Key factors governing uncertainty in the response to sunshade geoengineering from a comparison of the GeoMIP ensemble and a perturbed parameter ensemble

Peter J. Irvine[1], Olivier Boucher[2], Ben Kravitz[3], Kari Alterskjær[4], Jason N. S. Cole[5], Duoying Ji[6], Andy Jones[7], Daniel J. Lunt[8], John C. Moore[6], Helene Muri[4], Ulrike Niemeier[9], Alan Robock[10], Balwinder Singh[3], Simone Tilmes[11], Shingo Watanabe[12], Shuting Yang[13], and Jin-Ho Yoon[3]

[1]Institute for Advanced Sustainability Studies, Potsdam, Germany, [2]Laboratoire de Météorologie Dynamique, IPSL, CNRS/UPMC, Paris, France, [3]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington, USA, [4]Department of Geosciences, University of Oslo, Oslo, Norway, [5]Canadian Centre for Climate Modeling and Analysis, Environment Canada, Toronto, Ontario, Canada, [6]State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China, [7]Met Office Hadley Centre, Exeter, UK, [8]School of Geographical Sciences, University of Bristol, Bristol, UK, [9]Max Planck Institute for Meteorology, Hamburg, Germany, [10]Department of Environmental Sciences, Rutgers University, New Brunswick, New Jersey, USA, [11]National Center for Atmospheric Research, Boulder, Colorado, USA, [12]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan, [13]Danish Meteorological Institute, Copenhagen, Denmark

Abstract Climate model studies of the consequences of solar geoengineering are central to evaluating whether such approaches may help to reduce the harmful impacts of global warming. In this study we compare the sunshade solar geoengineering response of a perturbed parameter ensemble (PPE) of the Hadley Centre Coupled Model version 3 (HadCM3) with a multimodel ensemble (MME) by analyzing the G1 experiment from the Geoengineering Model Intercomparison Project (GeoMIP). The PPE only perturbed a small number of parameters and shares a common structure with the unperturbed HadCM3 model, and so the additional weight the PPE adds to the robustness of the common climate response features in the MME is minor. However, analysis of the PPE indicates some of the factors that drive the spread within the MME. We isolate the role of global mean temperature biases for both ensembles and find that these biases have little effect on the ensemble spread in the hydrological response but do reduce the spread in surface air temperature response, particularly at high latitudes. We investigate the role of the preindustrial climatology and find that biases here are likely a key source of ensemble spread at the zonal and grid cell level. The role of vegetation, and its response to elevated CO_2 concentrations through the CO_2 physiological effect and changes in plant productivity, is also investigated and proves to have a substantial effect on the terrestrial hydrological response to solar geoengineering and to be a major source of variation within the GeoMIP ensemble.

1. Introduction

Solar geoengineering has been proposed as a potential means of reducing the risks of global warming by reducing the absorption of sunlight by the planet and lowering temperatures [*Crutzen*, 2006; *Keith*, 2000]. If such solar geoengineering is indeed feasible, a key concern would be the effects it would have on the climate, i.e., whether it would actually reduce the harmful impacts of global warming. Research suggests that only field tests of solar geoengineering that are a substantial fraction of full-scale deployment could provide useful direct information on the climate outcomes [*MacMynowski et al.*, 2011; *Robock et al.*, 2010]. As the climate is naturally highly variable, even after a decade of observation of large-scale solar geoengineering deployment (e.g., more than a few tenths of a Wm^{-2} forcing globally), one would only be able to loosely bound regional climate changes and precise information on regional changes in climate could take a century or more to gather [*MacMynowski et al.*, 2011]. Thus, model projections are critical to assessing the potential consequences of proposed solar geoengineering schemes, and even if solar geoengineering were to be deployed, they would be central to assessing which observed weather events and decadal trends are attributable to solar geoengineering.

To make such projections of future climate, coupled ocean-atmosphere general circulation models (GCMs) are needed. GCMs have been used to assess the effectiveness and the potential climate consequences of a range of different proposed solar geoengineering techniques, e.g., sunshade geoengineering [*Govindasamy et al.*, 2003; *Lunt et al.*, 2008], sulfate aerosol injection [*Niemeier et al.*, 2011; *Rasch et al.*, 2008], cloud albedo modification [*Jones and Haywood*, 2012; *Latham et al.*, 2012], desert albedo geoengineering [*Irvine et al.*, 2011], and crop albedo modification [*Singarayer et al.*, 2009]. In these previous studies several different approaches to represent the proposed solar geoengineering techniques were adopted, different boundary conditions were chosen, and the data were not routinely made available, hence directly comparing the results of different studies was difficult. The Geoengineering Model Intercomparison Project (GeoMIP) has proposed a set of standard experiments and output to make such comparisons easier [*Kravitz et al.*, 2011], building on the work done for the Climate Model Intercomparison Project 5 (CMIP5) [*Taylor et al.*, 2012]. Many modeling groups have now run these standard GeoMIP experiments, enabling a comparison of their results and allowing the potential climate consequences of solar geoengineering and its uncertainties to be evaluated.

GeoMIP was envisioned as a way to create a multimodel ensemble (MME) of many different GCMs, but the strategy is equally applicable to perturbed parameter ensembles (PPE) of single models. Although the two approaches differ in important ways, both allow identification of common responses and also of disagreement in model response. They can both be used to project likely outcomes and give some measure of the uncertainty of such projections. Furthermore, they highlight the most important and sensitive processes that govern the response, thus guiding future studies and model development. MMEs are often "ensembles of opportunity," with no systematic selection process, and not all models within the ensemble perform equally well. However, it has been observed that the mean or median projections of an MME perform better than any single model when compared to observations for global metrics over a range of variables [*Reichler and Kim*, 2008]. PPEs on the other hand are typically generated by systematically perturbing a number of key physical parameters within a single GCM, and an arbitrary number of ensemble members can be generated, resources permitting [*Murphy et al.*, 2004; *Sanderson et al.*, 2008b; *Stainforth et al.*, 2005]. The systematic structure and larger ensemble size of PPEs has its advantages, but since all ensemble members retain the same model structure, common problems and responses, perhaps restricted to the specific model, can be present across the entire ensemble [*Collins et al.*, 2010; *Yokohata et al.*, 2012]. As has been shown, combining these two approaches to assess the uncertainties of and differences between model projections will lead to new insights [*Yokohata et al.*, 2012].

This study compares the results from the 12 member GeoMIP MME with a 20 member PPE of the Hadley Centre Coupled Model version 3 (HadCM3) GCM, for the sunshade geoengineering experiment G1 of GeoMIP [*Irvine et al.*, 2013; *Kravitz et al.*, 2011]. The G1 experiment consists of an instantaneous quadrupling of preindustrial CO_2 concentrations which is balanced by a uniform reduction of incoming solar radiation at the top of atmosphere, i.e., sunshade geoengineering (see *Kravitz et al.* [2011] for a full description and illustration of the experimental setup). This study aims to analyze the differences within and between the ensembles and to isolate the role of various factors that lead to these differences. The paper consists of the following sections: a methods section, a results section, a discussion, and a conclusion.

2. Methods

2.1. GeoMIP Multimodel Ensemble

In this study we use 12 models from the GeoMIP ensemble that completed the G1 sunshade experiment (*Kravitz et al.* [2013c, Table 1] lists the models that completed these simulations). The GeoMIP ensemble samples structural uncertainty in GCM projections of the response to solar geoengineering, i.e., the models differ in which processes are included and how they are parameterized. Within a given model structure, there are uncertainties over the parameter values to choose and this parametric uncertainty is what is explored in a PPE. The first key difference between the models of the GeoMIP ensemble is their horizontal and vertical resolutions of the atmospheric component, which range from 1° × 1° (latitude × longitude) to 2.5° × 3.75° (the resolution of HadCM3) horizontally and from 19 to 80 vertical levels. The vertical resolution affects, among other things, the fidelity of the representation of orography, boundary layer processes, and upper atmospheric dynamics [*Hardiman et al.*, 2012]. The atmospheric components of the

Table 1. A List of the Eight Parameters Perturbed in the PPE Experiment Conducted With HadCM3[a]

Parameter Name	Standard Value	Minimum Value	Maximum Value	Description
ENTCOEF (ratio)	3.0	0.6	9.0	Lateral entrainment rate coefficient
VF1 (m s^{-1})	1.0	0.5	2.0	Ice fall speed
CT (s^{-1})	1.0×10^{-4}	5.0×10^{-5}	4.0×10^{-4}	Cloud droplet to rain conversion rate
CW land/sea (kg m^{-3})	2.0×10^{-4} 5.0×10^{-5}	1.0×10^{-4} 2.0×10^{-5}	2.0×10^{-3} 5.0×10^{-4}	Cloud droplet to rain conversion threshold over land and sea
EACF (fraction)	0.5	0.47	0.65	Empirically adjusted cloud fraction at saturation
RHCRIT (fraction)	0.7	0.6	0.9	Threshold of relative humidity for cloud formation
ALPHAM (fraction)	0.5	0.47	0.65	Sea ice albedo at 0°C
VDIFF min/max (m^{2}s^{-1})	1.0–15.0	0.5–4.0	2.0–58.0	Background vertical diffusivity, which varies as a function of depth

[a]Shown are the parameter name with units in parentheses, value of the parameter in the standard configuration, the minimum and maximum for the parameter range, and a short description.

models contain many poorly understood and uncertain processes, such as clouds, aerosols, and aerosol-cloud interactions, which are represented differently in the different models. This affects their resultant forcing and feedback responses to greenhouse gases [*Andrews et al.*, 2012]. The ocean models also differ, affecting processes that control atmosphere-ocean phenomena such as ocean heat uptake, the El Niño–Southern Oscillation, and the Meridional Overturning Circulation [*Cheng et al.*, 2013; *Philip et al.*, 2010].

