Energy budget constraints on the time history of aerosol forcing and climate sensitivity

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access


It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1029/2020JD033622

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.
www.reading.ac.uk/centaur

CentAUR
Central Archive at the University of Reading
Reading’s research outputs online
Energy Budget Constraints on the Time History of Aerosol Forcing and Climate Sensitivity

C. J. Smith1,2, G. R. Harris3, M. D. Palmer4, N. Bellouin4, W. Collins4, G. Myhre5, M. Schulz6, J.-C. Golaz2, M. Ringer7, T. Storelvmo8, and P. M. Forster1

1Priestley International Centre for Climate, University of Leeds, Leeds, UK, 2International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, 3Met Office Hadley Centre, Exeter, UK, 4Department of Meteorology, University of Reading, Reading, UK, 5Center for International Climate and Environmental Research in Oslo (CICERO), Oslo, Norway, 6Norwegian Meteorological Institute, Oslo, Norway, 7Lawrence Livermore National Laboratory, Livermore, CA, USA, 8Department of Geosciences, University of Oslo, Norway

Abstract  An observationally constrained time series of historical aerosol effective radiative forcing (ERF) from 1750 to 2019 is developed in this study. We find that the time history of aerosol ERFs diagnosed in CMIP6 models exhibits considerable variation and explore how the time history of aerosol forcing influences the probability distributions of present-day aerosol forcing and emergent metrics such as climate sensitivity. Using a simple energy balance model, trained on CMIP6 climate models and constrained by observed near-surface warming and ocean heat uptake, we derive estimates for the historical aerosol forcing. We find 2005–2014 mean aerosol ERF to be −1.1 (−1.8 to −0.5) W m−2 relative to 1750. Assuming recently published historical emissions from fossil fuel and industrial sectors and biomass burning emissions from SSP2-4.5, aerosol ERF in 2019 is −0.9 (−1.5 to −0.4) W m−2. There is a modest recovery in aerosol forcing (+0.025 W m−2 decade−1) between 1980 and 2014. This analysis also gives a 5%–95% range of equilibrium climate sensitivity of 1.8°C–5.1°C (best estimate 3.1°C) with a transient climate response of 1.2°C–2.6°C (best estimate 1.8°C).

Plain Language Summary  There are two main human drivers of climate change: (a) Greenhouse gas emissions, which warm the planet; and (b) air pollution (aerosols) that offset some of this warming. Unfortunately, disentangling the effects of historical aerosol cooling is difficult based on the available observations. Therefore, we often use climate models to estimate how much aerosols have cooled the Earth since the start of the Industrial Revolution. Over the mid-to-late 20th Century, some climate models simulate less warming compared to 1850 than has been observed. This may be because aerosol cooling in some climate models is too strong. Our approach combines the relationships between aerosol emissions and their cooling effects on temperature from 11 climate models with simpler representations of the underlying physics. This simpler mathematical framework allows us to more fully account for uncertainty in both aerosol cooling and its effects on surface temperature and ocean heat uptake by running a much larger set of simulations. Our results suggest that the effect of aerosol cooling has only unwound slowly since 1980, and that it is difficult to determine how sensitive the climate is from this method.

1. Introduction

Aerosol effective radiative forcing (ERF) remains one of the most uncertain components of the present-day climate (Bellouin, Quaas, et al., 2020). Uncertainty in present-day forcing reduces our ability to confidently predict the future climate response to emissions (Forster et al., 2013) and the level of historical greenhouse gas warming masked by the cooling effect of aerosols (Samset et al., 2018). Aerosol forcing is the largest uncertainty governing future committed warming (Matthews & Zickfeld, 2012; Smith et al., 2019) and remaining carbon budgets consistent with Paris Agreement targets (Mengis & Matthews, 2020). In most future socio-economic scenarios, aerosol forcing is projected to become less negative over the 21st century (Gidden et al., 2019; Huppmann et al., 2018; Rogelj et al., 2018), promoting an increase in the rate of warming unless there is a concurrent reduction in greenhouse gas emissions (Shindell & Smith, 2019). The time history of aerosol ERF is a necessary input to many reduced-complexity climate models (Nicholls et al., 2020), which
in turn may be driven by simple emissions-to-forcing based relationships; these simple models find enormous utility when coupled to integrated assessment models (Huppmann et al., 2018).

Given its large uncertainty, aerosol forcing has remained an active research area. Several studies have quantified the aerosol ERF in the present day relative to pre-industrial based on observations, models, energy balance arguments, or a combination of approaches (Andrews & Forster, 2020; Bellouin, Quaas, et al., 2020; Boucher et al., 2013; Fiedler et al., 2019; Forest, 2018; Forest et al., 2002, 2006; Myhre, Shindell, et al., 2013b; Skeie et al., 2018; Smith et al., 2020; Zelinka et al., 2014). Fewer studies have attempted to diagnose a time series of historical aerosol forcing. Murphy et al. (2009) used observations of ocean heat uptake, surface temperature, and radiative forcing of long-lived greenhouse gases and volcanic eruptions since 1950 to determine the residual forcing, which was mostly attributed to aerosols. Skeie et al. (2011) and Lund et al. (2018) used chemistry-transport models with prescribed meteorology, evaluated at frequent time slice years since pre-industrial, to determine historical aerosol forcing since 1750. Shindell et al. (2013) used timeslices from 1850, 1930, 1980, and 2000 in full-complexity climate models to estimate the historical aerosol forcing. The most complete historical aerosol forcing time series, for 1750–2011, is given in Annex II (Prather et al., 2013) of the Intergovernmental Panel on Climate Change Working Group I Fifth Assessment Report (AR5), which takes in multiple lines of evidence including the Skeie et al. (2011) and Shindell et al. (2013) modeling studies.

