Unveiling Sulfate Aerosol Persistence as the Dominant Control of the Systematic Cooling Bias in CMIP6 Models: Quantification and Corrective Strategies

Jie Zhang[1,2*], Kalli Furtado[3*], Steven T. Turnock[4,5], Yixiong Lu[1,2], Tongwen Wu[1,2], Fang Zhang[1,2], Xiaoge Xin[1,2], Yuyun Liu[6]

1. *State Key Laboratory of Disaster Weather Science and Technology, CMA EMPC*
2. *CMA Earth System Modeling and Prediction Centre, China Meteorological Administration, Beijing, China*
3. *Centre for climate research Singapore, Singapore*
4. *Met Office Hadley Centre, Exeter, UK*
5. *University of Leeds Met Office Strategic (LUMOS) Research Group, University of Leeds, Leeds, UK*
6. *Center for Monsoon System Research, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*

Corresponding to: Jie ZHANG (jiezhang@cma.gov.cn) & Kalli Furtado (Kalli_FURTADO@nea.gov.sg)

Abstract

Including sophisticated aerosol schemes in the models of the sixth Coupled Model Inter-comparison Project (CMIP6) has not improved historical climate simulations. In particular, the models underestimate the surface air temperature anomaly (SATa) when anthropogenic sulfur emissions increased in 1960~1990, making the reliability of the CMIP6 projections questionable. Biases in cooling among the models are correlated with sulfate burden and the deposition of sulfur is the process responsible. Accordingly, we define a diagnostic tool, named Sulfur Assessment Metric for Earth system models (SAME), for model evaluation and improvement. We show that the SAME index determines the cooling biases. Reducing the biases to within the observational uncertainty is consistent with a physically plausible SAME of around 1.35 days, which is overestimated by almost all the CMIP6 models. Based on targeting a reduction of SAME, post-CMIP6 improvements to two models are shown to greatly improve SATa reproduction.

1. Introduction

Atmospheric aerosols have rapidly increased since the Industrial Revolution. Over this time period, the total aerosol effective radiative forcing (ERF) was dominated by the sulfate cooling effect, and offset a substantial portion of global-mean forcing from well-mixed greenhouse gases (IPCC, 2023). Without this historical aerosol ERF, the Paris Agreement's target of limiting global warming to 1.5^{o}C above pre-industrial levels would have already been missed in 2015 (Hienola et al., 2018). Similarly, stopping all present-day anthropogenic aerosol emissions is estimated to induce a global-mean surface heating of 0.5~1.1^{o}C (Samset et al., 2018). The year 2024 has been confirmed as the hottest year in human history, and was the first year to breach the 1.5^{o}C warming limit (Bevacqua et al., 2025). Moreover, recent years have seen temperature trends accelerate, which may be due to reductions in atmospheric aerosols, especially aerosols produced by commercial shipping (Hansen et al., 2025). Hence, it has been suggested that even small emissions in relatively pristine air have substantial effects, and better constraining the ability of global-climate models to predict aerosol effects may be crucial to obtaining reliable projections.

The observed temporal evolution of historical surface air temperature (SAT) is one of the major metrics used for evaluating the performance of climate models. However, the SAT anomalies in the CMIP6 models are systematically lower than was observed for the 1960~1990 period, whereas the CMIP5 models on average track the instrumental record quite well (e.g., Flynn and Mauritsen, 2020). The 1960~1990 period is referred to as the "pothole cooling period" (PHC) in our previous study (Zhang et al., 2021a), due to the 'pot-hole' shaped dip in SAT at that time, and in this study hereafter. The PHC period is coincident with the so-called Great Acceleration period, in which the human enterprise was boosted remarkably and led to global-scale impacts on Earth System functioning (Steffen et al., 2007). Recent studies hypothesized that aerosol forcing in CMIP6 is stronger than in CMIP5 and is responsible for the suppressed late 20th-century warming (Dittus et al., 2020; Smith and Forster, 2021).

The anomalous cooling points towards a problem with the sulfur cycle in recent

ESMs or the emissions data (Hardacre et al., 2021; Wang et al., 2021). Considering the importance of the sulfur cycle in historical aerosol ERF, we examine the sulfur related processes in eleven CMIP6 models with aerosol schemes in this study. All the models are forced with CMIP6 historical anthropogenic aerosol emissions (Hoesly et al., 2018), and therefore differences in their sulfate burdens are mainly due to different representations of the sulfur cycle in the models.

We will identify the key processes that determine sulfate-burden in these models, and introduce a simple index for measuring the level of activity in the sulfur cycle in the models on the global scale. This index (an effective diagnostic tool for global cycling of atmospheric sulfate) can be easily calculated from time series of global means only, without the need for complex diagnostics of the sulfur-cycle processes. We show that the index is strongly correlated with sulfate burden and anomalous cooling and has a clear physical interpretation that allows each model's sulfur cycle to be calibrated using historical temperature biases.

2. Model, data, and method

2.1 CMIP6 models and data

Table 1. Information of the eleven CMIP6 models with aerosol schemes.

Model	Country	Interactive Chemistry	Members	Reference
BCC-ESM1	China	Yes	3	Wu et al. (2020); Zhang et al. (2021b)
CESM2	US	No	11	Danabasoglu et al. (2020)
CESM2-FV2	US	No	3	Danabasoglu et al. (2020)
EC-Earth3-AerChem	European consortium	Yes	2	Döscher et al. (2021)
GFDL-ESM4	US	Yes	3	Dunne et al. (2020)
MIROC6	Japan	No	50	Tatebeet al. (2019)
MIROC-ES2L	Japan	No	30	Hajima et al.

				(2020)
MPI-ESM-1-2-HAM	Germany	Yes	3	Mauritsen et al. (2019)
MRI-ESM2-0	Japan	Yes	10	Yukimoto et al. (2019)
NorESM2-LM	Norway	Yes	3	Seland et al. (2020)
UKESM1-0-LL	UK	Yes	19	Sellar et al. (2019)

Eleven CMIP6 climate models with interactive aerosol schemes are utilized in this study, including seven models with interactive chemistry and four without (Table 1). The outputs from two CMIP6 experiments are used: (1) the historical experiment of climate change over the period 1850~2014, forced by time-varying external forcings that are based on observations of natural processes (e.g., solar activity, volcanic eruptions) and human-induced changes (e.g., greenhouse gas, aerosol emissions, land-use changes). All the available realizations for each model were used to minimize the uncertainty from internal variability in the climate system; (2) the 1pctCO2 simulations, in which CO_2 is gradually increased at a rate of 1% per year. The 1pctCO2 experiment is designed for studying model responses to CO_2 and is somewhat more realistic than rapidly increasing CO_2 such as in the abrupt-4×CO2 experiment.

