AGU ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters®

RESEARCH LETTER
10.1029/2021GL094948

Key Points:

- Coupled Model Intercomparison Project Phase 6 (CMIP6) models warm less in the multimodel mean than CMIP5 models in the 1960–2000 period
- Implied forcing from greenhouse gases and aerosols is lower (more negative) in CMIP6 than in CMIP5
- Stronger climate feedbacks in CMIP6 amplifies the cooling from aerosols

Supporting Information:

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

Correspondence to:

C. J. Smith,
c.j.smith1@leeds.ac.uk

Citation:

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

Received 21 JUN 2021
Accepted 19 AUG 2021

Suppressed Late-20th Century Warming in CMIP6 Models Explained by Forcing and Feedbacks

Christopher J. Smith[1,2] **and Piers M. Forster**[1]

[1]Priestley International Center for Climate, University of Leeds, Leeds, UK, [2]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Abstract For the 1960–2000 period, the latest generation of climate models (Coupled Model Intercomparison Project Phase 6 [CMIP6]) shows less global mean surface temperature change relative to pre-industrial than that seen in observations. In contrast, the previous generation of models (CMIP5) performed well over this period. It has been hypothesized that this suppressed late-20th Century warming seen in CMIP6 is caused by a stronger aerosol forcing. However, we find this to be only part of the story. Not only is the aerosol forcing marginally more negative in CMIP6 compared to CMIP5, the greenhouse gas forcing in CMIP6 is also weaker than in CMIP5. These forcing differences are amplified by differences in climate sensitivity between the CMIP5 and CMIP6 ensemble, which leads to both a stronger aerosol cooling over 1960–1990 and a stronger greenhouse gas induced warming from 1990, returning the warming post-2000 toward the observed level.

Plain Language Summary Climate models are our best tools for predicting how the climate will change in the future. Confidence in future projections relies on the ability to accurately simulate the past. Many of the latest climate models show less warming than observations around the 1960–2000 period, so understanding why is key to making more confident projections. Models respond to human and natural forcings such as greenhouse gases, air pollutants (aerosols), volcanic eruptions, and solar activity. We show that the latest models simulate a lower forcing from greenhouse gases and aerosols than older models, which is part of the reason why the climate does not warm in line with observations in the late 20th Century. The other part of the reason for the lower warming is that the newer models are more sensitive: they project a greater temperature change for a given amount of forcing. The higher sensitivity is found to result in more cooling from aerosols in the new models, but it opposes the lower forcing from greenhouse gases. The aerosol effects win out, causing suppressed warming in the late 20th Century.

1. Introduction

Climate responds to the effective radiative forcing (ERF), the perturbation in the Earth's energy balance caused by natural and anthropogenic activities. This is true both in climate models and in the real world. Directly measuring the ERF has so far eluded scientists, though observational evidence of increasing instantaneous radiative forcing (IRF), the main component of ERF, has been inferred from satellite observations (Kramer et al., 2021). Therefore, climate models with sufficiently accurate radiative transfer parameterisations (Pincus et al., 2020) currently remain our best tools to determine ERF.

One intriguing result of the latest generation of global climate models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) is the suppressed warming in the late 20th Century, approximately between 1960 and 2000, compared to observations and the older set of CMIP5 models (Flynn & Mauritsen, 2020). It has been hypothesized that aerosol forcing in CMIP6 may be stronger than in CMIP5 and responsible for cooler 20th Century projections (M. B. Andrews et al., 2020; Dittus et al., 2020; Flynn & Mauritsen, 2020; Gillett et al., 2021). However, it was found that temperature responses to aerosol forcing were stronger using CMIP5 emissions than CMIP6 emissions in the CanESM5 model (Fyfe et al., 2021), which opposes this hypothesis.

The strong aerosol forcing in the late 20th Century is followed by a period of rapid aerosol recovery (Smith et al., 2021) and rapid recent warming (Tokarska et al., 2020) in some CMIP6 models, such that the multi-model mean present-day warming is about in line with observations following a cool late 20th Century (Flynn & Mauritsen, 2020; Gillett et al., 2021). Rapid present-day aerosol unmasking in combination with

high transient climate sensitivity in many CMIP6 models leads to high levels of future warming projections (Tebaldi et al., 2021). If these projections are correct they have profound policy consequences, as scenarios designed to be consistent with 1.5°C (SSP1-1.9) and well-below 2°C (SSP1-2.6) Paris Agreement targets (O'Neill et al., 2016) far exceed these levels of warming (Tebaldi et al., 2021).

Not all CMIP6 models under-predict 20th Century warming, and other approaches are possible such as using climate emulators to constrain aerosol forcing (Smith et al., 2021). However, even in cases where observed warming is well represented in models, it can result from different combinations of forcing and response. Historical observed temperatures can be reproduced in climate models both with a high climate sensitivity/strong aerosol forcing combination and a low sensitivity/weak aerosol forcing combination (Shindell & Smith, 2019; Smith et al., 2018). As the strength of the aerosol forcing determines how much warming has been suppressed, and the climate sensitivity describes how much warming per unit forcing can be expected, both large and small amounts of future warming can be consistent with historical observations. The picture is further complicated by the changing patterns of sea-surface temperatures in the historical period (T. Andrews et al., 2018) which may modulate the global mean temperature change to a particular forcing signal (Gregory et al., 2020), and possibility for cloud feedbacks to change over time allowing high sensitivity, moderate aerosol forcing combinations to be consistent with historical warming (Bjordal et al., 2020). Efforts to narrow down the range of plausible climate sensitivity and aerosol forcing will enable us to better constrain future climate projections (Stevens et al., 2016).