Our goal is define an aerosol ERF time series from 1750 to 2019 that is consistent with energy balance constraints from observations; effectively to provide an update to AR5 Annex II that takes into account more recent evidence. Here we take a combination of the climate modeling and energy balance approaches. Under the Radiative Forcing (RFMIP) and Aerosol Chemistry (AerChemMIP) Model Intercomparison Projects, historically time-varying aerosol forcing can be diagnosed directly from CMIP6 models. However, in the multi-model mean, CMIP6 model simulations of global-mean surface air temperature (GSAT) are cooler than observations throughout the latter part of the 20th Century before recovering to near-present levels of warming today (Flynn & Mauritsen, 2020). One hypothesis is that aerosol forcing in the 20th Century may be too strong in some CMIP6 models, coupled with high transient climate response (TCR) that causes implausibly rapid recent warming in many models (Tokarska et al., 2020). Nevertheless, CMIP6 models remain an important line of evidence in determining historical aerosol forcing. Unlike for greenhouse gases, proxy records for aerosol forcing are sparse before widespread surface radiation measurements became available in the 1950s (Bellouin, Quaas, et al., 2020; Moseid et al., 2020). No global observations of aerosols were available until the satellite era (late 1970s), whereas CMIP6 models provide aerosol forcing estimates from 1850. Therefore, we use CMIP6 model forcing over the industrial era to inform our estimates of historical aerosol ERF, and “correct” for these responses by constraining the forcing estimates to observations of GSAT and Earth energy uptake (EEU).

2. Methods and Data

This section describes how the historical ERF time series are generated and how observational constraints are used with a simple energy-balance climate model to produce the best estimate and range of historical aerosol forcing estimates. A number of historical aerosol forcing time series are investigated. The primary focus of this study is an ensemble of timeseries generated from a simple relationship of global annual historical aerosol forcing estimates. Following the observational constraining, we refer to this time series as “CMIP6-constrained.” We also investigate replacing the CMIP6 emissions time series with scaled estimates of historical aerosol forcing from 11 CMIP6 models and one chemistry-transport model. This provides a total of 13 different historical aerosol forcing scenarios.

Throughout this paper, a probabilistic approach is taken, sampling 100,000 historical forcing timeseries per scenario with the same number of simple climate model configurations. Uncertainties in the non-aerosol components of historical forcing (greenhouse gases, land-use change, black carbon (BC) deposition on snow, aviation contrail and contrail cirrus, and natural forcings) are also taken into account. The resulting
GSAT and EEU time series from each ensemble member is compared to the observational constraints and a weighted posterior distribution produced of historical aerosol ERF.

2.1. Aerosol Effective Radiative Forcing Timeseries

2.1.1. CMIP6 Model Output

We start with 1850–2014 (or beyond) transient aerosol ERF derived from 11 GCMs (Figure 1; Table 1). Seven models were provided under RFMIP (Pincus et al., 2016), three under AerChemMIP (Collins et al., 2017), and one used a similar method to AerChemMIP but with Atmospheric Model Intercomparison Project (AMIP) sea-surface temperatures (SSTs) rather than model-diagnosed SSTs (Golaz et al., 2019). In all cases, aerosol ERF is diagnosed as the top-of-atmosphere radiation flux difference between parallel climate model experiments, one with time-varying aerosols, and one with pre-industrial aerosols (Table 2). In the RFMIP models, SSTs, sea ice, and non-aerosol forcings are given as a pre-industrial climatology in both the transient (CMIP6 name piClim-histaer) and control (piClim-control) experiments, with aerosols following the 1850–2014 (or 2100, in models running the SSP2-4.5 extension) emissions from CMIP6 in the transient run (Hoesly et al., 2018; van Marle et al., 2017). The pre-industrial control is a 30 yr climatology. In AerChemMIP models, we use historical (1850–2014) SSTs, sea-ice, and non-aerosol forcings with pre-industrial aerosols (histSST-pi-Aer) as the control. E3SM also followed this method (described in Golaz et al., 2019).

Figure 1 shows the aerosol ERF with respect to 1850 from CMIP6 models. Most models show a peak in negative aerosol forcing at some point between 1975 and 2010 before recovering in recent years.

2.1.2. Separation of Aerosol Components

For each CMIP6 model, the shortwave (SW) aerosol-radiation and aerosol-cloud interaction components of the ERF (ERFari_{SW} and ERFaci_{SW}) are determined using the Approximate Partial Radiative Perturbation (APRP) method (Taylor et al., 2007; Zelinka et al., 2014). The LW ERF from aerosol-cloud interactions (ERFaci_{LW}) was determined using the difference between all-sky and clear-sky forcing (difference in cloud radiative effect) with the LW ERF from aerosol-radiation interactions (ERFari_{LW}) estimated as the difference between ERF_{LW} and ERFaci_{LW}. The APRP is not exact and a small residual term arises that varies over time and by model (Figure S1), some of which is related to a small surface albedo adjustment (Ghan, 2013), but only the time-varying shapes and relative magnitudes of ERFari and ERFaci to each other are important for this decomposition.

For the RFMIP models, we calculate the APRP using the difference of each year of the piClim-histaer run against every year of the piClim-control run before averaging across the 30 piClim-control years to determine the ERFari and ERFaci from each year of 1850–2014 (or 2100). This method removes some nonlinearities.
Table 1
CMIP6 Models Providing Transient Historical Aerosol ERF Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>Long name</th>
<th>Model variants</th>
<th>Modeling protocol</th>
<th>Ensemble members</th>
<th>Time period</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CanESM5</td>
<td>Canadian Earth System Model, version 5.0.3</td>
<td>r1i1p2f1</td>
<td>RFMIP</td>
<td>3</td>
<td>1850–2100</td>
<td>Swart et al. (2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>r2i1p2f1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>r3i1p2f1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3SM-1-0</td>
<td>U.S. Department of Energy (DOE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Energy Exascale Earth System Model (E3SMv1)</td>
<td>3 ensemble member pairs (non-ESGF)</td>
<td>AMIP and AMIP with pre-industrial aerosols</td>
<td>3</td>
<td>1870–2014</td>
<td>Golaz et al. (2019)</td>
</tr>
<tr>
<td>GFDL-ESM4</td>
<td>Geophysical Fluid Dynamics Laboratory ESM4.1</td>
<td>r1i1p1f1</td>
<td>AerChemMIP</td>
<td>1</td>
<td>1850–2014</td>
<td>Dunne et al. (2020)</td>
</tr>
<tr>
<td>GFDL-CM4</td>
<td>Geophysical Fluid Dynamics Laboratory CM4.0</td>
<td>r1i1p1f1</td>
<td>RFMIP</td>
<td>2</td>
<td>1850–2100</td>
<td>Held et al. (2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>r3i1p1f1</td>
<td></td>
<td></td>
<td></td>
<td>Kelley et al. (2020)</td>
</tr>
<tr>
<td>GISS-E2-1-G</td>
<td>Goddard Institute for Space Studies ModelE 2.1-G</td>
<td>r1i1p1f2</td>
<td>RFMIP</td>
<td>1</td>
<td>1850–2100</td>
<td>Williams et al. (2018)</td>
</tr>
<tr>
<td>HadGEM3-GC31-LL</td>
<td>Met Office Hadley Center Global Coupled Model 3.1</td>
<td>r1i1p1f3</td>
<td>RFMIP</td>
<td>3</td>
<td>1850–2099</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>r2i1p1f3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>r3i1p1f3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSL-CM6A-LR</td>
<td>Institut Pierre Simon Laplace Climate Model 6A (low resolution)</td>
<td>r1i1p1f1</td>
<td>RFMIP</td>
<td>1</td>
<td>1850–2100</td>
<td>Boucher et al. (2020)</td>
</tr>
<tr>
<td>MIROC6</td>
<td>Model for Interdisciplinary Research on Climate, version 6</td>
<td>r1i1p1f1</td>
<td>RFMIP</td>
<td>3</td>
<td>1850–2100</td>
<td>Tatebe et al. (2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>r2i1p1f1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>r3i1p1f1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRI-ESM2-0</td>
<td>Meteorological Research Institute Earth System Model version 2.0</td>
<td>r1i1p1f1</td>
<td>AerChemMIP</td>
<td>1</td>
<td>1850–2014</td>
<td>Yukimoto et al. (2019)</td>
</tr>
<tr>
<td>NorESM2-LM</td>
<td>Norwegian Earth System Model, version 2</td>
<td>r1i1p2f1</td>
<td>RFMIP</td>
<td>3</td>
<td>1850–2100</td>
<td>Seland et al. (2020)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>r2i1p2f1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>r3i1p2f1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKESM1-0-LL</td>
<td>UK Earth System Model</td>
<td>r1i1p1f2</td>
<td>AerChemMIP</td>
<td>1</td>
<td>1850–2014</td>
<td>Sellar et al. (2019)</td>
</tr>
</tbody>
</table>