Model outputs used in this study comprise surface air temperature (SAT) and five key sulfur-cycle variables: sulfate burden (loadSO4), sulfate wet deposition and sulfate dry deposition, sulfur-dioxide (SO_2) wet deposition and SO_2 dry deposition. For these sulfur-cycle variables, the inter-member variability within the historical experiment is substantially smaller than that of SAT. The standard deviation of loadSO4 in PHC across the 11 CESM2 members is only 4% of its interannual variability, compared to approximately 21% for SAT. Similar results are also evident in the 19 UKESM1 members, where the standard deviation of loadSO4 is 3% of its interannual variability, versus 32% for SAT. Therefore, given the relatively small inter-member variability in sulfur-cycle variables compared to their interannual fluctuations and to SAT variability, we utilize the first realization of the historical simulations and neglect inter-member differences for these sulfur-related quantities.

The monthly mean SAT from the Met Office Hadley Centre/Climatic Research Unit global surface temperature (HadCRUT) data version 5 from 1850 to 2014 is used for model evaluations (Morice et al., 2021). Considering the lack of long-term reliable observations in polar regions, we focus on SAT changes between 60°S to 65°N and the 'global' mean is calculated as the area-weighted mean in this latitudinal belt.

2.2 The Sulfur Assessment Metric for ESMs (SAME) index

Atmospheric sulfate concentrations are determined by the emission and oxidation of sulfate precursors, as well as deposition processes. Together these processes make up the atmospheric part of the Earth's sulfur cycle. Anthropogenic SO_2 emissions are the major source of sulfate aerosol over land in polluted regions. Given that the same anthropogenic SO_2 emissions are used in all the CMIP6 models, most of the differences in simulated atmospheric sulfate concentrations occur due to oxidation of SO_2 and sulfate deposition processes. Since much of the loss of SO_2 occurs locally to its emission source by oxidation and deposition, faster SO_2 deposition is associated with weaker SO_2 oxidation. For sulfate itself, the faster the sulfate deposition rate, the less the sulfate will reside in the atmosphere. That is, both the SO_2 deposition and sulfate deposition are important for the sulfate concentrations in the atmosphere, directly or indirectly.

Here we define the **S**ulfur **A**ssessment **M**etric for **E**SMs (**SAME**) index. The SAME index is calculated as the ratio of the sulfate burden anomaly and sulfur deposition anomaly in PHC, relative to preindustrial period, to mitigate the influence of differing model climatology. Sulfur deposition encompasses the deposition fluxes of sulfate aerosol and its major precursor SO_2:

SAME = loadSO4a / (DSO4a + DSO2a)

where:

- loadSO4a is the total sulfate loading anomaly in the atmosphere,
- DSO4a denotes the total (wet plus dry) sulfate deposition anomaly, and
- DSO2a denotes the total (wet plus dry) SO_2 deposition anomaly during the

PHC period.

2.3 The transient Climate Response (TCR) index

The transient Climate Response (TCR) index is calculated as the mean SAT anomaly of a 1pctCO2 simulation in a 20-year period centered on year-number 70, by which a doubling CO_2 concentration has occurred. It is an important metric representing CO_2-related historical warming and has been widely used for model evaluations and comparisons (e.g., Bevacqua et al., 2025; O'neill et al., 2016).

3. Results

3.1 SAT and sulfate burden

The historical evolutions of near-global mean (60^{o}S to 65^{o}N) SATa in the eleven CMIP6 models with interactive aerosol schemes are shown in Fig. 1a. All the models tend to underestimate SATa since the 1930s. The anomalous cooling in CMIP6 model marked a notable departure from earlier model generations, which can effectively capture the instrumental SAT record with observation falling well within model spread (e.g., Flynn and Mauritsen, 2020; Hegerl, et al., 2007).

The cooling bias is most pronounced from 1960 to 1990, i.e., the PHC period. The SATa in PHC is about 0.34^{o}C in the observations. However, the multi-model mean (MMM) SATa in the models is about 0.3^{o}C lower with a large model spread. The SATa ranges from -0.24^{o}C in EC-Earth3-AerChem to 0.19^{o}C in GFDL-ESM4 and MIROC6. Significant cooling is evident across the mid- and high-latitudes of the Northern Hemisphere, as illustrated by the SATa map in Fig. 1a. The sudden drop in SATa in the early 1960s and 1990s may be due to the stronger model responses to large volcanic eruptions, Mount Agung in 1963 and Mount Pinatubo in 1991, than in the observations (Chylek et al., 2020). The anomalous cooling biases gets smaller later in the simulations, which is related to the generally high sensitivity of the models to GHG forcing (Smith and Forster, 2021).