In this paper, we estimate the historical ERF in 35 CMIP6 models using the methods of Forster et al. (2013), and compare these results directly with the implied ERF from 27 CMIP5 models in Forster et al. (2013). We also perform this forcing analysis for greenhouse gas, natural forcing, and aerosol attribution experiments and compare these results from CMIP6 to their CMIP5 equivalents, showing that the CMIP generations differ in their overall implied ERF and climate response.

2. Methods

2.1. Historical Implied ERF

We derive an estimate of the time-varying ERF from CMIP6 models following the method of Forster et al. (2013). Starting from the linearized Earth energy balance equation

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

we first obtain estimates of the climate feedback parameter λ derived from Zelinka et al. (2020); Zelinka (2021). These estimates of λ are derived from quadrupled CO_2 experiments in CMIP5 and CMIP6 models, determined as the regression slope of ΔN against ΔT of a 150-year integration of an *abrupt-4xCO2* run. In Equation 1, ΔN is the top-of-atmosphere (TOA) energy imbalance, ΔT is global-mean near-surface air temperature anomaly and ΔF is ERF. Δ represents anomalies with respect to a pre-industrial control climate state.

Next, we plug ΔN and ΔT from each model's coupled *historical* run to obtain a time-varying historical ΔF (Forster et al., 2013). To obtain ΔN and ΔT, the pre-industrial control run (*piControl*) is subtracted from the historical run year-wise, starting from the appropriate branch point in the *piControl* simulation to minimize the impact of any climate drift. We refer to ΔF obtained from this Forster et al. (2013) method as the "implied" ERF.

Our analysis is limited to those models that are analyzed by Zelinka et al. (2020); Zelinka (2021), broadly, those that have performed *abrupt-4xCO2* and *piControl* experiments, and have performed *historical* CMIP runs. As *piControl*, *historical* and *abrupt-4xCO2* experiments are core simulations in both CMIP5 and CMIP6 (Eyring et al., 2016; Taylor et al., 2012), data is available for a large number of models: 35 models in CMIP6 and 27 models in CMIP5 (Tables S1 and S2 provide a full list). For CMIP5 models, we use ΔN and ΔT estimates that were calculated from CMIP5 historical runs in Forster et al. (2013), and use climate feedback values from Zelinka et al. (2020); Zelinka (2021) that are also available for CMIP5 models to calculate ΔF.

2.2. Fixed-SST ERF

A more direct method to estimate historical ERF is obtained by running a climate model in atmosphere-only mode using climatological sea-surface temperatures (SSTs) and sea-ice distributions with time-varying historical forcings. In CMIP6, this is provided in the RFMIP *piClim-histall* experiment (Pincus et al., 2016) with a corresponding atmosphere-only preindustrial control run (*piClim-control*). As the heat capacity of the land and atmosphere is much smaller than that of the ocean, temperature changes induced by the forcing are much smaller, and $\Delta N \approx \Delta F$ (Forster et al., 2016; Hansen et al., 2005) and can be determined directly from the model. In fact, the small temperature change induced by the warming land surface does contribute and this method may underestimate true ΔF by 5%–15% (T. Andrews et al., 2021; Hansen et al., 2005; Smith et al., 2020). However, this "fixed-SST" ERF approach currently remains the best method to determine time-varying forcing (Forster et al., 2016). The biggest barrier to adoption of this method is the smaller subset of models that data is available from: eight in CMIP6 and with no equivalent experiment in CMIP5, compared to 35 and 27 for implied ERF in CMIP6 and CMIP5 respectively. Therefore, it is not possible to use this method to compare forcing between CMIPs or in an ensemble-wide manner, but we use it as a sense-check that the implied ERF calculated is reasonable.

2.3. Contributions to ERF From Greenhouse Gases, Aerosols and Natural Forcing

The implied ERF method is repeated for the *hist-GHG* (historical greenhouse gases only) and *hist-nat* (historical solar and volcanic forcing only) experiments from the Detection and Attribution Model Intercomparison Project (DAMIP; Gillett et al., 2016) to obtain ΔF_{GHG} and ΔF_{nat} from greenhouse gases and natural forcing respectively. Equivalent experiments exist in CMIP5 as *historicalGHG* and *historicalNat*. Residual anthropogenic forcing is estimated as $\Delta F_{\mathrm{other}} = \Delta F - \Delta F_{\mathrm{GHG}} - \Delta F_{\mathrm{nat}}$ and comprises aerosols, ozone and land-use change. In CMIP6, the *hist-aer* experiment (historical aerosols) is produced, from which we likewise obtain the implied ERF from aerosol forcing, ΔF_{aer}.

Both CMIP5 and CMIP6 have available historical attribution results from 13 models (Tables S1–S4). The 13-model DAMIP subsets are representative of the larger ensembles of CMIP5 and CMIP6 models in terms of their distributions of climate sensitivity.