Note. Runs extended beyond 2014 in RFMIP experiments followed an SSP2-4.5 forcing pathway.
Abbreviations: AerChemMIP, Aerosol Chemistry Model Intercomparison Project; RFMIP, Radiative Forcing Model Intercomparison Project.

Table 2
Experiment Definitions for the RFMIP and AerChemMIP Model Runs Used in This Study

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Experiment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFMIP</td>
<td>CMIP6 name piClim-histaer</td>
<td>CMIP6 name piClim-control</td>
</tr>
<tr>
<td></td>
<td>Pre-industrial SSTs and sea ice</td>
<td>Pre-industrial SSTs and sea ice</td>
</tr>
<tr>
<td></td>
<td>Pre-industrial non-aerosol forcing</td>
<td>Pre-industrial non-aerosol forcing</td>
</tr>
<tr>
<td></td>
<td>Historical aerosol forcing</td>
<td>Pre-industrial aerosol forcing</td>
</tr>
<tr>
<td>AerChemMIP</td>
<td>CMIP6 name histSST</td>
<td>CMIP6 name histSST-piAer</td>
</tr>
<tr>
<td></td>
<td>Historical SSTs and sea ice</td>
<td>Historical SSTs and sea ice</td>
</tr>
<tr>
<td></td>
<td>Historical non-aerosol forcing</td>
<td>Historical non-aerosol forcing</td>
</tr>
<tr>
<td></td>
<td>Historical aerosol forcing</td>
<td>Pre-industrial aerosol forcing</td>
</tr>
</tbody>
</table>

Abbreviations: AerChemMIP, Aerosol Chemistry Model Intercomparison Project; RFMIP, Radiative Forcing Model Intercomparison Project.
in the APRP (particularly in relation to the cloud fraction adjustment part of ERFaci) alongside minimizing the influence of internal variability. For the AerChemMIP models and E3SM-1-0, APRP was calculated using the parallel all-forcing and 1850-aerosol forcing AMIP ensemble members and averaged. In all cases where modeling groups provided more than one ensemble member, the APRP decomposition is calculated separately in each ensemble member and then averaged.

### 2.1.3. Forcing Emulator

Simple emissions-based relationships are then fit to the APRP-derived ERFari and ERFaci in each CMIP6 model:

\[
\text{ERFari} = \alpha_{SO2} E_{SO2} + \alpha_{BC} E_{BC} + \alpha_{OC} E_{OC},
\]

\[
\text{ERFaci} = -\beta \ln \left( \frac{1 + \frac{E_{SO2}}{s_{SO2}}}{\frac{E_{BC} + E_{OC}}{s_{BC+OC}}} \right),
\]

In Equations 1 and 2, \( E_{SO2}, E_{BC}, \) and \( E_{OC} \) refer to global annual total emissions in Tg yr\(^{-1}\) of SO\(_2\), BC, and organic carbon (OC), and \( \alpha_{SO2}, \alpha_{BC}, \alpha_{OC}, \beta, s_{SO2}, \) and \( s_{BC+OC} \) are scaling coefficients. \( \alpha \) values can be interpreted as the radiative efficiency of emission of each aerosol species. Strictly, Equations 1 and 2 represent radiative effects rather than radiative forcings. Radiative forcings are given by the difference of radiative effects in Equations 1 and 2 calculated between the emissions in the year of interest and the pre-industrial year (either 1850 or 1750).

Equation 1 follows from studies showing that ERFari scales linearly with emissions (Johnson et al., 2019; Lund et al., 2018; Mahajan et al., 2013; Rap et al., 2013). Equation 2 is an extension of the simple relationship of Stevens (2015) and is based on the understanding that the change in cloud albedo is logarithmic with sulphate burden, and that burden scales with emissions (Carslaw et al., 2013; Charlson et al., 1992). Including carbonaceous aerosol in Equation 2 represented by the sum of BC and OC emissions is useful as some CMIP6 models include the effects of BC and/or OC on the change in cloud condensation nuclei. The resulting aerosol-cloud interactions can be substantial, for example, a negative ERFaci to BC emissions in the MIROC6 model (Thornhill et al., 2021). Equation 2 is found to give a good heuristic approximation of global-mean ERFaci to a more sophisticated aerosol indirect effect model (Ghan et al., 2013) as shown in Smith, Forster, et al. (2018). The sum of BC and OC emissions is used following the original Ghan et al. (2013) aerosol indirect model which considers primary anthropogenic emissions to be BC + OC, and to limit the number of free parameters in Equation 2 to three.

