Sensitivity of historical climate simulations to uncertain aerosol forcing

Article

Published Version

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/10.1029/2019GL085806

Sensitivity of Historical Climate Simulations to Uncertain Aerosol Forcing

Andrea J. Dittus[1] [ID], Ed Hawkins[1] [ID], Laura J. Wilcox[1] [ID], Rowan T. Sutton[1] [ID], Christopher J. Smith[2] [ID], Martin B. Andrews[3] [ID], and Piers M. Forster[2] [ID]

[1]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK, [2]School of Earth and Environment, University of Leeds, Leeds, UK, [3]Met Office Hadley Centre, Exeter, UK

Abstract The relative importance of anthropogenic aerosol in decadal variations of historical climate is uncertain, largely due to uncertainty in aerosol radiative forcing. We analyze a novel large ensemble of simulations with HadGEM3-GC3.1 for 1850–2014, where anthropogenic aerosol and precursor emissions are scaled to sample a wide range of historical aerosol radiative forcing with present-day values ranging from –0.38 to –1.50 Wm^{-2}. Five ensemble members are run for each of five aerosol scaling factors. Decadal variations in surface temperatures are strongly sensitive to aerosol forcing, particularly between 1950 and 1980. Post-1980, trends are dominated by greenhouse gas forcing, with much lower sensitivity to aerosol emission differences. Most realizations with aerosol forcing more negative than about –1 Wm^{-2} simulate stronger cooling trends in the mid-20th century compared with observations, while the simulated warming post-1980 always exceeds observed warming, likely due to a warm bias in the transient climate response in HadGEM3-GC3.1.

Plain Language Summary Anthropogenic aerosols have an overall cooling effect on climate due to their interaction with incoming solar radiation and influence on cloud properties. Their emissions have offset some of the historical warming induced by increasing greenhouse gases. However, the magnitude of the cooling induced by anthropogenic aerosol remains poorly constrained. In this study, we use a state-of-the-art climate model, HadGEM3-GC3.1, driven by different levels of aerosol emissions. This experimental setup tests the sensitivity of simulated historical temperatures to the strength of aerosol forcing in a climate model, all other factors remaining equal. Our results show that the period from 1951 to 1980 is particularly sensitive to aerosol forcing, coinciding with a period of rapid increases in global aerosol emissions and observed cooling over many regions, while temperature trends from 1980 onwards are primarily driven by increases in greenhouse gas concentrations. The observed temperatures over 1951–1980 are best reproduced by simulations with lower aerosol emissions than the standard configuration, implying that this model responds too strongly to aerosol forcing. Concurrently, the simulated temperatures warm faster than observed temperatures from 1980 onwards, suggesting that this climate model also responds more strongly to greenhouse gas forcing than observations suggest.

1. Introduction

The magnitude of historical anthropogenic aerosol (AA) forcing remains very uncertain, despite the attention that has been devoted to this question (Booth et al., 2018; Bellouin et al., 2019; Kretzschmar et al., 2017; Myhre et al., 2013; Stevens, 2015). Model representations of aerosol processes are very diverse (Wilcox et al., 2015), resulting in a large range of simulated historical aerosol forcing in models participating in the Fifth Phase of the Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012), with the largest differences coming from aerosol-cloud interactions (Wilcox et al., 2015; Zelinka et al., 2014). Despite this diversity in aerosol processes and magnitude of simulated aerosol forcing across climate models, previous studies have shown the importance of including aerosol-cloud interactions for representing the historical evolution of global mean surface temperature (GMST, Ekman, 2014; Jones et al., 2013; Wilcox et al., 2013). Early results from CMIP6 (Eyring et al., 2016) suggest that understanding aerosol forcing and associated feedbacks in this new generation of models is key for interpreting historical simulations and future projections (Gettelman et al., 2019; Golaz et al., 2019). Beyond global mean temperatures, studies have also suggested that aerosol forcing plays an important role for multidecadal variability in Atlantic (Booth et al., 2012;

Watanabe & Tatebe, 2019) and Pacific (Smith et al., 2016) sea surface temperatures, monsoons (Krishnan et al., 2015; Polson et al., 2014), and droughts (Dong et al., 2014; Undorf et al., 2018). However, many of these processes remain poorly understood or uncertain, for example, the extent to which aerosol forcing has influenced variability in the Pacific (Oudar et al., 2018) or Atlantic (Kim et al., 2018; Robson et al., 2014; Zhang et al., 2013) remains an open question.

While the uncertainties surrounding aerosol forcing and its influence on climate are large, it is nevertheless clear that aerosol forcing has played an important role in the evolution of 20th and early 21st century climate. However, constraining aerosol forcing and its role in historical climate variability is critically important to better assess the impact of anticipated future reductions in aerosol precursor emissions (e.g., Riahi et al., 2017; O'Neill et al., 2016). Consequently, reduced aerosol forcing uncertainty will improve confidence in near-term climate projections at global and regional scales.

Existing multimodel ensemble experiment designs can make it difficult to separate the role of AA from other factors, such as differences in resolution and cloud parameterization, and to separate forcing uncertainty from response uncertainty. As part of the UK SMURPHS project, we have designed a climate model ensemble specifically to sample a wide range of historical aerosol forcing by scaling AA emissions by various factors in a CMIP6 generation climate model. This ensemble will enable us to examine the role of aerosol forcing uncertainty in climate variability in the absence of model structural diversity and, through the use of multiple ensemble members per scaling, provide the necessary signal to examine regional responses. This work introduces the ensemble and assesses the role of aerosol forcing for surface temperature variability and trends.

## 2. Simulations, Observations, and Methods

2.1. Ensemble Design

All simulations in the historical scaled aerosol emissions ensemble were conducted with the HadGEM3-GC3.1 global climate model at N96-ORCA1 resolution, corresponding to 135-km resolution in midlatitudes and 1° in the ocean (Kuhlbrodt et al., 2018; Williams et al., 2018). HadGEM3 uses the GLOMAP two-moment aerosol scheme that includes representations of aerosol effects on cloud albedo and lifetime (Mulcahy et al., 2018). The model version used here is a development version towards the United Kingdom's submission to CMIP6 (Andrews et al., 2020; Hardiman et al., 2019; see Supporting Information).