2.4. CanESM5 CMIP6 and CMIP5 Forcing Ensembles

Additionally, CanESM5 (a CMIP6 model) produced CMIP5-forced equivalents of all of the previous experiments, known as *historical-cmip5*, *hist-GHG-cmip5*, *hist-aer-cmip5*, and *hist-nat-cmip5*, as well as a *piControl-cmip5* run (Fyfe et al., 2021). This single-model ensemble allows a direct comparison between CMIP5 and CMIP6 forcings and climate responses in the same model. In other words, this set of experiments reduces the dimension of differences from two when comparing CMIP6 and CMIP5 ensembles (the differences in the forcings going into the model and the differences in how the models respond to these forcings) to one (only the differences in the forcings going into the model). Five ensemble members of each experiment were made available on ESGF and used in this analysis.

3. Results

3.1. Comparison of CMIP5 and CMIP6 Implied ERF and Temperatures

Figure 1a shows the differences in projected historical warming between CMIP5 (red) and CMIP6 (blue). Our results agree with those of Flynn and Mauritsen (2020) who show that CMIP6 is cooler than CMIP5 in the late 20th Century. For comparison, we plot observations from HadCRUT5 in Figure 1a using a baseline of 1850–1900, and again verify the results of Flynn and Mauritsen (2020) who show that CMIP5 multimodel mean warming is more in line with the observed warming than CMIP6. We do not adjust the CMIP5 and CMIP6 simulations, expressed as anomalies relative to the preindustrial, to this common baseline period. Doing so would increase the differences between the 20th Century warming further as the CMIP6 mean 1850–1900 warming is +0.03°C, whereas CMIP5 is −0.01°C. The differences in forcing and temperature between the CMIP generations are not statistically significant, but are large enough to be noticeable and drawn from a suitably large proportion of each ensemble to be meaningful.

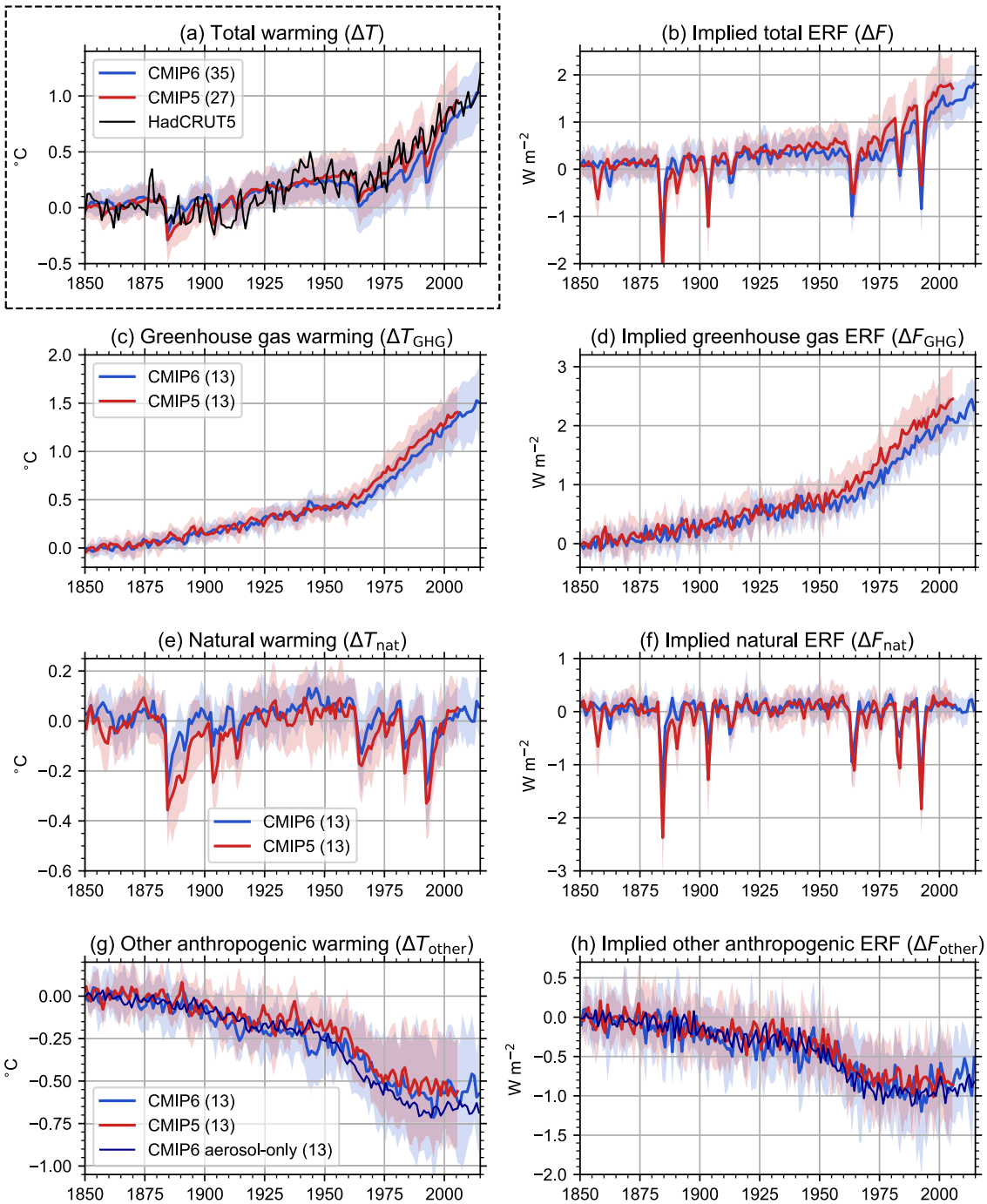

Figure 1. Historical warming (left column) and implied effective radiative forcing (ERF) (right column) from Coupled Model Intercomparison Project Phase 5 (CMIP5) (red) and CMIP6 (blue) models. Top row: total historical forcing, also showing temperature observations from HadCRUT5 (black), rebased to a 1850–1900 baseline as a proxy for preindustrial. Second row: greenhouse gas forcing. Third row: natural forcing. Bottom row: other anthropogenic forcing, with CMIP6 aerosol-only simulations shown in dark blue. Shaded regions are one standard deviation across models. Numbers in brackets in legends indicate the number of models from which data is available.