The representations of the land surface in the models differ, and these differences are summarized in Table S1 in the supporting information. All models include a representation of the CO_2 physiological effect (except EC-Earth), where the greater availability of ambient CO_2 causes the stomata on the plants to constrict, reducing stomatal conductance and suppressing transpiration. All models also include the CO_2 fertilization effect where the net primary productivity (NPP) of plants increases in response to the greater availability of CO_2, counteracting to some degree the reduction in transpiration from the CO_2 physiological effect [*Betts et al.*, 1997]. Most models include an interactive vegetation distribution with fixed vegetation distributions in the others (including HadCM3), although the Max Planck Institute-Earth System Model, which does not have dynamic vegetation, simulates changes in leaf area index (LAI), the ratio of leaf surface area to ground surface area. Nitrogen and other nutrient limitations reduce the extent to which increases in NPP in response to elevated CO_2 concentrations can be achieved; this photosynthesis downregulation is only represented in some models, and only the models which use the Community Land Model version 4 (CLM4) include a representation of the nitrogen cycle.

2.2. HadCM3 Perturbed Parameter Ensemble

The HadCM3 model was one of the CMIP3 generation of models and has a relatively low computational cost compared to the CMIP5 generation models [*Covey et al.*, 2003; *Gordon et al.*, 2000; *Solomon et al.*, 2007]. This relatively low computational cost has led to HadCM3 being used for a host of PPE applications, e.g., to investigate uncertainty in climate sensitivity [*Murphy et al.*, 2004; *Rowlands et al.*, 2012; *Sanderson et al.*, 2008b], to attribute weather extremes to climate change [*Stott et al.*, 2004], and to inform UK climate adaptation policy [*Murphy et al.*, 2009]. The PPE of HadCM3, which we use in this study, is described in full by *Irvine et al.* [2013] along with the methodology used to generate it and an assessment of its performance. Here we briefly describe the ensemble design and the parameters perturbed. For details of the ensemble performance, see *Irvine et al.* [2013].

The ensemble used here employs the fully coupled version of the model. We perturbed ocean parameters as well as atmospheric parameters, and we did not employ flux correction. *Irvine et al.* [2013] generated an initial ensemble of 200 members and retained those which were projected to remain within 2°C of the observed preindustrial global mean temperature on the basis of a 20 year preindustrial simulation. Twenty-one ensemble members were confirmed to remain within the target range, and six ensemble members were rejected at this stage, as their temperature change was greater than initially projected. After completing 150 years of a quadrupled CO_2 experiment, one further ensemble member was rejected, as it showed a nonlinear climate response of unchecked warming. The remaining 20 ensemble members form the PPE used in this study. The long spin-up requirement and resource constraints limited the size of the ensemble, which is far smaller than other PPEs of HadCM3 that have included thousands of members [*Stainforth et al.*, 2005].

Table 2. The Global Mean Response of the PPE and the GeoMIP Ensemble for a Number of Summary Variables[a]

	PPE	GeoMIP Ensemble
G1-piControl TOA radiative imbalance (W m^{-2})	0.10 ± 0.05 (0.00, 0.18)	0.01 ± 0.17 (-0.25, 0.24)
G1-piControl SAT (°C)	0.15 ± 0.05 (0.05, 0.24)	0.01 ± 0.25 (-0.30, 0.65)
G1-piControl precipitation (%)	-4.50 ± 0.18 (-4.83, -4.26)	-4.48 ± 1.34 (-6.51, -1.87)

[a]The results are presented as ensemble mean ± ensemble standard deviation, with the minimum and maximum of the ensemble response in parentheses. All variables are global means calculated from years 11 to 50 of the simulations.

The eight parameters perturbed in the PPE are shown in Table 1, which also appears in *Irvine et al.* [2013]. Only eight parameters were perturbed as a compromise between the goal of perturbing as many uncertain processes as possible and the goal of adequately sampling the multidimensional parameter space. If more parameters were perturbed, then a broader range of parametric uncertainty would be probed, but it would be harder to understand the response with the limited number of ensemble members it was possible to run. The parameter perturbations were sampled using a maximin latin hypercube approach, which aims to achieve a high degree of multivariate separation between sampled points in high-dimensional spaces. The six atmospheric parameters are those which were found to most strongly control climate sensitivity in HadCM3 [*Rougier et al.*, 2009; *Sanderson et al.*, 2008a; *Stainforth et al.*, 2005], and they affect convection, cloud formation, and precipitation formation. A single ocean parameter was chosen, the background ocean diffusivity parameter, which has been found to have the greatest effect on the transient climate response and on many other ocean properties in HadCM3 [*Brierley et al.*, 2010; *Collins et al.*, 2007]. The sea ice minimum albedo was also chosen, as it plays an important role in determining the control sea ice conditions and the response of sea ice to changed conditions. The land surface model in all members of the PPE remains as in the standard version of HadCM3 Met Office Surface Exchange Scheme (MOSES) 1 [*Cox et al.*, 1999], i.e., without dynamic vegetation or nutrient limitations for NPP.

2.3. Experimental Design

Three simulations are run for each model and PPE member: a preindustrial control run (piControl), a simulation with an instantaneous quadrupling of preindustrial CO_2 levels (abrupt4xCO_2) to ~1144 ppmv, and a simulation with an instantaneous quadrupling of CO_2 levels and a reduction in insolation (G1). The insolation reduction is chosen to achieve a top of atmosphere (TOA) radiative imbalance that is close to that of the preindustrial. This was achieved by testing whether the mean TOA radiative imbalance was within ± 0.1 Wm^{-2} of the piControl value over an initial 10 years of simulation, and if not, modifying the solar constant appropriately and repeating. This balanced insolation reduction was then used for a 50 year simulation of G1, which will be compared with the first 50 years of parallel simulations of abrupt4xCO2 and piControl. Five hundred years of preindustrial control were completed before these other experiments were started; as such, there is a small degree of drift in the climate of the piControl simulations for the members of the PPE, as the deep ocean is not fully in steady state [*Irvine et al.*, 2013]. All averages reported in this study are over years 11–50 unless otherwise stated. Although some of the GeoMIP ensemble members had multiple initial condition runs for each experiment, we use only one ensemble member from each model in this study to simplify the analysis. For more details on these experiments, *Kravitz et al.* [2013c] give an overview of the climate response in these simulations, *Tilmes et al.* [2013] analyze the precipitation response, and *Kravitz et al.* [2013b] analyze the hydrological response from an energetic perspective.

3. Results

3.1. The Role of Global Mean Temperature Biases in Ensemble Response

The GeoMIP ensemble average for the G1 experiment is in near-perfect radiative equilibrium in the global mean, with all individual ensemble members remaining within a few tenths of a Wm^{-2} of radiative balance. The same is true of members of the PPE, but the PPE mean is out of balance by +0.1 Wm^{-2}. See Table 2 for a summary of the ensemble response. Overall, the procedures applied by the modeling groups have been fairly successful in balancing the radiative forcing of abrupt4xCO_2 with an insolation reduction. As might be expected, the global mean temperature response of the ensembles reflects these TOA radiative imbalances, with the GeoMIP ensemble mean showing near-zero change and the PPE mean warming by

Table 3. Results Summarizing the Distribution of SAT Changes for the PPE and the GeoMIP Ensembles[a]

	HadCM3 PPE	GeoMIP Ensemble
G1-piControl land-ocean SAT (°C)	0.67 ± 0.08 (0.53, 0.79)	0.35 ± 0.18 (−0.02, 0.65)
G1-adjusted-piControl land-ocean SAT (°C)	0.58 ± 0.08 (0.45, 0.68)	0.35 ± 0.18 (0.00, 0.55)
G1-piControl 60–90°N SAT (°C)	1.32 ± 0.28 (0.80, 1.76)	0.95 ± 0.63 (0.06, 2.65)
G1-adjusted-piControl 60–90°N SAT (°C)	1.05 ± 0.24 (0.64, 1.47)	0.92 ± 0.28 (0.55, 1.41)
G1-piControl NH-SH SAT (°C)	0.49 ± 0.10 (0.27, 0.62)	0.17 ± 0.10 (−0.01, 0.36)
G1-adjusted-piControl NH-SH SAT (°C)	0.43 ± 0.10 (0.24, 0.58)	0.17 ± 0.09 (0.06, 0.34)

[a]The results are presented as ensemble mean ± ensemble standard deviation, with the minimum and maximum of the ensemble response in parentheses. All variables are means calculated from years 11 to 50 of the simulations.

0.15°C. However, individual members of the GeoMIP ensemble show some deviations, most notably Beijing Normal University Earth System Model (BNU-ESM) which shows a 0.65°C warming, whereas the PPE members are all within 0.1°C of the ensemble mean response.

These global mean surface air temperature (GMST) deviations seen in both ensembles are one source of systematic model spread that is artificial, i.e., this source of model disagreement is potentially avoidable. The climate system response to a radiative forcing perturbation can be conceptually divided into a rapid adjustment and a feedback response or a fast and slow response [*Andrews et al.*, 2010; *Bala et al.*, 2010]. The rapid adjustment is not associated with GMST and has been found to depend strongly on the nature of the forcing mechanism [*Andrews et al.*, 2010]. The feedback response, on the other hand, depends on the degree of GMST change and is largely independent of the forcing mechanism [*Andrews et al.*, 2010]. To investigate the role that GMST deviations in the G1 experiment have on the ensemble spread, we estimate the climate response to G1 if each model had perfectly maintained the preindustrial GMST, i.e., we remove the temperature-driven feedback response and investigate the rapid adjustment response only. Previous solar geoengineering experiments found a linear climate response to changes in insolation for a given CO_2 concentration [*Irvine et al.*, 2010; *Ricke et al.*, 2010]. We thus assume that the climate response of all variables to an insolation reduction varies linearly with the magnitude of this insolation reduction, i.e., starting from the abrupt4xCO2 climate and progressing along the vector G1-abrupt4xCO2 as the insolation reduction is increased. *Kravitz et al.* [2013b] investigated the rapid adjustment and feedback response of the G1 experiment in another way by separately analyzing the first year of simulation to represent the rapid adjustment and the last 40 years to represent a proxy for the feedback response.

