Response to reviewers
esd-2015-51
Original reviewer comments in normal typeface. **Responses in bold.**

In addition to reviewer comments, we have fixed a number of technical errors pointed out by the editor. We have attached a "tracked changes" version of the manuscript to point out our modifications.

Editor's comments:

* Eq6: Your silent assumption is y(0)=0, correct? In case yes, please make it explicit.

This is now explicitly mentioned.

* p14: 'all frequencies faster than' -> 'all frequencies lower than' ?

Yes, changed.

* p15: How do you get phase lags of 90°/45° rather than 0° for D->0?

This was a mistake, and it should be as omega goes to infinity, not 0. This has been fixed.

* Eq9: your integral to be replaced by \int_0^t y(t') dt'?

We acknowledge the sloppy notation and have changed the variable inside the integral.

* Line before Eq11: y=d to be replaced by 'y=d/(1+\tau s)'?
* Instead of Eqs 11+12 I receive from my top equation y = d / (1+\tau s +\beta e^{-sD} K) which looks quite different from the authors' expression.

These comments stem from the fact that we had not made our description clear when we were evaluating the full system, versus only the part of the system that characterizes the input-output relationship (i.e., excluding noise). We have addressed this more carefully, including additional sentences to make explicit which case is being discussed throughout the manuscript, as well as modifying the notation in the equations so there is no confusion.

* On p18 the authors are flipping back and forth with \omega carrying a unit vs being unit-less. Please stick to the physicists' tradition of letting it carry its unit. Furthermore, in the last line, there is a mistake with the unit of k_i .

We have added units to all instances of omega. We are unsure as to what our units error is for k_i in the last line of this page, but the calculation is correct:

w_gc=k_i*beta, and w_gc is chosen to be 0.2 rad/yr, so k_i=0.2/beta.

* In Eq 14, I could derive the expressions in the following order #1 -> #3 -> #2. Please check whether #2 is needed indeed.

In equation 14, the middle term is the direct interpretation of the quantities in the first term. K=kp+ki/s=kp-i*ki/omega, so Im(K)=-ki/omega, and Re(K)=kp. The third term then applies a trigonometric identity to the second term. Strictly speaking, #2 isn't necessary, but we feel that it is instructive to include (not every reader has memorized trigonometric identities for arctangent).

* p 21: colon between Eq1 & Eq2 is missing, merging both Eqs.

We have improved the spacing for this line so it is clear there are two equations.

* p23: k(y_2) -> -k(y_2) ?

Yes, thanks for pointing that out.

* on p25 I am confused by the Symbol 'S'. Is this S different from the previous S or not?

Thanks for pointing this out. This should be a different symbol, so we now call it Z.

Reviewer #1

How did you choose the values of beta and D and tau in fig 3?

These values were simply illustrative and (with the exception of the red lines, as stated) not meant to represent any particular physical system. We have added a clarification of this to the text.

Eq 13 why did you chose kp=ki*tau?

In looking at Equation 12, it would simplify the expression if one could make the factor (1+s*tau) cancel. This is accomplished via kp=ki*tau.

p. 28 section 3.6 why did you choose kp/ki=2.5?

Choosing kp/ki=2.5 is approximately what one would get for this system from equation 15, but in some sense, this is a design choice. Going through all of the details as to why someone would increase or decrease kp from the values obtained from equation 15 is a fairly detailed process requiring some experience in designing control systems. We initially included such discussions and decided that it reduced clarity so much that the extra precision added wasn't worth confusing the reader.

However, we now realize that not saying anything at all seems strange, so we have included a short note describing what we have just said here.

p. 31 section 3.7 why did you choose kp=ki for L0 T0 relationship?

See previous explanation.

p.36 1st line Antarctic insolation reduction reaches minus 16%

The reduction reaches 16%. A reduction reaching minus 16% would be an increase by 16%.

Tables 1 and 2: long term PIControl – how many years is long term? 500?

We realize this was misleading, because our simulations branched from a long-term preindustrial control run (available through the CMIP5 archive). We have removed the phrase "long term".

Reviewer #2

I think that the manuscript has considerably improved through the revision. And I only have a few further concerns specified below.

The authors now discuss more carefully the implications of their proposed control strategy for actually geoengineering the climate. This is in particular true for the introduction. I'm still not fully happy of the discussion in the final section. The authors still state that they had "demonstrated the ability to simultaneously manage multiple climate criteria using the common approach of changing solar irradiance". The common approach is to reduce solar irradiance. In one of the examples (2x2 with the CESM) it is clearly shown that the design goals can't be reached by reducing solar irradiance. Why not stating this very clearly instead of vaguely saying that "accomplishing the objectives with physically achievable mechanisms […] introduces additional complications". In particular as it is stated in the introduction now, that "it may not be possible to achieve all objectives due to physical constraints", it is nice to come back to this statement in the conclusions with an example.

We agree with this point and have added the reviewer's suggestion.