Table 1
Climate Sensitivity ($-1/\lambda$) Calculated From 150-Year Gregory Regressions (Gregory et al., 2004) in Coupled Model Intercomparison Project Phase 5 (CMIP5) and CMIP6 Models With Available Data Used in This Study

Experiment	CMIP6 (K W^{-1} m^2)	CMIP5 (K W^{-1} m^2)
historical	1.12 ± 0.37 (35 models)	0.99 ± 0.27 (27 models)
hist-GHG, hist-nat	1.13 ± 0.35 (13 models)	1.02 ± 0.32 (13 models)

Note. Original data from Zelinka et al. (2020). Uncertainty ranges are one standard deviation.

Figure 1b shows the implied ERF for CMIP5 and CMIP6 models. The implied ERF in CMIP6 is clearly lower than in CMIP5 from around 1940 to the end of the CMIP5 historical period in 2005.

3.2. Attribution of Implied ERF by Component

The attribution of the implied ERF and temperature into constituent parts reveals interesting differences. Figure 1c shows that warming response to GHG forcing is slightly lower in CMIP6 than CMIP5, but the implied GHG forcing is substantially lower (Figure 1d). A potential explanation is that the subset of 13 CMIP6 models producing the *hist-GHG* attribution experiment are more sensitive than the 13 CMIP5 models that produced *historicalGHG*. Higher climate sensitivity requires less forcing to produce the same temperature response. Similar differences in climate sensitivity between CMIP5 and CMIP6 are also present in the larger full *historical* ensembles (Table 1).

Figures 1e and 1f show small differences between natural forcings in CMIP5 and CMIP6. The negative forcing and the climate response to large volcanic eruptions is larger in CMIP5 than in CMIP6, though this may also be partly affected by the fact that preindustrial volcanic background aerosol is not consistently applied in CMIP5 (Fyfe et al., 2021), with some models specifying a zero volcanic background aerosol and some imposing a nonzero level based on a historical climatology. More volcanic cooling over the historical period can be expected from a zero volcanic baseline (Fyfe et al., 2021). Therefore, making robust conclusions on differences in natural forcings between CMIP5 and CMIP6 is difficult. However, other than in periods of high volcanic activity, differences in temperature response to natural forcings between CMIP5 and CMIP6 are small (Figure 1e), with CMIP5 tending to be marginally cooler in the ensemble mean. Figure 1g shows that the temperature response to other anthropogenic forcing is slightly more negative in CMIP6 than in CMIP5 in the late 20th Century. Using the CMIP6 *hist-aer* experiment (dark blue lines in Figures 1g and 1h) it is seen that aerosol forcing is the dominant contributor to the anthropogenic residual forcing. Differences between ΔF_{other} and ΔF_{aer} are mostly due to ozone (positive) and land-use change (negative), with ozone effects more dominant (Smith et al., 2020) and resulting in ΔT_{other} being slightly above ΔT_{aer} toward the end of the historical period (Figure 1g). Assuming that non-aerosol contributions to forcing are similar between CMIP5 and CMIP6, implied aerosol forcing and the climate response to aerosol forcing is marginally stronger in CMIP6 than in CMIP5 over most of the 20th Century, from ~1900 to 1990, before a post-1990 recovery in aerosol forcing in CMIP6 models brings it more into agreement with CMIP5 around 2000. Overall, CMIP6 and CMIP5 model temperatures start to diverge around 1940 (Figure 1a), where aerosol-induced cooling starts to become stronger in CMIP6 than in CMIP5 (Figure 1g), and this difference starts to become clear once the lower warming response to GHG forcing in CMIP6 from around 1960 emerges (Figure 1c).

Table 2 summarizes the implied ERF and temperatures in CMIP6 and CMIP5 using 1995–2005 means relative to pre-industrial. Here, it can be determined that CMIP6 models have a higher implied climate

Table 2
Implied Effective Radiative Forcing (ERF) and Temperatures in 1995–2005 Relative to Preindustrial Controls in Coupled Model Intercomparison Project Phase 5 (CMIP5) and CMIP6

	CMIP6			CMIP5		
Forcing	ΔF (W m^{-2})	ΔT (K)	KW^{-1} m^2	ΔF (Wm^2)	ΔT (K)	KW^{-1} m^2
historical	1.33 ± 0.38	0.67 ± 0.24	0.50	1.67 ± 0.57	0.80 ± 0.32	0.48
hist-GHG	1.98 ± 0.27	1.24 ± 0.26	0.63	2.28 ± 0.46	1.31 ± 0.24	0.57
hist-nat	0.12 ± 0.12	0.01 ± 0.09		0.15 ± 0.15	0.01 ± 0.10	
hist-otheranthro	0.82 ± 0.37	0.58 ± 0.31	0.70	0.83 ± 0.36	0.55 ± 0.31	0.66
hist-aer	0.93 ± 0.31	0.66 ± 0.27	0.71			

Note. Natural climate sensitivity is not calculated due to small signal-to-noise ratios. Uncertainty ranges are one standard deviation.

sensitivity ($\Delta T/\Delta F$) than CMIP5 models. This enables a lower implied GHG forcing in CMIP6 to elicit a similar temperature response to CMIP5, and a similar negative other anthropogenic forcing to elicit a stronger negative temperature response. We show in Figure S3 that in both CMIP5 and CMIP6, models with a higher historical climate sensitivity have weaker positive GHG implied ERF and stronger negative aerosol ERF. For GHG implied ERF the relationship is significant at the 5% level.

