more positive, and the aerosol forcing becomes more negative when using the $3-\lambda$ method. In general, the different forcing definitions give more different results the stronger the temperature response is.

The small underestimation of the CMIP6 linear temperature responses for the historical period seems to stem mainly from the response to aerosol forcing in some of the models (Figure S6 in Supporting Information S1). One reason for this could be a similar (possible state-dependence) effect for this negative forcing as we have with the small overestimation for the positive forcing. Another reason could be that assuming an efficacy of 1 for aerosol forcing does not work perfectly (Huusko et al., 2022; Modak & Bala, 2019; Richardson et al., 2019; Salvi et al., 2022). A spatially inhomogeneous forcing can lead to a pattern effect in the associated response to it, for example, by emphasizing the faster and stronger land responses more than in the response to CO_2 forcing. This could lead to responses deviating slightly from our linear response model, further causing small errors in our forcing estimates. Localized forcing can also trigger feedbacks differing from the global mean feedback, for example, due to differences in troposheric stability between the tropics and extratropics (Salvi et al., 2022).
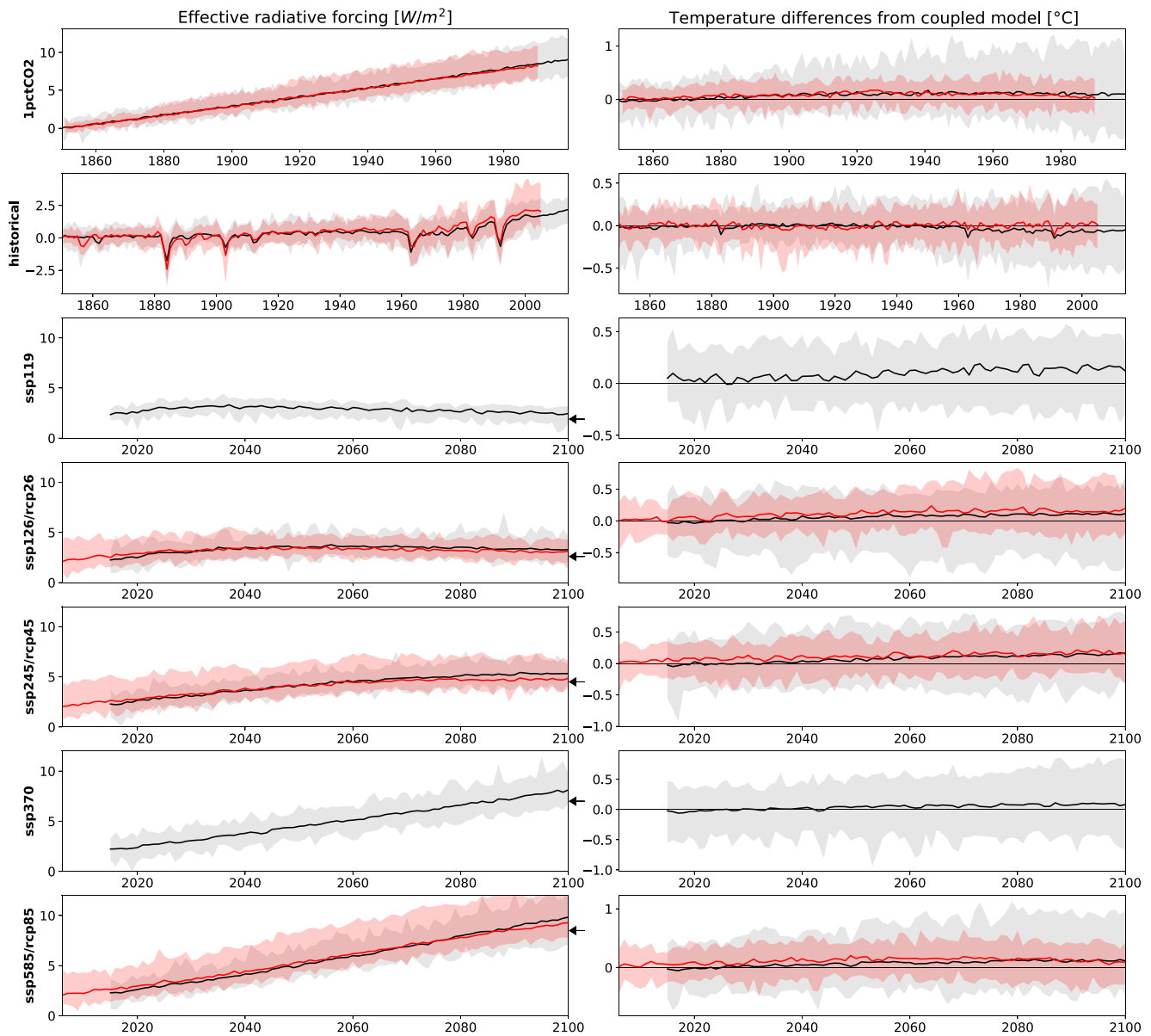
Figure 3. Left: Effective radiative forcing estimates using the $3-\lambda$ method from up to 50 models from CMIP6 (gray/black) and up to 21 models from CMIP5 (red). The shading shows the min and max values of the member mean forcing for each year, and the red/black curves the mean of all models. Each row shows an experiment denoted by the y-axis label. The black arrows denote the forcing levels 1.9, 2.6, 4.5, 7.0, and 8.5 W/m^2 in year 2100. Right: The difference between the temperature predicted by the linear temperature response and the global mean temperature from the coupled model.