I had suggested to remove Fig. 1 because I thought it is a banality that more or less reducing the irradiance would reduce the global temperature more or less. As the authors have argued against it I looked at Fig. 1 a bit more carefully than in the first review. I agree now that the Figure is interesting but in a very different sense which I think needs further discussion. I would have guessed that a smaller reduction of solar irradiance (as in the middle panel compare to the top panel) should increase the temperature almost globally. While a reduction seems to be simulated in some areas (like Northern Africa or

the North Atlantic) in other regions (e.g. polar regions and tropical oceans) the opposite is simulated. Am I missing something, here? If there is no error in my thinking or the figures I'd consider this an interesting example for non-linearity which would render the optimization difficult even with just one control parameter.

We agree with the reviewer that reducing irradiance will reduce global temperature. If comparing the results to a high CO2 world, all panels would show cooling, and we agree this is indeed a banality and doesn't need to be illustrated. However, we are comparing the results to a preindustrial control, meaning there would indeed be some residuals, as are shown in the figure. The slight differences in global mean temperature between the panels are not necessarily indicative of how different the regional patterns will be due to circulation changes, feedbacks, and other internal processes. However, the general latitudinal distribution of temperature (i.e., the equator-to-pole temperature gradient) is predictable, which is the purpose of this figure; indeed, the values of insolation reduction chosen for each panel were based on linear scaling of the temperature response to insolation. The reviewer has rightly pointed out something that we discuss repeatedly throughout the manuscript: carefully choosing one's objectives is important!

I'm still not totally happy with the use of the references to Kalidindi et al. (2014), Ferraro et al. (2014), and Niemeier et al. (2013). It sounds now like there was a consensus that different SRM methods would produce only regionally different responses. However, the latter two papers specifically argue also for global differences, even if one may consider them relatively small.

This is a good point. We have modified our description accordingly: "…although some global and regional effects, especially those due to stratospheric heating by the aerosols, are likely to differ between the two methods…"

Apparently also the other reviewer doesn't feel fully comfortable with reviewing the large part of the paper (Section 3) dealing with the design of the feedback strategy which albeit may seem the central part of this manuscript. It may hence be advisable to search for a specialist in this area for such a review.

The editor has taken care of this for us.

Manuscript prepared for Earth Syst. Dynam. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 18 March 2016

Geoengineering as a design problem

B. Kravitz[1], D. G. MacMartin[2], H. Wang[1], and P. J. Rasch[1]

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA

Correspondence to: B. Kravitz (ben.kravitz@pnnl.gov)

Abstract

Understanding the climate impacts of solar geoengineering is essential for evaluating its benefits and risks. Most previous simulations have prescribed a particular strategy and evaluated its modeled effects. Here we turn this approach around by first choosing example climate objectives and then designing a strategy to meet those objectives in climate models.

There are four essential criteria for designing a strategy: (i) an explicit specification of the objectives, (ii) defining what climate forcing agents to modify so the objectives are met, (iii) a method for managing uncertainties, and (iv) independent verification of the strategy in an evaluation model.

We demonstrate this design perspective through two multi-objective examples. First, changes in Arctic temperature and the position of tropical precipitation due to CO_2 increases are offset by adjusting high latitude insolation in each hemisphere independently. Second, three different latitude-dependent patterns of insolation are modified to offset CO_2-induced changes in global mean temperature, interhemispheric temperature asymmetry, and the equator-to-pole temperature gradient. In both examples, the "design" and "evaluation" models are state-of-the-art fully coupled atmosphere–ocean general circulation models.

1 Introduction

Geoengineering describes a set of technologies designed to offset some of the effects of anthropogenic climate change by deliberately intervening in the climate system. There are many proposed methods of *solar* geoengineering (methods of geoengineering that reduce incident shortwave radiation at the surface; all subsequent discussions of geoengineering specifically refer to solar geoengineering). Some of the most studied include introducing a layer of reflective sulfate aerosols into the stratosphere or brightening marine low clouds (e.g., Crutzen, 2006; Latham, 1990; NAS, 2015).

Many of the ongoing efforts in solar geoengineering research involve climate model simulations designed to ascertain the expected climate effects of various scenarios of geoengi-

neering (e.g., Kravitz et al., 2011, 2013b, 2015a). Many simulations focus on uniformly reducing solar irradiance or imposing a particular spatial pattern of forcing from stratospheric aerosols or cloud brightening. However, the expected climate effects depend not only on the amount of geoengineering, but also on the spatial pattern; both of these are, at least in part, design choices. Furthermore, the objectives of geoengineering may involve balancing multiple criteria, such as maintainting Arctic temperature without disrupting tropical precipitation (an example we explore below).

As an example, one of the results from geoengineering that is repeatedly discussed is that offsetting the global mean radiative forcing from a CO_2 increase by reducing total solar irradiance would result in an overcooling of the tropics and an undercooling of the poles (Govindasamy and Caldeira, 2000). This is largely due to the fact that CO_2 concentration is more or less evenly distributed in climate models, so CO_2 forcing has a much weaker latitude dependence than forcing from solar irradiance (Taylor et al., 2011). Kravitz et al. (2013a) showed that this pattern of temperature response is robust across all 12 models that simulated GeoMIP experiment G1, in which the radiative forcing from an abrupt quadrupling of the CO_2 concentration was offset by solar reduction (Kravitz et al., 2011). However, Fig. 1 illustrates that overcooling of the tropics and undercooling of the poles (top panel) is not a foregone conclusion, even if only one degree of freedom is varied – the amount of total solar irradiance reduction. One could easily reduce insolation less than in G1 so that no large region is overcooled (middle panel) or reduce insolation more than in G1 so that there is no residual warming (bottom panel).

