Best scale for detecting the effects of stratospheric sulphate aerosol geoengineering on surface temperature

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Earth's Future

RESEARCH ARTICLE

10.1029/2018EF000933

Best Scale for Detecting the Effects of Stratospheric Sulfate Aerosol Geoengineering on Surface Temperature

Y. T. E. Lo[1,2], Andrew J. Charlton-Perez[1], Eleanor J. Highwood[1], and Fraser C. Lott[3]

[1]Department of Meteorology, University of Reading, Reading, UK, [2]School of Geographical Sciences, University of Bristol, Bristol, UK, [3]Met Office Hadley Centre, Exeter, UK

Key Points:
- Large-scale spatiotemporal surface temperature patterns and the global temperature average are best for early geoengineering monitoring
- Reduced global coverage in in situ surface temperature observations increases the chance of detection in the first 5 years of implementation
- Challenging geoengineering detection in surface temperature on subcontinental scales

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. T. E. Lo,
eunice.lo@bristol.ac.uk

Citation:
Lo, Y. T. E., Charlton-Perez, A. J., Highwood, E. J., & Lott, F. C. (2018). Best scale for detecting the effects of stratospheric sulfate aerosol geoengineering on surface temperature. *Earth's Future*, *6*, 1660–1671. https://doi.org/10.1029/2018EF000933

Received 17 MAY 2018
Accepted 17 NOV 2018
Accepted article online 22 NOV 2018
Published online 15 DEC 2018

Abstract Stratospheric sulfate aerosol injection (SAI) has been proposed as a way to geoengineer climate. While swift global mean surface cooling is generally expected from tropical SAI, the regional impacts of such perturbation on near-surface air temperature (SAT) are projected to be spatially inhomogeneous. By using existing simulations from the Geoengineering Model Intercomparison Project G4 scenario, where 5 Tg/year of sulfur dioxide (SO_2) is injected into the tropical stratosphere to offset some of the warming in a midrange representative greenhouse gas concentration pathway (RCP4.5) between 2020 and 2070, we examine the regional detectability of the SAI surface cooling effect and attempt to find the best spatial scale for potential SAI monitoring. We use optimal fingerprint detection and attribution techniques to estimate the time horizon over which the SAI surface cooling effect would be detected after implementation in 2020 on subglobal scales, ranging from the near-global in situ observational coverage down to subcontinental regions. We show that using the spatiotemporal SAT pattern across the Northern and Southern extratropics and the Tropics, and across the Northern and Southern Hemispheres, as well as averaging SATs over the whole globe robustly result in successful SAI detection within 10 years of geoengineering implementation in a majority of the included plausible geoengineering realizations. However, detecting the SAI effect on SAT within the first decade of implementation would be more challenging on subcontinental scales.

Plain Language Summary It has been proposed that we could mimic explosive volcanic eruptions and inject sulfur dioxide gas into the upper atmosphere to cool Earth's surface. Climate models suggest that implementing this technology in the Tropics could lower global average temperature quickly, but it could cool some places more than the others. We use computer simulations of a midrange global warming scenario where this technology has been implemented, to estimate the time needed for its surface cooling effect to be detected over a range of regions of different sizes and locations. We find that we would have the highest chance to detect the cooling effect if we considered the space-time cooling pattern across the extratropics and the Tropics, or across the Northern and Southern Hemispheres, or if we averaged all surface temperature measurements over the globe. Smaller areas are less likely to show detectable surface cooling effects within the first decade of implementation of this technology.

1. Introduction

According to the Hadley Centre-Climatic Research Unit Version 4 (HadCRUT4) data set (Morice et al., 2012), 2015, 2016, and 2017 were the warmest years since records began in 1850 (Osborn, 2018; UK Met Office, 2018). In the face of dangerous climate change and insufficient current ambition to mitigate greenhouse gas emissions for achieving the international climate goal set in the Paris Agreement (Höhne et al., 2017; Rogelj et al., 2016), potential geoengineering methods that aim at deliberately cooling the climate have received increased attention (e.g., Chen & Xin, 2017; Parson, 2017).

Sulfate aerosol injection (SAI) into the stratosphere is potentially one of the most effective and affordable ways of geoengineering the climate (Royal Society Working Group, 2009). Mimicking large volcanic eruptions, SAI involves deliberate injections of stratospheric sulfate aerosols or their precursor, sulfur dioxide (SO_2), to increase Earth's albedo, thereby reducing the amount of incoming solar radiation and lowering surface temperatures. Using climate models, studies have shown that tropical SAI could effectively counteract global warming (e.g., Jones et al., 2010; 2016), although it could simultaneously bring unintended effects to the

Table 1
The Stratospheric Sulfate Aerosol Geoengineering Scheme, Stratospheric Ozone Chemistry Scheme, and the Number of Ensemble Members of the Climate Models Whose G4 Simulations are Used in This Study

Model	Stratospheric aerosols	Stratospheric ozone	Ensemble size
BNU-ESM	Prescribed AOD*	Prescribed	1
CanESM2	Prescribed AOD§	Prescribed	3
HadGEM2-ES	Global SO_2 injection	Prescribed	3
MIROC-ESM	Prescribed AOD*	Prescribed	1

Note. Models with an * prescribed a 25% scaling of the Sato et al. (1993) AOD distribution data set for the 1991 Mount Pinatubo eruption, wheras the model with a § prescribed AOD that was equivalent to 25% of their own Pinatubo simulations. HadGEM2-ES injected 5 Tg/year of SO_2 into the lower stratosphere globally to achieve a global distribution of stratospheric sulfate aerosols amid a lack of realistic stratospheric dynamics (Jones et al., 2010). None of the models have interactive stratospheric ozone chemistry (Eyring et al., 2013). AOD = aerosol optical depth; BNU-ESM = Beijing Normal University Earth System Model; HadGEM2-ES = Hadley Centre Global Environment Model, version 2 — Earth System; MIROC-ESM = Model for Interdisciplinary Research on Climate — Earth System Model; CanESM2 = Second Generation Canadian Earth System Model.

climate system (e.g., Aquila et al., 2014; Ferraro et al., 2014; Pitari et al., 2014), society (e.g., Preston, 2013; Svoboda et al., 2011), and human health (e.g., Eastham, 2015; Effiong & Neitzel, 2016).

Using a new detection and attribution technique, Bürger and Cubasch (2015) assessed the detectability of the effects of SAI on temperature and precipitation, in two Geoengineering Model Intercomparison Project (GeoMIP; Kravitz et al., 2011) scenarios where constant, tropical 5 Tg/year SO_2 injection (GeoMIP G4) and gradual SO_2 injection (GeoMIP G3) is implemented between 2020 and 2070 to offset global warming in the Representative Concentration Pathway 4.5 (RCP4.5; Thomson et al., 2011), respectively. They found that the global-scale temperature and precipitation signals of SAI would be detected after a few years of implementation in G4 and after a decade of implementation in G3. Using spatiotemporal information in their detection and attribution study yielded earlier SAI detectability. Lo et al. (2016) applied both Bürger & Cubasch's (2015) technique and a more conventional optimal fingerprint detection technique developed by Allen and Stott (2003) to G4, to estimate the time horizon between the start of 5 Tg/year SO_2 injection into the tropical lower stratosphere in 2020 and robust detection of its global mean cooling effect. More specifically, Lo et al. (2016) estimated the level of agreement between global mean near-surface air temperature (SAT) pseudo-observations in G4, and model-simulated SAT fingerprints of the SAI and RCP4.5 forcing. They concluded that regardless of the detection technique, the global mean cooling response to 5 Tg/year SAI would likely be robustly detected within 10 years of SAI implementation, that is, between 2020 and 2029, amid increasing greenhouse gas concentrations and internal variability.