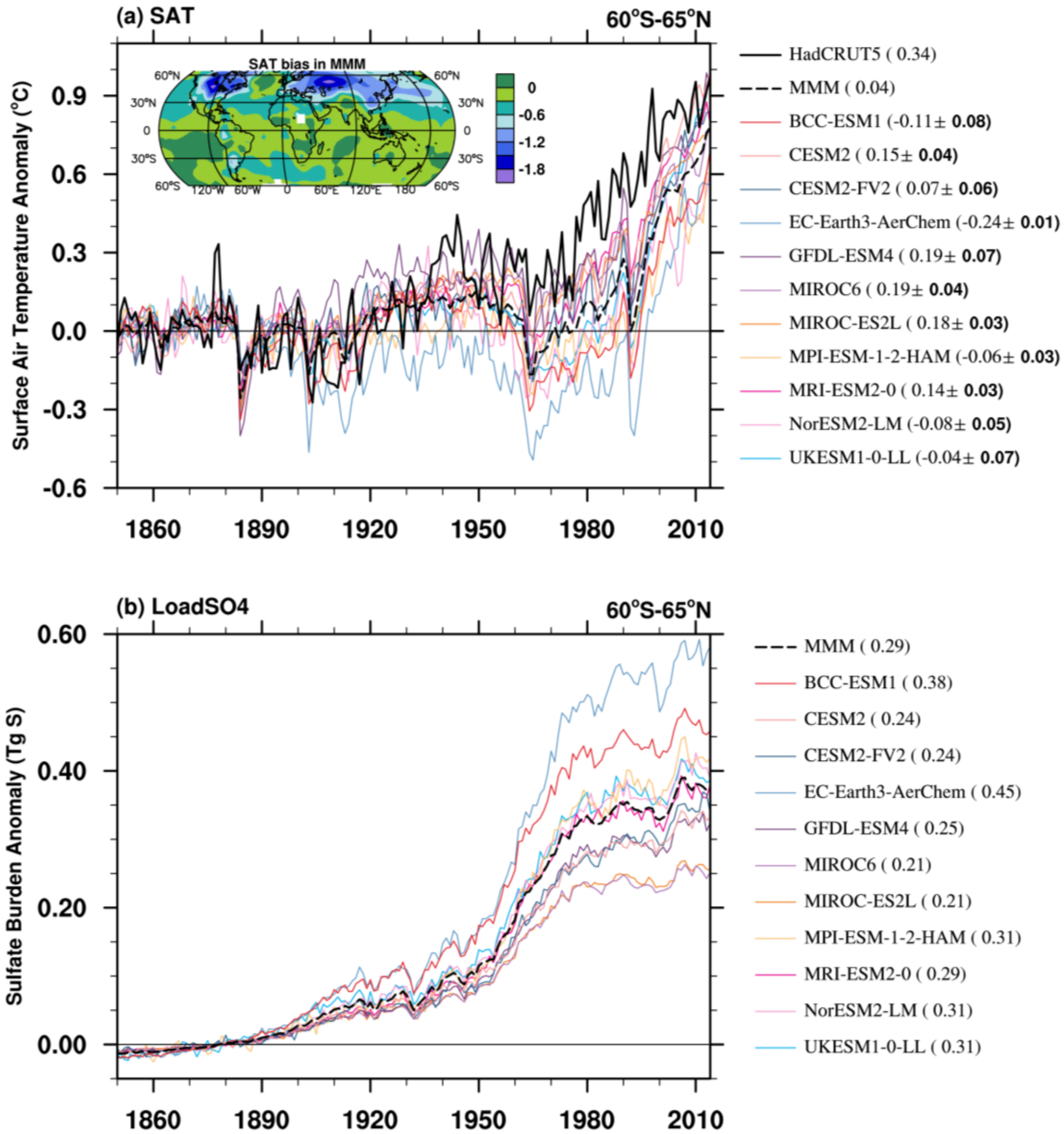

Figure 1. (a) Historical surface air temperature anomalies (SATa) relative to 1850~1900 mean for HadCRUT5 (thick black line), the ensemble mean for each CMIP6 model (solid color lines), and multi-model mean (MMM, dashed black line). The numbers in brackets are the mean results during the pothole period (1960~1990) together with the inter-member spread for each model. Units: °C. Panel (b) is the same as panel (a) but for sulfate burden anomalies for the first realization from each CMIP6 model (solid color lines) and MMM (dashed black line). Units: Tg S yr^{-1}.

The PHC coincides with increased anthropogenic emissions, particularly of sulfate precursors such as SO_2 (Zhang et al., 2021a). Global emissions of SO_2 grew steadily after the 1950s and peaked in the 1970s at 180Tg yr^{-1}, which is about 3.6 times the 1950s' emissions (Hoesly et al., 2018). As the main precursor of sulfate, the growing emission of SO_2 led to the accumulation of sulfate in the atmosphere, which interrupted a decades-long warming trend via the cooling effects of sulfate aerosols on climate, even though carbon-dioxide emissions continued to rise (Wilcox et al., 2013). Because of the emission control policies in Europe and North America (Hand et al., 2012; Vestreng et al., 2007), such as the Gothenburg Protocol (Eb, 1999) and the 1990 Clean Air Act Amendments in the U.S. (Likens et al., 2001), global anthropogenic SO_2 emissions were suppressed after the 1980s and SAT started to increase rapidly in the observations (Aas et al., 2019). However, anthropogenic SO_2 emission continued to increase over Asia due to industrial developments, although they have also decreased since 2006 in East Asia (Wang et al., 2021). Some of this decrease in SO_2 emissions at the beginning of the 21st century is not well represented in the CMIP6 emission inventory. But it is outside of the PHC period and the impact on SAT reproduction is beyond the scope of this paper.

In the 11 CMIP6 models, sulfate concentrations increased rapidly during the PHC period (Fig.1b). The intensified emission of anthropogenic SO_2 mainly comes from industries and the energy-transformation sector (e.g., Ohara et al., 2007; Vestreng et al., 2007). The SAT anomalies simulated by CMIP6 models are systematically lower than observations during the PHC period, indicating an excessively strong sulfate-induced cooling effect in CMIP6 models. The sulfate burden is the lowest in MIROC models (0.21 Tg S) with the smallest cooling bias（0.15^{o}C lower than HadCRUT5), and is doubled in EC-Earth3-AerChem (0.45 Tg S) with the largest cooling (0.58^{o}C lower than HadCRUT5). Generally, the models with higher sulfate burdens anomalies also tend to underestimate SAT anomalies the most. As shown in Fig. 2a, the correlation coefficient between anomalous sulfate burden and SAT during the PHC is -0.92, significant at the 1% level using a Student's *t*-test. Interactive chemistry may have an impact on sulfate formation and affect the sulfate aerosol burdens in the atmosphere

(Mulcahy et al., 2020). As shown in Fig.2a, models with interactive chemistry (color dots) seem to have higher sulfate burden anomaly and lower SATa than models without (color circles). However, the relationship between sulfate burden anomaly and SATa is consistent among models with and without interactive chemistry. That is, there is no obvious difference in relationship between sulfate burden anomaly and SATa for models with and without interactive chemistry.

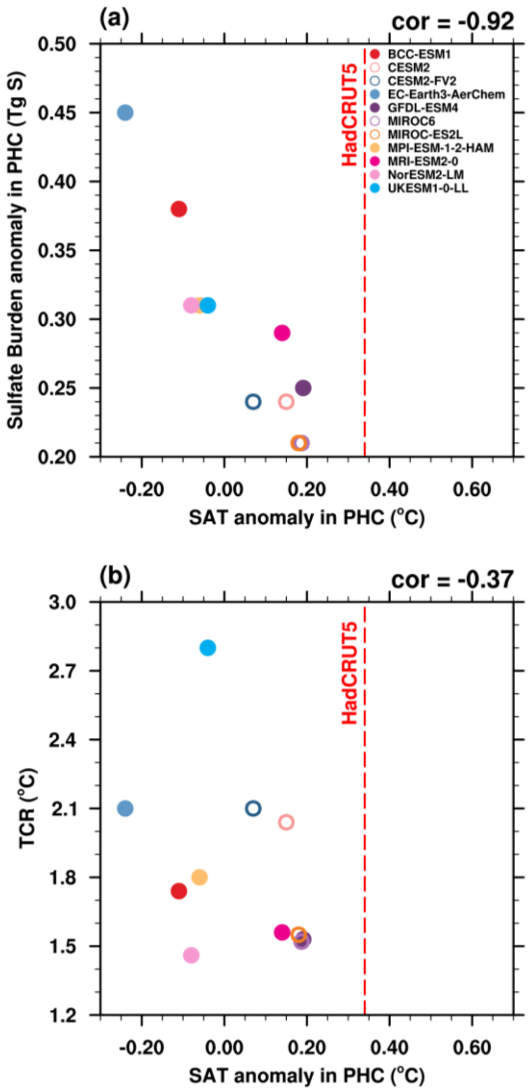

Figure 2. Scatter plots of SATa in PHC period (x-axis, °C) versus (a) sulfate burden anomaly in PHC period (y-axis, Tg S) in historical experiments, and (b) the transient climate response (TCR, °C) for each model calculated by 1pctCO2 experiments. The corresponding correlation coefficient (cor) is shown at the top-right corner of each panel. The anomalies are relative to 1850~1900 mean. Models with and without interactive chemistry are marked by color dots and color circles, respectively.