Figure 1 provides a simple illustration that many of the climate effects of geoengineering are design choices, presuming the ability to actually impose changes with specific characteristics. As such, statements about the climate effects of geoengineering *in general* are ill-posed; such statements require the context of specific climate objectives and an approach *designed* to meet them. A handful of studies have explored this idea of meeting climate objectives other than the oft-studied global mean temperature reduction. Ban-Weiss and Caldeira (2010) explored changes in the latitude of the solar geoengineering pattern and found that doing so could better offset the residual temperature changes in a G1-like exper-

iment. MacMartin et al. (2013) explored modifying insolation by latitude and season; they found that doing so greatly increased the range of achievable climates through insolation reduction, both globally and on a regional basis.

Climate model simulations suggest that many of the proposed methods of conducting solar geoengineering are likely to have both commonalities and differences in their climate effects (Crook et al., 2015; Kalidindi et al., 2014; Niemeier et al., 2013). Here we use the common idealized representation of reducing solar irradiance. This has been shown to be similar in global mean near-surface effects to simulations of stratospheric sulfate aerosols (Kalidindi et al., 2014), although ~~regional effects and~~ some global and regional effects, especially those due to stratospheric heating by the aerosols, are likely to differ between the two methods (Ferraro et al., 2014; Niemeier et al., 2013). In the present work, we reduce insolation as a function of latitude; while these exact patterns may not be achievable, the general characteristics of those patterns are broadly consistent with the types of variations that could be achieved via other means of geoengineering (e.g., stratospheric sulfate aerosols). Augmenting the discussion to other proposed methods of solar geoengineering adds additional degrees of freedom (for example, stratospheric aerosols include altitude and possibly particle composition as additional adjustable parameters), but these methods also include additional complications (e.g., atmospheric circulation imposes constraints on achievable latitudinal dependence). We discuss some of these issues in Sect. 6.

Our primary motivation in this study is to introduce a design perspective that can be used to more systematically evaluate some of the potentials and limitations of geoengineering. We do this by exploring two examples of geoengineering strategies designed to meet specific, multifaceted goals. For any strategy, achieving multifaceted goals can be accomplished via following a certain set of criteria:

1. an explicit definition of specific objectives of geoengineering;
2. determination of the particular degrees of freedom to be modified to meet the objectives;
3. a strategy for meeting the objectives in the presence of uncertainty;

4. verification of the designed strategy in a different evaluation model.

The examples we choose (Sect. 2) are not necessarily indicative of any particular objective that might be chosen, if there were ever a decision to engage in geoengineering in the future. Our purpose is simply to illustrate how, given an objective for geoengineering, a strategy to meet that objective might be designed.

Implicitly included in these four criteria is that it is necessary to determine the feasibility of the objectives. It may not be possible to achieve all objectives due to physical constraints on the climate system. Moreover, the space of possible climates may be further narrowed by technological limitations. As an example, it is not clear how stratospheric transport can be controlled, which may limit the spatial distribution of radiative forcing that is achievable via geoengineering with stratospheric sulfate aerosols. Our analyses inherently include the assumption that the radiative forcing is achievable.

In a system in which the relationships between adjustable climate parameters and the desired pattern of radiative forcing are well-characterized, one could optimize the relative contributions of the parameters such that the desired climate objectives are approximately met. This was the approach taken by Ban-Weiss and Caldeira (2010) and MacMartin et al. (2013), for example. In practice, even independent of uncertainties in the ability to achieve the desired climate system changes, there are substantial uncertainties in both the radiative forcing exerted by a change in insolation and the climate response to that radiative forcing (Stocker et al., 2013). In addition, the climate response is dependent upon the particular forcing agent; this concept was defined as *efficacy* by Hansen et al. (2005). Because climate models imperfectly represent the dynamical behavior of the real climate system and because climate observations are sparse, many of the uncertainties associated with understanding radiative forcing and climate response are difficult to reduce. Therefore, any deployment of solar geoengineering would require a method of managing these uncertainties to ensure that the chosen objectives of geoengineering are met as well as possible even in the presence of uncertainty.

One method of managing uncertainties is to use *explicit feedback*, in which geoengineering is regularly adjusted based on the observed climate state and how far it is from the

chosen objectives (Jarvis and Leedal, 2012; MacMartin et al., 2014b). Such techniques are well developed in the field of control theory (see Åström and Murray (2008) for a more thorough explanation). The use of explicit feedback has been demonstrated for several objectives, including reducing total solar irradiance to meet an objective defined in terms of global mean temperature (MacMartin et al., 2014b; Kravitz et al., 2014), reducing total solar irradiance to limit the rate of temperature change (MacMartin et al., 2014a), or injecting sulfur dioxide into the Arctic stratosphere to limit sea ice loss (Jackson et al., 2015). All of these previous studies involved modifying a single climate system feature (amount of solar geoengineering) to achieve a single climate objective. In all subsequent discussions, we refer to potentially modifiable climate system parameters as *degrees of freedom* in achieving climate objectives.