While detecting the global mean cooling effect of SAI is an important first step of SAI monitoring because global cooling is likely to be the primary aim of geoengineering, this diagnostic may not be best for early detection of the SAI surface cooling effect for monitoring purposes. This is because the impacts of SAI on SAT are projected to be spatially inhomogeneous (e.g., Yu et al., 2015). Indeed, the existing G4 simulations by Beijing Normal University Earth System Model (BNU-ESM; Ji et al., 2014), Second Generation Canadian Earth System Model (CanESM2; Chylek et al., 2011), Hadley Centre Global Environment Model, version 2 — Earth System (HadGEM2-ES; Bellouin et al., 2011), and Model for Interdisciplinary Research on Climate–Earth System Model (MIROC-ESM; Watanabe et al., 2011) that Lo et al. (2016) used in their global analysis all show spatially inhomogeneous SAT responses to SAI during 2020–2029 (Figures 1a–1d; please refer to Table 1 for the relevant details of these models). Although the strongest cooling effects of SAI in G4 are in the Arctic due to Arctic amplification in all of these models, statistically significant SAI-induced SAT changes are mainly found in low latitudes up to ∼ 60° (areas without hatching indicate significant changes at the 5% level). Furthermore, places such as Northeastern Canada, South Asia, and various parts over the ocean have vastly different SAT responses to SAI across the models. These inhomogeneous cooling effects of SAI raise the novel question of which spatial scale might be best for monitoring the effects of SAI on SAT early into deployment.

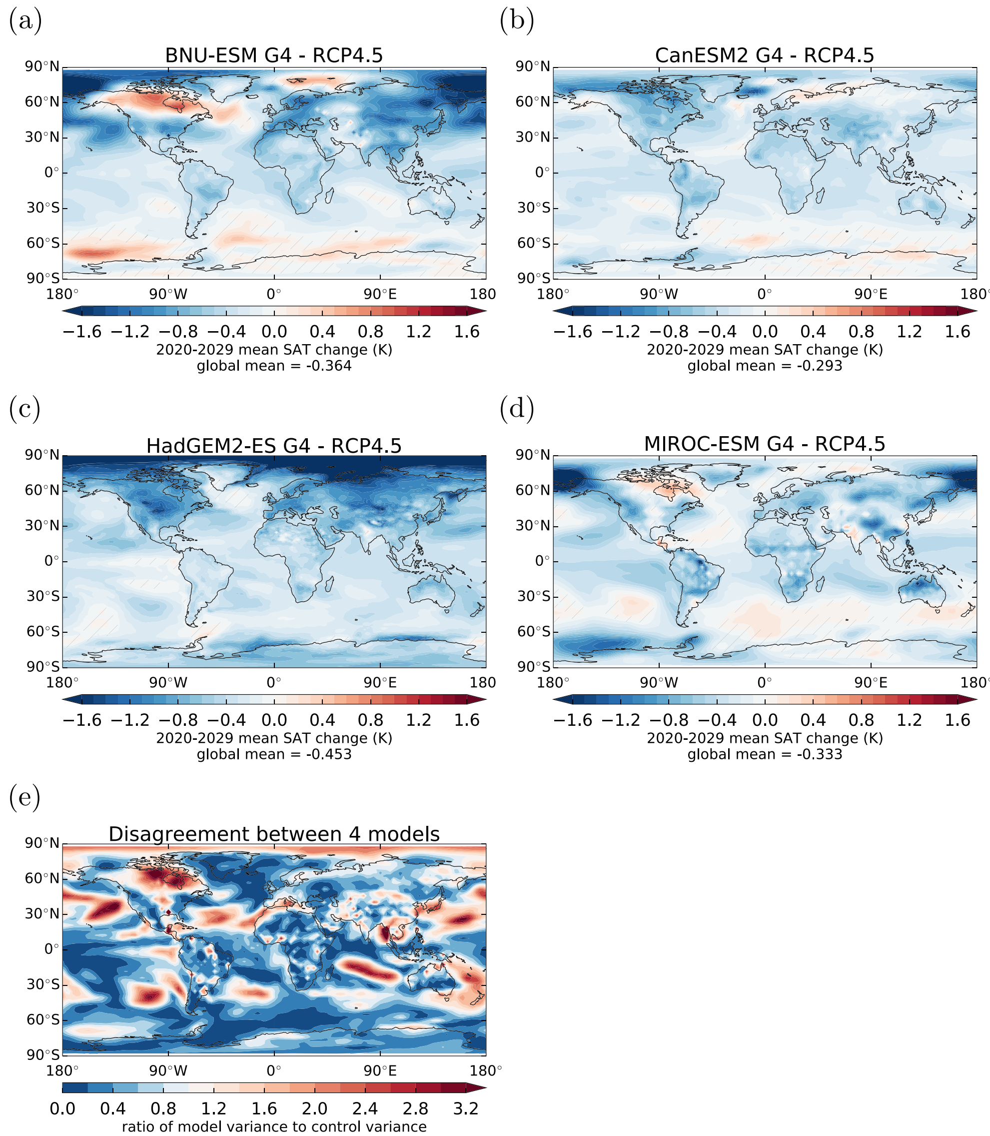

Figure 1. Spatial patterns of the ensemble mean SAT anomaly in G4 with respect to Representative Concentration Pathway 4.5, averaged over the first decade of sulfate aerosol injection deployment in G4 (2020–2029) in (a) BNU-ESM, (b) CanESM2, (c) HadGEM2-ES, and (d) MIROC-ESM. The corresponding global mean SAT changes are annotated under the color bars. Hatch marks indicate statistically insignificant changes from internal interdecadal variability at the 5% level, found with the Welch's *t* test (Welch, 1947). (e) The ratio of the variance across the cooling signals shown in (a)–(d) to that of multimodel internal interdecadal variability. BNU-ESM = Beijing Normal University Earth System Model; HadGEM2-ES = Hadley Centre Global Environment Model, version 2—Earth System; MIROC-ESM = Model for Interdisciplinary Research on Climate—Earth System Model; CanESM2 = Second Generation Canadian Earth System Model; SAT = near-surface air temperature; RCP4.5 = Representative Concentration Pathway 4.5.