5. Discussion

Future temperature projections from CMIP6 show stronger warming than the corresponding projections from CMIP5 (Tebaldi et al., 2021). The CMIP5 RCP scenarios have a different composition of greenhouse gases, aerosols and other forcers than the CMIP6 SSP scenarios, but they are designed such that they should reach approximately the same forcing levels by the end of the 21st century (Gidden et al., 2019). However, Wyser et al. (2020) shows that at least half of the temperature increase from CMIP5 to CMIP6 for the model EC-Earth3-Veg is due to the increase in the ERF, and in particular the greenhouse gas concentrations. Chapter 4 of the IPCC's Sixth Assessment Report Working Group 1 (Lee et al., 2021) shows that ERF is substantially higher for CMIP6 SSPs that are nominally the same forcing as CMIP5 RCPs (SSP1-2.6 vs. RCP2.6 e.g.,), and comes to a similar conclusion, namely that the increase in forcing contributes to about half of the temperature increase in CMIP6 models compared to CMIP5 models with the other half attributed to the increase in climate sensitivity. The IPCC
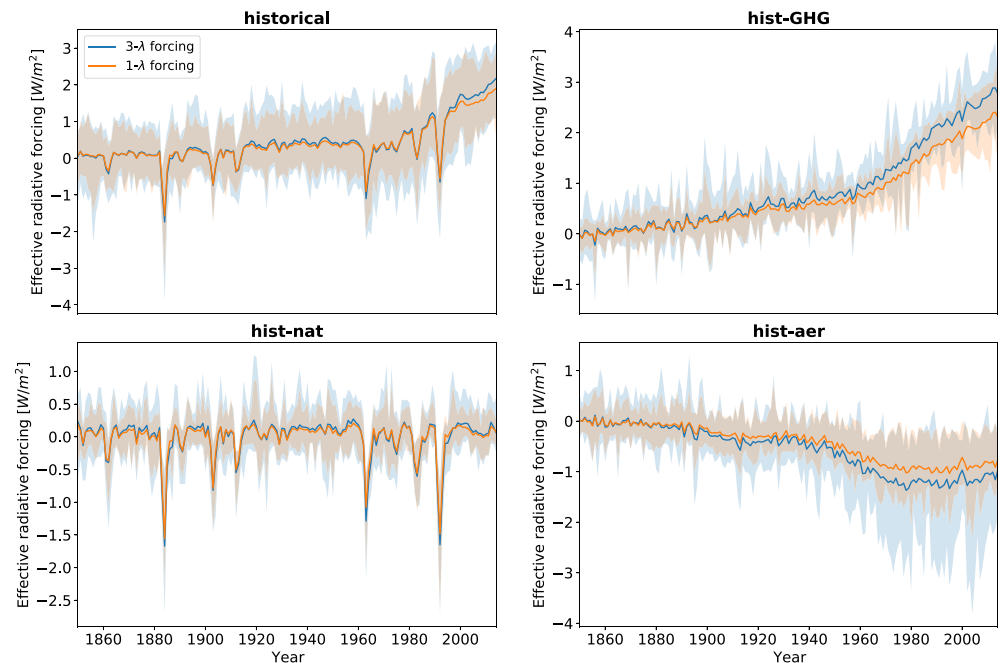
Figure 4. The effective radiative forcing for the historical, hist-GHG, hist-nat and hist-aer experiments, computed using both the $1-\lambda$ (orange) and the $3-\lambda$ method (blue). The shading shows the min and max values of the model ensemble, and the solid curves the model means.

approach used calibrated climate emulators to arrive at their conclusion, while we show that this systematic difference between CMIP5 and CMIP6 exists also when computing the forcing model by model.