Although these past studies were instrumental in developing applications of explicit feedback for geoengineering, their applicability is limited in that they do not address the potential for multifaceted geoengineering goals. Offsetting multiple independent features of climate change requires modifying multiple simultaneous degrees of freedom. Ensuring that those climate objectives are met in the presence of uncertainty requires explicit feedback. The present study is the first to combine these two aspects, illustrating some of the potentials and limitations associated with designing geoengineering strategies.

Addressing Criterion 4 requires a two-stage process, as was illustrated by Kravitz et al. (2014). We illustrate this procedure and explore its consequences using two independently developed models designed to simulate Earth's climate. These models are imperfect approximations of Earth and of each other. The explicit feedback strategy is first analyzed in a design model: in this model, numerous tests are permitted to fully characterize the dynamics of the climate system with and without coupling to the explicit feedback. After the strategy is designed, it is then implemented in an evaluation model; this verifies that the strategy does not depend on a highly-accurate description of the dynamics of the design model, as the dynamics will not be identical between the design and evaluation models. For the design model, we use the Community Earth System Model (CESM) 1.0.2 (Hurrell et al., 2013), a fully coupled atmosphere–ocean general circulation model (AOGCM) that

participated in the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2012). For the evaluation model, we use the Goddard Institute for Space Studies (GISS) ModelE2 (Schmidt et al., 2014), another fully coupled AOGCM that participated in CMIP5. This two-model approach captures the fact that in a real deployment, the situation would require designing a model-based strategy that works in actual deployment. Of course, these two models may not represent the differences between models and reality. Nevertheless, this process is both illustrative and provides additional confidence beyond a demonstration in which the strategy was designed and implemented in the same model. If there were ever a deployment of geoengineering, the design process would presumably incorporate information from a wide range of climate models.

We illustrate the design approach through two examples; one regionally-focused and the other globally-focused, described in Sect. 2. Section 3 describes in detail the procedure for designing a feedback algorithm, including a discussion of "system identification" simulations used to estimate the relevant dynamics of the design model. The results from the design and evaluation models for the two examples are discussed in Sects. 4 and 5. Section 6 includes a discussion of the present study, including some of the differences in this process if one were studying stratospheric aerosols or marine cloud brightening rather than using idealized latitude-dependent solar reductions.

2 Strategy

Here we illustrate the nature of geoengineering as a design problem through two examples, which we will call 2×2 and 3×3, indicating the number of inputs (degrees of freedom that are modified) and outputs (climate objectives). The first of these examples focuses on countering Arctic warming that would occur under CO_2 increases (a regional objective) while seeking to minimize shifts in tropical precipitation that would occur due to both CO_2 increases and if only high-latitude Northern Hemisphere insolation was adjusted (Haywood et al., 2013). The second design problem considers a more global perspective on geoengineering, but rather than only considering global mean temperature, the feedback design

compensates for both the relative overcooling of the tropics (or undercooling of the poles) apparent in Fig. 1a and the temperature difference between the two hemispheres that can in turn lead to shifts in tropical precipitation (largely characterized by the Intertropical Convergence Zone, or ITCZ). In both cases we evaluate strategies in the presence of a $1\,\%\,\mathrm{yr}^{-1}$ increase in the CO_2 concentration (abbreviated 1pctCO2).

The 2×2 case is motivated by Arctic warming, which is a strong driver of Arctic sea ice loss (Serreze et al., 2007), permafrost thaw (e.g., Schaefer et al., 2011), and other impacts (e.g., Bintanja and Selten, 2014). Arctic insolation reductions could offset some Arctic warming (Caldeira and Wood, 2008; Robock et al., 2008; MacCracken et al., 2013; Tilmes et al., 2014; Jackson et al., 2015), but only cooling the Arctic would tend to shift the ITCZ toward the warmer hemisphere (e.g., Broccoli et al., 2006). Concomitant changes in Antarctic insolation are unlikely to substantially affect Arctic temperature but could be used to offset the changes in tropical precipitation caused by CO_2 and Arctic insolation reductions.