The differences in implied climate sensitivity between the historical, GHG, and non-GHG anthropogenic experiments may hint at different efficacies to differing forcing agents (Marvel et al., 2016). This is difficult to conclude, as the different forcers have different trajectories (greenhouse gas forcing is monotonically increasing whereas non-GHG anthropogenic forcers are more stable and in the latter case closer to equilibrium; compare the historical climate sensitivities to $4\times CO_2$ in Table 1), and the 11-year analysis period is short.

3.3. Comparison of Implied ERF to Fixed-SST ERF

In Figure 2, the implied ERF and fixed-SST ERF are compared for the eight CMIP6 models where data is available. Taking ensemble mean values for the implied ERF, it can be seen that half of the models (GISS-E2-1-G, MIROC6, CNRM-CM6-1, and CanESM5) show good agreement between the implied ERF and the fixed-SST ERF, whereas the other half (HadGEM3-GC31-LL, IPSL-CM6A-LR, GFDL-CM4, and NorESM2-LM) show distinctly lower estimates for implied ERF.

The implied ERF is calculated based on a time-invariant climate feedback λ. In many climate models, λ is not constant, becoming less negative (feedbacks increasing in strength) over time (Armour, 2017; Rugenstein et al., 2020). We show that models where agreement between the implied and fixed-SST ERF is less good are those with more curvature in their Gregory plots implying a stronger time-dependence of the climate feedback parameter (Text S2, Figures S1 and S2). However, while non-constancy of λ leads to low-biased estimates of ERF using this implied method, it does capture the trend of forcing projected by the fixed-SST method, further evidenced by high correlation coefficients between the two methods (Figure S1), and is appropriate to use the implied ERF method providing CMIP5 models behave similarly. We believe a similar curvature of λ exists in CMIP5 models as discussed in Section 4.

3.4. Comparison of CMIP5 and CMIP6 Forcing in CanESM5

Fyfe et al. (2021) showed temperature projections from CanESM5 forced with both CMIP6 and CMIP5 forcings. They showed that CMIP5 forcing resulted in cooler late-20th Century temperature projections than CMIP6 forcing in CanESM5, and that this difference was due to aerosol forcing. This result is in contrast to the multimodel CMIP5 and CMIP6 means shown in Figure 1. CanESM5 has the advantage that it shows relatively little time-dependence of climate feedback parameter and does not severely underestimate the fixed-SST ERF using the implied ERF method (Table S6, Figure 2d).

We reproduce the results of Fyfe et al. (2021) for temperature contributions and also show the implied forcing from CMIP5 and CMIP6 forcing agents in CanESM5 (Figure 3). Curiously, and in agreement with Fyfe et al. (2021), stronger aerosol forcing is implied with the CMIP5 forcing than the CMIP6 forcing, which is also the major contributor to lower overall implied forcing during the late 20th Century in the CMIP5 forcing. This results in suppressed warming in CMIP5 forcing simulations in this model, whereas CMIP6 forcing follows observed temperatures well until 2000 (Figure 3b). Additionally in CanESM5, large differences are present in the temperature response to natural forcings that contribute to the overall 20th Century suppressed CMIP5 warming, due in part to the CanESM5 CMIP5 pre-industrial control setup that used no background volcanic forcing (Fyfe et al., 2021). While parallel CMIP5 forcing experiments in more than one CMIP6 model are required to make more general conclusions, it may not necessarily be the case that aerosol forcing itself is stronger in CMIP6 models, but the CMIP6 model generation responds to aerosol forcing differently.