We calculate an adjustment factor C that depends on the GMST values for the different experiments as follows:

$$C = \frac{(\text{abrupt4xCO2}-\text{piControl})}{(\text{abrupt4xCO2}-\text{G1})}$$

Starting from the GMST of the abrupt4xCO2 experiment and subtracting the anomaly between the G1 and abrupt4xCO2 experiments, multiplied by our adjustment factor C, will return the preindustrial global mean temperature. The mean adjustment factor for the GeoMIP ensemble is 1.0 (i.e., no change) with a standard deviation of 0.06. BNU-ESM at 1.14 and Goddard Institute for Space Studies (GISS)-E2-R at 0.90 are the two extreme cases in the GeoMIP ensemble, which were too warm and too cold, respectively. The PPE has a mean adjustment factor of 1.03 with a standard deviation of 0.01. We calculate the adjustment factor separately for every ensemble member and apply this linear scaling to generate an adjusted climate response for G1 (from now on G1-adjusted) and apply this to every variable. This G1-adjusted experiment will be used for the remainder of this study unless otherwise stated.

The pattern of surface air temperature (SAT) changes arising from a certain radiative forcing perturbation depend on the spatial pattern and other properties of the radiative forcing, but they also depend on the magnitude of the resulting GMST response, as positive feedbacks specific to certain locations can amplify the local response [*Armour et al.*, 2013]. Table 3 summarizes some changes to the distribution of SAT changes for the two ensembles for the G1 and G1-adjusted experiments covering Arctic warming, which affects many aspects of the climate of the Northern Hemisphere [*Solomon et al.*, 2007], the interhemispheric temperature difference which controls the location of the Intertropical Convergence Zone (ITCZ) [*Haywood*

Table 4. Precipitation Responses for the PPE and GeoMIP Ensembles for the G1 and G1-Adjusted Simulations[a]

	HadCM3 PPE	GeoMIP Ensemble
G1-piControl precipitation (%)	−4.50 ± 0.18 (−4.83, −4.26)	−4.48 ± 1.34 (−6.51, −1.87)
G1-adjusted-piControl precipitation (%)	−4.79 ± 0.17 (−5.00, −4.43)	−4.46 ± 1.13 (−6.47, −2.59)
G1-piControl ocean precipitation (%)	−5.00 ± 0.31 (−5.60, −4.39)	−4.25 ± 1.06 (−5.66, −2.30)
G1-adjusted-piControl ocean precipitation (%)	−5.30 ± 0.35 (−5.30, −4.71)	−4.34 ± 0.90 (−5.42, −2.39)
G1-piControl land precipitation (%)	−2.55 ± 1.34 (−5.50, −0.40)	−5.05 ± 3.63 (−12.22, −0.53)
G1-adjusted-piControl land precipitation (%)	−2.77 ± 1.36 (−5.87, −0.88)	−5.01 ± 3.47 (−12.53, −1.87)

[a]The results are presented as ensemble mean ± ensemble standard deviation, with the minimum and maximum of the ensemble response in parentheses. All variables are means calculated from years 11 to 50 of the simulations.

et al., 2013; Zeng, 2003], and the land-sea temperature difference, which affects the distribution of precipitation between the oceans and the continents [Bala et al., 2008]. Almost all models show greater warming at high northern latitudes than elsewhere, due to the latitudinal differences between the solar and CO_2 forcing [Govindasamy et al., 2003; Kravitz et al., 2013a; Lunt et al., 2008]. In G1-adjusted both ensembles show almost the same Arctic mean warming and spread; the linear GMST scaling has approximately halved the GeoMIP ensemble standard deviation. This indicates that GMST plays a particularly strong role in the Arctic response to geoengineering, likely due to the action of positive feedbacks such as the melting of ice and snow [Moore et al., 2014]. All models show either no change or a relative warming of the Northern Hemisphere for G1-piControl, likely due to a combination of the warming of the northern landmasses and the action of stronger positive feedbacks at high northern latitudes than at high southern latitudes. The same is true for the land-sea temperature difference. For both of these measures, the GMST adjustment does not greatly change the response. For both of these measures, for G1 and G1-adjusted, almost all PPE members show a greater change than any of the GeoMIP ensemble members, i.e., a greater warming of the land relative to the ocean and of the Northern Hemisphere relative to the Southern Hemisphere.

The strength of the hydrological cycle also depends on the GMST, with simulations and observations showing that the global mean precipitation increases and patterns of precipitation minus evaporation (P − E) intensify as the planet warms [Andrews et al., 2010; Chou et al., 2009]. Thus, the GMST deviations from the piControl will alter the patterns of hydrological change and could lead to greater model disagreement than would be the case had the models perfectly restored the piControl TOA radiative balance. Table 4 shows the global mean, land mean, and ocean mean precipitation anomalies from the piControl for both ensembles for the G1 and G1-adjusted experiments. As most models show some degree of deviation from the piControl GMST for G1, the G1-adjusted response decreases the standard deviation of the global mean precipitation response of the GeoMIP ensemble by 16% but does not greatly affect the already low standard deviation of the PPE. A similar response is seen for the ocean mean and land mean where there is little change in the mean precipitation but a reduction of PPE ensemble standard deviation. GMST deviations thus play a substantial role in the intraensemble spread for the GeoMIP ensemble hydrological response but do not explain all of the differences within the GeoMIP ensemble. The hydrological changes seen in the G1 experiment are described in detail by Tilmes et al. [2013].

Figure 1 shows the zonal mean standard deviation and ensemble range for G1 and G1-adjusted for SAT, evaporation and precipitation, for both ensembles. The GeoMIP ensemble has a larger range and standard deviation for all the variables displayed in these figures than the PPE ensemble. For SAT, G1-adjusted has a substantially reduced ensemble range and standard deviation of SAT compared to G1 for the GeoMIP ensemble. The differences between the range and standard deviation of G1 and G1-adjusted are smaller for the PPE, which had much smaller differences between the GMSTs of its members. The adjustment has a much smaller effect on evaporation and precipitation, reducing the spread in the GeoMIP ensemble somewhat, but it has little effect on the PPE. Figure 2 shows a similar slight reduction in the zonal mean standard deviation and range of terrestrial evaporation and precipitation for the GeoMIP ensemble after adjustment. Although, while the adjustment decreases the range and standard deviation of the GeoMIP land mean SAT response at most latitudes, it *increases* these around the equator. Thus, the GMST deviations between G1 and piControl were actually suppressing the differences between the GeoMIP SAT responses over tropical land regions, indicating that other factors than global mean SAT error are driving the terrestrial response.
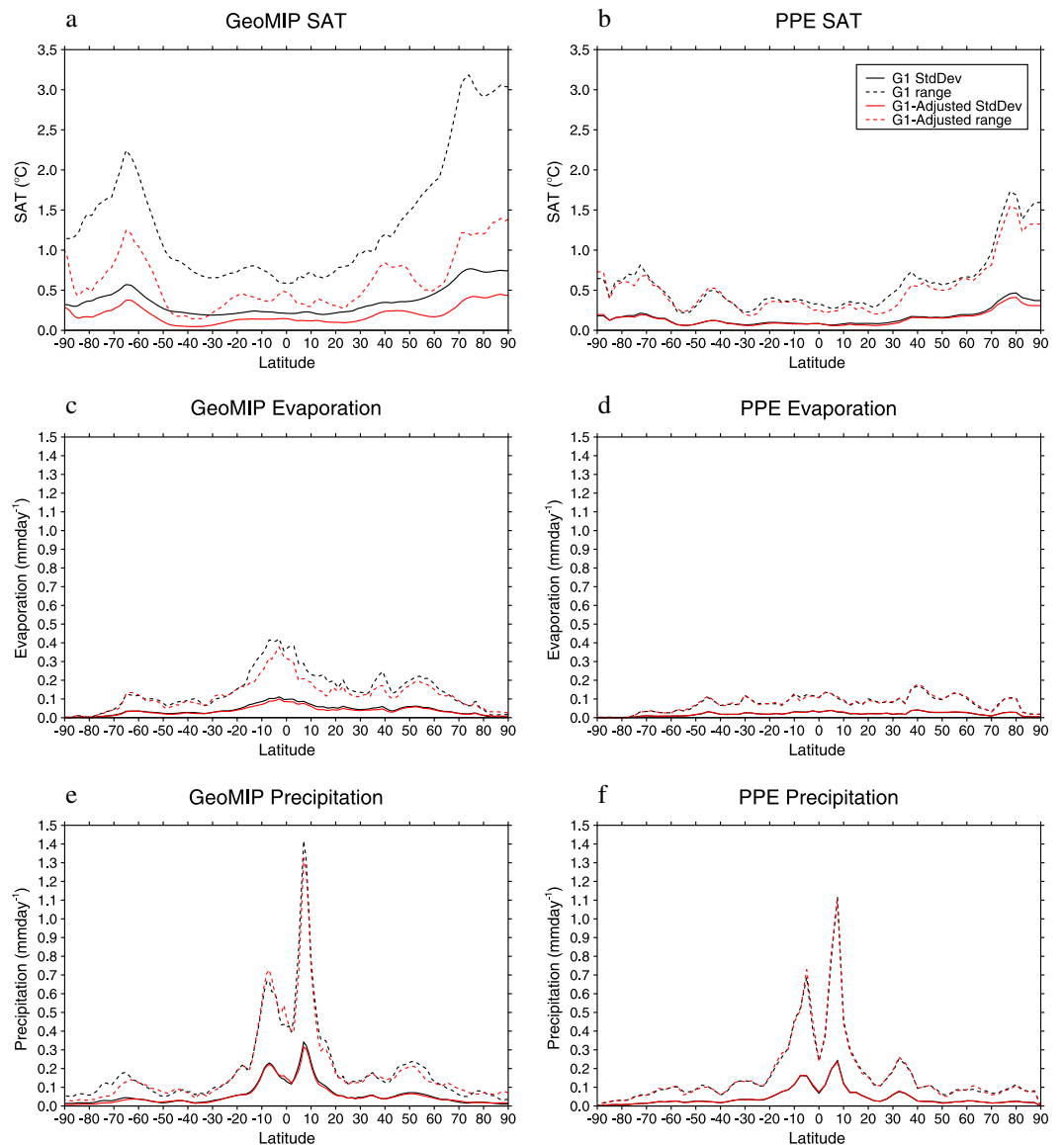
Figure 1. Standard deviation and range of the G1 and G1-adjusted zonal mean response for (a and b) SAT, (c and d) total evaporation, and (e and f) precipitation for the GeoMIP ensemble and PPE.