Greenhouse gases (GHGs) also show a rapidly increasing trend in the PHC period. However, TCR, which can generally indicate the impact of GHGs, is insignificantly correlated with SAT anomalies in CMIP6 models and the correlation coefficient is even negative (Fig.2b). Therefore, the biases of atmospheric sulfate burden and the associated sulfate aerosol cooling effects play an essential role in the anomalous-cooling biases in the CMIP6 models.

3.2 Sulfur deposition and a metric for the global sulfur cycle diagnostic (SAME)

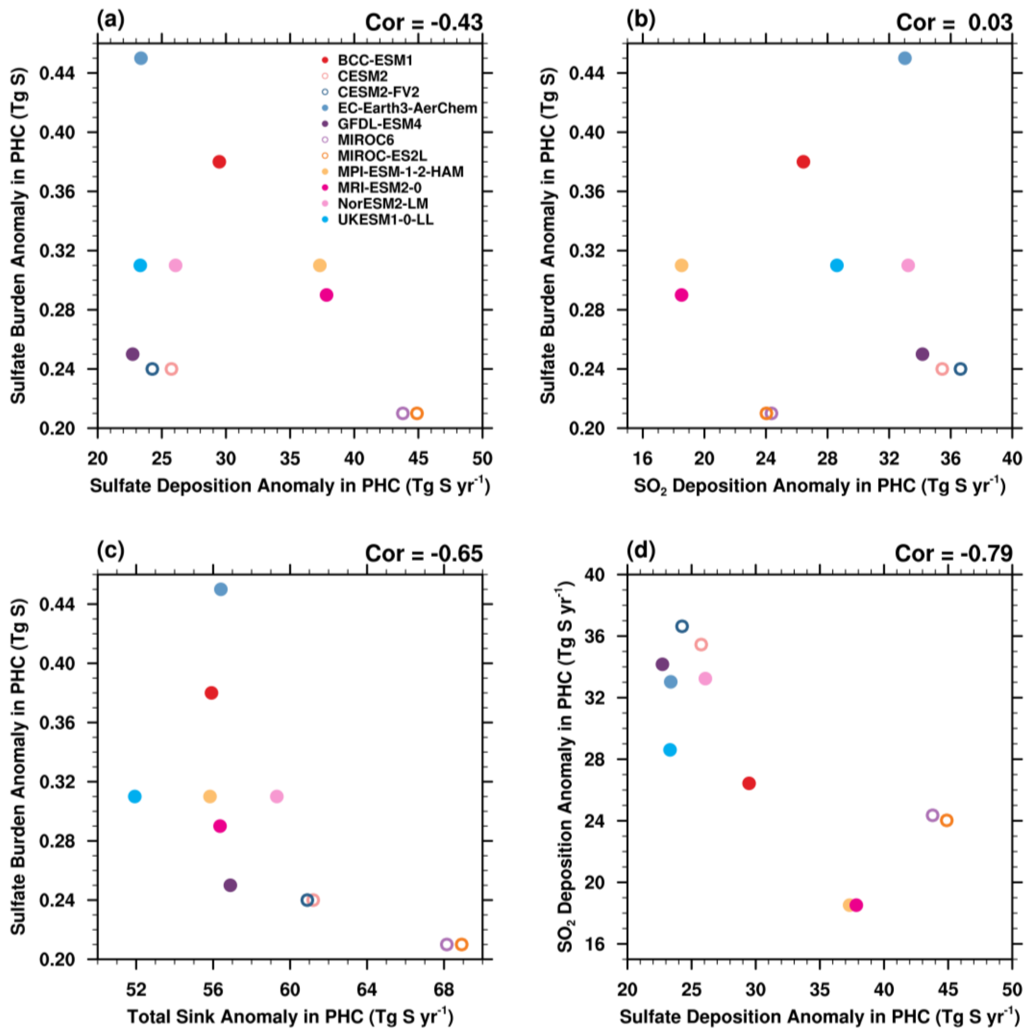

Figure 3. (a) Sulfate deposition anomaly, (b) SO_2 deposition anomaly, and (c) total sulfur sink (sulfate and SO_2 deposition) anomaly versus sulfate burden anomaly in PHC period (Tg S, y-axis). (d) Sulfate deposition anomaly (x-axis) versus SO_2 deposition anomaly (y-axis) in PHC period. Unit for deposition anomaly is Tg S yr^{-1}.

Fig. 3 shows comparisons of the global mean total sulfate burdens versus sinks of sulfur as anomalies in the PHC period relative to a baseline period of 1850~1900. As shown in Fig.3a, the sulfate burden anomaly is negatively correlated with sulfate deposition anomaly. However, the correlation is not statistically significant, partly attributable to a subset of five models characterized by low sulfate deposition and low sulfate burden. These models prevent the robustness of a robust linear fit derived from the remaining models. There is no clear statistical relationship between sulfate burden anomaly and SO_2 deposition anomaly (Fig. 3b). However, the correlation between sulfate burden anomaly and total sulfur sink (deposition of sulfate and SO_2) anomaly increases to -0.65, significant at the 5% level using a Student's t-test (Fig.3c). Notably, within the subset of five models exibiting both low sulfate deposition and low sulfate burden, most display higher SO_2 deposition in relative to the ensemble mean. The high SO_2 deposition compensates for their low sulfate deposition, making the total sulfur deposition magnitude sufficiently to sustain a significant correlation with sulfate burden. We can hypothesize that in these 5 models' oxidation of precursors to sulfate proceeds slower than in the other models. This is reflected by their larger SO_2 deposition rates, and leads to less sulfate in the atmosphere. That is, both the sulfate deposition and the SO_2 deposition (via its relationship with oxidation rates) are responsible for the sulfate burden anomalies, although the relative ratio of both deposition processes is different among the models. Further examination indicates that the anomalous SO_2 deposition rate among the models is highly negatively correlated with the anomalous sulfate deposition rate with correlation coefficient of -0.79 (Fig.3d). The total sulfur sink is examined and discussed hereafter.