More concretely, the two inputs in the 2×2 system are changes in Arctic insolation and Antarctic insolation (Fig. 2a), where for the former we choose insolation reductions from 60–90° N, and for the latter, 60–90° S. (This choice is neither "optimal" in any sense, nor necessarily achievable, but sufficient to illustrate the design strategy.) One of the outputs is change in Arctic temperature, defined as an area-weighted average of surface air temperature in the region spanning the Arctic (66 $2/3$–90° N); this is affected by changes in Arctic insolation and is relatively unaffected by changes in Antarctic insolation. The other output characterizes the latitudinal displacement of zonally averaged precipitation P by defining

a precipitation centroid (χ):

$$\chi \approx \frac{\int_{-\pi/2}^{\pi/2} P \cdot \psi \, \mathrm{d}A}{\int_{-\pi/2}^{\pi/2} P \, \mathrm{d}A} \tag{1}$$

where ψ is latitude (integration over $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is the entire latitude range of 90° S to 90° N). A is area-weighted latitude, i.e.,

$$\mathrm{d}A = \cos(\psi)\,\mathrm{d}\psi \Rightarrow A = \int_{-\pi/2}^{\pi/2} \cos(\psi)\,\mathrm{d}\psi = 2.$$

A quantitative representation of the dynamic relationships between the inputs and outputs, known as the *influence matrix*, is given in Sect. 3.6.

The 3×3 case considers a more global objective. An increase in CO_2 would increase global mean temperature (abbreviated T_0). The Northern Hemisphere would warm more under CO_2 increases than the Southern Hemisphere (Kang et al., 2015), which influences ITCZ location and tropical precipitation patterns (e.g., Marshall et al., 2014). Also, because of the various mechanisms associated with poleward heat transport and polar amplification (e.g., Holland and Bitz, 2003), high latitudes would warm more than low latitudes. Reducing total solar irradiance could offset changes in T_0, but due to the different latitudinal patterns of CO_2 warming and insolation reduction, there would still be residual changes in both the differential Northern vs. Southern Hemisphere warming and the equator-to-pole temperature gradient (Fig. 1, see also Caldeira and Wood, 2008; Kravitz et al., 2013a). However, these residual patterns could be offset by choosing different patterns of insolation reduction beyond a globally-uniform reduction.

As metrics for these, we define T_1 and T_2 as the linear and quadratic meridional-dependence of zonal-mean temperature $T(\psi)$:

$$
\begin{aligned}
T_0 &= \frac{1}{A} \int_{-\pi/2}^{\pi/2} T(\psi)\, \mathrm{d}A \\
T_1 &= \frac{1}{A} \int_{-\pi/2}^{\pi/2} T(\psi) \sin\psi\, \mathrm{d}A \\
T_2 &= \frac{1}{A} \int_{-\pi/2}^{\pi/2} T(\psi) \frac{1}{2}(3\sin^2\psi - 1)\, \mathrm{d}A.
\end{aligned}
\tag{2}
$$

These equations are defined by the projection of $T(\psi)$ onto the first three Legendre polynomial functions (constant, linear, and quadratic) of $\sin(\psi)$, abbreviated L_0, L_1, and L_2 (Fig. 2b):

$$
\begin{aligned}
L_0 &= 1 \\
L_1 &= \sin(\psi) \\
L_2 &= \frac{1}{2}(3\sin^2(\psi) - 1).
\end{aligned}
\tag{3}
$$

These correspond to the first three terms of a polynomial expansion of the zonal-mean temperature. Similarly, we define the inputs as a a reduction in insolation with latidudinal dependence L_0, L_1, and L_2; these are similar basis functions to those used by Ban-Weiss and Caldeira (2010) and MacMartin et al. (2013). For simplicity, we subsequently refer to the three patterns of solar reduction given in Fig. 2b as L_0, L_1, and L_2. Additional terms could be considered, but there is a clear physical mechanism underlying the influence between these three inputs and three outputs; we discuss the importance of this physical linkage in Sect. 6. Note that changes in L_0 were conducted by all models participating in GeoMIP experiment G1 described previously (Kravitz et al., 2011). The functions in Eq. (3)

are orthogonal, which will be useful in designing feedback strategies (discussed further in Sect. 3.4).

All of these simulations are conducted using the method of explicit feedback, as described by MacMartin et al. (2014b) and Kravitz et al. (2014, 2015b). Section 3 below is devoted to a discussion of how one determines a feedback algorithm that will effectively meet these goals.

3 Designing a multivariate feedback strategy

3.1 Overview and motivation

While the previous section introduced the idea of choosing multiple spatial degrees of freedom to balance multiple criteria, this section is concerned with how to choose the amplitude of each of these degrees of freedom *as a function of time* so that the desired climate objectives are met despite uncertainty. With perfect climate models, this process would be straightforward, but in actuality, the amount of each degree of freedom would need to be continually adjusted in response to observations, increasing or decreasing as appropriate to avoid under- or over-compensating relative to specified goals. This adjustment in response to observations is a feedback process, and is an essential element of any plausible geoengineering deployment strategy. With proper design, this adjustment process will converge to the chosen objectives for a wide range of uncertainty in the expected climate response.

Feedback design (design of the explicit feedback algorithm) requires some information about the system response to an input. This information is provided by the design model, and feedback is then used to bridge the gap between the modeled response and the real-world response if this design were implemented.