The SMURPHS ensemble was designed to sample a plausible range of historical aerosol forcing (Booth et al., 2018), with present-day AA effective radiative forcing (ERF) ranging from –0.38 to –1.50 Wm^{-2}, which spans most of the 95% confidence interval presented in IPCC AR5 (Boucher et al., 2013). The targeted aerosol forcings were achieved by applying a constant scaling factor in space and time to the standard historical CMIP6 AA and precursor emissions (Hoesly et al., 2018), namely, organic and black carbon (fossil and biofuel) and sulfur dioxide (SO_2) emissions. All other forcing agents follow historical CMIP6 emissions (Meinshausen et al., 2017). An illustration of the effect of scaling the SO_2 emissions is shown in Figure 1a, with regional time series of aerosol and precursor emissions and the spatial pattern of AA optical depth trends shown in Figures S2 and S3, respectively. The scalings were chosen such that the targeted present-day aerosol forcings are roughly equally spaced:

- 0.2× scaling to give –0.38 Wm^{-2},
- 0.4× scaling to give –0.60 Wm^{-2},
- 0.7× scaling to give –0.93 Wm^{-2},
- 1.0× scaling to give –1.17 W/m^2 (standard emissions), and
- 1.5× scaling to give –1.50 Wm^{-2}.

The forcing values for each scaling ensemble were determined from five equivalent atmosphere-only time-slice runs for year 2014 with preindustrial sea surface temperatures of 10-year length each (one for each scaling factor).

Five historical simulations were performed for each scaling over the period 1850–2014. The same five initial conditions were used for each scaling subensemble, the first four corresponding to the initial conditions used for the historical simulations with HadGEM3-GC3.1 for CMIP6 (Andrews et al., 2020). These were well
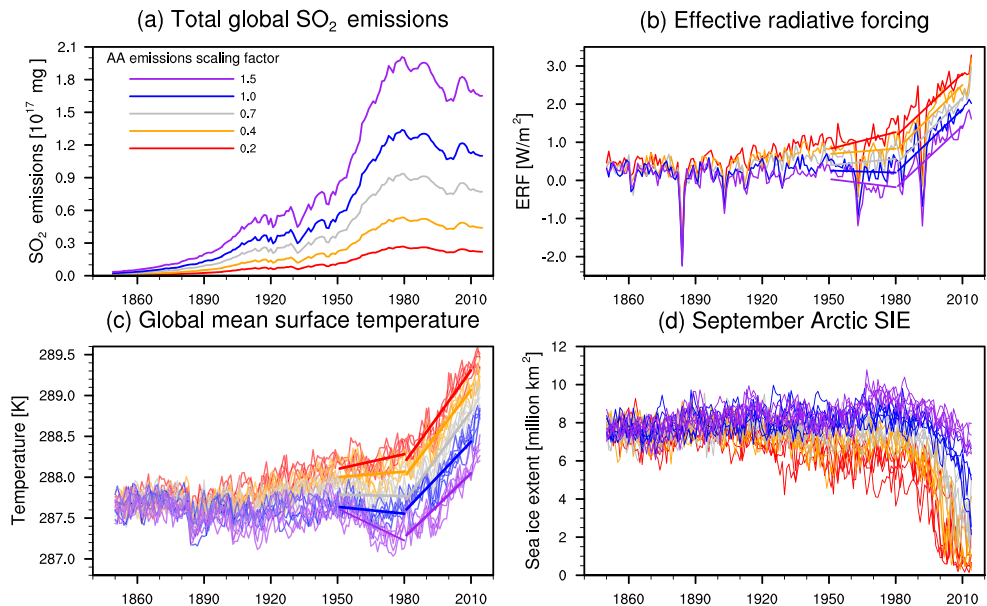
Figure 1. (a) Scaled global SO$_2$ emissions used in the ensemble. OC and BC were similarly scaled but are not shown. (b) Historical ERF time series for each of the aerosol scalings. (c) GMST time series for each of the simulations, color coded by the anthropogenic aerosol scaling factor. The thick lines represent the linear trend fitted to each five-member ensemble mean for the periods 1951–1980 and 1981–2010, respectively. (d) September Arctic sea ice extent in each ensemble member.

spaced in a preindustrial control simulation and sample different phases of internal variability in both the Pacific and Atlantic (Sellar et al., 2020). The fifth ensemble member was selected independently according to the same criteria. We recommend only analyzing data from 1900, as introducing the scaled aerosols in 1850 when aerosol precursor emissions are small but non-zero (Hoesly et al., 2018) produces a small initial drift in the climate system. Most of this drift has stopped by 1900 (Figure S4). In addition to the 25 coupled simulations, fixed SST runs were performed with scaled AA to estimate the ERF time series associated with these simulations following the RFMIP protocol (Figure 1b, Pincus et al., 2016).

2.2. Observations and Simulated Temperatures

The observational data sets used in this study are the 21 June 2019 release of version 2.0 of the infilled Had-CRUT4 data set, updated with HadSST4 (1850–present, Cowtan & Way, 2014, https://www-users.york.ac.uk/~kdc3/papers/coverage2013/series.html) and GISTEMP v4 (1880–present, Lenssen et al., 2019). Following Cowtan et al. (2015), surface air temperatures and sea surface temperatures from the model were blended in order to generate temperatures comparable with the observational products. As in Cowtan et al. (2015), air temperatures were used over land and grid cell fractions containing sea ice, while sea surface temperatures were used over ocean grid cells. To calculate the pattern correlations between observations and model simulations, the model data were regridded to the resolution of the observational data set using bilinear interpolation. All data were smoothed using an area-weighted 5-point smoothing algorithm before calculating the pattern correlations. The CMIP6 HadGEM3-GC3.1 preindustrial control run is used to estimate unforced variability (Menary et al., 2018).