\( \alpha_{SO2}, \alpha_{BC}, \alpha_{OC}, \beta, s_{SO2}, \) and \( s_{BC+OC} \) parameters in Equations 1 and 2 are estimated using a least squares curve fit of each CMIP6 model’s ERFari and ERFaci (Table 3). A multi-model mean emulation is performed where the ERFari and ERFaci from the 11 models is averaged and Equations 1 and 2 applied. The multi-model mean \( \alpha \) coefficients \((-2.5, +28.5, \) and \(-8.5 \) mW m\(^{-2}\) (Tg yr\(^{-1}\))\(^{-1}\) for SO\(_2\), BC, and OC respectively) are of similar magnitudes to the radiative efficiencies from Aerocom models (Myhre, Samset, et al., 2019) for SO\(_2\) and BC, and a little stronger for OC here. The radiative efficiency coefficients are negative for BC and positive for OC in the IPSL-CM6A-LR model. However, using coefficients derived from the AerChemMIP single-forcing experiments for BC, OC, and SO\(_2\) in IPSL-CM6A-LR-INCA gives a much less good fit to the historical aerosol forcing in IPSL-CM6A-LR than our fitted coefficients, and while the fitted coefficients may not be representative of the true physical behavior in this model, allowing these values as part of the prior sampling allows for a larger diversity of aerosol forcing time series. We do not attribute a nitrate forcing to avoid overfitting and because most models do not include the effects of nitrogen compounds on aerosol formation. In reality, nitrate formation may compete with sulphate formation for available ammonium. Model evidence suggests this is of limited importance historically (Thornhill et al., 2021), but may become more important in future scenarios where the nitrate to sulphate emissions ratio is projected to increase (Bellouin et al., 2011; Hauglustaine et al., 2014).

The best-fit values for \( s_{SO2} \) and \( s_{BC+OC} \) span several orders of magnitude. We treat these terms as shape parameters, describing how linear or logarithmic the change in ERFaci is with increasing anthropogenic SO\(_2\) and carbonaceous aerosol emissions. With large \( s_{SO2} \) and \( s_{OC+BC} \) values, a linear response in ERFaci to
emissions is exhibited (from the Taylor expansion of \( \ln(1 + x) \approx x \) for small \( x \)). The degree of logarithmic behavior that ERFaci exhibits to emissions differ considerably between GCMs (Wilcox et al., 2015), and the possibility that ERFaci may be linear with emissions in some CMIP5 models was discussed in Booth et al. (2018).

Figure 2 shows the emulated fits to each CMIP6 model from Equations 1 and 2 as colored curves with the APRP-derived forcing from the GCMs in gray using an 1850 reference. The gray curves in Figure 2 show that most models show a peak in ERFari that is weakening in recent years, whereas ERFaci is approximately constant up to 2014 or shows a slower weakening trend. It can be seen that the emulated relationships (colored curves) give good representations of each component of the aerosol forcing in each model. We extrapolate these CMIP6 model-specific emulations back to 1750 in each model, resulting in a small positive forcing in 1750 relative to 1850. Where models do not provide an SSP2-4.5 future projection, we also extend these time series forward using Equations 1 and 2. Finally, we re-base all emulated time series to a 1750 reference (Figure 3), where the impacts of the different shapes for historical aerosol forcing due to different parameter combinations are more clearly seen. One notable feature in all time series is an increased forcing between 2014 and 2015 in the emulated curves owing to a 16% reduction in global SO\(_2\) emissions from the CMIP6 historical to SSP2-4.5 data sets over 1 year.

### 2.1.4. Ensemble Generation

To simulate time-varying aerosol forcing we take probabilistic ensembles for both the magnitude and the shape of the historical aerosol forcing. To generate historical shapes, we take 100,000 samples of \( \alpha_{SO_2}, \alpha_{BC}, \alpha_{OC}, \beta_{SO_2}, \) and \( \beta_{BC,OC} \) based on their distributions from the 11 participating GCMs (Table 3). A joint kernel-density estimate of the \( \alpha \) coefficients is used to derive a distribution which is then sampled from for ERFari (Figures 4a–4c). Accounting for correlation between the coefficients maintains the connection that different aerosol species are often co-emitted. For ERFaci, we take \( \ln\left(\alpha_{SO_2}\right) \) and \( \ln\left(\alpha_{BC,OC}\right) \) from each model and derive a joint kernel distribution from these two parameters (Figure 4d). Logarithms of the values in Table 3 are used because of the total ERFaci in Equation 2 has a logarithmic relationship to emissions and individual model estimates.
Figure 2. Simple emissions-based fits from Equations 1 and 2 to ERFari (left), ERFaci (center), and total aerosol forcing (right) for 11 CMIP6 models. Gray curves show the relevant forcing components from each GCM diagnosed using the APRP method. APRP, Approximate Partial Radiative Perturbation; ERF, effective radiative forcing.
of these parameters span several orders of magnitude. $\beta$ is not a degree of freedom in this setup because the resultant ERFaci time series will be scaled as described below.

We combine the sampled $\alpha$ and $s$ coefficients with 100,000 samples of the absolute values of the 1850 to 2005–2015 ERFari and ERFaci from the process-based assessment in Figure 8 of Bellouin, Quaas, et al. (2020); (hereafter the "Ringberg assessment"), using the distributions that do not account for energy budget constraints. We run the emissions emulator in Equations 1 and 2 using a update of the historical CEDS emissions to 2019 (O’Rourke et al., 2020) for energy and industrial sectors and BB4CMIP (for biomass burning) under a historical + SSP2-4.5 assumption (the emissions are much less sensitive to the choice of scenario for biomass burning than for energy and industry). Critically, the updated CEDS emissions account for phenomena such as an earlier and more gradual reduction in SO$_2$ emissions from China than were in the original CMIP6 data set (Paulot et al., 2018), as well as other corrections, and avoids arbitrarily choosing a scenario for post-2014 emissions. Differences in these time series are plotted in Figure S2.

This process rescales the $\alpha$ coefficients and selects $\beta$ in each ensemble member consistent with the present-day ERFaci. These 100,000 sampled time series are then rebased to a 1750 baseline by subtracting the 1750 forcing from the 1850 forcing, producing 100,000 candidate historical time series of aerosol forcing for the period 1750–2019 that differ in shape and magnitude (Figure 5).