We take a regional approach similar to that taken in conventional regional detection and attribution studies (e.g., Bindoff et al., 2013; Stott et al., 2010), to detect the subglobal SAI effects on SAT in the GeoMIP G4 scenario. We first extend Lo et al.'s (2016) work to the near-global, low-resolution HadCRUT4 observational network that covers about 84% of the Earth's surface (gray shading in Figure 2) to investigate how SAI monitoring using this widely cited, in situ observational data set might differ from SAI monitoring under full global coverage similar to that delivered by high-resolution remote sensing products. We then extend our study to the hemispheric, latitudinal, continental, and regional scales, as well as spatiotemporal detection diagnostics, in an attempt to address the novel scientific question of "which spatial scale would be best for detecting the SAI effects in SAT during the first decade of geoengineering implementation?" We suggest the answer to this question has important implications for future SAI monitoring, should it be needed.

10.1029/2018EF000933

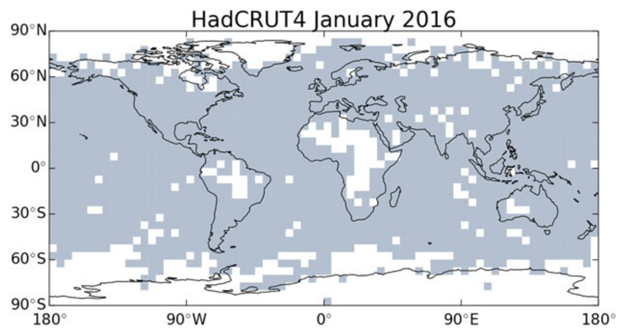

Figure 2. Missing data (white grid cells) in the HadCRUT4 network as of January 2016. Unsampled regions are concentrated in the high latitudes, Africa and South America. HadCRUT4 = Hadley Centre-Climatic Research Unit Version 4.

2. Methods

2.1. Optimal Fingerprint Detection in G4

The total least squares (TLS) detection and attribution model (Allen & Stott, 2003) is used in this study, as it was in Lo et al. (2016). In short, this model quantifies the level of agreement between observed changes in the real world and patterns of externally forced changes simulated by a climate model, through multivariate TLS linear regression. The climate system's internal (unforced) variability is the null hypothesis, and detection of an externally forced climate signal is claimed when the observed changes in climate cannot be ascribed to natural internal variability alone at a certain confidence level.

Applying the TLS detection model to the G4 scenario, which, by design, includes an RCP4.5 base climate and the forcing from the deliberately injected stratospheric sulfate aerosols, the detection model can be written as

$$\mathbf{y} = (\mathbf{x}_{\mathrm{RCP}} - \nu_{\mathrm{RCP}})\beta_{\mathrm{RCP}} + (\mathbf{x}_{\mathrm{SAI}} - \nu_{\mathrm{SAI}})\beta_{\mathrm{SAI}} + \nu_0 \quad (1)$$

where $\mathbf{y}$ is the observations, $\mathbf{x}_{\mathrm{RCP}}$ and $\mathbf{x}_{\mathrm{SAI}}$ are the model-simulated fingerprints of the RCP4.5 and geoengineering aerosol forcing, ν_{RCP} and ν_{SAI} are the sampling uncertainty in the corresponding fingerprints, β_{RCP} and β_{SAI} are the corresponding scaling factors to be estimated, and ν_0 is unforced climate variability.

Just as in Lo et al. (2016), $\mathbf{y}$ is unknown here because geoengineering by means of 5 Tg/year SO_2 injection on top of RCP4.5 is only hypothetical. Therefore, individual G4 realizations from the four climate models included in Figure 1 are used in turn to represent plausible observed subglobal SAT trajectories from 2020. These pseudo-observations are regressed against subglobal RCP4.5 and G4 SAT fingerprints that are simulated by the rest of the climate models, individually or in the form of multimodel average. We follow Lo et al.'s (2016) linear transformation approach for deriving β_{SAI} from the scaling factors of the RCP4.5 and G4 fingerprints. Cross comparing all available pseudo-observations and climate models results in 32 pseudo-observation model comparisons. These comparisons are studied in all of the spatial scales presented in this article.

For hypothesis testing, ν_0 is estimated from the preindustrial control simulations of a six-model ensemble (section 2.2), following Lo et al.'s (2016) choice. This ensemble includes Beijing Normal University Earth System Model, CanESM2, Commonwealth Scientific and Industrial Research Organisation Mk3L climate system model version 1.2 (CSIRO-Mk3L-1-2 Phipps et al., 2011), Goddard Institute for Space Studies ModelE2 Russell (GISS-E2-R; Schmidt et al., 2014), Hadley Centre Global Environmental Model Version 2, and MIROC-ESM. These models are chosen because the power spectra of their preindustrial temperature variability is comparable to those of the standardized 1981–2015 HadCRUT4 and the GISS Surface Temperature Analysis time series, on the time and spatial scales of interest.

The observations and model-simulated fingerprints are projected onto κ leading modes of variability, or empirical orthogonal functions (EOFs), before regression, following the standard optimization procedure detailed in Allen and Stott (2003); κ is chosen via a residual consistency test, which compares the weighted sum of squared regression residuals to the model-simulated noise variance via an F test (Allen & Stott, 2003). For each pseudo-observation model comparison in each studied spatial scale, we choose the highest EOF at which the F test probability is within the 5–95% range and the corresponding estimated scaling factor varies little with the number of EOFs retained. It is important that a substantial fraction of the original data is explained in this truncated space. Figure S1 in the supporting information shows that more than 50% of the variance of the original G4 fingerprint (as simulated by individual climate models or multiple models) is explained in the range of chosen EOFs in representative regions of the studied spatial scales or detection diagnostics. Other studied regions or diagnostics have similarly high fractions of explained variance. This suggests meaningful detection results in section 3.

The TLS detection and attribution model takes into account sampling uncertainty in the fingerprints, in this study this is ν_{RCP} and ν_{SAI}. Since ensemble averaging in the fingerprints (see section 2.2) reduces uncertainty, we scale the RCP4.5 and G4 fingerprints with the square root of the number of ensemble members included to match the noise variance in each fingerprint to that in the observations. The estimated scaling factors, β_{RCP}

and β_{G4}, are also scaled accordingly before linear transformation (see above). Please refer to Allen and Stott (2003) for more details regarding noise in the fingerprints.

Equation 1 is solved, and the best estimate scaling factors are found, by minimizing the sum of squared perpendicular distances from the best fit two-dimensional plane to the κ noise-contaminated observations and simulated points, following the methodology detailed in Allen and Stott (2003). The uncertainty associated with the best estimate scaling factors is found by defining a set of points on a two-dimensional sphere whose radius equals the critical value corresponding to the 90th percentile of an F distribution. For each of these points, the corresponding scaling factors are estimated. The minimum and maximum of these scaling factors form the 90% confidence interval for the best estimate.

In order to estimate the time horizon over which the regional SAT response to SAI would be detected after deployment in 2020 in G4, we lengthen $\mathbf{y}$, $\mathbf{x}_{RCP}$, $\mathbf{x}_{G4}$ (the G4 fingerprint), and segments of ν_0 progressively by 5 years until $\mathbf{x}_{SAI}$ is detected at the 10% significance level, that is, when β_{SAI} and its two-tailed confidence level (its 5 to 95 percentile) do not include 0. The first year at which the SAI cooling effect would be detected is recorded as the *SAI detection horizon*, and this metric is estimated at 5-year resolution in each pseudo-observation model comparison on each studied spatial scale. We focus on the detectability of the SAI surface cooling effect in the first 10 years of implementation, as this timescale is thought to be most policy relevant. The following section outlines the data preprocessing procedure for each included experiment.