Our results confirm that the ERF is indeed increasing more for CMIP6 than for CMIP5 during the 21st century. From the 1pctCO2 experiment we note that given the same increase in CO_2 concentrations, the ERF is similar in CMIP5 and CMIP6 models. For the 4×CO2 forcing we find a small mean increase in forcing of 0.25 W/m^2 from CMIP5 to CMIP6 models (see Figure S5 in Supporting Information S1). Hence the radiative efficiency of CO_2 is unchanged or slightly higher in CMIP6 models. This suggests that the higher year-2100 CO_2 concentrations in CMIP6 compared to similar CMIP5 scenarios (Meinshausen et al., 2020) is important for explaining the increase in ERF. Other forcing agents also differ substantially between comparable RCPs and SSPs, but for the strongest scenarios SSP5-8.5/RCP8.5 the change in other major agents such as methane and sulfur contribute to a lower forcing in SSP5-8.5 compared to RCP8.5 (Tebaldi et al., 2021), hence the increased CO_2 concentrations is necessary for explaining the net increase in forcing.

Despite the higher temperature increase in the future scenarios, the historical temperature increase is less in CMIP6 than in CMIP5 (Flynn & Mauritsen, 2020; Smith & Forster, 2021). Smith and Forster (2021) explain this as a combination of stronger feedbacks and lower historical forcing, from both aerosols and greenhouse gases (GHGs). Our new $3-\lambda$ historical forcing estimates diverge from the $1-\lambda$ forcing used in Smith and Forster (2021) only after 1995, and is hence not changing their conclusions.

The higher temperature increase for CMIP6 during the 21st century is also partly explained by the increase in climate sensitivity, as evidenced by less negative global feedbacks, increased equilibrium climate sensitivity and transient climate response (Flynn & Mauritsen, 2020; Zelinka et al., 2020). The climate sensitivity does not tell the full story of how much warming we can expect at all times, and an increased sensitivity may not necessarily contribute to increased temperature responses during the historical period. In addition to forcing differences, the pattern of warming also matters, as some regions more effectively radiate out excess energy than others, hence modulating the effective global climate sensitivity with time (the *pattern effect*). During the late 20th century, the warming pattern in the tropical Equatorial Pacific has led to lower estimates of the effective climate sensitivity, (e.g., Andrews et al., 2018, 2022; Zhou et al., 2016). This pattern is not expected to persist in the future, likely causing the effective climate sensitivity to increase in the near future compared to estimates based on observation of the last decades.

Normalizing the abrupt-4xCO2 responses by our estimated forcing yields also a measure of the model sensitivity per unit forcing, which may be helpful for understanding if temperature responses are stronger because of a high climate sensitivity or a higher forcing. We find that CMIP6 models have an average response slightly stronger than CMIP5 models, and a larger model spread (Figure S7 in Supporting Information S1). Hence our results confirm that CMIP6 models are overall more sensitive, but the relative role of the climate sensitivity for explaining higher temperature responses is highly model dependent.

6. Conclusions

We show that the $3-\lambda$ estimation method can both approximate fixed-SST forcing corrected for land temperatures, and reproduce GCM temperatures. The $3-\lambda$ ERF is therefore a useful alternative whenever fixed-SST estimates are not available. Our comparison of the CMIP5 and CMIP6 ensembles show that CMIP6 forcing is on average weaker for the historical period, but increases more and becomes stronger than CMIP5 forcing during the 21st century, likely due to the higher CO_2 concentrations. In addition to the stronger forcing, CMIP6 models are overall more sensitive, and we encourage more models to make fixed-SST estimates and run RCP forcing (such as Fyfe et al. (2021)) to better quantify the relative roles of forcing and sensitivity in explaining different temperature responses.

Data Availability Statement

The models used are listed in the Supporting Information and the original CMIP6 data sets are available at https://esgf-node.llnl.gov/projects/cmip6/. Processed data and code are permanently stored in https://doi.org/10.5281/zenodo.7687534 (Fredriksen, 2023).

References

Andrews, T., Bodas-Salcedo, A., Gregory, J. M., Dong, Y., Armour, K. C., Paynter, D., et al. (2022). On the effect of historical SST patterns on radiative feedback. *Journal of Geophysical Research: Atmospheres*, *127*(18), e2022JD036675. https://doi.org/10.1029/2022JD036675

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., et al. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, *45*(16), 8490–8499. https://doi.org/10.1029/2018GL078887

Andrews, T., Smith, C. J., Myhre, G., Forster, P. M., Chadwick, R., & Ackerley, D. (2021). Effective radiative forcing in a GCM with fixed surface temperatures. *Journal of Geophysical Research: Atmospheres*, *126*(4), e2020JD033880. https://doi.org/10.1029/2020JD033880

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

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

Etminan, M., Myhre, G., Highwood, E. J., & Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, *43*(24), 12614–12623. https://doi.org/10.1002/2016GL071930

Flynn, C. M., & Mauritsen, T. (2020). On the climate sensitivity and historical warming evolution in recent coupled model ensembles. *Atmospheric Chemistry and Physics*, *20*(13), 7829–7842. https://doi.org/10.5194/acp-20-7829-2020