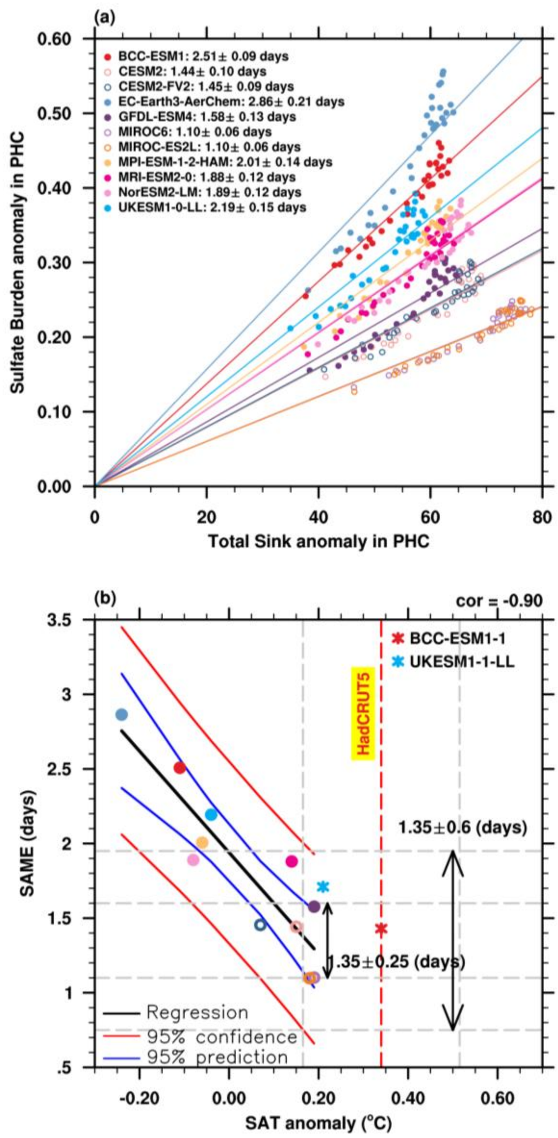

Figure 4. (a) Scatter plots of yearly total sulfur sink anomaly (x-axis, Tg S yr^{-1}) versus sulfate burden anomaly (y-axis, Tg S) in PHC period in relative to 1850~1900 mean. Number in legend shows the mean and standard deviation of ratio between sulfate burden anomaly and total sulfur sink anomaly in PHC period, defined as SAME, units: days. (b) The mean SATa (°C, x-axis) versus SAME (days, y-axis) in PHC period for each model. The black solid line is the linear fitting. The blue and red solid lines are the 95% confidence interval (CI) and 95% prediction interval (PI), respectively. SAT anomaly in HadCRUT5 and its 0.175°C boundaries are shown by the red dashed line and parallel gray dashed lines, respectively. The red and blue asterisks are the results in the two post-CMIP6 models BCC-ESM1-1 and UKESM1-1-LL, respectively.

Considering the importance of anomalous total sulfur sink to sulfate burden, Fig.4a examines their relationship during the PHC period in each model. Generally, the anomalous sulfate burden and total sink are positively correlated and co-vary almost linearly in all the models. The ratio between anomalous sulfate burden and total sulfur sink is defined as the SAME index in Section 2.2. The mean SAME in PHC ranges from 1.1 days in MIROC models to 2.86 days in EC-Earth3-AerChem. The SAME is generally longer in models with interactive chemistry (color dot) than without (color circle).

The standard deviation of SAME for each model in PHC ranges from 0.03 to 0.12 days, about 3.0% of the mean SAME. That is, although the sulfate burden increased significantly in the PHC period, the SAME hardly changed. This is an important sign that SAME is a robust index for evaluating the sulfur cycle in model development. Our finding that a single value of SAME is capable of characterizing the anomalous cooling for each model, makes it a convenient target for model tuning. Focusing on a single, representative parameter can make tuning more efficient and help to reduce the computation cost, especially when the model resolutions become relatively high. Moreover, because SAME has a clear physical interpretation as a globally defined efficiency factor for sulfur removal processes, tuning based on SAME can give confidence that SAT biases are reduced for a 'right' (i.e., physically sensible) reason.

3.3 The recommended SAME value

Tuning based on SAME requires an empirical best-estimate SAME value to aim for. Therefore, a further question is how to estimate the reasonable values for SAME. Here we try to constrain the SAME using the SATa in observations. Fig. 4b shows the SAME and SATa in PHC in each model. The SATa is highly correlated with SAME with a correlation coefficient of -0.90. The SAT anomaly in PHC is 0.34^{o}C, shown by the vertical red dash line (HadCRUT5). Considering the internal variability in the climate system and the uncertainty in observation, the observed uncertainty is suggested to be 0.175^{o}C (the vertical gray dash line parallel with the red dash line). The observed uncertainty is estimated as the standard deviation of observed annual mean globally averaged SAT in HadCRUT5 from 1850 to 2014 after removing the least squares linear trend. We calculate the linear fitting between SATa and SAME (black line in Fig. 4b), the 95% confidence interval (CI, blue curves), and the 95% prediction interval (PI, red curves), respectively. SATa in seven CMIP6 models falls beyond the observational range, which in the remaining four models, SATa closely approaches the lower bound of observation, giving a range of SAME between 1.1 to 1.58 days. That is, the SO_2 oxidation or deposition terms in CMIP6 models may need to be modified.

Here we use this metric to modify the sulfur cycle in BCC-ESM1, more specifically we quadruple the SO_2 dry deposition over land and multiply the SO_2 dry deposition over the ocean by 1.5. This effect is similar to that in UKESM1-0-LL by modifying SO_2 dry deposition parameterization (Hardacre et al., 2021; Mulcahy et al., 2023). The impact of changes to the SO_2 dry deposition parameterization in UKESM1-0-LL is an increase of SO_2 dry deposition by a factor of 2 to 4. As shown by the red asterisk in Fig.4b, the SAME reduced from 2.51 days to 1.43 days in updated BCC-ESM1 (BCC-ESM1-1), falling right within the PI constraint. The new SAME index is 57% of its previous values. Accordingly, the SATa in PHC is 0.34^{o}C, falling within the observational range from 0.165^{o}C to 0.515^{o}C. We also examine the SAME in UKESM1-1-LL with modified SO_2 dry deposition parameterization. The SAME is shortened from 2.19 days to 1.71 days, falling within the CI constraint. Accordingly, the SATa in PHC period increases by about 0.25^{o}C.

Given that most models underestimate SATa relative to observations, extrapolating SAME values for SATa exceeding the observation (0.34^{o}C) becomes highly uncertain. Result from BCC-ESM1-1 suggests that the rate of decrease in SAME predicted by the regression line may not hold for SATa values above the observed lower bound (0.165 °C). Therefore, we recommend a central SAME estimate of 1.35 days by the linear fitting at the observed lower bound. Critically, this value carries inherent uncertainties that must be quantified:

- The 95% confidence interval (CI) of ±0.25 days (i.e., 1.10–1.60 days).