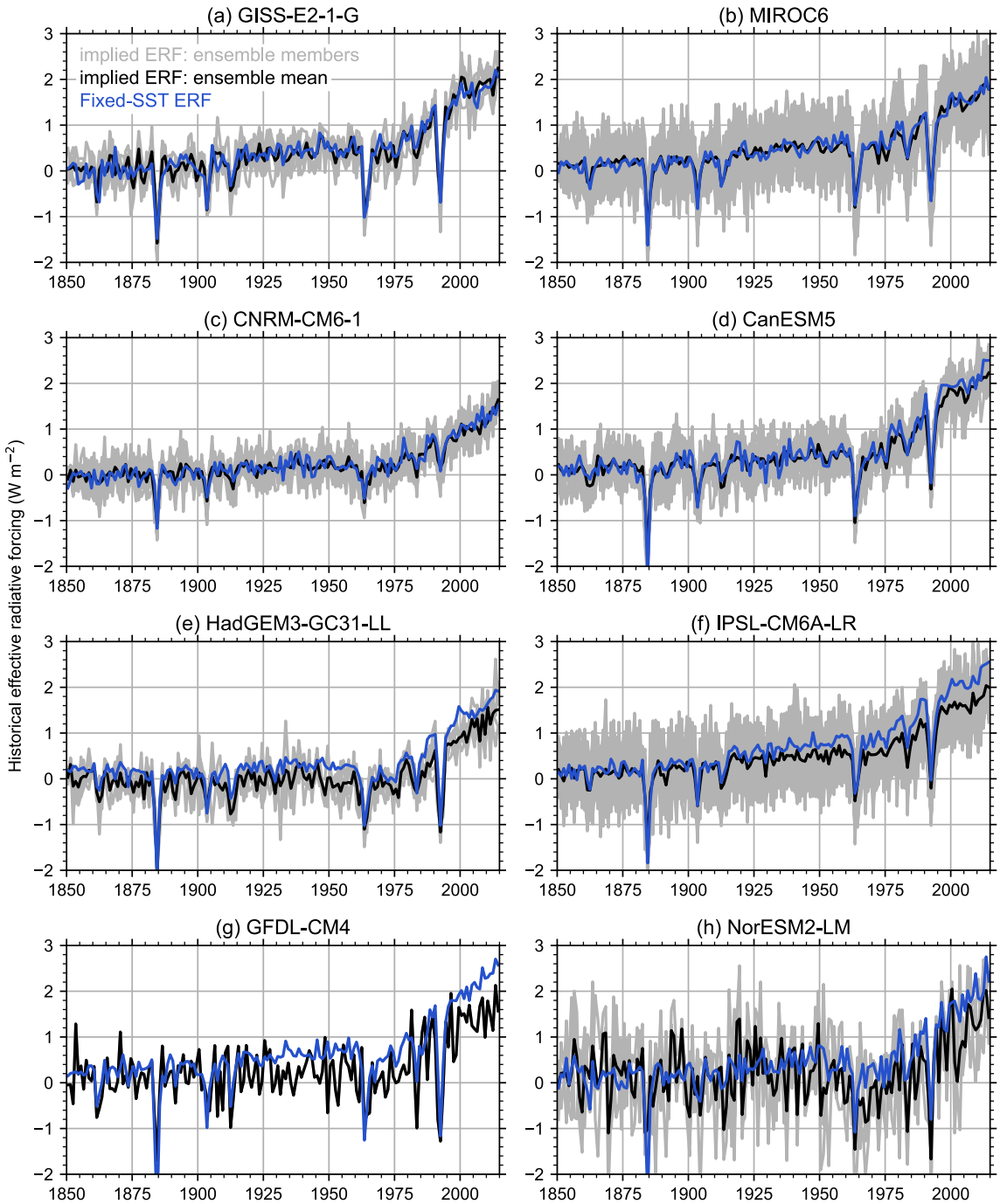

Figure 2. Comparison of the implied effective radiative forcing (ERF) (individual ensemble members in gray with ensemble mean in black) with the fixed-sea-surface temperature (SST) ERF (blue) for eight Coupled Model Intercomparison Project Phase 6 (CMIP6) models that performed both experiments. Models are ordered from best to worst in terms of their correspondence between fixed-SST ERF and implied ERF (Table S6).

4. Discussion and Conclusions

We offer an explanation to the cooler temperature projections in CMIP6 models compared to CMIP5 over the 1960–2000 period. Greenhouse gas forcing is weaker (less positive) and aerosol forcing slightly stronger (more negative) in CMIP6 compared to CMIP5. This leads to lower total historical forcing in CMIP6 compared to CMIP5. While the increasing climate sensitivity in CMIP6 models compared to CMIP5 mostly (but not quite) cancels out the lower implied GHG forcing, it accentuates the stronger negative aerosol forcing.