### 2.1.5. Scaled Model Forcing

Alongside the emissions-based aerosol forcing time series, we repeat the analyses using scaled historical ERFari and ERFaci from each CMIP6 model, where the shapes of the historical forcing derived from the APRP method (Figure 2) in each model are fixed, but the pre-industrial to present-day magnitudes are allowed to vary. We also use RFari and RFaci time series generated from the Oslo-CTM3 chemistry transport model (Lund et al., 2018, 2019) for 1750–2020 under historical + SSP2-4.5. In these 11 CMIP6 models plus Oslo-CTM3, the magnitude of 1850 to 2005–2015 forcing is allowed to vary according to the Ringberg assessment distributions, and the generated time series are extrapolated backwards to a 1750 baseline. Unlike the emissions time series run with the forcing emulator, the historical shapes of ERFari and ERFaci from these 12 model estimates are fixed.
Non-Aerosol Forcing Time Series

The non-aerosol component forcings are also generated from a 100,000-member Monte Carlo ensemble. As they are either less uncertain or smaller in magnitude (or both) than the aerosol forcing they are not the main focus of this paper but are generated to provide a consistent view of the total ERF, including uncertainty estimates, so that energy budget constraints can be applied. Detailed information is provided in Text S1 with a summary of the key data sources in Table 4.

Figure 4. Joint distributions of (a) $\alpha_{SO_2}$ and $\alpha_{BC}$, (b) $\alpha_{SO_2}$ and $\alpha_{OC}$, (c) $\alpha_{BC}$ and $\alpha_{OC}$, (d) ln$(s_{SO_2})$ and ln$(s_{BC+OC})$. The grayscale 2D hexbin histogram of points represents a density of the 100,000 drawn sample sets and the colored points are fits to each CMIP6 model.
2.3. Simple Climate Model

We use our 100,000 member ensemble of aerosol and non-aerosol forcings and run them in a two-layer energy balance model, including efficacy of deep-ocean heat uptake (Geoffroy, Saint-Martin, Bellon, et al., 2013; Geoffroy, Saint-Martin, Olivié, et al., 2013; Held et al., 2010). To perform this many simulations precludes the use of a comprehensive GCM and similar constrained Monte Carlo ensemble methods using reduced-complexity models have been done previously (Meinshausen et al., 2009; Smith, Forster, et al., 2018). The structural uncertainty associated with the choice of simple climate model has been found to have limited impact on global mean temperature projections in historical simulations (Nicholls et al., 2020). We choose to use the two-layer model due to its computational efficiency, inclusion of both EEU and GSAT, and its proven ability as a useful emulator of complex GCMs (Palmer et al., 2018). We use the formulation from the two Geoffroy et al. papers, hereafter G13a and G13b. The two-layer model can be written as

![Figure 5](image_url)

**Figure 5.** Sampled ERFari, ERFaci, and total aerosol ERF time series before constraint. Overplotted are three individual ensemble members (colored lines) that have present-day total aerosol forcing close to the ensemble median with different time histories. The 5%–95% and 16%–84% ranges of the ensemble are shown as light and dark gray bands with the ensemble median in black. ERF, effective radiative forcing.

### Table 4

**Non-Aerosol Forcing Present-Day Uncertainties and Key References**

<table>
<thead>
<tr>
<th>Forcing type</th>
<th>Description</th>
<th>Key references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-mixed greenhouse gases</td>
<td>Concentrations to radiative forcing; tropospheric adjustments</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Etminan et al. (2016), Gidden et al. (2019), Hodnebrog, Aamaas, et al. (2020); Hodnebrog, Myhre, et al. (2020); Meinshausen et al. (2017, 2020); Smith et al. (2020) and Smith, Kramer, et al. (2018)</td>
</tr>
<tr>
<td>Ozone</td>
<td>Analysis of CMIP6 models; precursor emissions to forcing</td>
<td>Skeie et al. (2020); Smith, Forster, et al. (2018) and Stevenson et al. (2013)</td>
</tr>
<tr>
<td>Other anthropogenic</td>
<td>Land use forcing (including irrigation); black carbon on snow; aviation contrails; stratospheric water vapor from methane oxidation</td>
<td>Bond et al. (2013), Ghimire et al. (2014), Lee et al. (2020), Myhre, Shindell, et al. (2013) and Sherwood et al. (2018)</td>
</tr>
<tr>
<td>Volcanic</td>
<td>Forcing from stratospheric aerosol optical depth</td>
<td>Global Volcanism Program (2013), Gregory et al. (2016); Kovilakam et al. (2020); Larson &amp; Portmann (2016) and Toohey &amp; Sigl (2017)</td>
</tr>
<tr>
<td>Solar</td>
<td>Total solar irradiance; tropospheric adjustments</td>
<td>Gray et al. (2009); Matthes et al. (2017); Smith, Kramer, et al. (2018) and Vieira et al., (2011)</td>
</tr>
</tbody>
</table>
\[ C_{\text{mix}} \frac{dT_{\text{mix}}}{dt} = F + \lambda T_{\text{mix}} - \epsilon \gamma (T_{\text{mix}} - T_{\text{deep}}), \]  
\[ \epsilon C_{\text{deep}} \frac{dT_{\text{deep}}}{dt} = \epsilon \gamma (T_{\text{mix}} - T_{\text{deep}}), \]

where \( C_{\text{mix}} \) and \( C_{\text{deep}} \) (W yr m\(^{-2}\) K\(^{-1}\)) are the heat capacities of the ocean mixed layer and deep ocean, \( T_{\text{mix}} \) and \( T_{\text{deep}} \) (K) are the respective layer temperature anomalies, \( \lambda \) (W m\(^{-2}\) K\(^{-1}\)) is the climate feedback parameter (using the convention that negative values indicate stabilizing feedbacks), \( F \) (W m\(^{-2}\)) is the ERF, \( \epsilon \) (dimensionless) is the efficacy of deep ocean heat uptake and \( \gamma \) (W m\(^{-2}\) K\(^{-1}\)) the heat transport between the two layers. The relatively small heat capacity of the land surface and atmosphere compared to the ocean means that it can be neglected and the GSAT anomaly is given by \( T_{\text{mix}} \).