2.2. Data Preprocessing

The detectability of the average SAT response to SAI over the near-global HadCRUT4 network, the global scale except the polar regions (60° N to 60° S), the Northern and Southern Hemispheres, the Arctic (90–60° N), Northern midlatitudes (60–25° N), Tropics (25° N to 25° S), Southern midlatitudes (25–60° S) and Antarctica (60–90° S), as well as five continental areas and 12 SREX regions (Hewitson et al., 2014) are estimated in this study. On all of these scales the detection diagnostic is the time evolution of annual mean area mean SAT anomalies that start in 2020, relative to the 2006–2019 mean. Additionally, detection is attempted in the spatiotemporal SAT pattern across the Northern and Southern Hemispheres; the Northern extratropics, Tropics, and the Southern extratropics; as well as several subregions within the five continental scale areas.

Single, existing G4 realizations from the four climate models included in Figure 1 are used as pseudo-observations in turn. The two fingerprints are the RCP4.5 and G4 ensemble mean annual mean area mean SAT time series, or spatiotemporal SAT patterns where applicable, generated by the individual climate models that are not used to represent the pseudo-observations, or the average of these models. All of these time series consist of 14-year moving trends, a technique that is equivalent to smoothing the time series with the 14-year $C1$ filter originally suggested by Bürger and Cubasch (2015). This means climate in a year is estimated from the trend over its preceding 14 years. A 5-year fingerprint spanning 2020–2024 thus has its last data point (corresponding to year 2024) estimated from the trend over 2010–2023, whereas a 10-year fingerprint spanning 2020–2029 has its last data point (corresponding to year 2029) estimated from the trend over 2015–2028. Lo et al. (2016) found that 14-year trends worked best with the TLS algorithm for global mean SAI detection during the first decade of implementation. Using 14-year trends with the TLS detection algorithm here allows direct comparison of our subglobal detection results with Lo et al.'s (2016) global mean results.

Model output on the spatial scales of interest are extracted in the following ways prior to spatial and temporal averaging and filtering. For the near-global HadCRUT4 experiment, all model SAT output are first regridded onto HadCRUT4's 5° by 5° grid. Throughout this study, area-weighted regridding is performed where applicable. The unsampled grid cells in HadCRUT4 in January 2016 (colored in white in Figure 2) are then given zero weight when area-weighted global average SAT is calculated.

For the hemispheric and latitudinal experiments, model output is also regridded onto the 5° by 5° grid. Average SATs over the hemispheric and latitudinal areas are then calculated separately, without masking of the HadCRUT4 unsampled regions. We do not consider the HadCRUT4 unsampled regions on spatial scales smaller than the global scale, in order to identify the best scales for early SAI monitoring for design of future monitoring systems that may not be constrained by the current in situ observational coverage.

For the continental and SREX regional experiments, model SAT output is regridded onto CSIRO-Mk3L-1-2's 3.2° by 5.6° latitude/longitude grid instead, as this resolution is the coarsest in the six-model ensemble used for estimating internal variability. Unifying model resolution here ensures the same grid cells within the continental areas and regions of interest are extracted across the pseudo-observations, fingerprints, and multimodel

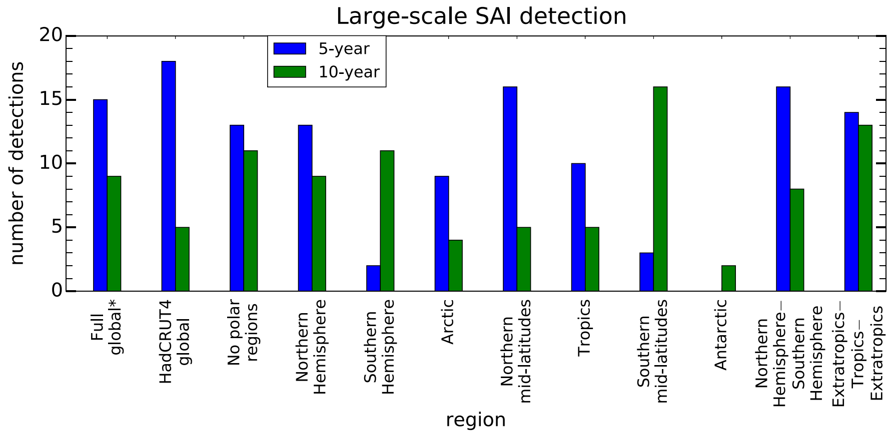

Figure 3. The number of successful within-a-decade SAI detections in near-surface air temperature time series or spatial-temporal series beginning in 2020 in each large-scale experiment, including the full global experiment studied in Lo et al. (2016) (denoted by an asterisk [*] in figure). The same 32 pseudo-observation model comparisons are made in each experiment. Blue color represents the number of 5-year detections (detections by 2025), while green color represents the number of 10-year detections (detections by 2030). HadCRUT4 = Hadley Centre-Climatic Research Unit Version 4; SAI = sulfate aerosol injection.

preindustrial control simulations for fair comparison. Area-weighted continental or regional average SATs are then computed.

Two global scale and five continental scale spatiotemporal detection diagnostics are considered in this study. In the Northern Hemisphere-Southern Hemisphere diagnostic, the smoothed, annual mean hemispheric mean (ensemble mean as well for the fingerprints) SAT anomalies with respect to the corresponding 2006–2019 mean are computed for each hemisphere, according to the data preprocessing procedure described above. Individual hemispheric time series are then sorted into spatiotemporal pseudo-observations and fingerprints as input to the TLS detection algorithm. The equivalent is done for the Northern extratropics-Tropics-Southern extratropics diagnostic, which consists of three latitudinal bands spanning 90–25° N, 25° N to 25° S and 25–90° S, and all continental scale areas, which consist of three to four SREX regions. In other words, the spatiotemporal detection diagnostics have an additional dimension of the number of subareas within a larger area. They thus contain information about the spatial SAT contrast between neighbouring subareas in each larger area, in addition to the temporal SAT evolution within each subarea.

Finally, unforced preindustrial control simulations from the six-model ensemble mentioned in section 2.1 are used to estimate internal variability, following Lo et al.'s (2016) choice. This means internal variability in a certain detection experiment is the pooled estimate from the six models and remains the same in all 32 pseudo-observation model comparisons. Each segment of the preprocessed control simulations is then treated the same way as the corresponding pseudo-observations according to the experiment, as described in the previous paragraphs.

3. Results

Throughout section 3, we present the total number of examined pseudo-observation model comparisons in which the surface cooling effect of SAI would be detected during the first 5 and 10 years of SAI implementation in G4 (i.e., 2020–2024 and 2020–2029, respectively) in the studied spatial areas and detection diagnostics. The total number of 5- and 10-year SAI detections is referred to as *the number of within-a-decade* (*WAD*) *SAI detections* hereafter. We use the number of WAD SAI detections over a certain area or diagnostic as a measure of the efficacy of the area or diagnostic for effective SAI monitoring in SAT.