3. Results

3.1. Simulated Global Mean Changes

As expected, the evolution of GMST is highly sensitive to the aerosol scaling factor applied to AA emissions, with low aerosol simulations warming more rapidly than simulations with high aerosol forcing (Figure 1c), starting to diverge around the 1880s. The evolution of GMST closely follows that of the radiative forcing time series. The period from the 1950s to the late 1970s, which coincides with the rapid increase of SO$_2$ emissions over Europe and North America, is highly sensitive to the magnitude of aerosol forcing (Figures 1a and S2). This is followed by a period of rapid warming from the 1980s onwards for all scaling factors, suggesting that temperature trends are driven primarily by greenhouse gas forcing from this period onwards, consistent
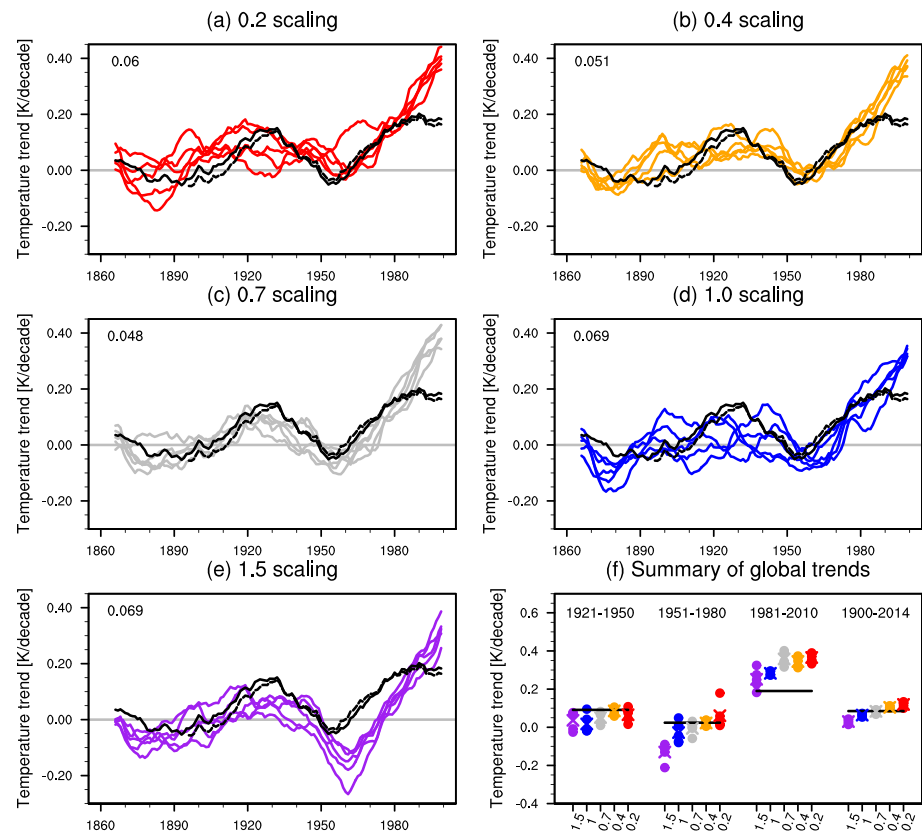
Figure 2. (a–e) 31-year running trends are shown for each scaling factor, indexed to the central year of the 31-year trend. Two observational data sets, Cowtan and Way (2014) and GISTEMP v4 (Lenssen et al., 2019), are shown in black. The numbers in the top left-hand corner correspond to the Euclidean distance between the individual ensemble members and the Cowtan and Way (2014) observations, averaged over all ensemble members and the entire period. Panel (f) shows the simulated trends for each scaling over four different time periods. Crosses indicate the ensemble mean trend while the horizontal bar corresponds to observations from Cowtan and Way (2014).

with findings from previous studies(Wilcox et al., 2013). Feedbacks in high latitudes mean that changes in sea ice are likely to be important in explaining the pattern of temperature changes. Sea ice extent is strongly affected by aerosol forcing and associated temperature changes (Figure 1d). In the 0.2 scaling ensemble, the decline in September sea ice begins in the early 1900s. In the 1.5 scaling ensemble, Arctic sea ice steadily increases until the 1960s when the increase in sea ice accelerates for a decade or two, after which sea ice starts to decline in these simulations as well, coinciding with the period of rapid warming discussed above. It is worth noting that the 0.2 scaling ensemble is nearly ice free by the end of the simulations, which may be important in explaining the pattern of temperature changes discussed later.

Comparing 31-year running trends in GMST to observations reveals that all simulations overestimate the warming rates since the 1980s, regardless of aerosol scalings (Figure 2), confirming that warming in this period is not primarily driven by changes in AA emissions. This overestimation of the observed trends suggests that the temperature response to greenhouse gas forcing may be too strong in HadGEM3-GC3.1. At the same time, the aerosol-driven cooling ending in the late 1970s is too strong in a majority of the realizations from the 1.5, 1.0, and 0.7 ensembles, though there are some realizations in each of the 1.0 and 0.7 ensembles that are consistent with observations (Figure 2f). It is further worth noting that the strongest 31-year cooling trend ends approximately 5 years earlier in the observations compared with the model simulations, which may mask some of the discrepancy between simulated and observed trends in Figure 2f. The annual mean time series of GMST compared with observations confirm these discrepancies in the rate of change of GMST, but the comparison with observations is sensitive to the choice of base period for the calculation of anomalies (Figures S5 and S6).

Importantly, no scaling is consistent with observations for all periods. The best performing aerosol scalings overall are the 0.4 and 0.7 ensembles (Figures 2a– 2e), based on the Euclidean distance between the modeled running trends and observations averaged across all five members. While these results suggest that weaker aerosol forcings are more compatible with the observed historical record, a thorough constraint on historical aerosol forcing is not possible with this method due to the likely concurrent warm bias in the transient climate response (TCR), a metric often used to characterize transient warming in climate models. It is usually defined as the temperature at the time of CO_2 doubling in a 1% CO_2 simulation, sometimes referred to as the "true TCR" in the literature (Adams & Dessler, 2019). To determine the best performing aerosol scalings, taking into account the likely concurrent TCR bias, we derive an estimate of observed and simulated historical TCR using the method outlined in Gregory and Forster (2008), where it is assumed that the temperature change is proportional to the change in forcing. The historical TCR is estimated using the following equation:

$$TCR_{hist} = \Delta T \frac{ERF_{2 \times CO2}}{\Delta ERF}. \tag{1}$$

We use the model's documented $ERF_{2 \times CO2} = 4.05$ W/m^2 (Andrews et al., 2019) and use ΔT and ΔERF estimated from the model simulations over the period 1981–2010. If we assume that the modeled change in ERF over the period 1981–2010 is a good estimate of the real-world ERF, an estimate of TCR_{hist} for the real-word can be obtained by using the observed temperature change over the same period. We use the standard deviation in 30-year trends from the model control run to provide an estimate of the role of internal variability in contributing to uncertainty in ΔT and hence TCR_{hist}. Due to relatively invariant global aerosol emissions over this time period (Figure 1a), the change in ERF over the period 1981–2010 is not very sensitive to aerosol scaling (Figures 1b and S7b). However, the TCR_{hist} estimates obtained are sensitive to the aerosol scaling, with the low aerosol simulations corresponding to a higher TCR_{hist} estimate (Figure S7a). This indicates a nonlinearity in the temperature response (ΔT) to a change in forcing in the 1981–2010 period, implying a change in either the climate feedback parameter or the rate of ocean heat uptake(Gregory et al., 2015). These changes are likely related to differences in the spatial patterns discussed in the next section (Andrews et al., 2015) and may also point to a role for feedbacks in high latitudes (Figure 1d). Estimates of TCR_{hist} are typically lower than true TCR, primarily due to the influence of internal variability, spatial patterns of warming, and observational coverage issues (Adams & Dessler, 2019). As discussed earlier, model temperatures have been blended in this study to avoid the latter issues. Assuming a zero-layer model and neglecting upper ocean energy storage likely contributes to this discrepancy (Gregory et al., 2015; Jiménez-de-la Cuesta & Mauritsen, 2019).
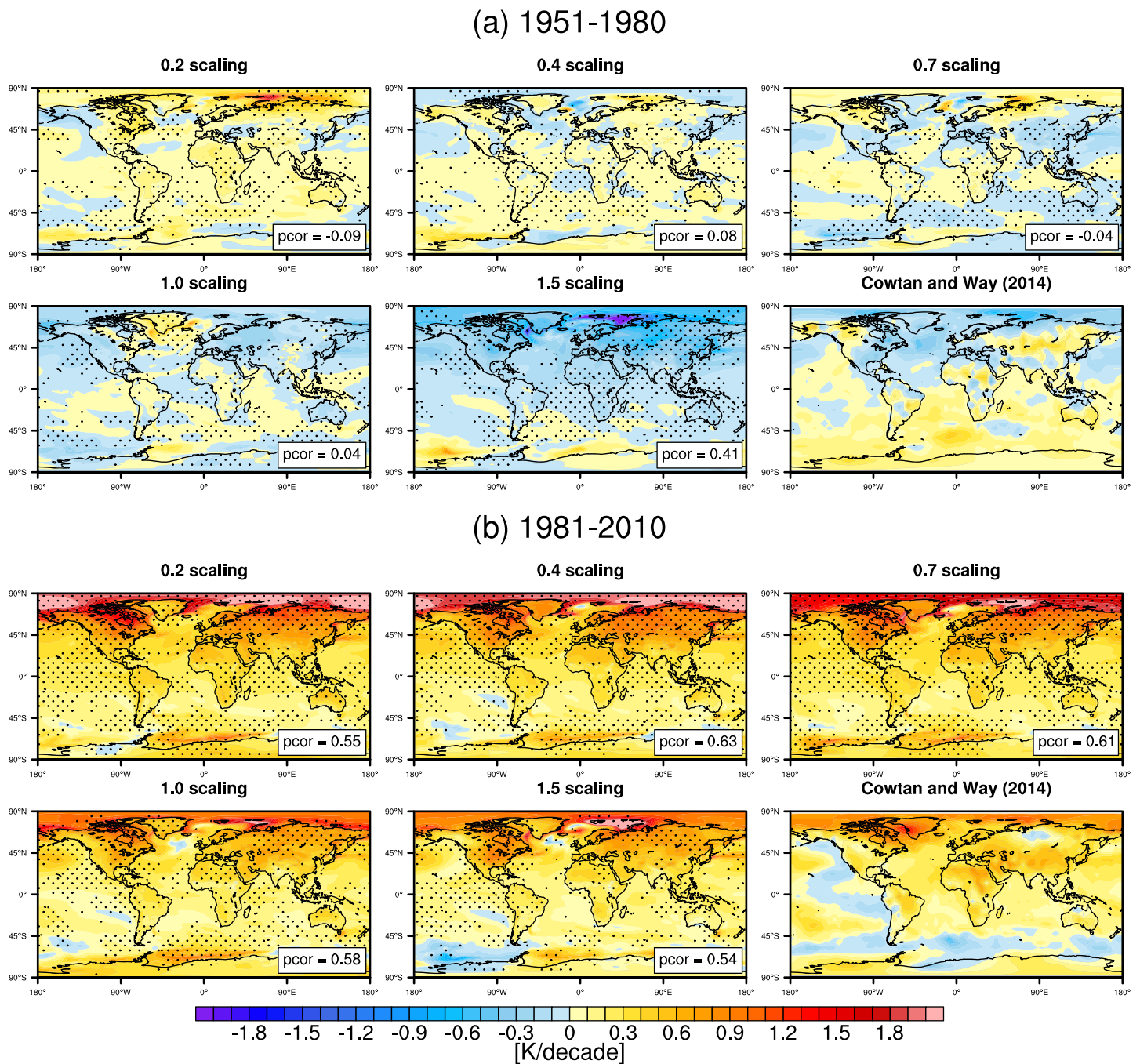
Using this method, the observation-based estimate of TCR_{hist} is 1.37 ± 0.58 K ($\pm 2\sigma$ from the model control run as stated above), while the model TCR_{hist} estimate is 2.38 ± 0.79 K ($\pm 2\sigma$ estimated across all ensemble members) if all scalings are considered equally likely. If the two largest aerosol scalings are excluded due to their lower agreement with the observed record (Figure 2), the TCR_{hist} estimate becomes 2.63 ± 0.47 K. These results suggest that a substantial warm TCR bias is likely, though the uncertainty due to internal variability and the nonlinearity in the response to the change in forcing is large (indicated by the uncertainties provided above, also Figure S7).

$$\Delta ERF = \frac{\Delta T \cdot ERF_{2 \times CO2}}{TCR_{hist}} \tag{2}$$

Rearranging Equation (1) to give (2) and using the observational TCR estimate previously derived, the estimated change in ERF for the period 1951–1980 from observations is approximately 0.2 Wm^{-2}, but uncertainty is large (Figure S7b). Figure S7b shows that even when a possible TCR bias is accounted for, the 0.4 and 0.7 scaling ensembles still give the best agreement with observations.