3.2. Changes in Precipitation and Evaporation Over Land and Ocean

We compare the zonal mean land and ocean hydrological response to G1-adjusted minus piControl of the two ensembles in Figures 3 and 4. Over the oceans, the balance of CO_2 and solar forcing leads to a reduction in evaporation peaking at a value of around 5% across the tropical oceans, where the insolation reduction is greatest, but with a large range of responses at high latitudes where sea ice changes may be playing an important role. The oceanic precipitation response generally shows a reduction across the tropics but one which is highly heterogeneous, with a lesser reduction in the subtropics and a consistent reduction at midlatitudes. There are very large differences in the pattern of tropical oceanic precipitation change within the GeoMIP ensemble, whereas all members of the PPE show a similar, strong decrease south of the equator. This distinctive reduction in precipitation south of the equator may be tied to the relatively strong warming of the Northern Hemisphere in all members of the PPE, as hemispheric temperature differences are simulated to lead to a shift of the ITCZ toward the warmer hemisphere [*Haywood et al.*, 2013; *Zeng*, 2003].
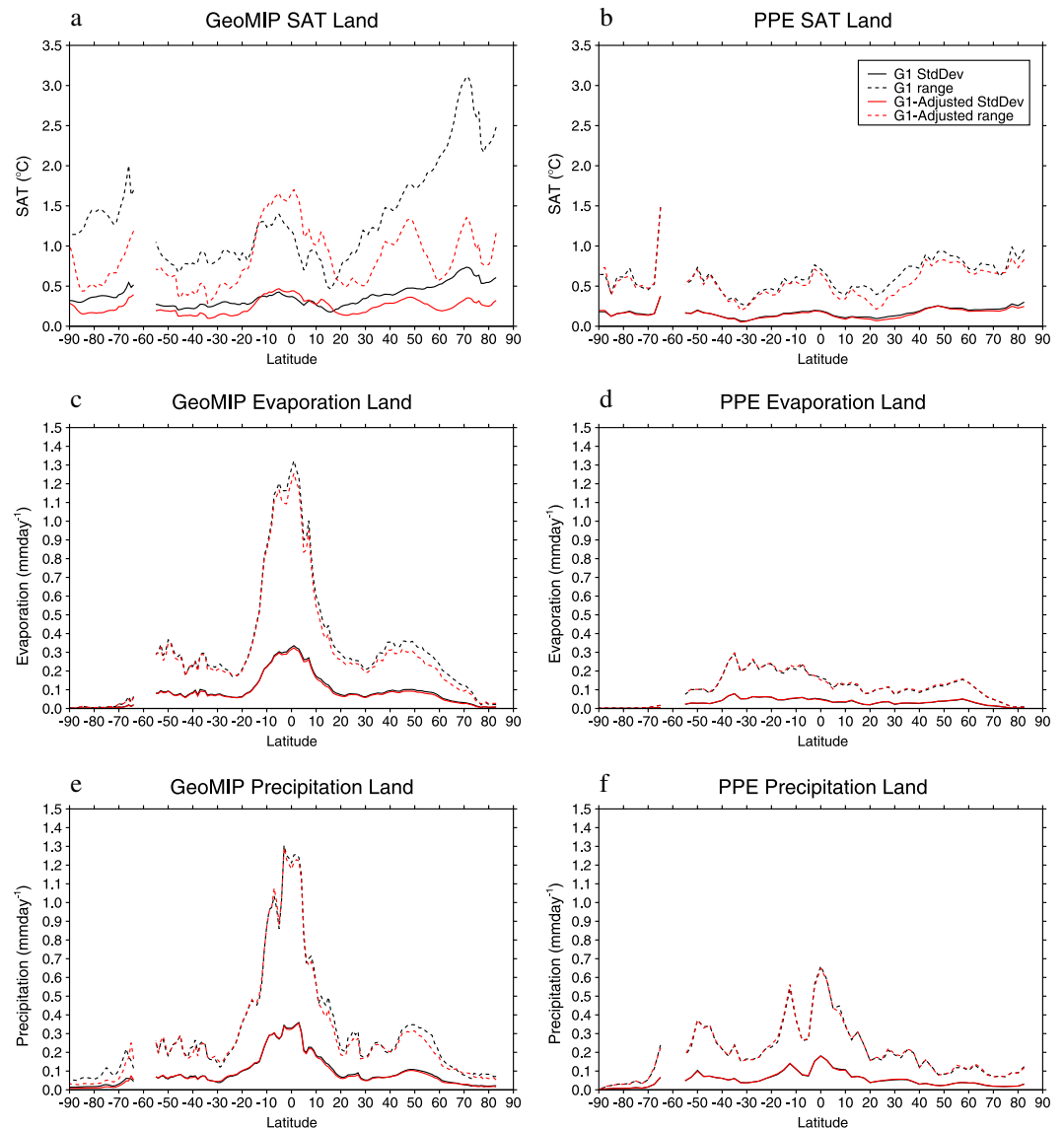
Figure 2. Standard deviation and range of the G1 and G1-adjusted zonal, land mean response for (a and b) SAT, (c and d) total evaporation, and (e and f) precipitation for the GeoMIP ensemble and PPE.

The zonal mean land hydrology response shows larger interensemble and intraensemble differences than the ocean response (see Figures 3 and 4). All members of the PPE show a consistent ~15% reduction in continental evaporation around the equator, whereas the GeoMIP ensemble shows equatorial responses ranging from a slight increase in evaporation for BNU-ESM to a ~25% reduction for GISS-E2-R, with the other models spanning this entire range. The zonal mean land precipitation response of different ensemble members differs much more than the ocean response in both ensembles. However, both ensembles show a general reduction across the tropics, a weaker reduction in the subtropics, and a clear reduction at northern midlatitudes. The magnitude of the tropical and midlatitude reduction in land precipitation correlates with the land evaporation reduction. This is clearest at the northern midlatitudes where the PPE shows a broadly consistent ~ 8% reduction in precipitation, and it can be seen that the GeoMIP ensemble members with the greatest evaporation reductions also show the greatest precipitation reductions [*Kravitz et al.*, 2013b].

3.3. Role of piControl Climatology in the Ensemble Response

All climate models show some degree of bias when compared to observational data, and these are expected to be persistent across different simulations [*Solomon et al.*, 2007]. Models have biases in the particular
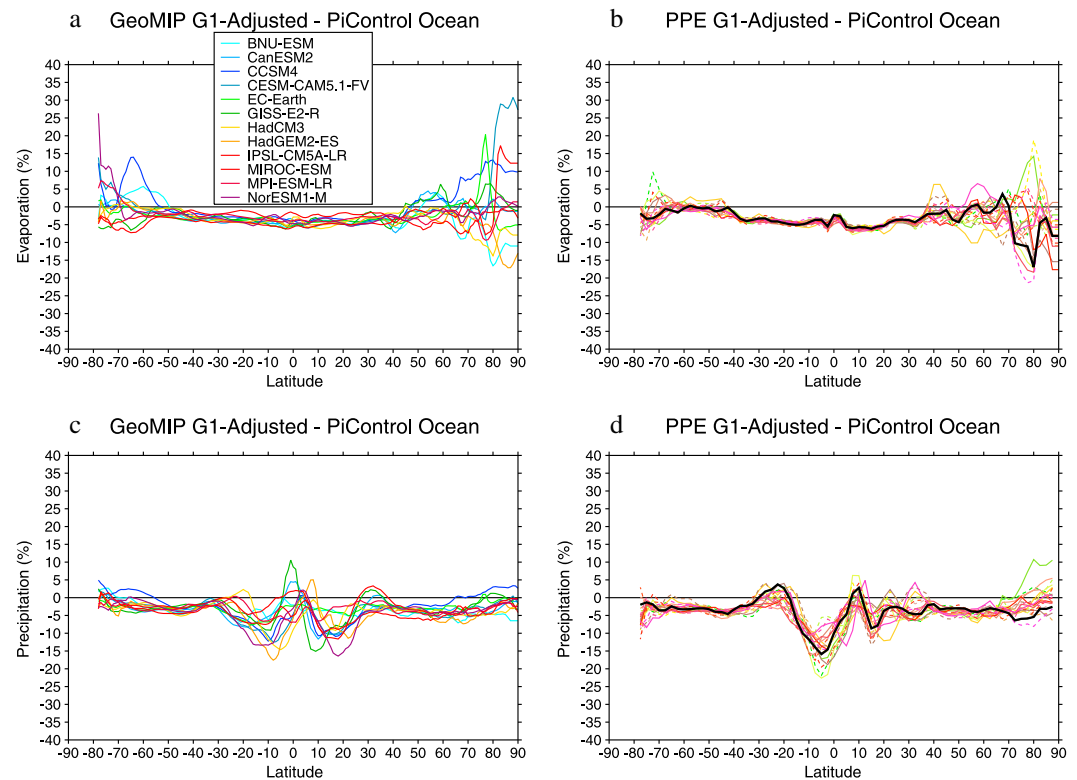
Figure 3. G1-adjusted-piControl zonal, ocean mean of evaporation and precipitation change for the individual members of the (a and c) GeoMIP MME and (b and d) PPE, respectively. The solid black line shows the standard HadCM3 model. Means are calculated from years 11 to 50 of the simulations.

locations of certain key-observed features of the climate, such as the location of the ITCZ in a given season and the locations of greatest surface ocean temperature variability in the region affected by El Niño–Southern Oscillation [*Power et al.*, 2013]. When model projections are compared, as in the zonal mean plots shown in Figures 3 and 4, these biases in the locations of key features may be responsible for some of the disagreement between the model projections, e.g., if the models showed a consistent northward shift of a certain feature, there could be an apparently large degree of disagreement in a zonal mean anomaly plot. To investigate the role of the preindustrial climatology on the response to G1, we sorted the models' grid cells from lowest to highest according to their piControl values for the variable of interest and then calculated the average anomalies for each percentile band of this piControl distribution. In the next section the role of vegetation is examined, and to test the role of vegetation in mediating the climate response to G1, the data are sorted as above but by the preindustrial net primary productivity (NPP). However, land regions with low levels of NPP, below 1% of the grid cell with the highest value, are excluded and are grouped together as a "low NPP" region. As not all GeoMIP models and none of the PPE members produced results for NPP, the ensemble mean of the preindustrial NPP from the GeoMIP ensemble is used as the basis for these calculations. The distribution of NPP should be similar enough across the ensemble to ensure that the regions of high and low productivity occur in similar regions. Figures S1 and S2 show maps of the preindustrial climatology, and Figure S3 shows averages for each of the percentile bands of the preindustrial climatology for SAT, precipitation, evaporation, and NPP.