3.1. HadCRUT4, Hemispheric, and Latitudinal Detection

Figure 3 shows the number of 5- (blue color) and 10-year detections (green color) in the HadCRUT4 coverage, the global average except the polar regions (60° N to 60 ° S), the Northern and Southern Hemispheres, as well as the Arctic (90–60° N), Northern midlatitudes (60–25° N), Tropics (25° N to 25° S), Southern midlatitudes (25–60° S) and Antarctica (60–90° S). For easy comparison, the number of WAD SAI detections in Lo et al.'s (2016) full global experiment in the same 32 pseudo-observation model comparisons is also shown in Figure 3. Note that Lo et al. (2016) examined 44 pseudo-observation model comparisons in total. We removed CSIRO-Mk3L-1-2, the climate model that modeled SAI by reducing the solar constant (Bürger & Cubasch, 2015), from their pseudo-observation model comparisons and added eight comparisons that use the multimodel mean as fingerprints (section 2.2) to arrive at the shown results for the full global experiment.

Using the HadCRUT4 near-global coverage, the cooling effect of SAI would be detected within 10 years of SAI implementation in 23 of the 32 comparisons. This result is similar to the 24 found for the full global average, indicating that both the SAT averages in the HadCRUT4 data set and the full global coverage that could be derived from high-resolution satellite remote sensing products (e.g., Jackson & Wick, 2010) would be similarly efficacious for effective SAI monitoring in a scenario like G4.

Nonetheless, three more 5-year SAI detections are found in the HadCRUT4 coverage than the full global coverage. This suggests that the limited resolution and coverage of our current in situ observational network would increase the likelihood of detecting the SAI cooling effect after just 5 years of SAI implementation. Factors such as the lack of in situ temperature measurements in the polar regions where internal variability is high and thus rejection of the null hypothesis (section 2.1) is challenging, and the exclusion of the Arctic Ocean, Northeastern Canada and parts of South Asia where the difference in model responses to sulfate aerosols is large (Figure 1e) may have contributed to the increased number of 5-year detections in the HadCRUT4 coverage. However, the relatively low number of 5-year detections, 13, found for 60° N to 60° S suggests that exclusion of the polar regions alone cannot explain the increased 5-year detectability in the HadCRUT4 coverage.

The Northern and Southern Hemispheres give very different numbers of WAD SAI detections, with the Northern Hemisphere giving 22 (of which 13 are 5-year detections) and the Southern Hemisphere giving only 13 (of which 2 are 5-year detections). These results indicate that the Northern Hemisphere would be more efficacious than the Southern Hemisphere for early SAI monitoring in SAT in the event of 5 Tg/year SAI. The large contrast in the WAD detectability of the cooling effect of SAI in G4 between the two hemispheres may be attributed to the fact that the Northern Hemisphere has a higher proportion of land, which responds quicker to climate forcing and thus provides a larger cooling signal for SAI detection in the first decade of deployment than the ocean (Figure 1).

On the latitudinal scale, the Northern midlatitudes and the Southern midlatitudes give the highest numbers of WAD SAI detections (21 and 19), respectively (Figure 3). This means that although Figure 1 has shown that the 2020–2029 mean SAI cooling that is statistically significant from internal variability is mainly confined to the Tropics, the midlatitudes outperforms the Tropics for early SAI detection with the optimal fingerprint technique. This is because the temporal pattern of the SAI fingerprint over 2020–2029 is consistent with that of the pseudo-observations in more comparisons in the midlatitudes than the other latitudinal bands.

The Antarctic and Arctic give the lowest numbers of WAD SAI detections (2 and 13, respectively) among the latitudinal bands. The very low WAD SAI detectability in the Antarctic may be attributed to the statistically insignificant model responses to SAI in the region (Figures 1a–1d), whereas the low Arctic WAD SAI detectability may be explained by the large model difference in the Arctic cooling response to SAI (Figure 1e). All in all, our latitudinal detection results suggest considerable challenge in detecting the SAI surface cooling effect on the latitudinal scale in the event of 5 Tg/year SO_2 injection in a midrange warming scenario, with the polar regions being the least efficacious for early SAI monitoring.

3.2. Global Scale Spatiotemporal Detection

In this section, we examine two global scale spatiotemporal detection diagnostics to investigate whether the contrast in the SAT response to 5 Tg/year SO_2 injection between the Northern Hemisphere and the Southern Hemisphere, and between the extratropics and the Tropics (Figure 1) would increase SAI detectability on the global scale in the first decade of implementation in G4. Considering the SAT contrast between the Northern and Southern Hemispheres in addition to the temporal SAT evolution in each hemisphere gives a total of 24 WAD detections, of which 16 happen in the first 5 years of SAI implementation (Figure 3). Although these results are similar to that of the full global average, they show an increase in the 5-year detectability from the hemispheric experiments above. This means the hemispheric contrast in the SAT response to the sulfate aerosols provides useful information for SAI detection and monitoring in the first 5 years of implementation.

We found in the previous section that the WAD SAI detectability varies with latitude. Here we show that the latitudinal contrast in the SAT response to SAI would be of great use to monitor the global scale effect

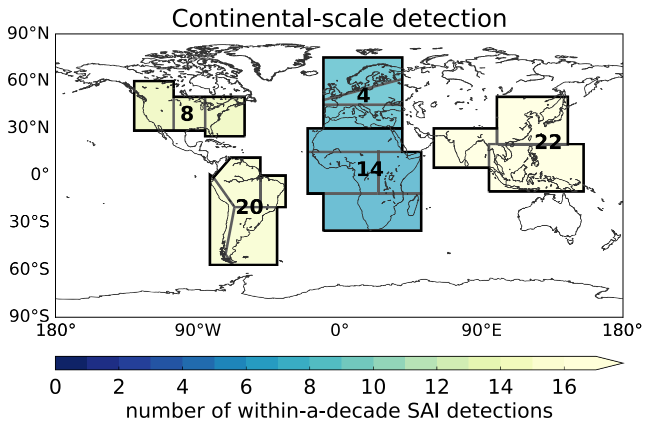

Figure 4. The number of WAD SAI detections (out of 32 comparisons) in different continental scale areas. The bold black lines indicate the boundaries of these areas. The gray lines within the areas indicate the boundaries between the subregions (SREX regions) that constitute the continental scale areas. The shading of each continental scale area shows the number of WAD SAI detections when the temporal evolution of the near-surface air temperature averaged over the area is used as the detection diagnostic. The number annotated in bold in each area represents the number of WAD SAI detections obtained from using the temporal evolution of the spatial near-surface air temperature pattern across the constituting subregions as the detection diagnostic. SAI = sulfate aerosol injection; WAD = within a decade.

of SAI on SAT early into implementation, as a total of 27 WAD SAI detections are found with the Northern extratropics-Tropics-Southern extratropics spatiotemporal diagnostic (Figure 3). This high number of WAD SAI detections suggests that this global scale spatiotemporal SAT pattern should be considered for effective SAI monitoring in the event of 5 Tg/year SO_2 injection.