Specifying exactly what information is needed to design the feedback algorithms is not immediately obvious. We begin this section with a discussion of dynamic modeling for feedback design (Sect. 3.2), followed by a brief introduction to the design of the feedback algorithm (Sect. 3.3), focused primarily on single-input single-output (SISO) design. The main

focus of this paper involves balancing multiple criteria using multiple degrees of freedom; this multivariate feedback will be designed using sequential closure of SISO feedbacks, introduced in Sect. 3.4. The discussion of feedback algorithm design motivates what model information is needed to determine the input/output relationships for each example. This information is obtained through "system identification" simulations, described in Sect. 3.5. Sections 3.6 and 3.7 then describe both the input/output system identification and multivariate feedback design for the 2×2 and 3×3 examples, respectively. Further details on feedback algorithm design for geoengineering can be found in MacMartin et al. (2014b); an accessible text covering feedback design more broadly is Åström and Murray (2008). A reader only interested in the results and not the design of the feedback algorithms can skip to Sect. 4. All simulations and analyses in this section are conducted with the design model CESM 1.0.2.

3.2 Dynamic modeling for feedback design

The feedback algorithm defines the rule by which the "input" (e.g., solar reduction) is adjusted in response to observations of the "output" (e.g., difference between measured and desired global mean temperature). The design of this algorithm starts with a dynamic model of the input/output behavior of the system. This dynamic model does not describe how the entire climate state responds to a perturbation in the input signal, but specifically the response of the output signal. We use the term *dynamic* to indicate that this model includes transient behavior and not just the equilibrium response. We assume that this process can be reasonably approximated by a linear relationship, and that nonlinear effects are small enough that they are managed by the feedback algorithm, which provides robustness to uncertainty. As we will show later, this assumption is not detrimental to meeting our chosen objectives, although it is potentially problematic for other objectives (Sect. 6).

A general linear dynamic input/output relationship can be described by a convolution equation in the time-domain. However, many of the expressions we wish to evaluate are greatly simplified when expressed in the frequency domain, because convolution is replaced by multiplication, and coupled differential equations in the time-domain become algebraic

relationships in the frequency domain. A time-domain equation $f(t)$, where t is time, can be represented in the frequency domain via the Laplace Transform:

$$F(s) = \int_0^\infty e^{-st} f(t)\,\mathrm{d}t \tag{4}$$

where $s = i\omega$, and ω is (angular) frequency.

In illustrating feedback design guidelines in the next subsection, it is convenient to consider a first-order linear (i.e., first-order autoregressive) description of the input–output relationship, including a time-delay D. This is the simplest non-trivial dynamical system. Although we would not necessarily expect this system to match the dynamics of the actual climate system at all frequencies, it is sufficient for illustration. With $y(t)$ as the climate output signal (e.g., temperature change) and $u(t)$ as the input signal (e.g., solar reduction), then

$$\tau\dot{y} = -y(t) + \beta u(t - D) + d(t) \tag{5}$$

for some coefficient β, where τ is an e-folding time constant (as used here, in years) and $\dot{y}$ indicates the time derivative of y. Including an explicit time delay of D years is necessary here as our simulations adjust forcing for the next year based on the average climate output over the previous year; each of these choices contributes on average a half-year delay (MacMartin et al., 2014b). In addition to the response to $u(t)$, the signal $y(t)$ will also include effects both from natural variability and from anthropogenic climate change; these are captured above through the exogenous input $d(t)$. ~~With~~ For the purposes of characterizing the input-output response, we eliminate sources of variability in output that are not associated with the input (i.e., noise) by setting $d = 0$~~(to characterize the input/output response), then~~. Then for an abrupt change in the input $u(t)$ from zero to one at time $t = 0$, ~~$y(t)$~~ $y_u(t)$ (where the subscript indicates that noise is not included) for $t \geq D$ is given by

$$y_u(t+D) = \beta(1 - e^{-t/\tau}). \quad (6)$$

where $y(0) = 0$.

Taking the Laplace transform of Eq. (6) and dividing by the Laplace transform of the input $\boldsymbol{u}(t)$, the response ~~y~~ y_u to the input u can equivalently be characterized in the frequency domain as ~~$y(s) = G(s)u(s)$~~ $y_u(s) = G(s)u(s)$ through the transfer function

$$G(s) = e^{-sD}\frac{\beta}{1+s\tau}. \quad (7)$$

where again $d(s)$ is omitted in this expression, as the transfer function describes only the changes in output that are linearly related to changes in input.

At any frequency ω, the complex number $G(s) = G(i\omega)$ can be described by its magnitude $\|G(i\omega)\|$ and phase $\phi(G(i\omega)) = \tan^{-1}(\mathrm{Im}(G)/\mathrm{Re}(G))$. Note that because $||e^{-sD}|| = 1$, the time delay adds phase lag but does not change the magnitude. The magnitude and phase of $G(i\omega)$ are shown for several parameter values in Fig. 3, providing a graphical representation of the frequency response of the transfer function (also called a Bode plot). Red lines are roughly consistent with the relationship identified between Arctic insolation and Arctic temperature in the 2×2 design example that follows in subsequent sections. Different values of β scale the magnitude at all frequencies but do not change the phase. The time constant τ determines the range of frequencies over which the system response is *quasi-static* (roughly the same magnitude as the equilibrium response, indicated by the flat part of the curves in Fig. 3). τ is an e-folding timescale, so the quasi-static response is approximately characterized by all frequencies ~~faster~~ lower than $1/(3\tau)\,\mathrm{rad\,yr}^{-1}$. The phase contribution from the term $1 + s\tau$ in the denominator transitions from zero to $-90°$, contributing $-45°$ at frequency $\omega = 1/\tau$. Time delay contributes substantial phase lag at high frequencies.