All available 44 CMIP6 models that published both abrupt-4xCO\(_2\) and piControl simulations to the Earth System Grid Federation as of July 2, 2020 were used to tune the two-layer model using the method set out in G13a and G13b. Models that were available on different resolutions (e.g., NorESM2-LM and NorESM2-MM), and physical and Earth system models from the same group (e.g., CNRM-CM6-1 and CNRM-ESM2-1) were treated as separate models, but different physics versions of the same model were not (e.g., r1i1p1f1 and r1i1p3f1 from GISS-E2-1-G). The two-layer model is tuned to the GSAT and TOA radiation imbalance of each CMIP6 model's abrupt-4xCO\(_2\) run. There are five free parameters in the G13b model: \( \gamma, \epsilon, \lambda, C_{\text{mix}}, \) and \( C_{\text{deep}} \). Radiative forcing from a quadrupling of CO\(_2\) \( (F_{4x}) \) is also calibrated using this method with the abrupt-4xCO\(_2\) experiments. Table S1 sets out the parameters for each model and Figure S3 shows the temperature evolution in CMIP6 models for the model output and the simulated two-layer model fits for abrupt-4xCO\(_2\). The equilibrium climate sensitivity (ECS) is an emergent parameter from this model and is calculated as \( F_{4x} / 2\lambda \). The inclusion of the efficacy of ocean heat uptake in the two-layer model leads to different and often larger estimates of climate sensitivity than from a “Gregory regression” of TOA energy imbalance against global mean temperature anomaly for the first 150 yr of abrupt-4xCO\(_2\) (the so-called “effective” climate sensitivity, EffCS). The strengthening of climate feedbacks over time as temperatures approach equilibrium in abrupt-4xCO\(_2\) experiments results in the long-term ECS being in the region of 10%–30% larger than EffCS (Rugenstein et al., 2020). The version of the two-layer model that includes efficacy of ocean heat uptake captures this effect somewhat, and we cautiously refer to our derived climate sensitivity as ECS for this reason.

A joint kernel density estimate distribution of the six input parameters to the two-layer model are sampled based on the values resulting from the 44 CMIP6 model tunings (marginal distributions for each parameter are shown in Figure S4). The joint distribution allows for correlation between model parameters to be taken into account when sampling (Table S2).

Internal variability in GSAT is simulated by analyzing the piControl run in all available models (49 as of July 2, 2020). Global mean temperatures from the piControl simulations are detrended to remove any residual drift with the residuals around the long term trend taken to be the internal variability. The autocovariance matrix of these residuals in each model is constructed by regressing the time series with itself at lag 1 yr, lag 2 yr, ..., lag \( n \) years where \( n \) is the length of the model's piControl simulation. This input is then used as the covariance of a multivariate Gaussian distribution that governs the time-dependent internal variability of temperatures in each model. For each ensemble member simulated, one of the 49 CMIP6 models is selected at random, and 270 yr (1750–2019) of internal variability generated based on the underlying distribution in the selected GCM. This allows for a more realistic construction of internal variability than a memoryless process noting that in reality events such as ENSO tend to cluster warm and cool years, and also provides for the recreation of low-frequency long-term internal variability that is present in some models such as CNRM-ESM2-1 (Figure S5).

### 2.4. Observational Constraints

We use GSAT estimates from 1850 to 2019 and total Earth system energy uptake observations from 1971 to 2018 to constrain our simulations. To estimate GSAT we use the Cowtan & Way (2014) data set of infilled global-mean surface temperatures (GMSTs; near-surface temperatures over land and sea-ice and SSTs over
open ocean) for 1850–2019, multiplied by a time-varying ratio of GSAT/GMST from CMIP5 models under the historical and RCP8.5 pathways from 1861 to 2014 (Richardson et al., 2016). This ratio converges toward 1.08 for the recent past (Rogelj et al., 2019) and we extend this ratio forward to 2019. Both observations and model output are rebased to the 1850–1900 mean as a proxy for pre-industrial following the IPCC Special Report on 1.5°C. This results in a central estimate of GSAT warming of 1.10°C for 2010–2019 relative to 1850–1900 with a warming trend of 0.30°C per decade since 2010.

For observations of total EEU, we use data from the Global Climate Observing System (GCOS) observational data set that includes estimates from ocean heat uptake, cryosphere, atmosphere, and land surface (Von Schuckmann et al., 2020). The ocean has absorbed 89% of the total energy uptake in the Earth system since 1960 owing to its larger thermal capacity compared to other components of the Earth system. The two-layer model only tracks heat uptake into the ocean, but to be physically consistent with the total observed EEU we assume that the heat uptake of the atmosphere, cryosphere, and land is taken into account in the heat uptake of the mixed layer of the ocean. The GCOS data set extends back to 1960, but we focus on the period from 1971 to 2018 due to the limited coverage of deep ocean temperature observations before the advent of expendable bathythermographs (XBTs) in the late 1960s (Palmer, 2017). Our constraint is based on the agreement of EEU calculated in the two-layer model with the total EEU from 1971 to 2018 in GCOS of 358 ± 37 ZJ (1 s.d.).

Following the running of the two-layer model, each of the 100,000 ensemble members is assigned a weight $w_i$ based on how well it reproduces GSAT and EEU observations. The weighting is based on the model weighting technique of Knutti et al. (2017):

$$w_i = \exp \left( -\frac{r_{\text{GSAT},i}^2}{\sigma_{\text{GSAT},D}^2} - \frac{r_{\text{EEU},i}^2}{\sigma_{\text{EEU},D}^2} \right),$$  

(5)

where $r_{X,i}$ is a measure of how well the model reproduces observations for variable $X$ for ensemble member $i$, and $\sigma_{X,D}$ is the “radius of model quality” (Sanderson et al., 2015). The radius of model quality is a subjective choice and for both GSAT and EEU we use assessed uncertainties. For GSAT, $r_{\text{GSAT},i}$ represents the root-mean-square error of each ensemble member’s simulated temperature compared to observations and $\sigma_{\text{GSAT},D} = 0.12°C$ based on the “likely” (>66%) range of GMST from the 1850–1900 period to 2006–2015 in IPCC Special Report on Global Warming of 1.5°C. For EEU, we use $r_{\text{EEU},i}$ as the difference in EEU between the model ensemble member and the GCOS best estimate of 358 ZJ (1971–2018) and use the GCOS uncertainty of $\sigma_{\text{EEU},D} = 37$ ZJ. Unlike in Knutti et al. (2017) we do not downweight similar ensemble members.