3.3. Continental and Continental Scale Spatiotemporal Detection

We examine five continental scale areas (bounded by bold back lines in Figure 4) in this section. They roughly represent the contiguous United States (130–60° W, 60–25° N), South America (82–34° W, 11.4° N to 56.7° S), Europe (10° W to 40° E, 75–30° N), Africa (20° W to 52° E, 30° N to 35° S) and South and East Asia (60–155° E, 50° N to 10° S). Areas in high latitudes are not considered here given the low Arctic and Antarctic WAD SAI detectabilities found in section 3.1.

Each of the continental scale areas consists of three to four SREX regions defined in Hewitson et al. (2014), whose boundaries are indicated by the gray lines in Figure 4. Shading in each continental scale area in Figure 4 indicates the number of WAD SAI detections found in the mean SAT over the corresponding continental scale area, whereas the bold number annotated in each area indicates the number of WAD SAI detections found in the spatiotemporal SAT diagnostic (section 2.2) within the corresponding continental scale area. The breakdown of these results, and those from the other experiments, into the numbers of 5- and 10-year detections can be found in Table 2.

South and East Asia gives the highest number of WAD SAI detections, 16; whereas Africa gives the lowest number of WAD detections, 7, when the time evolution of the average SAT over the continental scale areas is used as the detection diagnostic. These results suggest that while we may have a higher chance in monitoring the average SAT response to 5 Tg/year SO_2 injection such as that hypothesized in G4 in South and East Asia than in the Contiguous United States, South America, Europe, and Africa during the first decade of SAI implementation, the continental scale may not be as efficacious as the larger scales for SAI monitoring in mean SAT. This is mainly because averaging SATs over the smaller, continental scale areas results in larger climate noise than averaging over larger areas, leading to a smaller signal-to-noise ratio for effective SAI detection.

With the use of the spatiotemporal diagnostic, however, more WAD SAI detections are found in South America, Africa and South and East Asia (rows denoted with "(ST)" in Table 2 show detection results found with the spatiotemporal diagnostics). These results suggest that additional spatial information across neighboring subcontinental regions would likely be useful for improving the WAD SAI detectability on the continental scale.

However, additional spatial information does not result in higher WAD SAI detectability in the contiguous United States and Europe. This may be attributed to the fact that all of the SREX regions within the contiguous United States and Europe have relatively large internal variability, and only averaging is sufficient in reducing this noise to enable early SAI detection. Our results therefore imply that the optimal detection diagnostic for effective SAI monitoring in SAT would be area dependent, and it would be imprudent to regard any particular diagnostic as the one-size-fits-all approach to effective SAI monitoring without further investigation.

3.4. Regional Detection

We now proceed to look at the regional scale. SAI detection is attempted in the time series of annual mean regional mean SAT anomalies since 2020 in 12 SREX regions: West North America (130–105° W, 60–28.6° N), Central North America (105–85° W, 50–28.6° N), East North America (85–60° W, 50–25° N), Central America (118.3–68.8° W, 28.6° N to 1.2° S), Amazon (79.9–50° W, 11.4° N to 20° S), North Europe (10° W to 40° E, 75–48° N), Mediterranean (10° W to 40° E, 45–30° N), Southern Africa (10° W to 52° E, 11.4–35° S), South Asia (60–100° E, 30–5° N), East Asia (100–145° E, 50–20° N), Southeast Asia (95–155° E, 20° N to 10° S) and Pacific Islands region (155–210° E, 5° N to 5° S). Please refer to Hewitson et al. (2014) for the vertex locations of these polygonal regions.

Table 2
The Number of 5-year, 10-year, and WAD Sulfate Aerosol Injection Detections Found in the 32 Comparisons in Each Experiment

Region	5-year (/32)	10-year (/32)	WAD (/32)
Idealized global*	15	9	24
HadCRUT4	18	5	23
No polar regions	13	11	23
Northern Hemisphere	13	9	22
Southern Hemisphere	2	11	13
Arctic	9	4	13
Northern midlatitudes	16	5	21
Tropics	10	5	15
Southern midlatitudes	3	16	19
Antarctic	0	2	2
Northern Hemisphere – Southern Hemisphere (ST)	16	8	24
Extratropics – Tropics – Extratropics (ST)	14	13	27
Contiguous United States	7	7	14
South America	6	9	15
Europe	6	2	8
Africa	4	3	7
South & East Asia	12	4	16
Contiguous United States (ST)	6	2	8
South America (ST)	5	15	20
Europe (ST)	2	2	4
Africa (ST)	10	4	14
South & East Asia (ST)	14	8	22
West North America	3	1	4
Central North America	1	8	9
East North America	2	3	5
Central America	1	3	4
Amazon	6	5	11
North Europe	4	0	4
Mediterranean	1	3	4
Southern Africa	3	6	9
South Asia	6	4	10
East Asia	7	16	23
Southeast Asia	7	2	9
Pacific Islands region	8	4	12

Note. ST denotes a spatiotemporal detection diagnostic over the stated area. HadCRUT4 = Hadley Centre-Climatic Research Unit Version 4; WAD = within a decade.

Shading in Figure 5 shows the number of WAD SAI detections in the 12 SREX regions. These regions are chosen because they are regions of the world that are relatively populated, vulnerable to climate change, or politically important in climate policymaking, and they lie in latitudes where relatively high numbers of WAD SAI detections were found in section 3.1. Except in North Europe, the cooling response to SAI in all of the studied regions emerge from climate noise during the first decade of SAI implementation in G4 in at least three of the four included models (Figure S2; please refer to Text S1 in the supporting information for details; Cleveland, 1981; Hawkins & Sutton, 2012). This demonstrates that the SAI signal-to-noise ratio is reasonably large for WAD detection in a vast majority of the included SREX regions. Caution should be taken with regard to the results for North Europe.

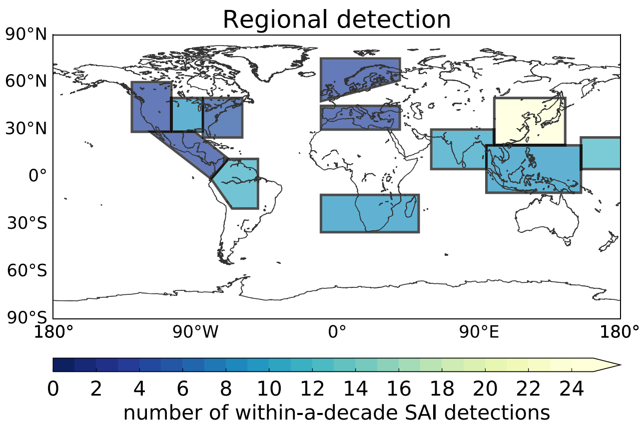

Figure 5. The number of within-a-decade SAI detections (out of 32 comparisons) in different SREX regions. SAI = sulfate aerosol injection.

The highest number of regional WAD SAI detections found is 23 in East Asia. East Asia stands out from the rest of the SREX regions in terms of efficacy for SAI monitoring in the first decade of implementation likely because its SAT response to SAI has similar temporal patterns between the pseudo-observations and model-simulated fingerprints. Nevertheless, although similarly high numbers of WAD detections are found in East Asia and in the global averages, far fewer, 7, 5-year detections are found in East Asia. This demonstrates that the regionalscale is less efficacious for SAI monitoring in the first 5 years of implementation than the global scale. The rest of the studied SREX regions have only 4 (e.g. Central America) to 12 (Pacific Islands region) WAD SAI detections (Figure 5 and Table 2). Owing large climate noise as a result of averaging over the small SREX areas, detection of the SAI cooling signal within the first decade of 5 Tg/year SO_2 injection is generally difficult on the regional scale. For this reason, we do not examine the remainder of the SREX regions defined in Hewitson et al. (2014).