A semi-infinite diffusion model has been shown by MacMynowski et al. (2011) to more accurately capture the response of the global mean temperature to a uniform solar reduction

in the HadCM3L general circulation model; this information was used by MacMartin et al. (2014b) to design feedback strategies. The corresponding transfer function is

$$G_d = e^{-sD} \frac{\beta_d}{1 + (s\tau_d)^{1/2}} \tag{8}$$

where the subscript d indicates the diffusion model. Figure 4 compares the first order linear model with $\beta = 0.447$, $\tau = 1.946$, $D = 1$ (these values are used in describing the 2×2 case in Sect. 3.6) and the semi-infinite diffusion model where we choose $\beta_\mathrm{d} = 0.732$, $\tau_\mathrm{d} = 4.063$, and $D = 1$ to give the same magnitude and phase of the transfer function at $\omega = 0.2\,\mathrm{rad\,yr^{-1}}$. The value of $\beta_\mathrm{d} = 0.732$ corresponds to an equilibrium climate sensitivity of 2.71 °C, which is similar to the climate sensitivity of HadCM3L (MacMynowski et al., 2011). A value of $\tau = 4.063$ years corresponds to a rise time to $1/e$ of the equilibrium value of 6.339 years (MacMynowski et al., 2011). This is somewhat less than the value obtained by MacMynowski et al. (2011) (Fig. 3), but this value is not unreasonable in characterizing the frequency response of climate models in general (Caldeira and Myhrvold, 2013). As would be expected from Eqs. (7) and (8), the first order linear model with no time delay asymptotes to a phase lag of 90° as $\omega \to 0$ $\omega \to \infty$, and the semi-infinite diffusion model asymptotes to 45°. For reasons that will be clear in the next subsection, designing feedback strategies does not require knowledge of the system dynamics at all frequencies. Figure 4 illustrates that there are multiple possible system representations that could have the same transfer function magnitude and phase at a single frequency.

3.3 Single-input, single-output (SISO) feedback design

We now consider the design of the feedback algorithm, using the model in Eq. (7) for illustration. As was done by MacMartin et al. (2014b) and Kravitz et al. (2014, 2015b), we choose proportional-integral control. This choice (or its augmented counterpart of proportional-integral-derivative control) is ubiquitous in control theory and is a standard "first attempt" when designing a feedback algorithm. As we will show, proportional-integral control is sufficient for our purposes. In the continuous time domain, proportional-integral control is rep-

resented as

$$u(t) = -k_p y(t) - k_i \int_0^{\infty} y^t y(t\sigma)\, dt\sigma \tag{9}$$

where k_p and k_i are the proportional and integral gains, respectively, collectively called *control gains*. The negative sign associated with $y(t)$ is included by convention. $y(t)$ represents the departure of the system state (e.g., temperature) from the reference point at any given time, that is, the goal is to minimize y. Taking the Laplace transform, this can be represented in the frequency domain as $u(s) = -K(s)y(s)$ through the transfer function

$$K(s) = k_p + k_i/s = \frac{k_p s + k_i}{s}. \tag{10}$$

The full system is now described in the frequency domain by $y(s) = G(s)u(s) + d(s)$~~, where d describes the part of y~~; again, $G(s)$ (the transfer function) describes the portion of the output $y(s)$ that is related to the input $u(s)$, and $d(s)$ is the noise, or the part of $y(s)$ that is due to sources other than the input ~~u~~ $u(s)$ (e.g., climate change and natural variability)~~, and our~~. The feedback algorithm is described by $u(s) = -K(s)y(s)$. In the absence of feedback ($K(s) = 0$) then $y = d$. With feedback,

$$y(s) = \frac{1}{1 + G(s)K(s)} d(s). \tag{11}$$

The characteristics of the system with feedback thus depend only on the product $G(s)K(s)$. This product is referred to as the *loop transfer function*; with the simple model in Eq. (7) and proportional-integral control, its frequency response is the product of Equations 7 and 10:

$$G(s)K(s) = e^{-Ds} \frac{\beta}{1 + s\tau} \cdot \frac{k_p s + k_i}{s}. \tag{12}$$

These two equations illustrate a substantial advantage of working in the frequency domain, as the equivalent time-domain formulation would be much more complicated and would provide less insight.