3.2. Spatial Patterns of Temperature Change

Maps of the trend in the ensemble means for each scaling for 1951–1980 and 1981–2010 are shown in Figure 3. In the 1951–1980 period, the aerosol scaling makes a large difference to the simulated spatial patterns, with the low aerosol scalings showing widespread warming over this period, while the higher aerosol scalings show widespread cooling. The comparison with observed trends shows that the model simulated patterns are too spatially uniform and are unable to simulate the interhemispheric temperature gradient seen in the observations. The 1.5 scaling comes closest to the observed pattern; however, the simulated cooling in

AGU
100
ADVANCING EARTH
AND SPACE SCIENCE

**Geophysical Research Letters**

10.1029/2019GL085806

Figure 3. Surface air temperature trends in the five-member ensemble means for each scaling for 1951–1980 (top panels) and 1981–2010 (bottom panels), with the observed trend shown in the bottom right-hand panels. Average pattern correlations (pcor) between observations and model simulations are also given for each scaling (calculated for each realization individually then averaged across five members). Stippling indicates areas where the observed trends lie outside the range of simulated trends. The trend patterns for each individual realization are shown in Figures S9 and S10.

the northern high latitudes is too widespread and too strong compared with observations. The pattern is robust across all members of the 1.5 scaling ensemble (Figure S9). Note that the pattern correlations shown in Figure 3 are calculated for individual ensemble members, then averaged, as the individual ensemble members are a combination of forced response and internal variability, like the observations, and there is considerable variability across individual ensemble members (Figure S9).
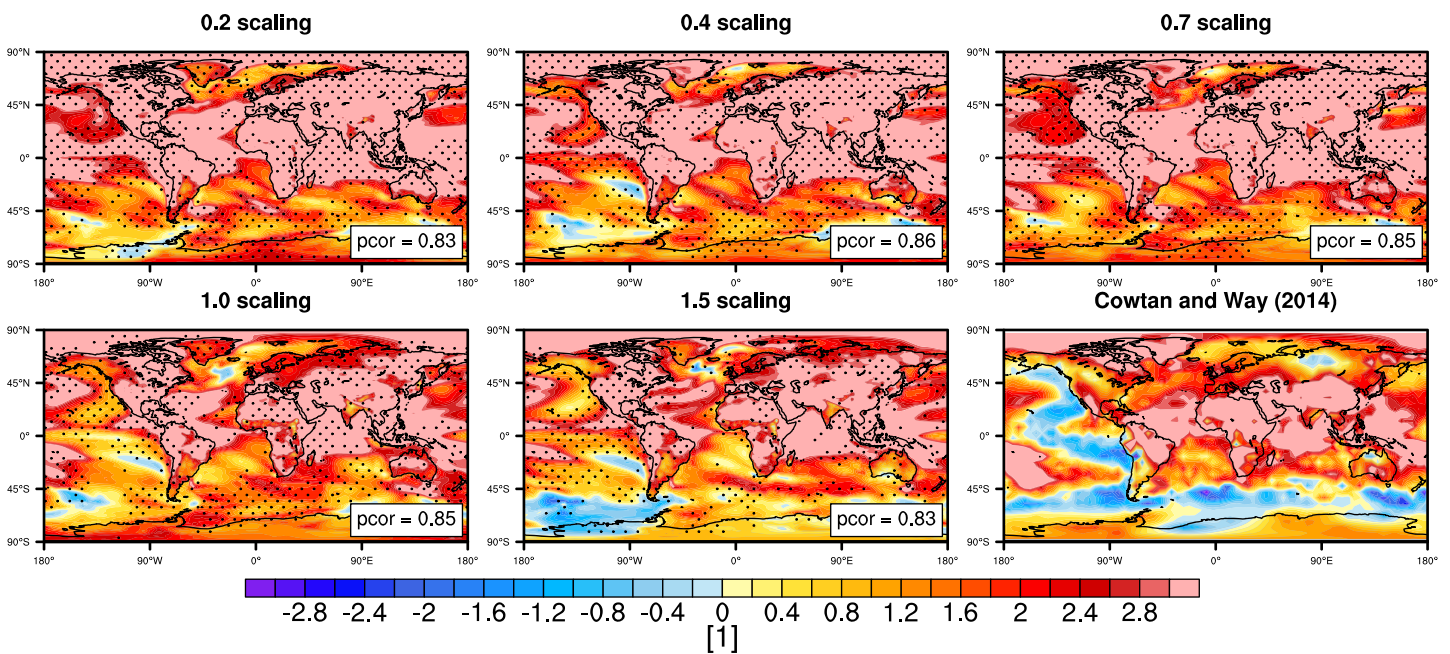
From 1981–2010, the simulated spatial patterns are more similar across scalings, consistent with the smaller aerosol influence on trends in this period. Differences include enhanced polar amplification in the low