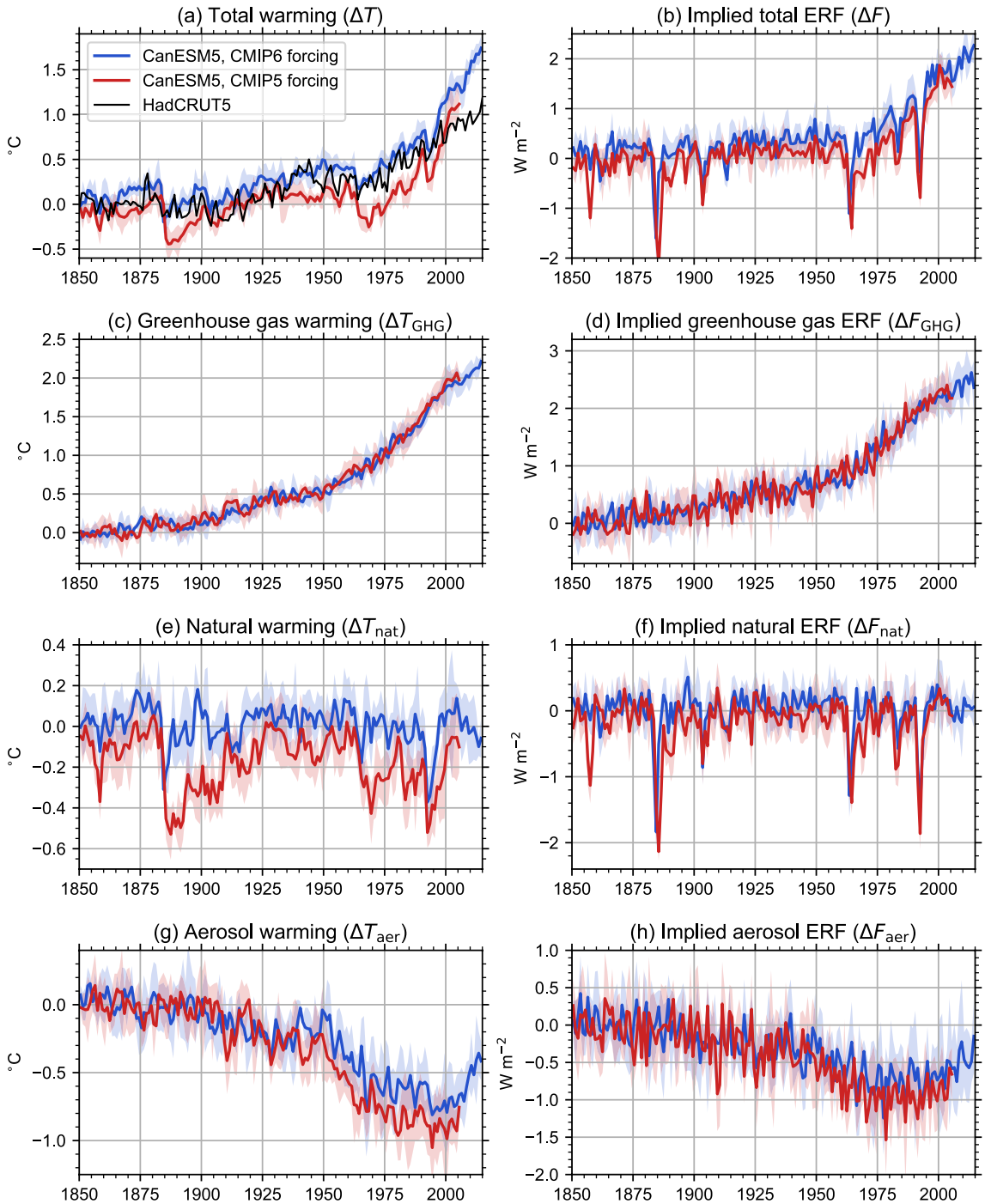

Figure 3. Implied effective radiative forcing (ERF) (left) and temperature anomalies (right) from the CanESM5 ensemble for both CMIP5 forcing (red) and Coupled Model Intercomparison Project Phase 6 (CMIP6) forcing (blue). Shown are the response to all historical forcings (top), greenhouse gas forcing (second row), natural forcing (third row), and aerosol forcing (bottom row). Shaded regions are one standard deviation across the five ensemble members.

Furthermore, there is a negative and significant relationship between climate sensitivity (the reciprocal of climate feedback) and implied GHG forcing, and a positive (though not significant) relationship between climate sensitivity and non-GHG anthropogenic (mostly aerosol-driven) ERF (Figure S3). Therefore, the resultant suppressed warming in CMIP6 over 1960–2000 is a combination of forcing and feedback effects.

Our forcing results are based on the implied ERF, and rely on the assumption that the differences in implied ERF between CMIP5 and CMIP6 adequately represent the true ERF. For CMIP6 we show that the implied ERF is close to the fixed-SST ERF in some models and under-predicts the fixed-SST ERF in others, showing that models with a greater time-dependence of climate feedbacks are those where ERF is under-predicted. It is not possible to produce a similar comparison in CMIP5 models as fixed-SST historical ERF experiments were not performed. While acknowledging the approximate nature of the implied ERF due to the reliance on a constant λ assumption, for useful comparisons between implied ERF in CMIP5 and CMIP6 it is sufficient to determine that a similar time-dependence of λ is present across the two CMIP generations. Zelinka et al. (2020, their Figure S11) shows that CMIP5 models have slightly more time-dependence in their climate feedbacks than CMIP6 models, suggesting that all other factors being equal, differences between implied and fixed-SST ERF in CMIP5 models would be greater than in CMIP6 models. Our results show that CMIP5 models have a higher implied ERF than CMIP6, the opposite of what would be expected from the action of the curvature in λ alone. Another line of indirect evidence for similar time-dependence in CMIP5 and CMIP6 exists from using the efficacy of deep-ocean heat uptake ε (Held et al., 2010; Winton et al., 2010) as a proxy for curvature in λ. When *abrupt-4xCO2* runs are calibrated to a two-layer simple climate model (Geoffroy et al., 2013), the multimodel mean efficacy of deep-ocean heat uptake is similar in CMIP5 ($\varepsilon = 1.28$) (Geoffroy et al., 2013) and CMIP6 ($\varepsilon = 1.29$) (Smith et al., 2021). Therefore, we tentatively rule out differences in λ curvature as a confounding factor in any CMIP5 to CMIP6 difference. Even in those models where implied ERF and fixed-SST ERF diverge in Figure 2, implied ERF matches the general trend of fixed-SST ERF and we conclude that implied ERF is a good-enough approximation to compare forcing behavior across CMIPs.