Figure 5 shows anomalies between G1-adjusted and the preindustrial control globally and over land on the basis of the piControl distribution. The changes in SAT show a cooling of the regions which were warmest in the preindustrial and a warming of the coolest regions, which is due to the overcooling of the tropics and undercooling of the poles, which has been noted for sunshade geoengineering previously [*Govindasamy et al.*, 2003; *Kravitz et al.*, 2013a; *Lunt et al.*, 2008]. There is some disagreement over the magnitude of this trend across the GeoMIP ensemble, although the absolute magnitude of these changes is small relative to those for abrupt4xCO$_2$ (Figure S4), and there is good agreement across the PPE.
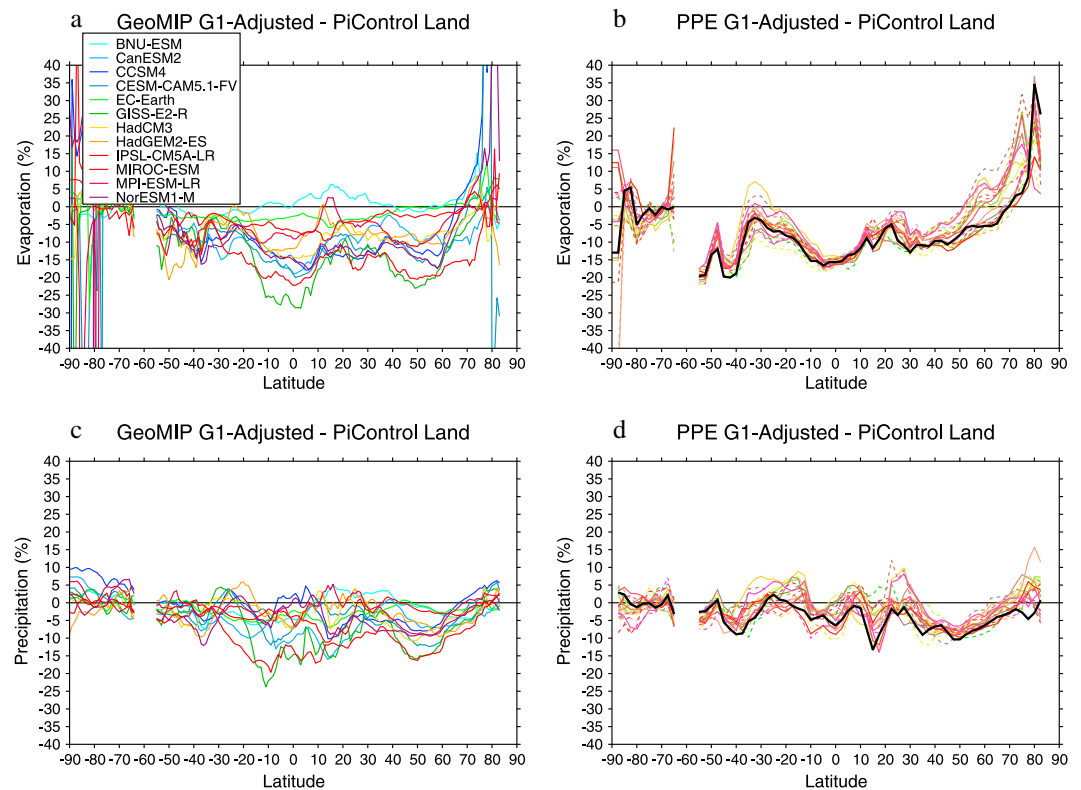
Figure 4. G1-adjusted-piControl zonal, land mean of evaporation and precipitation change for the (a and c) GeoMIP MME and (b and d) for the PPE, respectively. The solid black line shows the standard HadCM3 model. Means are calculated from years 11 to 50 of the simulations.

G1-adjusted shows a global mean reduction in the hydrological cycle with precipitation and evaporation reduced by around 4.5% on average in both ensembles, but the GeoMIP ensemble has an ensemble standard deviation more than 5 times greater than the PPE. This same pattern is reproduced globally in regions with greater than average preindustrial evaporation, i.e., in the upper 50% of the distribution, where a reduction of roughly 5% is shown for both ensembles with a much greater spread seen in the GeoMIP ensemble (Figure 5). For the lower 50%, there is either relatively less evaporation or an increase in evaporation. However, even though the relative changes are large for the lowest evaporation regions, the absolute changes are small due to the substantially lower evaporation in these regions in the preindustrial (Figure S3). This increase in evaporation in low-evaporation regions is likely due to the warming of high latitudes. Similar results are found for changes in precipitation, although the reduction in precipitation is more consistent between regions of high and low preindustrial precipitation. The changes in precipitation over land are similar to the global results but with a greater intraensemble spread, which is to be expected given the differences in the representation of land and boundary layer processes across the GeoMIP ensemble, the smaller number of grid cells, and the greater heterogeneity of the preindustrial climatology over land. However, the evaporation changes over land differ from the global results, showing a reduction in evaporation that is greater for high values of preindustrial evaporation, rather than flat for most values as in the global case. This suggests that an additional process is acting to suppress evaporation over land that does not act over the ocean, such as the CO_2 physiological effect [*Fyfe et al.*, 2013; *Tilmes et al.*, 2013].

3.4. Role of Vegetation Distribution in the Ensemble Response

In the G1 experiment the global mean radiative balance is maintained despite the elevated CO_2 concentrations, but the effect of CO_2 on plants remains. The direct effect of the response of plants to elevated CO_2 has been shown to have a considerable influence on regional climate and hydrology in model simulations through the suppression of transpiration due to reduced stomatal conductance [*Betts et al.*, 2007; *Boucher et al.*, 2009;
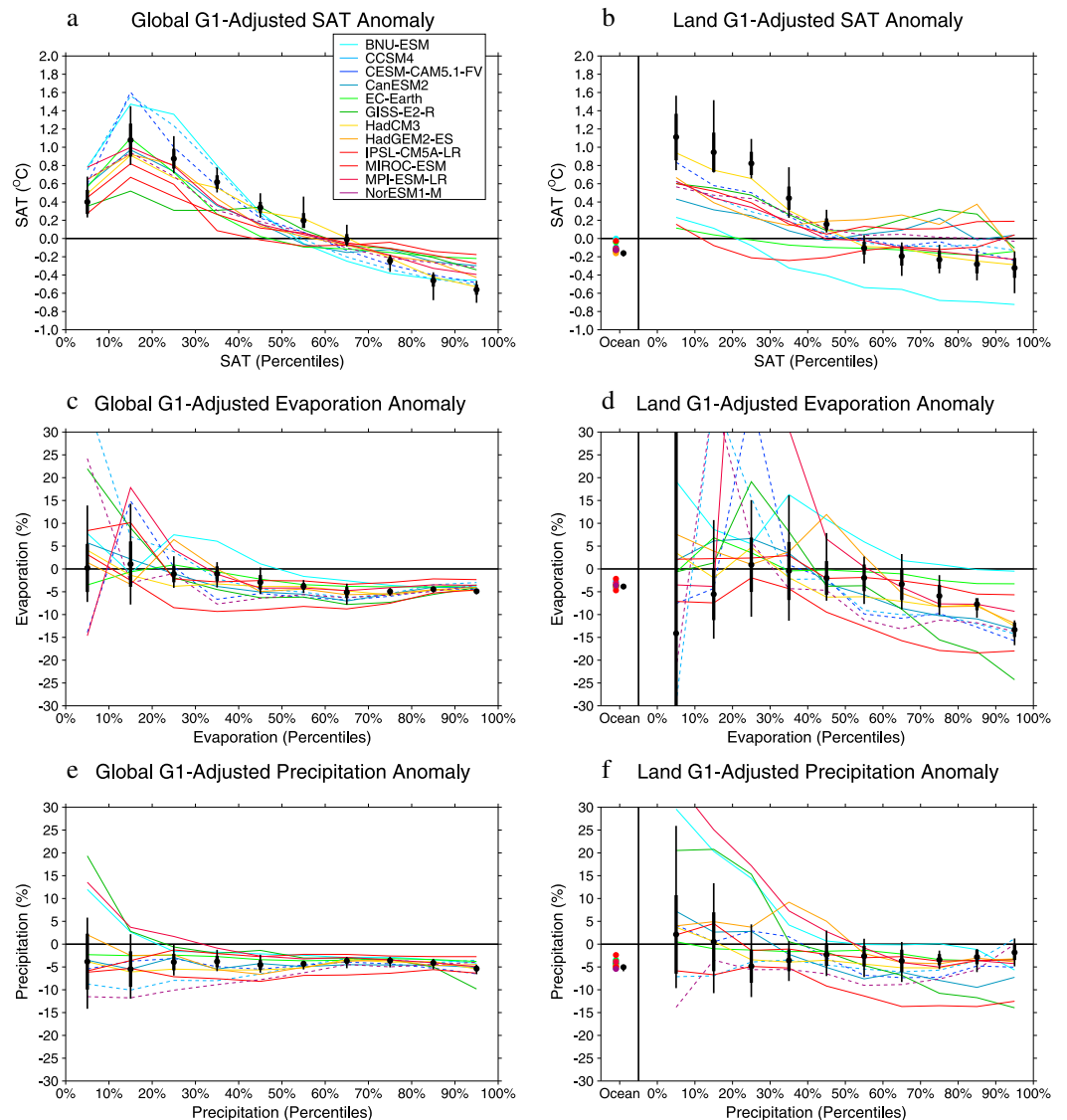
Figure 5. G1-adjusted-piControl anomaly plotted on the basis of the percentiles of the piControl distribution of each of the respective variables, calculated separately for each ensemble member. Plotted are the results for the global and land-only distributions of (a and b) SAT, (c and d) evaporation, and (e and f) precipitation. The PPE results are shown in black with the ensemble mean plotted as a large point, the standard deviation as a thick vertical line, and the range as a thin vertical line. The three GCMs which employ the Community Land Model (CLM) are shown with dashed lines. For the land distribution plots, the ocean mean results are also shown.