(a) 1951-1980
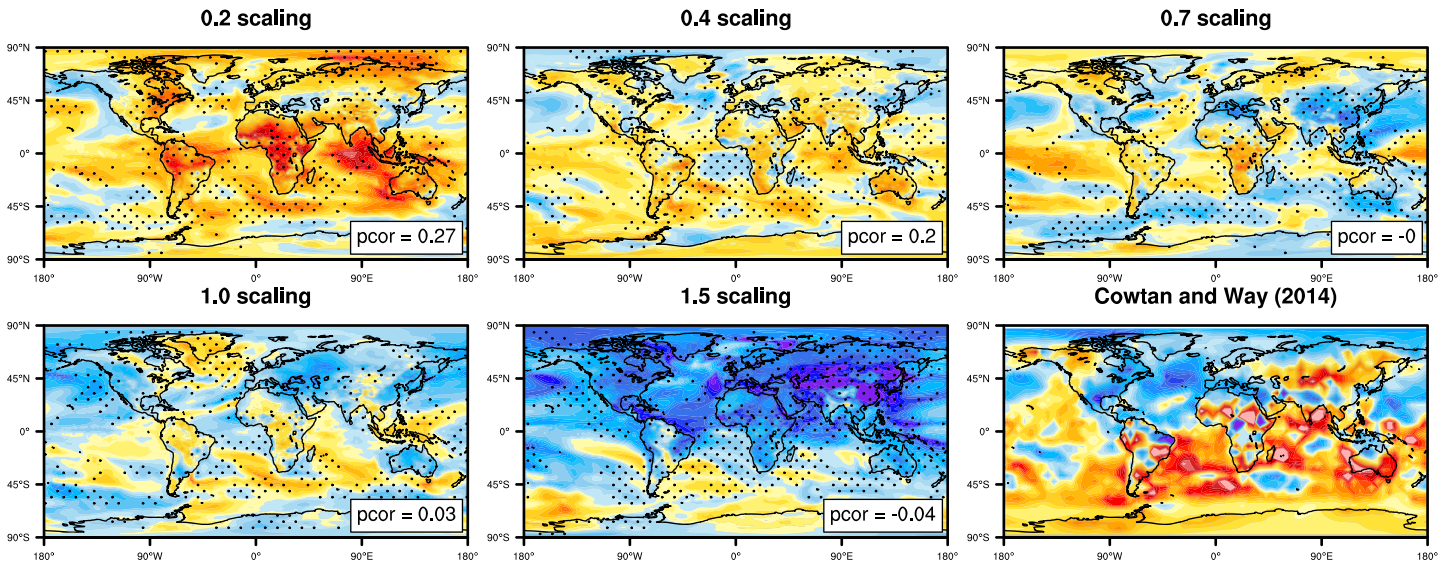
(b) 1981-2010

Figure 4. Signal-to-noise maps for the two periods considered, 1951 to 1980 and 1981 to 2010, respectively. Signal is defined as the trend in the ensemble mean, while noise is defined as the standard deviation across 31-year trends in the control run. Average pattern correlations (pcor) between observations and model simulations are also given for each scaling (calculated for each realization individually then averaged across four members). Stippling indicates areas where the observed signal-to-noise lies outside the simulated range.

aerosol scalings, likely related to the decreases in sea ice shown in Figure 1d. The high aerosol scaling (1.5) simulations also show some areas with cooling trends over this period, notably in the Pacific sector of the Southern Ocean and areas of the subtropical Pacific, with some common features with the observed trends in the Pacific over that period. Some of these areas of cooling can also be seen for some of the other scalings to a lesser extent. These results suggest that AA forcing (possibly in combination with volcanic forcing) may have contributed to the observed cooling in the Pacific associated with the negative phase of the Pacific

Decadal Oscillation. However, these results alone are not sufficient evidence of a role for external forcing in Pacific variability and warrant further investigation.

Signal-to-noise maps provide a different perspective to trend maps (Figure 4), as they highlight areas where the forced response is particularly large relative to internal variability. Here, signal is defined as the ensemble mean trend, and noise is estimated as the standard deviation of rolling 30-year trends from the corresponding 500-year control run (Figure S8). Due to the lack of reliable estimates of the range of trends that could have occurred due to internal variability in the observations, the model noise estimate is used for the observations. Figure 4 shows that over the 1951–1980 period, high signal-to-noise trends are found in the Northern Hemisphere midlatitudes in the high-aerosol simulations, consistent with the expected spatial pattern in response to strong aerosol forcing. Over the same period in the 0.2 ensemble, the strongest signal (associated with warming trends) is found in the tropics and subtropics, consistent with expectations from a spatially uniform forcing agent such as greenhouse gas forcing. The signal-to-noise maps clearly show a different balance of forcings between the low- and high-aerosol simulations in the 1951–1980 period and further suggest that greenhouse gas forcing is the main driver of temperature changes over the 1981–2010 period for all scalings.

The observation-based signal-to-noise map for 1951–1980 suggests a competing influence of greenhouse gas-induced warming in the Southern Hemisphere (resembling the 0.2 scaling) and aerosol-induced cooling over the Northern Hemisphere (similar to the 1.5 ensemble). However, as we have no way of estimating the variability in observed signal-to-noise, it is important to recognize that the observed signal-to-noise pattern is a combination of forced signal and a particular realization of internal variability. In the ensemble mean, none of the subensembles capture this competing influence of aerosol forcing and greenhouse gas forcing seen in the observations for the Northern versus Southern Hemisphere. While the 1.5 scaling ensemble trend patterns agree better with observations for the 1951–1980 period, the signal-to-noise patterns do not compare well, providing additional evidence that aerosol forcing in the 1.5 scaling ensemble is likely unrealistically strong. The ensemble members most similar to the observations over 1951–1980 in signal-to-noise are ensemble members 2 from the 0.2 ensemble followed by ensemble member 3 from the 1.0 ensemble (Figures S9 and S11), showing that internal variability is large.

4. Discussion and Conclusions

The ensemble introduced in this paper provides a novel data set for investigating the role of aerosol forcing in historical climate variability. By scaling emissions of AA and their precursors, a wide range of the uncertainty in AA forcing from IPCC AR5 is sampled within this ensemble. Five ensemble members are run for five different aerosol scaling ensembles, providing a good balance between sampling internal variability and uncertainty in aerosol forcing (from -1.5 W/m^2 to -0.38 Wm^{-2}).