4. Discussion and Conclusions

By assuming 5 Tg/year SO_2 injection into the tropical stratosphere on top of an RCP4.5 base climate as described in the GeoMIP G4 scenario (Kravitz et al., 2011), we estimated the detectability of the SAT response to stratospheric SAI on subglobal scales. Lo et al. (2016) concluded that the global mean cooling effect of 5 Tg/year SAI would be robustly detected 10 years into geoengineering implementation using TLS optimal fingerprint techniques. We applied the conventional detection technique they employed and extended their detection analysis to smaller spatial scales, in an attempt to find the best spatial scale for effective SAT monitoring in SAT.

We compared the average SAT responses in the near-global HadCRUT4 coverage (Figure 2), the Northern and Southern Hemispheres, the global scale excluding the polar regions, five latitudinal bands, five continental scale areas, as well as 12 SREX regions (Hewitson et al., 2014) with their corresponding assumed SAT pseudo-observations. Furthermore, we investigated how additional information of the spatial contrast in the SAT response across the Northern and Southern Hemispheres, the extratropics and the Tropics, and several SREX regions would affect SAI detection on the global and continental scales.

Owing to the spatially inhomogeneous SAT changes projected in G4 relative to the RCP4.5 base climate, the different model responses to SAI in G4, the difference in the statistical significance of these responses, and the wide range of surface areas and amount of information considered among the detection experiments, the resulting numbers of WAD SAI detections vary substantially across the detection experiments, even though the same pseudo-observation model comparisons were made throughout.

So which spatial scale would be best for detecting the SAI effect in SAT during the first decade of geoengineering implementation? By using the number of WAD SAI detections in a total of 32 comparisons as a measure of the early detectability of SAI in a certain region or detection diagnostic, (in descending order) the Northern extratropics-Tropics-Southern extratropics spatiotemporal pattern, the Northern Hemisphere-Southern Hemisphere spatiotemporal pattern, and the full global average would be best for SAI monitoring during the first decade of SAI implementation in the G4 scenario (Table 2). These results indicate that the spatiotemporal and temporal SAT patterns on the global scale would be best for effective SAI monitoring in a scenario like G4. Our results suggest considerable challenge in monitoring the regional effects of SAI on SATs in a scenario like G4.

Nevertheless, WAD SAI detection in SAT is not very likely (>90% probability) even on the global scale. There is a 84% probability of successful SAI detection in the first decade of implementation in the Northern extratropics-Tropics-Southern extratropics spatiotemporal SAT pattern, a probability that is lower than the 90% confidence level used for detection against internal variability in individual pseudo-observation model comparisons in this study. Despite the foreseeably strong incentive in detecting the surface cooling effects on SAI in the event of deployment (because surface cooling is likely to be the aim of SAI), SAT may not be the best climate variable for early SAI monitoring. The vertical temperature profile or top-of-atmosphere shortwave radiation may provide a stronger signal for early SAI monitoring. Future research could look into detecting the SAI signal in these variables.

Future research could also examine other time periods within the first decade of SAI implementation, as SAI-forced climate responses would likely vary with time. Since SAI hypothetically starts in 2020 in the G4 scenario, we postulated that detection of the SAI effects would begin in 2020. In reality, there is little limitation on the time period the observations and fingerprints in a detection study span. Lo et al. (2016) examined a global mean diagnostic beginning in 2000 for detecting the SAI signal in G4. Although this predeployment diagnostic was found to be less efficacious than the equivalent diagnostic beginning in 2020 in terms of early SAI monitoring, future work could use diagnostics that span, for example, 2025–2029 to investigate whether the SAI-forced response in other periods would result in a high probability of successful detection.

Since geoengineering by means of SAI has not been implemented in the real world, our study relied on the hypothetical G4 scenario and climate model simulations. We chose G4 because a plausible way of future SAI monitoring and control would be to robustly detect the climate signal of constant SAI (such as that applied in G4) before altering the injection rate or location to meet climate goals, if necessary. Nevertheless, other ways of SAI implementation have been discussed in geoengineering studies. For example, increasing amounts of SO_2 could be injected to keep top-of-atmosphere radiative forcing constant (e.g. GeoMIP G3; Kravitz et al., 2011), time-varying SO_2 injections could be made at several independent locations to meet multiple climate objectives simultaneously (Kravitz et al., 2017; MacMartin et al., 2017), and SO_2 could be injected into seasonally varying areas to achieve more zonally uniform shortwave radiative forcing (Laakso et al., 2017). As already illustrated by Bürger and Cubasch (2015) with the difference between GeoMIP G3 and G4, different injection strategies would result in different temporal and spatial patterns of the geoengineering climate signal and, therefore, different detectabilities of the signal. We do not expect our results to hold for all SAI scenarios but conclude that they are indicative of the best scale for monitoring SAI in SAT within the first decade since deployment, should constant SO_2 injection be implemented as hypothesized in G4.

Acknowledgments
The CMIP5 and GeoMIP output can be obtained from the Earth System Grid Federation (http://esgf.llnl.gov). The Optimal Fingerprint package used in this study can be obtained from Environment Canada (see https://www.wcrp-climate.org/ictp-2014-tutorials). All data preprocessing was done with the Python library, Iris (http://scitools.org.uk/iris/docs/latest/). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank all participants of the Geoengineering Model Intercomparison Project and their model development teams, CLIVAR/WCRP Working Group on Coupled Modeling for endorsing GeoMIP, and the scientists managing the Earth System Grid data nodes who have assisted with making GeoMIP output available. We thank Environment Canada for the release of and technical support for the Environment Canada's Optimal Fingerprint package. Y. T. Eunice Lo was funded by the University of Reading (ref GS15-108). Fraser C. Lott was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra.

References

Allen, M. R., & Stott, P. A. (2003). Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dynamics*, *21*(5-6), 477–491.

Aquila, V., Garfinkel, C. I., Newman, P. A., Oman, L. D., & Waugh, D. W. (2014). Modifications of the quasi-biennial oscillation by a geoengineering perturbation of the stratospheric aerosol layer. *Geophysical Research Letters*, *41*, 1738–1744. https://doi.org/10.1002/2013GL058818

Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., Hinton, T. J., et al. (2011). The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, *4*(3), 723–757.

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., et al. (2013). Detection and attribution of climate change: From global to regional. In T. F. Stocker (Ed.), *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, pp. 867–952.

Bürger, G., & Cubasch, U. (2015). The detectability of climate engineering. *Journal of Geophysical Research: Atmospheres*, *120*, 11,404–11,418. https://doi.org/10.1002/2015JD023954

Chen, Y., & Xin, Y. (2017). Implications of geoengineering under the 1.5° C target: Analysis and policy suggestions. *Advances in Climate Change Research*, *8*(2), 123–129.