- The wider 95% prediction interval (PI) of ±0.6 days (i.e., 0.75–1.95 days).

The substantial difference between the CI and PI ranges underscores the challenge in precisely constraining SAME. We advise using the PI for applications requiring robustness against individual model deviations.

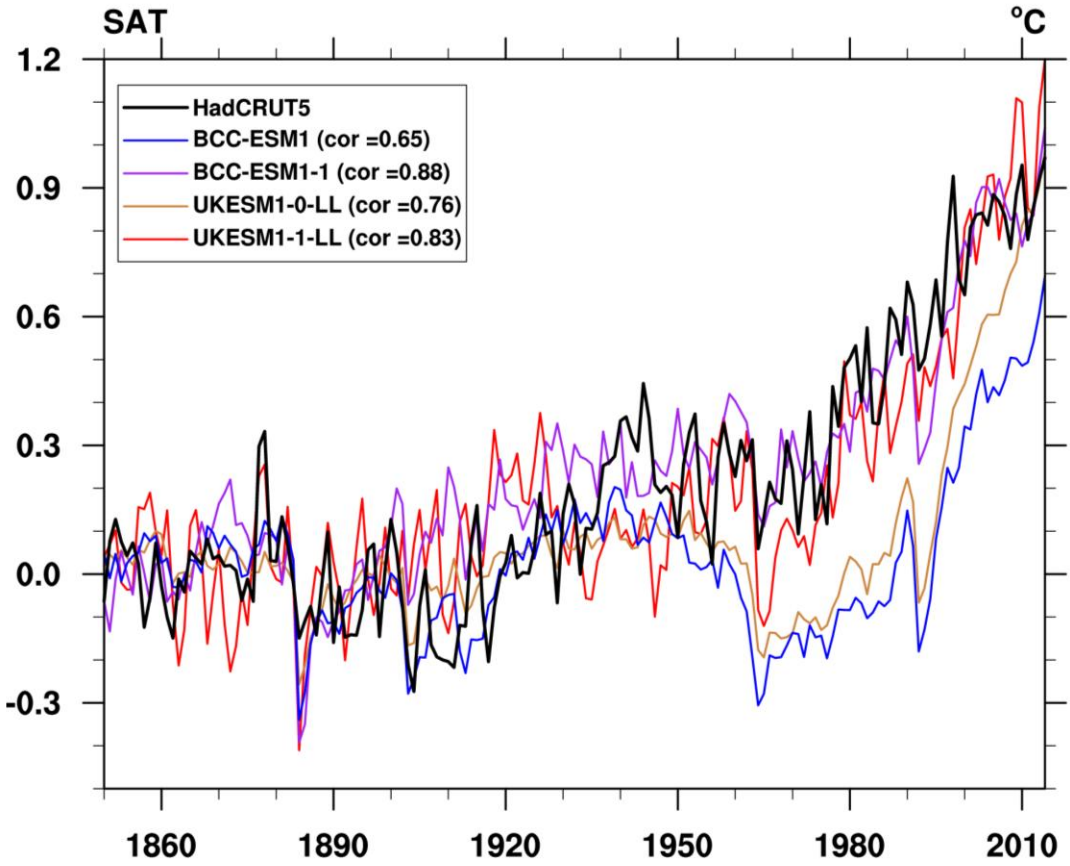

Figure 5. Evolutions of SAT anomalies relative to 1850~1900 mean for HadCRUT5, BCC-ESM models, and UKESM models. The numbers in legend are the corresponding correlation coefficients with HadCRUT5.

As demonstrated by the global mean SATa in BCC-ESM and UKESM models (Fig.5), both models on average tracked the instrumental record quite well with statistically higher correlation coefficients with observation (HadCRUT5). That is, improvements in sulfur deposition parameterizations, which reduced the SAME index, improved the representation of historical surface temperature variation.

4. Discussion: Sulfate lifetime in CMIP6 models and the two post-CMIP6 models

Table 2 Sulfate burden, sulfate depositions, and sulfate lifetime in CMIP6 models, BCC-ESM1-1 and UKESM1-1-LL in PHC period.

model name	**Sulfate burden (Tg S)**	**Sulfate Deposition (Tg S yr-1)**		**Sulfate lifetime (days)**
		DrySO4	**WetSO4**	
BCC-ESM1	0.59	2.07	43.78	4.70
CESM2	0.40	6.14	35.13	3.54
CESM-FV2	0.43	5.92	32.60	4.07
EC-Earth3-AerChem	0.75	1.29	40.39	6.57
GFDL-ESM4	0.46	7.58	31.68	4.28
MIROC6	0.33	7.50	61.29	1.75
MIROC-ES2L	0.33	5.67	67.42	1.65
MPI-ESM-1-2-HAM	0.74	2.41	69.41	3.76
MRI-ESM2-0	0.53	0.75	56.96	3.35
NorESM2-LM	0.52	6.39	40.33	4.06
UKESM1-0-LL	0.63	7.00	34.79	5.50
BCC-ESM1-1	0.48	1.34	19.2	8.53
UKESM1-1-LL	0.52	5.57	27.34	5.77

Generally, the SAME metric is used to facilitate model tuning of the sulfate burden, ensuring that models do not overestimating the sulfate cooling effect over the historical period, as was the case in CMIP6 and is a current concern for model performance in the upcoming CMIP7 experiments. SAME is relative to sulfate lifetime but it is calculated by the anomalous changes and also considers SO_2 deposition.

Sulfate lifetime is critical for validating the model's physical realism of sulfate cycle. Therefore, sulfate lifetime should be further examined to ensure model credibility. Here we calculated sulfate lifetime in PHC as the ratio of sulfate burden to total sulfate deposition (wet plus dry). As shown in Table 2, sulfate lifetime in CMIP6 models ranges from 1.65 days in MIROC-ES2L to 6.57 days in EC-Earth3-AerChem. The mean sulfate lifetime is 3.93 days, consistent with previous literatures, particularly the mean value of 4.12 days in AeroCom models with standard deviation of 18% (Textor et al., 2006). The wide sulfate lifetime range in CMIP6 models is attributed to variations in both sulfate burden (0.33 to 0.75 Tg S) and deposition rates (0.75 to 7.58 Tg S yr^{-1} for dry deposition, and 31.68 to 69.41 Tg S yr^{-1} for wet deposition).