Results show strong sensitivity of global temperatures to aerosol forcing, with differences in GMST due to aerosol forcing emerging as early as the 1900s. The period from 1951–1980 is particularly sensitive to aerosol forcing, while temperature trends since 1980 are primarily greenhouse gas driven. Ensemble members with weak historical aerosol forcing better match the observed global mean trends, suggesting that aerosol forcing is too strong in the standard configuration of this model. The simulated rapid warming post-1980 exceeds observed trends, suggesting a potential positive bias in the TCR in this model. Using a simple relationship between forcing and response, and assuming that the modeled trend in ERF is a realistic estimate of the real-world ERF trend over this period, we find that the historical TCR estimated from HadGEM3-GC3.1 simulations with scaled aerosol emissions is substantially larger than the estimate derived from observations. However, other plausible explanations for the mismatch in recent observed and simulated trends have also been suggested (e.g., Gregory et al., 2020). We further show that the simulations with reduced aerosol emissions (corresponding to a present-day aerosol forcing of -0.6 and -0.93 Wm^{-2} for the 0.4 and 0.7 scaling ensembles, respectively) better reproduce the observed historical record and that this result appears largely independent of this model's TCR under the assumptions used here.

Our results further suggest a mismatch between the observed and simulated spatial patterns of surface temperature trends in HadGEM3-GC3.1. While it is possible that the specific realization of internal variability that occurred in the real world is not sampled in our model ensemble due to an insufficient number of ensemble members, another possibility is that the simulated temperature response is indeed too spatially

uniform. Without additional ensemble members, it is difficult to distinguish between the two, but the individual ensemble members do highlight that there is considerable variability in the trends and signal-to-noise patterns. However, very few members generate an interhemispheric gradient as seen in observations, suggesting that there may also be biases in the model response to aerosol forcing. Investigating the processes responsible is beyond the scope of this study, but possibilities include biases in the atmospheric transport of aerosol, aerosol lifetime in the atmosphere, and in atmospheric and ocean heat transport.

Our results compare well with previous studies using intermediate complexity Earth System Models that have attempted to constrain aerosol forcing by scaling AA emissions (e.g., Forest, 2018; Libardoni et al., 2018). However, sampling structural and parametric uncertainty results in much wider interval of credible historical aerosol forcing (e.g., Regayre et al., 2018). The ensemble presented here is the first with a fully coupled global climate model to sample a wide range in historical aerosol forcing by scaling emissions within a single climate model framework. It allows an investigation of the role of aerosol forcing for many aspects of global climate variability, setting aside the model structural differences that often make such investigations challenging. We have shown that reducing the aerosol forcing in historical simulations improves agreement with the observed record in HadGEM3-GC3.1. We encourage modeling centers to perform similar experiments to gain a better understanding of the sensitivity of these results to other model characteristics.

References

Adams, B. K., & Dessler, A. E. (2019). Estimating transient climate response in a large-ensemble global climate model simulation. *Geophysical Research Letters*, *46*, 311–317. https://doi.org/10.1029/2018GL080714

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., et al. (2019). Forcings, feedbacks, and climate sensitivity in hadgem3-gc3.1 and ukesm1. *Journal of Advances in Modeling Earth Systems*, *11*, 4377–4394. https://doi.org/10.1029/2019MS001866

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *Journal of Climate*, *28*(4), 1630–1648. https://doi.org/10.1175/jcli-d-14-00545.1

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

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., WatsonParris, D., et al. (2019). Bounding global aerosol radiative forcing of climate change. *Reviews of Geophysics*, *58*, e2019RG000660. https://doi.org/10.1029/2019RG000660

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., & Bellouin, N. (2012). Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, *484*, 228–232. https://doi.org/10.1038/nature10946

Booth, B. B. B., Harris, G. R., Jones, A., Wilcox, L., Hawcroft, M., & Carslaw, K. S. (2018). Comments on "Rethinking the Lower Bound on Aerosol Radiative Forcing". *Journal of Climate*, *31*(22), 9407–9412. https://doi.org/10.1175/JCLI-D-17-0369.1

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). Clouds and aerosols. In T. F. Stocker, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Cowtan, K., Hausfather, Z., Hawkins, E., Jacobs, P., Mann, M. E., Miller, S. K., et al. (2015). Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophysical Research Letters*, *42*, 6526–6534. https://doi.org/10.1002/2015GL064888

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. https://doi.org/10.1002/qj.2297

Dong, B., Sutton, R. T., Highwood, E., & Wilcox, L. (2014). The impacts of European and Asian anthropogenic sulfur dioxide emissions on Sahel rainfall. *Journal of Climate*, *27*(18), 7000–7017. https://doi.org/10.1175/jcli-d-13-00769.1

Ekman, A. M. L. (2014). Do sophisticated parameterizations of aerosol-cloud interactions in CMIP5 models improve the representation of recent observed temperature trends? *Journal of Geophysical Research: Atmospheres*, *119*, 817–832. https://doi.org/10.1002/2013JD020511

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Forest, C. E. (2018). Inferred net aerosol forcing based on historical climate changes: A review. *Current Climate Change Reports*, *4*(1), 11–22. https://doi.org/10.1007/s40641-018-0085-2

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High climate sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters*, *46*, 8329–8337. https://doi.org/10.1029/2019GL083978

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*, 2089–2129. https://doi.org/10.1029/2018MS001603

Gregory, J. M., Andrews, T., Ceppi, P., Mauritsen, T., & Webb, M. J. (2020). How accurately can the climate sensitivity to CO2 be estimated from historical climate change? *Climate Dynamics*, *54*(3), 129–157.https://doi.org/10.1007/s00382-019-04991-y

Gregory, J. M., Andrews, T., & Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO2. *Philosophical Transactions of the Royal Society A*, *373*(2054), 20140417. https://doi.org/10.1098/rsta.2014.0417

Gregory, J. M., & Forster, P. M. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research*, *113*, D23105. https://doi.org/10.1029/2008JD010405

Hardiman, S. C., Andrews, M. B., Andrews, T., Bushell, A. C., Dunstone, N. J., Dyson, H., et al. (2019). The impact of prescribed ozone in climate projections run with HadGEM3-GC3.1. *Journal of Advances in Modeling Earth Systems*, *11*, 3443–3453. https://doi.org/10.1029/2019MS001714

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, *11*(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Jiménez-de-la Cuesta, D., & Mauritsen, T. (2019). Emergent constraints on Earth's transient and equilibrium response to doubled CO2 from post-1970s global warming. *Nature Geoscience*, *12*, 902–905. https://doi.org/10.1038/s41561-019-0463-y