One limitation of the attribution experiments is the number of available models to perform the analysis: 13 out of 35 for CMIP6 and 13 out of 27 for CMIP5. Therefore, our conclusions about the sources of the differences between CMIP5 and CMIP6 forcing and temperature response could be made more robust with the addition of more models running the DAMIP *hist-GHG*, *hist-nat*, and *hist-aer* experiments. As shown by the single-model ensemble from CanESM5, it is also unclear whether differences in the forcing and temperature response between CMIP5 and CMIP6 really are due to changes in the forcing datasets or in the model responses. More CMIP6 models running the additional CMIP5 forcing experiments will help determine whether forcing and temperature differences between CMIP5 and CMIP6 are a consequence of the models or the forcing. Nevertheless, the all-historical forcing differences between CMIP5 and CMIP6 are clear to see, with a clear difference in the implied forcing in the late 20th Century between CMIP5 and CMIP6. CMIP6 has a substantially lower total forcing in the late 20th Century, impacting the temperature response.

This work reiterates the importance of getting the ERF in models correct, and more models diagnosing the ERF using the fixed-SST method by running the RFMIP *piClim-histall* experiment would enable more analysis of historical forcing to be undertaken, potentially setting a valuable precedent for comparison with future phases of CMIP models. Forcing drives temperature response, and if forcing is not correct, temperature projections in the past, and the future, will be biased.

This work supports the use of post-1980 warming to constrain climate sensitivity (Flynn & Mauritsen, 2020; Tokarska et al., 2020) as this trend is both sensitive to climate feedbacks and less sensitive to aerosol forcing. However, it also suggests that the 1960–2000 trend could be potentially useful for constraining aerosol forcing.

Data Availability Statement

Data and code are available at https://doi.org/10.5281/zenodo.5154111.

Acknowledgments
The authors acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP. They thank the climate modelling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP and ESGF. They thank Mark Zelinka for making climate feedback data from CMIP6 and CMIP5 models available (Zelinka, 2021). C. J. Smith was supported by a NERC/IIASA Collaborative Research Fellowship (NE/T009381/1). P. M. Forsterwas supported by the European Union's Horizon 2020 Research and Innovation Programme under grant no. 820829 (CONSTRAIN).

References

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., et al. (2020). Historical simulations with HadGEM3-GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, *12*(6), e2019MS001995. https://doi.org/10.1029/2019MS001995

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

Bjordal, J., Storelvmo, T., Alterskjær, K., & Carlsen, T. (2020). Equilibrium climate sensitivity above 5°C plausible due to state-dependent cloud feedback. *Nature Geoscience*, *13*(11), 718–721. https://doi.org/10.1038/s41561-020-00649-1

Dittus, A. J., Hawkins, E., Wilcox, L. J., Sutton, R. T., Smith, C. J., Andrews, M. B., & Forster, P. M. (2020). Sensitivity of historical climate simulations to uncertain aerosol forcing. *Geophysical Research Letters*, *47*(13), e2019GL085806. https://doi.org/10.1029/2019gl085806

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

Flynn, C. M., & Mauritsen, T. (2020). On the climate sensitivity and historical warming evolution in recent coupled model ensembles. *Atmospheric Chemistry and Physics*, *20*, 7829–7842. https://doi.org/10.5194/acp-2019-1175

Forster, P., 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*, *118*(3), 1139–1150. https://doi.org/10.1002/jgrd.50174

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

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*, *118*(23). https://doi.org/10.1073/pnas.2016549118

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., & Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *Journal of Climate*, *26*(6), 1859–1876. https://doi.org/10.1175/JCLI-D-12-00196.1

Gillett, N. P., Kirchmeier-Young, M., Ribes, A., Shiogama, H., Hegerl, G. C., Knutti, R., et al. (2021). Constraining human contributions to observed warming since the pre-industrial period. *Nature Climate Change*, *11*(3), 207–212. https://doi.org/10.1038/s41558-020-00965-9

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

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

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

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

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), 2418–2427. https://doi.org/10.1175/2009JCLI3466.1

Kramer, R. J., He, H., Soden, B. J., Oreopoulos, L., Myhre, G., Forster, P. M., & Smith, C. J. (2021). Observational evidence of increasing global radiative forcing. *Geophysical Research Letters*, *48*(7), e2020GL091585. https://doi.org/10.1029/2020GL091585

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

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., & Sanderson, B. M. (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., Buehler, S. A., Brath, M., Crevoisier, C., Jamil, O., Franklin Evans, K., & Tellier, Y. (2020). Benchmark calculations of radiative forcing by greenhouse gases. *Journal of Geophysical Research: Atmospheres*, *125*(23), e2020JD033483. https://doi.org/10.1029/2020JD033483

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

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

Shindell, D., & Smith, C. J. (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

Smith, C., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., & Regayre, L. A. (2018). FAIR v1.3: A simple emissions-based impulse response and carbon cycle model. *Geoscientific Model Development*, *11*(6), 2273–2297. https://doi.org/10.5194/gmd-11-2273-2018

Smith, C., Harris, G., Palmer, M., Bellouin, N., Collins, W., Myhre, G., et al. (2021). Energy budget constraints on the time history of aerosol forcing and climate sensitivity. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033622. https://doi.org/10.1029/2020JD033622

Smith, C., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., & Forster, P. M. (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

Stevens, B., Sherwood, S. C., Bony, S., & Webb, M. J. (2016). Prospects for narrowing bounds on earth's equilibrium climate sensitivity. *Earth's Future*, *4*(11), 512–522. https://doi.org/10.1002/2016EF000376

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., & Ziehn, T. (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

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

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

Zelinka, M. D. (2021). *Tables of ECS, effective radiative forcing, and radiative feedbacks from Zelinka et al GRL (2020), v1.9*. Retrieved from https://github.com/mzelinka/cmip56_forcing_feedback_ecs

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