Fyfe et al., 2013] and in previous studies was found to make up a considerable fraction of the total climate response when comparing G1-like experiments to piControl experiments [*Fyfe et al.*, 2013; *Tilmes et al.*, 2013].

In the first instance, elevated CO_2 concentrations lead to a reduction in stomatal conductance as the plant acts to reduce water loss, and hence transpiration, while still gathering sufficient CO_2 to photosynthesize; this results in an increase of the plant's water-use efficiency, the water loss per unit of net primary productivity (NPP) [*Farquhar et al.*, 1989; *Franks et al.*, 2013]. The greater availability of CO_2 for photosynthesis and the improved water-use efficiency of photosynthesis lead to an increase in net primary productivity in appropriate conditions [*Farquhar*, 1997]. This increased NPP is realized as a combination of increased leaf-level carbon accumulation and an increased number of leaves, i.e., a greater leaf area index (LAI) measured as the fraction of leaf area to surface area [*Donohue et al.*, 2013; *Franks et al.*, 2013]. This increase in NPP and LAI leads to an increase in transpiration that can partly or wholly offset those initial reductions from the

direct CO_2 physiological effect depending on conditions [*Betts et al.*, 1997; *Donohue et al.*, 2013]. The net effect of the direct CO2 physiological effect and changes in net primary productivity on stomatal conductance can be represented qualitatively as

$$g_{rel} \approx NPP_{rel}/C_{rel}$$

Where g is the stomatal conductance, C is the carbon dioxide concentration, and the subscript "rel" indicates that these values are relative to some reference case [*Franks et al.*, 2013]. This means that if CO_2 levels quadrupled and NPP doubled, one would expect an approximate halving of stomatal conductance which would produce a similar effect on transpiration absent any other changes [*Franks et al.*, 2013]. Arid regions are most likely to see a large increase in NPP as the increased water-use efficiency reduces the key limitation in those regions [*Donohue et al.*, 2013], while densely vegetated regions will get a diminishing return from increased investment in LAI and other limitations may be more important. This implies that outside of arid regions, the vegetation response across the GeoMIP ensemble is likely to differ a lot depending on how great the NPP response is, which will depend on the model's formulation. A critical difference is whether or not nitrogen limitations are considered; *Jones et al.* [2013] noted that the GeoMIP models that include a nitrogen cycle showed a far smaller increase in NPP in response to elevated CO_2 concentrations than those without a nitrogen limitation (see also *Kravitz et al.* [2013a, Figure 1] for more on the wide range of NPP responses across the GeoMIP ensemble).

The effects of the vegetation response to CO_2 on climate cannot be isolated without additional simulations, but the dependence of the climate response on vegetation activity can be investigated. Regions with a higher NPP will have a greater rate of transpiration and may be expected to show the greatest response to elevated CO_2 concentrations. Figure 6 shows the absolute response of SAT, evaporation and precipitation, and the percentage change of evaporation and precipitation, for G1-adjusted as a function of the terrestrial NPP. CO_2 affects transpiration but this variable was not available for all models, so changes in evaporation will be analyzed as these include transpiration changes. The evaporation anomaly for G1-adjusted minus piControl shows a reduction for most models across all vegetated regions, with the greatest reductions occurring in high NPP regions, both in absolute and percentage change terms. The GeoMIP ensemble spread is greatest for the highest percentiles of the NPP distribution despite the much lower initial evaporation for the lower percentiles, which should act to exaggerate any percentage differences between the models as was seen for the global and land mean changes in Figure 5. This wide intraensemble spread of the GeoMIP ensemble for the high NPP regions can be contrasted with the far narrower intraensemble spread of the PPE for the highest NPP regions. It seems likely that these differences in response are connected to differences in the CO_2 physiological response for the following reasons: all members of the PPE have the same CO_2 physiological response, the three GCMs which employ the Community Land Model (CLM) all show a very similar response (Community Earth System Model with the Community Atmosphere Model version 5.1 (CESM-CAM5.1-FV), Community Climate System Model version 4 (CCSM4) and the Norwegian Earth System Model (NorESM1-M) all shown with dashed lines), and EC-Earth which has no CO_2 physiological effect shows a flat response across the NPP distribution. Significantly, the Hadley Centre Global Environmental Model version 2 (HadGEM2) also shows a similar response to the PPE and they share the same land model MOSES, albeit different versions of it. However, the GeoMIP ensemble has a wider range of boundary layer process representations, and members differ in a number of other important ways; additionally, HadCM3 and HadGEM2 as well as CESM-CAM5 and CCSM4 are closely related models that were developed in the same research institutes which may explain some of their similarity [*Knutti et al.*, 2013].

For the GeoMIP ensemble, the SAT response to G1-adjusted is a warming of a few tenths of a degree Celsius for both the low NPP regions and across the rest of the distribution but with a wider GeoMIP ensemble spread for the highest NPP regions. That there is a mix of high- and low-latitude regions within the NPP percentile bands likely explains the flatness of this temperature response (see Figure S1 for a map of the NPP percentile bands). The precipitation and evaporation anomalies are similar, with a greater reduction and wider GeoMIP ensemble spread for high NPP regions indicating that the CO_2 physiological effect may be playing some role in controlling the terrestrial precipitation response, but this is less clearly the case than for evaporation. This more muted response is likely because precipitation is less determined by local factors than evaporation, with a greater role in determining the precipitation response played by changes to low-level convergence, remote changes in oceanic source evaporation, and changes in atmospheric stability.
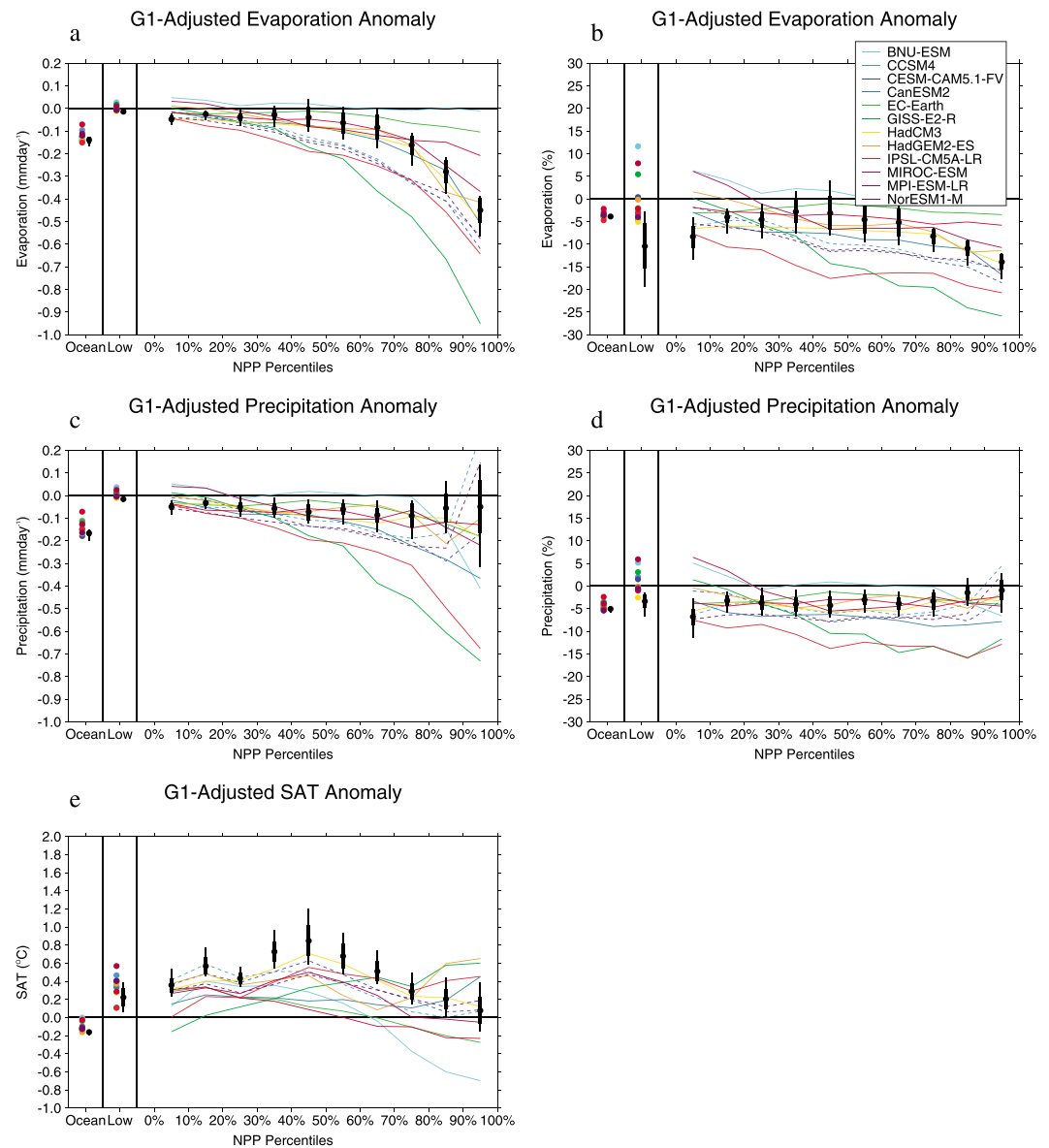
Figure 6. G1-adjusted-piControl anomaly plotted on the basis of the percentiles of the piControl GeoMIP ensemble mean distribution of NPP, excluding land regions with a value of NPP less than 1% of the maximum recorded NPP. Plotted are the absolute and relative changes in (a and b) evaporation and (c and d) precipitation, along with the absolute changes in (e) SAT. The PPE results are shown in black with the ensemble mean plotted as a large point, the standard deviation as a thick vertical line, and the range as a thin vertical line. The three GCMs which employ the Community Land Model (CLM) are shown with dashed lines. The ocean mean and the mean of the regions of low NPP excluded from the calculation of the distribution are also shown.