Sulfate lifetimes in the two post-CMIP6 models, 8.53 days in BCC-ESM1-1 and 5.77 days in UKESM1-1-LL, are generally longer than those of their CMIP6 versions. The longer sulfate lifetimes in the two post-CMIP6 models may be due to lower SO_2 in these revised models but also could be due to physical climate changes (e.g., temperatures, clouds, rainfall). Compared to prior lifetime measures reported in the literature and considering the range of lifetimes found in recent models, the sulfate lifetimes in BCC-ESM1-1 and UKESM1-1-LL also appear reasonable (e.g., Charlson et al, 1992; Kristiansen et al. 2012; Textor et al., 2006).

5. Conclusions

Aerosol cooling effect is considered to be the second most important anthropogenic forcing over the 20^{th} Century. Our study, based on the 11 CMIP6 models with aerosol schemes, demonstrates that the anomalous cooling bias in the PHC period is closely related to the sulfate burden changes in the atmosphere. Sulfate burden in the models, and hence the strength of the anomalous cooling, is determined by sulfur deposition. We introduce a metric, called the SAME index, which incorporates the effects of sulfur removal processes on sulfate concentration. The index is highly correlated with cooling, and can be used to constrain sulfur removal processes in models, on a global scale.

A constraint on SAME by observed SATa, is used to inform a choice of tunable parameters for model depositions. Modifying sulfur deposition properties leads to an improved SAME in BCC-ESM1-1 and UKESM1-1-LL, as well as SATa reproductions. The optimal target value of SAME is 1.35 days with uncertainty range of ±0.25 days by 95% CI and ±0.6 days by 95% PI. Sulfate lifetime is critical for validating the model's physical realism and should be further examined to ensure model credibility.

Given that CMIP6 models overestimate the cooling effects of sulfate during the PHC period, when emissions were rising, it is reasonable to assume that they will underestimate the rate of warming during periods in climate projections. This has potential implications for the use of CMIP6 in scenarios that incorporate clean-air measures to inform the Paris Agreement goals of limiting warming to below 2 or 1.5^{o}C, i.e., SSP1-2.6 and SSP1-1.9 in CMIP6 (O'neill et al., 2016). To improve the reliability of projections, sulfur cycle processes in models should be improved. The SAME metric introduced in this paper provides a physically meaningful measure of the activity of the sulfur cycle, at a global scale, which we have shown can be used to improve modeled sulfur processes.

Code availability

All data processing codes are available if a request is sent to the corresponding authors.

Data availability

The HadCRUT5 dataset is accessible through Met Office Hadley Centre observations database (https://www.metoffice.gov.uk/hadobs/hadcrut5/). All the model data can be freely downloaded from the Earth System Grid Federation (ESGF) nodes (https://aims2.llnl.gov/search/cmip6/).

Author contributions

The main ideas were formulated by J.Z. and K.F. J.Z. wrote the original draft. The results were supervised by K.F. and S.T.T. All the authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare no competing financial and/or non-financial interests.

Acknowledgements

We acknowledge all data developers, their managers, and funding agencies for the datasets used in this study, whose work and support are essential to us.

Financial support

This work was jointly supported by the National Natural Science Foundation of China (Grant no. 42341202 and 42230608) and the UK–China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund.

References

Aas, W., Mortier, A., Bowersox, V., Cherian, R., Faluvegi, G., Fagerli, H., Hand, J., Klimont, Z., Galy-Lacaux, C., Lehmann, C. M. B., Myhre, C. L., Myhre, G., Olivié, D., Sato, K., Quaas, J., Rao, P. S. P., Schulz, M., Shindell, D., Skeie, R. B., Stein, A., Takemura, T., Tsyro, S., Vet, R., and Xu, X.: Global and regional trends of atmospheric sulfur, Scientific Reports, 9, 953, 10.1038/s41598-018-37304-0, 2019.

Bevacqua, E., Schleussner, C., and Zscheischler, J.: A year above 1.5 °C signals that Earth is most probably within the 20-year period that will reach the Paris Agreement limit, Nature Climate Change, 1-4, 10.1038/s41558-025-02246-9, 2025.

Chylek, P., Folland, C., Klett, J. D., and Dubey, M. K.: CMIP5 Climate Models Overestimate Cooling by Volcanic Aerosols, Geophysical Research Letters, 47, e2020GL087047, https://doi.org/10.1029/2020GL087047, 2020.

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Syst., 12, 35, 10.1029/2019ms001916, 2020.

Dittus, A. J., Hawkins, E., Wilcox, L. J., Sutton, R. T., Smith, C. J., Andrews, M. B., and Forster, P. M.: Sensitivity of Historical Climate Simulations to Uncertain Aerosol Forcing, Geophysical Research Letters, 47, e2019GL085806, 10.1029/2019gl085806, 2020.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L. P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J. P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6, Geosci. Model Dev. Discuss., 2021, 1-90, 10.5194/gmd-2020-446, 2021.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P. G., Griffies, S. M., Guo, H., Hallberg, R. W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P. C. D., Nikonov, S., Paynter, D. J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B. G., Robinson, T., Schwarzkopf, D. M., Sentman, L. T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, Y., and Zhao, M.: The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall Coupled Model Description and Simulation Characteristics, J. Adv. Model. Earth Syst., 12, 10.1029/2019ms002015, 2020.

EB, U.: Protocol to Abate Acidification, Eutrophication and Ground-level Ozone, 1999.

Flynn, C. M. and Mauritsen, T.: On the climate sensitivity and historical warming evolution in recent coupled model ensembles, Atmos. Chem. Phys., 20, 7829-7842, 10.5194/acp-20-7829-2020, 2020.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks, Geoscientific Model Development, 13, 2197-2244, 10.5194/gmd-13-2197-2020, 2020.

Hand, J. L., Schichtel, B. A., Malm, W. C., and Pitchford, M. L.: Particulate sulfate ion concentration and SO₂ emission trends in the United States from the early 1990s through 2010, Atmos. Chem. Phys., 12, 10353-10365, 10.5194/acp-12-10353-2012, 2012.

Hansen, J., Kharecha, P., Sato, M., Tselioudis, G., Kelly, J., Bauer, S., Ruedy, R., Jeong, E., Jin, Q., Rignot, E., Velicogna, I., Schoeberl, M., Schuckmann, K., Amponsem, J., Cao, J., Keskinen, A., Li, J., and Pokela, A.: Global Warming Has Accelerated: Are the United Nations and the Public Well-Informed?, Environment: Science and Policy for Sustainable Development, 67, 6-44, 10.1080/00139157.2025.2434494, 2025.

Hardacre, C., Mulcahy, J. P., Pope, R. J., Jones, C. G., Rumbold, S. T., Li, C., Johnson, C., and Turnock, S. T.: Evaluation of SO₂, SO₄²⁻ and an updated SO₂ dry deposition parameterization in the United Kingdom Earth System Model, Atmospheric Chemistry and Physics, 21, 18465-18497, 10.5194/acp-21-18465-2021, 2021.