Jones, G. S., Stott, P. A., & Christidis, N. (2013). Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, *118*, 4001–4024. https://doi.org/10.1002/jgrd.50239

Kim, W. M., Yeager, S. G., & Danabasoglu, G. (2018). Key role of internal ocean dynamics in Atlantic multidecadal variability during the last half century. *Geophysical Research Letters*, *45*, 13,449–13,457. https://doi.org/10.1029/2018gl080474

Kretzschmar, J., Salzmann, M., Mülmenstädt, J., Boucher, O., & Quaas, J. (2017). Comment on "Rethinking the Lower Bound on Aerosol Radiative Forcing". *Journal of Climate*, *30*(16), 6579–6584. https://doi.org/10.1175/JCLI-D-16-0668.1

Krishnan, R., Sabin, T. P., Vellore, R., Mujumdar, M., Sanjay, J., Goswami, B. N., et al. (2015). Deciphering the desiccation trend of the South Asian monsoon hydroclimate in a warming world. *Climate Dynamics*, *47*(3-4), 1007–1027. https://doi.org/10.1007/s00382-015-2886-5

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., et al. (2018). The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *Journal of Advances in Modeling Earth Systems*, *10*, 2665–2888. https://doi.org/10.1029/2018MS001370

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the uncertainty model in the Goddard Institute for Space Studies Surface Temperature (GISTEMP) analysis. *Journal of Geophysical Research: Atmospheres*, *124*, 6307–6326. https://doi.org/10.1029/2018JD029522

Libardoni, A. G., Forest, C. E., Sokolov, A. P., & Monier, E. (2018). Baseline evaluation of the impact of updates to the MIT Earth System Model on its model parameter estimates. *Geoscientific Model Development*, *11*(8), 3313–3325. https://doi.org/10.5194/gmd-11-3313-2018

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*(5), 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017

Menary, M. B., Kuhlbrodt, T., Ridley, J., Andrews, M. B., Dimdore-Miles, O. B., Deshayes, J., et al. (2018). Preindustrial Control Simulations With HadGEM3-GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, *10*, 3049–3075. https://doi.org/10.1029/2018ms001495

Mulcahy, J. P., Jones, C., Sellar, A., Johnson, B., Boutle, I. A., Jones, A., et al. (2018). Improved aerosol processes and effective radiative forcing in HadGEM3 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *10*, 2786–2805. https://doi.org/10.1029/2018MS001464

Myhre, G., Shindell, D., Bron, F.-M., Collins, W., Fuglestvedt, J., Huang, J., et al. (2013). Anthropogenic and natural radiative forcing. *Climate change 2013 - the physical science basis on Climate Change, Intergovernmental Panel*. Cambridge: Cambridge University Press.

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

Oudar, T., Kushner, P. J., Fyfe, J. C., Sigmond, M. (2018). No Impact of Anthropogenic Aerosols on Early 21st Century Global Temperature Trends in a Large Initial-Condition Ensemble. *Geophysical Research Letters*, *45*, 9245–9252. https://doi.org/10.1029/2018gl078841

Pincus, R., Forster, P. M., & Stevens, B. (2016). The Radiative Forcing Model Intercomparison Project (RFMIP): Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(9), 3447–3460. https://doi.org/10.5194/gmd-9-3447-2016

Polson, D., Bollasina, M., Hegerl, G. C., & Wilcox, L. J. (2014). Decreased monsoon precipitation in the Northern Hemisphere due to anthropogenic aerosols. *Geophysical Research Letters*, *41*, 6023–6029. https://doi.org/10.1002/2014gl060811

Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M. H., Booth, B. B. B., et al. (2018). Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF. *Atmospheric Chemistry and Physics*, *18*(13), 9975–10,006. https://doi.org/10.5194/acp-18-9975-2018

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. (2017). The Shared Socioeconomic Pathways and their energy land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Robson, J., Sutton, R., & Smith, D. (2014). Decadal predictions of the cooling and freshening of the North Atlantic in the 1960s and the role of ocean circulation. *Climate Dynamics*, *42*(9-10), 2353–2365. https://doi.org/10.1007/s00382-014-2115-7

Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., et al. (2020). Implementation of UK Earth system models for CMIP6. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001946. https://doi.org/10.1029/2019MS001946

Smith, D. M., Booth, B. BB, Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., et al. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*(10), 936–940. https://doi.org/10.1038/nclimate3058

Stevens, B. (2015). Rethinking the lower bound on aerosol radiative forcing. *Journal of Climate*, *28*(12), 4794–4819. https://doi.org/10.1175/JCLI-D-14-00656.1

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

Undorf, S., Polson, D., Bollasina, M. A., Ming, Y., Schurer, A., & Hegerl, G. C. (2018). Detectable impact of local and remote anthropogenic aerosols on the 20th century changes of West African and South Asian monsoon precipitation. *Journal of Geophysical Research: Atmospheres*, *123*, 4871–4889. https://doi.org/10.1029/2017jd027711

Watanabe, M., & Tatebe, H. (2019). Reconciling roles of sulphate aerosol forcing and internal variability in Atlantic multidecadal climate changes. *Climate Dynamics*, *53*, 4651–4665. https://doi.org/10.1007/s00382-019-04811-3

Wilcox, L. J., Highwood, E. J., & Carslaw, K. S. (2015). Quantifying sources of inter-model diversity in the cloud albedo effect. *Geophysical Research Letters*, *42*, 1568–1575. https://doi.org/10.1002/2015GL063301

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

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, *10*, 357–380. https://doi.org/10.1002/2017ms001115

Zelinka, M. D., Andrews, T., Forster, P. M., & Taylor, K. E. (2014). Quantifying components of aerosol-cloud-radiation interactions in climate models. *Journal of Geophysical Research: Atmospheres*, *119*, 7599–7615. https://doi.org/10.1002/2014JD021710

Zhang, R., Delworth, T. L., Sutton, R., Hodson, D. L. R., Dixon, K. W., Held, I. M., et al. (2013). Have aerosols caused the observed Atlantic multidecadal variability? *Journal of the Atmospheric Sciences*, *70*, 1135–1144. https://doi.org/10.1175/jas-d-12-0331.1