Following calculation of each $w_i$, the ensemble weights are normalized such that $\sum w_i = 1$. Although subjective, our choices for the goodness-of-fit to the constraints ensure that GSAT and EEU have approximately equal influence on the total weighting given to each ensemble member (Figure S6). An ensemble member will receive a high weight if it is close to both observed GSAT and observed EEU resulting in fewer high-weight ensemble members than using a single constraint only. Results using just one constraint are reported in the supporting information.

3. Results

Figure 6 shows the aerosol ERF time series that best fit the observational constraints of GSAT and EEU for the CMIP6-constrained ensemble plus 12 climate models. Also shown are the projections of GSAT and EEU with the applied ensemble weighting. The weighted 5%–95% range from the CMIP6-constrained time series using both GSAT and EEU as constraints is shown as a gray band with the weighted mean as a gray line. The 1750–2019 aerosol ERF is determined to be $-0.90$ W m$^{-2}$ ($-1.55$ to $-0.35$ W m$^{-2}$ 5%–95% range), comprised from $-0.31$ ($-0.62$ to $-0.08$) W m$^{-2}$ for ER Farr and $-0.59$ ($-1.18$ to $-0.10$) W m$^{-2}$ for ER Faci. The central estimate of total aerosol forcing is equal to the $-0.9$ ($-1.9$ to $-0.1$) W m$^{-2}$ assessed for 1750–2011 in AR5, with a narrower “very likely” (>90%) range in this study.
With the CMIP6-constrained time series, aerosol ERF exhibits a slight recovery between 1980 and 2014 of +0.025 W m\(^{-2}\) decade\(^{-1}\). This is a lower aerosol recovery than seven of the 11 CMIP6 models, although the constrained 5\%–95\% range is wide (−0.074 to +0.111 W m\(^{-2}\) decade\(^{-1}\)) and includes the means from all but the UKESM1-0-LL model (Figure 7). These results indicate that a rapid aerosol forcing recovery is unlikely and not consistent with the energy budget constraints, but whether aerosol forcing has been strengthening, weakening, or stable in recent decades is not conclusive.

Figure 8 uses the GSAT and EEU constraints to show the present-day distributions of aerosol forcing. Alongside this, we use the ensemble weights to calculate distributions of ECS and TCR from the ensemble given the two-layer model parameter distributions and ensemble weights. To calculate TCR we take the approach of Jiménez-de-la-Cuesta and Mauritsen (2019) noting the TCR is approximately \(\frac{F_{\alpha}}{2(\lambda + \epsilon)}\). The mean, 68\% and 90\% range for these parameters along with their unconstrained (prior) distributions are shown in Table 5. All 13 historical time series are shown in Table S4 for both GSAT and EEU constraints, and in Tables S5 and S6 using only GSAT or EEU respectively.

A diversity of constrained present-day aerosol ERF distribution shapes is possible for the aerosol forcing from each model’s historical time evolution. In particular, there are a group of models (the two from...
Our results are consistent with the conclusions of previous studies that have attempted to constrain present-day aerosol forcing based on energy balance arguments. We find that very large negative values of pre-industrial to present-day aerosol ERF are inconsistent with observed warming and total Earth system energy gain (Andrews & Forster, 2020; Forest, 2018; Forest et al., 2002, 2006; Skeie et al., 2018; Smith, Forster, et al., 2018). Our best estimate 1750–2019 aerosol forcing of −0.90 W m\(^{-2}\) is similar to recent observationally constrained studies that put aerosol forcing in the −0.8 to −0.9 W m\(^{-2}\) range (Andrews & Forster, 2020; Andrews & Forster, 2020; Forest, 2018; Forest et al., 2002, 2006; Skeie et al., 2018; Smith, Forster, et al., 2018).

It is likely that regional effects are significant and aerosols emitted from different source regions affect the global energy balance in different ways which is not captured in a global emissions to forcing relationship (Kretzschmar et al., 2017). Including how global forcing changes to regional aerosol emissions could be an improvement to the forcing emulator, although may be difficult to implement for aerosol-cloud interactions in marine stratocumulus cloud decks which may be thousands of kilometers from the emissions source (Regayre et al., 2014). However, we show for the 11 GCMs that our relationship is trained on, the forcing emulator works well (Figure 2) and is useful as a first-order global relationship that incorporates sufficient flexibility in its forcing response to emissions.

4. Discussion and Conclusions

Comprehensive climate models are the best tools available for determining global aerosol forcing where other spatially complete lines of evidence do not exist, such as prior to the satellite era (~1980). However, model-derived aerosol forcing depends on a chain of processes, and ultimately on spatially resolved aerosol emissions time series that are themselves uncertain (Hoesly et al., 2018). It is not possible to determine here whether the tendency for some models to project strong aerosol forcing in the second half of the 20th Century and/or too much of a recent aerosol recovery, at least compared to what would be implied by the observational constraints, is due to model processes or uncertainties in the emissions.

It is likely that regional effects are significant and aerosols emitted from different source regions affect the grid of linear aerosol forcing trends for 1980–2014 simulated by the 100,000 member Monte Carlo ensemble (thin black histogram) and weighted after application of observational constraints in the CMIP6-constrained time series (thick gray histogram). The mean of these distributions is shown as gray and black lines above the histograms with 5 and 95 percentiles of the constrained and unconstrained distributions. The trends of each CMIP6 model’s aerosol forcing are shown as colored lines, calculated as a 35 yr regression from 1980 to 2014 and error bars showing 5%–95% confidence ranges in the slope of the regression. Numbers are provided in Table S3.

SMITH ET AL.
Skeie et al., 2018; Smith, Forster, et al., 2018, and the review of 19 inverse estimates in Forest (2018) of −0.77 W m\(^{-2}\). We also find that the most likely shape of recent (1980–2014) aerosol forcing is approximately constant or with a slightly positive trend, which is in line with reanalysis-derived estimates of aerosol RFari and RFaci (Bellouin, Davies, et al., 2020), and a rapid recovery in aerosol forcing is not likely.