Chylek, P., Li, J., Dubey, M. K., Wang, M., & Lesins, G. (2011). *Atmospheric Chemistry and Physics Discussions*, *11*(8), 22893–22907. Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model (CanESM2).

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, *35*(1), 54.

Eastham, S. D. (2015). Human health impacts of high altitude emissions. Massachusetts Institute of Technology.

Effiong, U., & Neitzel, R. L. (2016). Assessing the direct occupational and public health impacts of solar radiation management with stratospheric aerosols. *Environmental Health*, *15*(1), 7.

Eyring, V., Arblaster, J. M., Cionni, I., Sedláček, J., Perlwitz, J., Young, P. J., et al. (2013). Long-term ozone changes and associated climate impacts in CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, *118*, 5029–5060. https://doi.org/10.1002/jgrd.50316

Ferraro, A. J., Highwood, E. J., & Charlton-Perez, A. J. (2014). Weakened tropical circulation and reduced precipitation in response to geoengineering. *Environmental Research Letters*, *9*(1), 014001.

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*, L01702. https://doi.org/10.1029/2011GL050087

Hewitson, B. C., Carter, T. R., Giorgi, F., Jones, R. G., Kwon, W. T., & van Aalst, M. (2014). Regional context. In V. R. Barros, et al. (Ed.), *Climate change 2014: Impacts, adaptation, and vulnerability. Part b: Regional aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1133–1820). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Höhne, N., Kuramochi, T., Warnecke, C., Röser, F., Fekete, H., Hagemann, M., et al. (2017). The paris agreement: Resolving the inconsistency between global goals and national contributions. *Climate Policy*, *17*(1), 16–32.

Jackson, D. L., & Wick, G. A. (2010). Near-surface air temperature retrieval derived from AMSU-A and sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, *27*(10), 1769–1776. https://doi.org/10.1175/2010JTECHA1414.1

Ji, D., Wang, L., Feng, J., Wu, Q., Cheng, H., Zhang, Q., et al. (2014). Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1. *Geoscientific Model Development*, *7*(5), 2039–2064.

Jones, A., Haywood, J., Boucher, O., Kravitz, B., & Robock, A. (2010). Geoengineering by stratospheric SO2 injection: Results from the Met Office HadGEM2 climate model and comparison with the Goddard Institute for Space Studies ModelE. *Atmospheric Chemistry and Physics*, *10*(13), 5999–6006.

Jones, A. C., Haywood, J. M., & Jones, A. (2016). Climatic impacts of stratospheric geoengineering with sulfate, black carbon and titania injection. *Atmospheric Chemistry and Physics*, *16*(5), 2843–2862.

Kravitz, B., MacMartin, D. G., Mills, M. J., Richter, J. H., Tilmes, S., Lamarque, J. F., et al. (2017). First simulations of designing stratospheric sulfate aerosol geoengineering to meet multiple simultaneous climate objectives. *Journal of Geophysical Research: Atmospheres*, *122*, 12,616–12,634. https://doi.org/10.1002/2017JD026874

Kravitz, B., Robock, A., Boucher, O., Schmidt, H., Taylor, K. E., Stenchikov, G., & Schulz, M. (2011). The Geoengineering Model Intercomparison Project (GeoMIP). *Atmospheric Science Letters*, *12*(2), 162–167.

Laakso, A., Korhonen, H., Romakkaniemi, S., & Kokkola, H. (2017). Radiative and climate effects of stratospheric sulfur geoengineering using seasonally varying injection areas. *Atmospheric Chemistry and Physics*, *17*(11), 6957.

Lo, Y. T. E., Charlton-Perez, A. J., Lott, F. C., & Highwood, E. J. (2016). Detecting sulphate aerosol geoengineering with different methods. *Scientific reports*, *6*, 39169.

MacMartin, D. G., Kravitz, B., Tilmes, S., Richter, J. H., Mills, M. J., Lamarque, J. F., et al. (2017). The climate response to stratospheric aerosol geoengineering can be tailored using multiple injection locations. *Journal of Geophysical Research: Atmospheres*, *122*, 12,574–12,590. https://doi.org/10.1002/2017JD026868

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, *117*, D08101. https://doi.org/10.1029/2011JD017187

Osborn, T. (2018). Tim Osborn: HadCRUT4 global temperature graphs. Climate Research Unit, University of East Anglia https://crudata.uea.ac.uk/˜timo/diag/tempdiag.htm [Online; accessed 27-February-2018].

Parson, E. A. (2017). Opinion: Climate policymakers and assessments must get serious about climate engineering. *Proceedings of the National Academy of Sciences*, *114*(35), 9227–9230.

Phipps, S. J., Rotstayn, L. D., Gordon, H. B., Roberts, J. L., Hirst, A. C., & Budd, W. F. (2011). The CSIRO Mk3L climate system model version 1.0–part 1: Description and evaluation. *Geoscientific Model Development*, *4*(2), 483–509.

Pitari, G., Aquila, V., Kravitz, B., Robock, A., Watanabe, S., Cionni, I., et al. (2014). Stratopsheric ozone response to sulfate geoengineering: Results from the Geoengineering Model Intercomparison Project (GeoMIP). *Journal of Geophysical Research: Atmospheres*, *119*, 2629–2653. https://doi.org/10.1002/2013JD020566

Preston, C. J. (2013). Ethics and geoengineering: Reviewing the moral issues raised by solar radiation management and carbon dioxide removal. *Wiley Interdisciplinary Reviews: Climate Change*, *4*(1), 23–37.

Rogelj, J., Den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., et al. (2016). Paris agreement climate proposals need a boost to keep warming well below 2°C. *Nature*, *534*(7609), 631–639.

Royal Society Working Group (2009). *Geoengineering the climate: Science, governance and uncertainty* (pp. 1–81). London: Royal Society.

Sato, M., Hansen, J. E., McCormick, M. P., & Pollack, J. B. (1993). Stratospheric aerosol optical depths, 1850–1990. *Journal of Geophysical Research*, *98*(D12), 22987–22994.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., et al. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, *6*, 141–184. https://doi.org/10.1002/2013MS000265

Stott, P. A., Gillett, N. P., Hegerl, G. C., Karoly, D. J., Stone, D. A., Zhang, X., & Zwiers, F. (2010). Detection and attribution of climate change: A regional perspective. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(2), 192–211.

Svoboda, T., Keller, K., Goes, M., & Tuana, N. (2011). Sulfate aerosol geoengineering: The question of justice. *Public Affairs Quarterly*, *25*(3), 157–179.

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., et al. (2011). RCP4.5: A pathway for stabilization of radiative forcing by 2100. *Climatic change*, *109*(1-2), 77–94.

UK Met Office (2018). Met Office Hadley Centre observations datasets, HadCRUT.4.6.0.0 (current version). https://www.metoffice.gov.uk/news/releases/2018/2017-temperature-announcement

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takcmura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*(4), 845–872.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28–35.

Yu, X., Moore, J. C., Cui, X., Rinke, A., Ji, D., Kravitz, B., & Yoon, J. (2015). Impacts, effectiveness and regional inequalities of the GeoMIP G1 to G4 solar radiation management scenarios. *Global and Planetary Change*, *129*, 10–22.