4. Discussion

The PPE of HadCM3 used in this study reproduces many of the features of the climate response to solar geoengineering seen in other studies of the G1 experiment [*Kravitz et al.*, 2013a; *Schmidt et al.*, 2012; *Tilmes et al.*, 2013]. However, the PPE produces a narrower range of responses to G1 than the GeoMIP ensemble does, with most members showing a similar pattern of change but differing in the magnitude of change. The PPE used in this study only perturbed a limited number of parameters, primarily those that are important for determining the climate sensitivity of HadCM3, resulting in a range of climate sensitivities of between 3.3 and 5.3°C [*Irvine et al.*, 2013]. This choice was made on the assumption that those processes most important for determining the magnitude of the climate change response would also be important for determining the response to solar geoengineering. However, as *Kravitz et al.* [2013b] and *Schmidt et al.* [2012]

note, the agreement on the climate response to G1 is much greater in the GeoMIP ensemble than for the climate response to abrupt4xCO$_2$ due to the much smaller change in SAT and the mostly unchanged GMST which means that the large differences between the models' temperature feedback response is not realized in G1 but plays a large role in abrupt4xCO$_2$. PPEs are limited by their common structure and share many biases with the standard model on which they are based [*Collins et al.*, 2010], and this ensemble clearly shows underdispersion in the range of responses to the G1 experiment [*Irvine et al.*, 2013].

Comparing the GeoMIP and PPE ensembles, it was possible to infer some of the uncertain processes that are responsible for differences within and between the ensembles. First, we note that some of the GeoMIP and PPE members did not perfectly balance the CO$_2$ and solar forcings in G1, so there was some degree of temperature change compared to the piControl. We found that by correcting for this GMST deviation, the spread of regional SAT responses, particularly at high latitudes, was reduced. However, this adjustment did not substantially alter the ensemble spread of the hydrological responses. Second, we assessed the role of the piControl climatology in determining the response to G1. At the global level of analysis, we found that there was a fractional reduction in evaporation and precipitation from piControl that was broadly consistent between regions of high and low precipitation and evaporation. An exception to this is found at high latitudes and hence in regions of low evaporation and precipitation, where the imbalance between solar and CO$_2$ forcing led to a rise in temperature and a consequent rise in evaporation and precipitation. This consistency in the reduction in the intensity of precipitation and evaporation suggests that some of the differences in the modeled precipitation response seen in other studies of the GeoMIP ensemble may be arising due to differences in the piControl distribution rather than through differences in the response of the models to the forcing. Third, we assessed the role of terrestrial vegetation on the response to G1. The greatest fractional reductions in evaporation occurred in highly vegetated regions where transpiration is responsible for a large fraction of the total evaporation. There was a large degree of intraensemble spread in this response as the degree to which increases in NPP offset the initial CO$_2$ physiological effect differs greatly between the land models used in the GeoMIP ensemble. This finding was reinforced by the similarity of this response for the three models which shared the CLM4 land surface model and included a nitrogen cycle (NorESM1-M, CCSM4, and CESM-CAM5.1-FV) and separately for the PPE of HadCM3 and HadGEM2-ES which share different versions of the MOSES land model. It is worth noting that all climate models have structural and other similarities, and this is particularly true for models that are produced by the same research center, as HadCM3 and HadGEM2-ES and CCSM4 and CESM-CAM5.1-FV are [*Knutti et al.*, 2013]; this means that other structural similarities, such as the treatment of the boundary layer, may give rise to their similar behavior. As terrestrial precipitation is more strongly affected by remote processes than terrestrial evaporation [*Gimeno et al.*, 2010], we did not find as clear an indication for the role of vegetation processes for this variable.

This study has highlighted a number of uncertainties by comparing a small, targeted PPE with the GeoMIP ensemble, but the PPE did not perturb all relevant physical processes. This limited both its usefulness as a means of investigating the uncertainty in the climate response to solar geoengineering and for identifying the underlying reasons for these uncertainties. However, we believe that this approach of comparing a PPE to an MME is a useful one, and future studies could build on this one to improve the understanding of the uncertainty in the response to solar geoengineering. The findings of this study suggest that while parameters with the largest effect on processes determining climate sensitivity are important, other uncertain processes are also important for determining the response to the G1 solar geoengineering experiment, as temperatures are stabilized, and there is only a limited role for temperature feedbacks. Future studies generating targeted PPEs for solar geoengineering should also include parameters that affect climate dynamics, the control distribution of precipitation, and the CO$_2$ physiological effect [*Booth et al.*, 2012]. The magnitude of the NPP response, which counteracts the initial CO$_2$ physiological effect to some extent, is a particularly significant uncertainty, as it is critical to determining the terrestrial carbon cycle response and is connected to the broader surface hydrological response. Understanding the response of vegetation and surface hydrology to the combined effects of elevated CO$_2$ and climate change is critical to understanding the implications of solar geoengineering.

5. Conclusions

Multimodel ensembles and perturbed parameter ensembles are complementary approaches for assessing the range of potential climate responses to solar geoengineering, and combining these approaches

can provide new insights into the climate response to sunshade geoengineering. Here we found similar responses for G1-piControl in the PPE to those already found in the GeoMIP ensemble but noted that the PPE showed a much narrower range of responses, particularly in terms of its land mean hydrology, land versus ocean and Northern versus Southern Hemisphere temperature changes, and its pattern of tropical precipitation change. We also found that adjusting the G1 response, so that all ensemble members reproduced the piControl global mean SAT, narrowed the range of SAT responses in both ensembles but had little effect on the spread of the terrestrial responses. The central role of the piControl climatology in determining the hydrological response to solar geoengineering was shown, indicating that biases in the piControl climatology could be a major source of model disagreement at the zonal and grid cell level. The terrestrial hydrological response to solar geoengineering was affected by the CO_2 physiological effect and the increase of NPP, the strength of which differs markedly across the GeoMIP ensemble. Uncertainty in the magnitude of this reduction in transpiration appears to be a major source of model disagreement over the terrestrial hydrological response to solar geoengineering.

References

Andrews, T., P. M. Forster, O. Boucher, N. Bellouin, and A. Jones (2010), Precipitation, radiative forcing and global temperature change, *Geophys. Res. Lett.*, *37*, L14701, doi:10.1029/2010GL043991.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor (2012), Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere–ocean climate models, *Geophys. Res. Lett.*, *39*, L09712, doi:10.1029/2012GL051607.

Armour, K. C., C. M. Bitz, and G. H. Roe (2013), Time-varying climate sensitivity from regional feedbacks, *J. Clim.*, *26*(13), 4518–4534, doi:10.1175/JCLI-D-12-00544.1.

Bala, G., P. B. Duffy, and K. E. Taylor (2008), Impact of geoengineering schemes on the global hydrological cycle, *Proc. Natl. Acad. Sci. U.S.A.*, *105*(22), 7664–7669, doi:10.1073/pnas.0711648105.

Bala, G., K. Caldeira, and R. Nemani (2010), Fast versus slow response in climate change: Implications for the global hydrological cycle, *Clim. Dyn.*, *35*(2), 423–434, doi:10.1007/s00382-009-0583-y.

Betts, R. A., P. M. Cox, S. E. Lee, and F. I. Woodward (1997), Contrasting physiological and structural vegetation feedbacks in climate change simulations, *Nature*, *387*(6635), 796–799.

Betts, R. A., et al. (2007), Projected increase in continental runoff due to plant responses to increasing carbon dioxide, *Nature*, *448*(7157), 1037–1035, doi:10.1038/nature06045.

Booth, B. B. B., C. D. Jones, M. Collins, I. J. Totterdell, P. M. Cox, S. Sitch, C. Huntingford, R. A. Betts, G. R. Harris, and J. Lloyd (2012), High sensitivity of future global warming to land carbon cycle processes, *Environ. Res. Lett.*, *7*(2), 024002, doi:10.1088/1748-9326/7/2/024002.

Boucher, O., A. Jones, and R. A. Betts (2009), Climate response to the physiological impact of carbon dioxide on plants in the Met Office Unified Model HadCM3, *Clim. Dyn.*, *32*(2–3), 237–249, doi:10.1007/s00382-008-0459-6.

Brierley, C. M., M. Collins, and A. J. Thorpe (2010), The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model, *Clim. Dyn.*, *34*(2–3), 325–343, doi:10.1007/s00382-008-0486-3.

Cheng, W., J. C. H. Chiang, and D. Zhang (2013), Atlantic Meridional Overturning Circulation (AMOC) in CMIP5 models: RCP and historical simulations, *J. Clim.*, doi:10.1175/JCLI-D-12-00496.1.

Chou, C., J. D. Neelin, C.-A. Chen, and J.-Y. Tu (2009), Evaluating the "Rich-Get-Richer" mechanism in tropical precipitation change under global warming, *J. Clim.*, *22*(8), 1982–2005, doi:10.1175/2008JCLI2471.1.

Collins, M., C. M. Brierley, M. MacVean, B. B. B. Booth, and G. R. Harris (2007), The sensitivity of the rate of transient climate change to ocean physics perturbations, *J. Clim.*, *20*(10), 2315–2320, doi:10.1175/jcli4116.1.

Collins, M., B. Booth, B. Bhaskaran, G. Harris, J. Murphy, D. Sexton, and M. Webb (2010), Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi-model ensembles, *Clim. Dyn.*, *36*(9–10), 1737–1766, doi:10.1007/s00382-010-0808-0.

Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor (2003), An overview of results from the Coupled Model Intercomparison Project, *Global Planet. Change*, *37*(1–2), 103–133, doi:10.1016/s0921-8181(02)00193-5.

Cox, P. M., R. A. Betts, C. B. Bunton, R. L. H. Essery, P. R. Rowntree, and J. Smith (1999), The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, *Clim. Dyn.*, *15*(3), 183–203.

Crutzen, P. J. (2006), Albedo enhancement by stratospheric sulfur injections: A contribution to resolve a policy dilemma?, *Clim. Change*, *77*(3–4), 211–219, doi:10.1007/s10584-006-9101-y.