**Table 5**  
*Ensemble Percentiles for Aerosol Forcing, ECS and TCR, for the CMIP6-Informed Constrained Distributions*

<table>
<thead>
<tr>
<th>Time period</th>
<th>5%</th>
<th>16%</th>
<th>Mean</th>
<th>84%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total aerosol ERF (W m(^{-2}))</td>
<td>1750–2019</td>
<td>−1.56</td>
<td>−1.26</td>
<td>−0.90</td>
<td>−0.54</td>
</tr>
<tr>
<td></td>
<td>1750–2010</td>
<td>−1.78</td>
<td>−1.50</td>
<td>−1.10</td>
<td>−0.70</td>
</tr>
<tr>
<td>ERFari (W m(^{-2}))</td>
<td>1750–2019</td>
<td>−0.62</td>
<td>−0.47</td>
<td>−0.31</td>
<td>−0.15</td>
</tr>
<tr>
<td></td>
<td>1750–2010</td>
<td>−0.77</td>
<td>−0.59</td>
<td>−0.40</td>
<td>−0.21</td>
</tr>
<tr>
<td>ERFaci (W m(^{-2}))</td>
<td>1750–2019</td>
<td>−1.18</td>
<td>−0.93</td>
<td>−0.59</td>
<td>−0.26</td>
</tr>
<tr>
<td></td>
<td>1750–2010</td>
<td>−1.36</td>
<td>−1.08</td>
<td>−0.69</td>
<td>−0.31</td>
</tr>
<tr>
<td>ECS (°C)</td>
<td>Constrained</td>
<td>1.76</td>
<td>2.15</td>
<td>3.10</td>
<td>3.94</td>
</tr>
<tr>
<td>TCR (°C)</td>
<td>Constrained</td>
<td>1.24</td>
<td>1.43</td>
<td>1.83</td>
<td>2.22</td>
</tr>
</tbody>
</table>


Abbreviations: ECS, equilibrium climate sensitivity; ERF, effective radiative forcing; TCR, transient climate response.
The shape of historical aerosol forcing time series - whether from our simple emulator or provided by a particular CMIP6 model - does not provide a constraint on the overall 1750–2019 aerosol forcing. However, the choice of aerosol forcing data set matters less for constraining ECS and TCR than it does for the shape or magnitude of present-day forcing (Figure 8). It is difficult to constrain the upper bound of ECS due to the heavy tail of the prior distribution, and 95th percentile values of ECS range from 4.3°C to 6.0°C depending on the aerosol forcing time series used (Table S4). Other studies also show that these constrained climate sensitivity distributions are sensitive to the priors used (Sherwood et al., 2020). For example, our prior sample space informed by CMIP6 models has very few ensemble members with ECS < 1.5°C and TCR < 1.0°C, both lower bounds of the respective “likely” range in AR5, and it is possible that this area of the distribution is undersampled. Note that we do not perform a full Bayesian analysis in this paper. However, our ECS estimate of 3.1°C (1.8°C–5.1°C) is in line with, although with wider uncertainty, than the Bayesian estimate of 3.2 (2.3°C–4.7°C) in Sherwood et al. (2020), which takes into account several lines of evidence in their assessment.

The two-layer model used in this study includes the efficacy of deep ocean heat uptake, which can partially account for the “pattern effect” (Andrews et al., 2018; Sherwood et al., 2020) in which evolving patterns of sea surface temperature change can influence estimates of climate feedback as warming approaches equilibrium. In the notation of Sherwood et al. (2020) the total effective climate feedback can be written \( \lambda - \Delta \lambda \) with \( \Delta \lambda \) the contribution from the pattern effect. Across the sample of CMIP6 two-layer model calibrations in this study, the pattern effect \( \Delta \lambda \) varies from \(-0.1\) to \(+0.6\) W m\(^{-2}\) K\(^{-1}\) (mean \(+0.2\) W m\(^{-2}\) K\(^{-1}\)) between 1980 and 2050 (assuming SSP2-4.5 forcing), similar to previous studies (e.g., Armour 2017). This forced contribution to the pattern effect arises through the efficacy of deep ocean heat uptake in the two-layer model, and occurs where \( \epsilon > 1 \) (Geoffroy, Saint-Martin, Bellon, et al., 2013) as it is in the majority of CMIP6 model calibrations. The pattern effect means that historically, the effective climate feedback tends to be slightly weaker than the equilibrium feedback, \( \lambda \). However, this historical pattern effect is not as large as that determined from observed SSTs from AMIP models (Andrews et al., 2018) of about \(+0.5\) W m\(^{-2}\) K\(^{-1}\), as the historical pattern effect also includes an unforced component that is related to internal variability (Zhou et al., 2021). If we included the unforced component of the pattern effect through a further adjustment to \( \Delta \lambda \), it is likely our derived historical aerosol forcing would be weaker. However, applying this historical pattern effect approach to our projections is not straightforward. Calculating \( \Delta \lambda \) this way requires a 30 yr moving window regression (Andrews et al., 2018), and historical simulations in CMIP6 are provided only to 2014 meaning \( \Delta \lambda \) can only be estimated until around 2000. \( \Delta \lambda \) is sensitive to volcanic eruptions and aerosol forcing (Gregory et al., 2020), and the historical SST record is only one realization. We therefore do not take the historical approach, but account for the pattern effect through the sampling of deep ocean heat uptake efficacy, and for internal variability through the temporal autocorrelation of piControl runs.

Our results suggest that the limited number of CMIP6 models considered here have a stronger aerosol forcing than may have actually occurred during the 20th Century and this effect may be responsible for the modest warming in the CMIP6 ensemble mean over this time period (0.24 ± 0.22°C for 1961–1990 relative to 1850–1900 in CMIP6 models compared to reconstructed GSAT observations of 0.39 ± 0.06°C over the same period; Figure S7). The diagnosis of historical aerosol forcing in more CMIP6 models to confirm or disprove this would be welcomed. Inclusion of uncertainties in historical emissions would be useful to determine whether this is a factor in suppression of warming. Re-running of GCMs with updated emissions inventories could determine how sensitive models are to emissions uncertainties, and the importance of regional effects. The time history of aerosol forcing and its present-day magnitude both constrain key climate system uncertainties such as climate sensitivity and the rate of recent warming (Tanaka & Raddatz, 2011). Reducing uncertainty in both will reduce uncertainty in climate projections.

**Data Availability Statement**

All code and data used in this paper are available from https://doi.org/10.5281/zenodo.4624765. The two-layer climate model and the APRO code is part of the climateforcing python package (v0.1.0) available from https://doi.org/10.5281/zenodo.4621059 and https://pypi.org/project/climateforcing/. Cowtan & Way temperature observations are available at https://www-users.york.ac.uk/~kdc3/papers/coverage2013/had4_krig_annual_v2_0_0.txt (accessed November 24, 2020). GCOS Earth system energy observations

References