Hegerl, G.C., F. W. Zwiers, P. Braconnot, N.P. Gillett, Y. Luo, J.A. Marengo Orsini, N. Nicholls, J.E. Penner and P.A. Stott, 2007: Understanding and Attributing Climate Change. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hienola, A., Partanen, A.-I., Pietikäinen, J.-P., O'Donnell, D., Korhonen, H., Matthews, H. D., and Laaksonen, A.: The impact of aerosol emissions on the 1.5 °C pathways, Environmental Research Letters, 13, 044011, 10.1088/1748-9326/aab1b2, 2018.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J. I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369-408, 10.5194/gmd-11-369-2018, 2018.

IPCC. Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 10.1017/9781009157896, 2023.

Likens, G. E., Butler, T. J., and Buso, D. C.: Long- and short-term changes in sulfate deposition: Effects of the 1990 Clean Air Act Amendments, Biogeochemistry, 52, 1-11, 10.1023/a:1026563400336, 2001.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Flaeschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimenez-de-la-Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Moebis, B., Muellner, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2, J. Adv. Model. Earth Syst., 11, 998-1038, 10.1029/2018ms001400, 2019.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research-Atmospheres, 126, 10.1029/2019jd032361, 2021.

Mulcahy, J. P., Jones, C. G., Rumbold, S. T., Kuhlbrodt, T., Dittus, A. J., Blockley, E. W., Yool, A., Walton, J., Hardacre, C., Andrews, T., Bodas-Salcedo, A., Stringer, M., de Mora, L., Harris, P., Hill, R., Kelley, D., Robertson, E., and Tang, Y.: UKESM1.1: development and evaluation of an updated configuration of the UK Earth System Model, Geosci. Model Dev., 16, 1569-1600, 10.5194/gmd-16-1569-2023, 2023.

Mulcahy, J. P., Johnson, C., Jones, C. G., Povey, A. C., Scott, C. E., Sellar, A., Turnock, S. T., Woodhouse, M. T., Abraham, N. L., Andrews, M. B., Bellouin, N., Browse, J., Carslaw, K. S., Dalvi, M., Folberth, G. A., Glover, M., Grosvenor, D. P., Hardacre, C., Hill, R., Johnson, B., Jones, A., Kipling, Z., Mann, G., Mollard, J., O'Connor, F. M., Palmieri, J., Reddington, C., Rumbold, S. T., Richardson, M., Schutgens, N. A. J., Stier, P., Stringer, M., Tang, Y., Walton, J., Woodward, S., and Yool, A.: Description and evaluation of aerosol in UKESM1 and HadGEM3-GC3.1 CMIP6 historical simulations, Geoscientific Model Development, 13, 6383-6423, 10.5194/gmd-13-6383-2020, 2020.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geosci. Model Dev., 9, 3461-3482, 10.5194/gmd-9-3461-2016, 2016.

Ohara, T., Akimoto, H., Kurokawa, J., Horii, N., Yamaji, K., Yan, X., and Hayasaka, T.: An Asian emission inventory of anthropogenic emission sources for the period 1980-2020, Atmospheric Chemistry and Physics, 7, 4419-4444, 10.5194/acp-7-4419-2007, 2007.

Samset, B. H., Sand, M., Smith, C. J., Bauer, S. E., Forster, P. M., Fuglestvedt, J. S., Osprey, S., and Schleussner, C.-F.: Climate Impacts From a Removal of Anthropogenic Aerosol Emissions, Geophysical Research Letters, 45, 1020-1029, https://doi.org/10.1002/2017GL076079, 2018.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y. C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations, Geosci. Model Dev., 13, 6165-6200, 10.5194/gmd-13-6165-2020, 2020.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the UK Earth System Model, J. Adv. Model. Earth Syst., 11, 4513-4558, 10.1029/2019ms001739, 2019.

Smith, C. J. and Forster, P. M.: Suppressed Late-20th Century Warming in CMIP6 Models Explained by Forcing and Feedbacks, Geophysical Research Letters, 48, 10.1029/2021gl094948, 2021.

Steffen, W., Crutzen, P. J., and McNeill, J. R.: The Anthropocene: Are humans now overwhelming the great forces of nature, Ambio, 36, 614-621, 10.1579/0044-7447(2007)36[614:taahno]2.0.co;2, 2007.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'Ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M.: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, Geoscientific Model Development, 12, 2727-2765, 10.5194/gmd-12-2727-2019, 2019.

Vestreng, V., Myhre, G., Fagerli, H., Reis, S., and Tarrasón, L.: Twenty-five years of continuous sulphur dioxide emission reduction in Europe, Atmos. Chem. Phys., 7, 3663-3681, 10.5194/acp-7-3663-2007, 2007.

Wang, Z., Lin, L., Xu, Y., Che, H., Zhang, X., Zhang, H., Dong, W., Wang, C., Gui, K., and Xie, B.: Incorrect Asian aerosols affecting the attribution and projection of regional climate change in CMIP6 models, Npj Climate and Atmospheric Science, 4, 10.1038/s41612-020-00159-2, 2021.

Wilcox, L. J., Highwood, E. J., and Dunstone, N. J.: The influence of anthropogenic aerosol on multi-decadal variations of historical global climate, Environmental Research Letters, 8, 10.1088/1748-9326/8/2/024033, 2013.

Wu, T., Zhang, F., Zhang, J., Jie, W., Zhang, Y., Wu, F., Li, L., Yan, J., Liu, X., Lu, X., Tan, H., Zhang, L., Wang, J., and Hu, A.: Beijing Climate Center Earth System Model version 1 (BCC-ESM1): model description and evaluation of aerosol simulations, Geosci. Model Dev., 13, 977-1005, 10.5194/gmd-13-977-2020, 2020.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M.: The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component, Journal of the Meteorological Society of Japan, 97, 931-965, 10.2151/jmsj.2019-051, 2019.

Zhang, J., Furtado, K., Turnock, S. T., Mulcahy, J. P., Wilcox, L. J., Booth, B. B., Sexton, D., Wu, T., Zhang, F., and Liu, Q.: The role of anthropogenic aerosols in the anomalous cooling from 1960 to 1990 in the CMIP6 Earth System Models, Atmos. Chem. Phys. Discuss., 2021, 1-39, 10.5194/acp-2021-570, 2021a.

Zhang, J., Wu, T., Zhang, F., Furtado, K., Xin, X., Shi, X., Li, J., Chu, M., Zhang, L., Liu, Q., Yan, J., Wei, M., and Ma, Q.: BCC-ESM1 Model Datasets for the CMIP6 Aerosol Chemistry Model Intercomparison Project (AerChemMIP), Advances in Atmospheric Sciences, 38, 317-328, 10.1007/s00376-020-0151-2, 2021b.