Donohue, R. J., M. L. Roderick, T. R. McVicar, and G. D. Farquhar (2013), CO2 fertilisation has increased maximum foliage cover across the globe's warm, arid environments, *Geophys. Res. Lett.*, *40*, 3031–3035, doi:10.1002/grl.50563.

Farquhar, G. D. (1997), Carbon dioxide and vegetation, *Science*, *278*(5342), 1411, doi:10.1126/science.278.5342.1411.

Farquhar, G. D., J. R. Ehleringer, and K. T. Hubick (1989), Carbon isotope discrimination and photosynthesis, *Annu. Rev. Plant Biol.*, *40*(1), 503–537.

Franks, P. J., et al. (2013), Sensitivity of plants to changing atmospheric CO2 concentration: From the geological past to the next century, *New Phytol.*, *197*(4), 1077–1094, doi:10.1111/nph.12104.

Fyfe, J. C., J. N. S. Cole, V. K. Arora, and J. F. Scinocca (2013), Biogeochemical carbon coupling influences global precipitation in geoengineering experiments, *Geophys. Res. Lett.*, *40*, 651–655, doi:10.1002/grl.50166.

Gimeno, L., A. Drumond, R. Nieto, R. M. Trigo, and A. Stohl (2010), On the origin of continental precipitation, *Geophys. Res. Lett.*, *37*, L13804, doi:10.1029/2010GL043712.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dyn.*, *16*(2–3), 147–168.

Govindasamy, B., K. Caldeira, and P. B. Duffy (2003), Geoengineering Earth's radiation balance to mitigate climate change from a quadrupling of CO2, *Global Planet. Change*, *37*(1–2), 157–168, doi:10.1016/s0921-8181(02)00195-9.

Hardiman, S. C., N. Butchart, T. J. Hinton, S. M. Osprey, and L. J. Gray (2012), The effect of a well-resolved stratosphere on surface climate: Differences between CMIP5 simulations with high and low top versions of the Met Office Climate Model, *J. Clim.*, *25*(20), 7083–7099, doi:10.1175/JCLI-D-11-00579.1.

Haywood, J. M., A. Jones, N. Bellouin, and D. Stephenson (2013), Asymmetric forcing from stratospheric aerosols impacts Sahelian rainfall, *Nature Clim. Change*, *3*, 660–665, doi:10.1038/nclimate1857.

Irvine, P. J., A. J. Ridgwell, and D. J. Lunt (2010), Assessing the regional disparities in geoengineering impacts, *Geophys. Res. Lett.*, *37*, L18702, doi:10.1029/2010GL044447.

Irvine, P. J., A. J. Ridgwell, and D. J. Lunt (2011), Climatic effects of surface albedo geoengineering, *J. Geophys. Res.*, *116*, D24112, doi:10.1029/2011JD016281.

Irvine, P. J., L. J. Gregoire, D. J. Lunt, and P. J. Valdes (2013), An efficient method to generate a perturbed parameter ensemble of a fully coupled AOGCM without flux-adjustment, *Geosci. Model Dev.*, *6*(5), 1447–1462, doi:10.5194/gmd-6-1447-2013.

Jones, A., and J. M. Haywood (2012), Sea-spray geoengineering in the HadGEM2-ES earth-system model: Radiative impact and climate response, *Atmos. Chem. Phys.*, *12*(22), 10,887–10,898, doi:10.5194/acp-12-10887-2012.

Jones, A., et al. (2013), The impact of abrupt suspension of solar radiation management (termination effect) in experiment G2 of the Geoengineering Model Intercomparison Project (GeoMIP), *J. Geophys. Res. Atmos.*, *118*, 9743–9752, doi:10.1002/jgrd.50762.

Keith, D. W. (2000), Geoengineering the climate: History and prospect, *Annu. Rev. Energy Env.*, *25*, 245–284.

Knutti, R., D. Masson, and A. Gettelman (2013), Climate model genealogy: Generation CMIP5 and how we got there, *Geophys. Res. Lett.*, *40*, 1194–1199, doi:10.1002/grl.50256.

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

Kravitz, B., et al. (2013a), Climate model response from the Geoengineering Model Intercomparison Project (GeoMIP), *J. Geophys. Res. Atmos.*, *118*, 8320–8332, doi:10.1002/jgrd.50646.

Kravitz, B., et al. (2013b), An energetic perspective on hydrological cycle changes in the Geoengineering Model Intercomparison Project, *J. Geophys. Res. Atmos.*, *118*, 13,087–13,102, doi:10.1002/2013JD020502.

Kravitz, B., A. Robock, and P. Irvine (2013c), Robust results from climate model simulations of geoengineering, *Eos Trans. AGU*, *94*(33), 292–292, doi:10.1002/2013EO330005.

Latham, J., et al. (2012), Marine cloud brightening, *Philos. Trans. R. Soc., A*, *370*(1974), 4217–4262, doi:10.1098/rsta.2012.0086.

Lunt, D. J., A. Ridgwell, P. J. Valdes, and A. Seale (2008), "Sunshade World": A fully coupled GCM evaluation of the climatic impacts of geoengineering, *Geophys. Res. Lett.*, *35*, L12710, doi:10.1029/2008GL033674.

MacMynowski, D. G., D. W. Keith, K. Caldeira, and H.-J. Shin (2011), Can we test geoengineering?, *Energy Environ. Sci.*, *4*(12), 5044–5052.

Moore, J. C., et al. (2014), Arctic sea ice and atmospheric circulation under the GeoMIP G1 scenario, *J. Geophys. Res. Atmos.*, *119*, 567–583, doi:10.1002/2013JD021060.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, and M. Collins (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, *430*(7001), 768–772, doi:10.1038/nature02771.

Murphy, J. M., D. M. H. Sexton, G. J. Jenkins, B. B. B. Booth, C. C. Brown, R. T. Clark, M. Collins, G. R. Harris, E. J. Kendon, and R. A. Betts (2009), UK climate projections science report: Climate change projections, Met Office Hadley Centre, Exeter, UK. 192 pp.

Niemeier, U., H. Schmidt, and C. Timmreck (2011), The dependency of geoengineered sulfate aerosol on the emission strategy, *Atmos. Sci. Lett.*, *12*(2), 189–194, doi:10.1002/asl.304.

Philip, S. Y., M. Collins, G. J. van Oldenborgh, and B. van den Hurk (2010), The role of atmosphere and ocean physical processes in ENSO in a perturbed physics coupled climate model, *Ocean Sci.*, *6*(2), 441–459.

Power, S., F. Delage, C. Chung, G. Kociuba, and K. Keay (2013), Robust twenty-first-century projections of El[thinsp]Nino and related precipitation variability, *Nature*, *502*, 541–545, doi:10.1038/nature12580.

Rasch, P. J., S. Tilmes, R. P. Turco, A. Robock, L. Oman, C. C. Chen, G. L. Stenchikov, and R. R. Garcia (2008), An overview of geoengineering of climate using stratospheric sulphate aerosols, *Philos. Trans. R. Soc., A*, *366*(1882), 4007–4037, doi:10.1098/rsta.2008.0131.

Reichler, T., and J. Kim (2008), Supplement to how well do coupled models simulate today's climate?, *Bull. Am. Meteorol. Soc.*, *89*(3), S1–S6, doi:10.1175/BAMS-89-3-Reichler.

Ricke, K. L., M. G. Morgan, and M. R. Allen (2010), Regional climate response to solar-radiation management, *Nat. Geosci.*, *3*(8), 537–541, doi:10.1038/ngeo915.

Robock, A., M. Bunzl, B. Kravitz, and G. L. Stenchikov (2010), A test for geoengineering?, *Science*, *327*(5965), 530–531, doi:10.1126/science.1186237.

Rougier, J., D. M. H. Sexton, J. M. Murphy, and D. Stainforth (2009), Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments, *J. Clim.*, *22*(13), 3540–3557, doi:10.1175/2008jcli2533.1.

Rowlands, D. J., D. J. Frame, D. Ackerley, T. Aina, B. B. Booth, C. Christensen, M. Collins, N. Faull, C. E. Forest, and B. S. Grandey (2012), Broad range of 2050 warming from an observationally constrained large climate model ensemble, *Nat. Geosci.*, *5*(4), 256–260.

Sanderson, B. M., et al. (2008a), Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes, *J. Clim.*, *21*(11), 2384–2400, doi:10.1175/2008jcli1869.1.

Sanderson, B. M., C. Piani, W. J. Ingram, D. A. Stone, and M. R. Allen (2008b), Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations, *Clim. Dyn.*, *30*(2–3), 175–190, doi:10.1007/s00382-007-0280-7.

Schmidt, H., et al. (2012), Solar irradiance reduction to counteract radiative forcing from a quadrupling of CO2: Climate responses simulated by four earth system models, *Earth Syst. Dyn.*, *3*(1), 63–78, doi:10.5194/esd-3-63-2012.

Singarayer, J. S., A. Ridgwell, and P. Irvine (2009), Assessing the benefits of crop albedo bio-geoengineering, *Environ. Res. Lett.*, *4*(4), 8, doi:10.1088/1748-9326/4/4/045110.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (2007), Climate change 2007: The physical science basis, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by D. Q. S. Solomon et al., pp. 996, Cambridge Univ. Press, Cambridge.

Stainforth, D. A., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, *433*(7024), 403–406, doi:10.1038/nature03301.

Stott, P. A., D. A. Stone, and M. R. Allen (2004), Human contribution to the European heatwave of 2003, *Nature*, *432*(7017), 610–614, doi:10.1038/nature03089.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tilmes, S., et al. (2013), The hydrological impact of geoengineering in the Geoengineering Model Intercomparison Project (GeoMIP), *J. Geophys. Res. Atmos.*, *118*, 11,036–11,058, doi:10.1002/jgrd.50868.

Yokohata, T., J. Annan, M. Collins, C. Jackson, M. Tobis, M. Webb, and J. Hargreaves (2012), Reliability of multi-model and structurally different single-model ensembles, *Clim. Dyn.*, *39*(3), 599–616, doi:10.1007/s00382-011-1203-1.

Zeng, N. (2003), Drought in the Sahel, *Science*, *302*(5647), 999–1000, doi:10.1126/science.1090849.