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., et al. (2021). The Earth's energy budget, climate feedbacks, and climate sensitivity (Eds.). In *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* (pp. 923–1054). Cambridge University Press. https://doi.org/10.1017/9781009157896.009

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1139–1150. https://doi.org/10.1002/jgrd.50174

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., et al. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *Journal of Geophysical Research: Atmospheres*, *121*(20), 12460–12475. https://doi.org/10.1002/2016JD025320

Fredriksen, H.-B. (2023). *Hegebf/CMIP6-forcing: Code and data used for the paper "21st century scenario forcing increases more for CMIP6 than CMIP5 models" (v1.0.0)*. Zenodo. https://doi.org/10.5281/zenodo.7687534

Fredriksen, H.-B., Rugenstein, M., & Graversen, R. (2021). Estimating radiative forcing with a nonconstant feedback parameter and linear response. *Journal of Geophysical Research: Atmospheres*, *126*(24), e2020JD034145. https://doi.org/10.1029/2020JD034145

Fyfe, J. C., Kharin, V. V., Santer, B. D., Cole, J. N. S., & Gillett, N. P. (2021). Significant impact of forcing uncertainty in a large ensemble of climate model simulations. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(23), e2016549118. https://doi.org/10.1073/pnas.2016549118

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., et al. (2019). Global emissions pathways under different socio-economic 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

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., & Webb, M. J. (2020). How accurately can the climate sensitivity to CO_2 be estimated from historical climate change? *Climate Dynamics*, *54*(1–2), 129–157. https://doi.org/10.1007/s00382-019-04991-y

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., & Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3), L03205. https://doi.org/10.1029/2003GL018747

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., & Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Research*, *110*(D18), D18104. https://doi.org/10.1029/2005JD005776

Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W., & Zelinka, M. (2022). Climate simulations: Recognize the 'hot model' problem. *Nature*, *605*(7908), 26–29. https://doi.org/10.1038/d41586-022-01192-2

Huusko, L., Modak, A., & Mauritsen, T. (2022). Stronger response to the aerosol indirect effect due to cooling in remote regions. *Geophysical Research Letters*, *49*(21), e2022GL101184. https://doi.org/10.1029/2022gl101184

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., et al. (2021). Future global climate: Scenario-based projections and near-term information (Eds.). In *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* (pp. 553–672). Cambridge University Press. https://doi.org/10.1017/9781009157896.006

Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., et al. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, *13*(8), 3571–3605. https://doi.org/10.5194/gmd-13-3571-2020

Modak, A., & Bala, G. (2019). Efficacy of black carbon aerosols: The role of shortwave cloud feedback. *Environmental Research Letters*, *14*(8), 084029. https://doi.org/10.1088/1748-9326/ab21e7

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747–756. https://doi.org/10.1038/nature08823

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

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

Richardson, T. B., Forster, P. M., Smith, C. J., Maycock, A. C., Wood, T., Andrews, T., et al. (2019). Efficacy of climate forcings in PDRMIP models. *Journal of Geophysical Research: Atmospheres*, *124*(23), 12824–12844. https://doi.org/10.1029/2019JD030581

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., et al. (2020). Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophysical Research Letters*, *47*(4), e2019GL083898. https://doi.org/10.1029/2019GL083898

Salvi, P., Ceppi, P., & Gregory, J. M. (2022). Interpreting differences in radiative feedbacks from aerosols versus greenhouse gases. *Geophysical Research Letters*, *49*(8), e2022GL097766. https://doi.org/10.1029/2022GL097766

Senior, C. A., & Mitchell, J. F. B. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, *27*(17), 2685–2688. https://doi.org/10.1029/2000GL011373

Smith, C. J., & Forster, P. M. (2021). Suppressed late-20th Century warming in CMIP6 models explained by forcing and feedbacks. *Geophysical Research Letters*, *48*(19), e2021GL094948. https://doi.org/10.1029/2021GL094948

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

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., et al. (2021). Climate model projections from the scenario model intercomparison project (ScenarioMIP) of CMIP6. *Earth System Dynamics*, *12*(1), 253–293. https://doi.org/10.5194/esd-12-253-2021

Winton, M., Takahashi, K., & Held, I. M. (2010). Importance of ocean heat uptake efficacy to transient climate change. *Journal of Climate*, *23*(9), 2333–2344. https://doi.org/10.1175/2009JCLI3139.1

Wyser, K., Kjellström, E., Koenigk, T., Martins, H., & Döscher, R. (2020). Warmer climate projections in EC-Earth3-veg: The role of changes in the greenhouse gas concentrations from CMIP5 to CMIP6. *Environmental Research Letters*, *15*(5), 054020. https://doi.org/10.1088/1748-9326/ab81c2

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

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

References From the Supporting Information

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

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

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