There are three critical observations: (1) at very low frequencies ($\omega \ll \omega_{\text{gc}}$ where ω_{gc} is a fixed value defined below), $G(i\omega)K(i\omega)$ is large for any non-zero $G(s)$. This means that feedback achieves the goal of maintaining $\boldsymbol{y}(s)$ small, and further, that it is not necessary to know the dynamics of the climate at low frequencies to successfully design a feedback algorithm. (2) At very high frequencies ($\omega \gg \omega_{\text{gc}}$), $G(i\omega)K(i\omega)$ is small, and thus $y(s)$ is unchanged by the feedback; again, it is not necessary to know the dynamics of the climate at very high frequencies because there is no significant input signal at high frequencies. (3) If at some frequency ω we had $G(i\omega)K(i\omega) = -1$, then the system would be unstable (an unbounded response to a disturbance at that precise frequency), and if $G(i\omega)K(i\omega)$ is close to -1 at some frequency, then the coupled feedback system results in amplifying $d(s)$ at that frequency. The key take-away from this final observation is that $K(s)$ should be designed to manage the characteristics of the loop transfer function at frequencies where the magnitude of $G(i\omega)K(i\omega)$ is close to unity.

The frequency where the magnitude $\|G(i\omega)K(i\omega)\| = 1$ is called the *loop crossover frequency*, denoted ω_{gc}. This is approximately equal to the *bandwidth* of the system, which describes how rapidly the feedback loop responds to differences between the observed and desired states, with $1/\omega_{\text{gc}}$ being roughly the time-constant for system convergence (see Fig. 7). At this frequency, the distance from the point -1 in the complex plane can be characterized by the *phase margin*, defined as the difference between the phase of $G(i\omega_{\text{gc}})K(i\omega_{\text{gc}})$ and $-180°$, which we denote Φ_{pm}; this quantity approximately characterizes the closest distance between $G(i\omega)K(i\omega)$ and -1 for any frequency. Small phase margin implies a lack of robustness to uncertainty in the model $\mathbf{G}(s)$. Note that since the feedback operates on the observed (or simulated in our case) climate signal, it will act not only on the climate response to anthropogenic greenhouse gases, but also on natural climate variability. Small phase margin also implies high amplification of natural climate variability at frequencies near ω_{gc} (see Eq. 11), with oscillatory "ringing" in the time-domain response (Fig. 7; also see MacMartin et al., 2014b). Phase margin thus both gives an indication of how robust the system is to modeling errors and how much amplification there is of natural variability.

With proportional-integral control, the control gains k_i and k_p are design parameters; their choice is related to the bandwidth and the phase margin. A higher choice of bandwidth (faster convergence time) typically makes it more difficult to achieve a desired phase margin. Feedback design thus inherently involves trade-offs. Somewhat arbitrarily, we aim for a convergence time-constant of roughly 5 years, corresponding to $\omega_\mathrm{gc} \approx 0.2\,\mathrm{rad\,yr}^{-1}$, which is chosen as a reasonable trade-off to give fast enough convergence without excessive response to natural variability nor unacceptable robustness. We discuss the trade-offs in more detail at the end of this subsection.

We now outline a process for determining choices for k_p and k_i that yield convergence to the desired climate objectives for the system despite uncertainty in $G(s)$. For several choices of k_i and k_p, the Bode plots in Fig. 5 provide graphical representations of the frequency response of the loop transfer function, characterized by the magnitude (upper panel) and phase (lower panel) of the complex number $G(i\omega)K(i\omega)$ as a function of frequency ω. The loop crossover frequency $\omega_\mathrm{gc} \approx 0.2\,\mathrm{rad\,yr}^{-1}$ is indicated by dashed lines in Fig. 5.

First consider a pure integral control ($k_\mathrm{p} = 0$, black line in Fig. 5). At low frequencies ($\omega \ll 1$ rad yr^{-1}), pure integral control means that $G(s)K(s)$ is large, so the feedback loop results in good "performance" in the sense of the chosen variable meeting its specified objective. Higher values of k_i lead to higher bandwidth (larger values of ω_gc) and faster convergence (smaller values of $1/\omega_\mathrm{gc}$). However, the integral term adds 90° phase lag from the phase of $1/(i\omega)$. With pure integral control, the phase margin can thus be poor due to the combined phase lag from the time delay and the system dynamics (recall that the factor $\tau s + 1$ in the denominator leads to 90° phase lag at high frequencies).

Adding the proportional gain k_p (red line in Fig. 5) increases the phase margin. For example, from Eq. (12), choosing $k_\mathrm{p} = \tau k_\mathrm{i}$ results in

$$G(s)K(s) = k_\mathrm{i}\beta\frac{1}{s}e^{-Ds}. \tag{13}$$

With no delay ($D = 0$), this would have 90° phase margin, no amplification of natural climate variability, and a bandwidth $\omega_\mathrm{gc} = k_\mathrm{i}\beta$. As noted previously, to achieve a convergence time-constant of roughly 5 years, we choose $\omega_\mathrm{gc} \approx 0.2$ rad yr^{-1} and thus choose $k_\mathrm{i} = 0.2/\beta$. This

choice for k_i and k_p corresponds to the blue lines in Fig. 5. With one year time delay ($D = 1$), this gives a phase margin of $\pi/2 - \omega_\mathrm{gc} D$ or 79°. Decreasing the proportional gain ($k_\mathrm{p} < \tau k_\mathrm{i}$) would reduce the response to high-frequency climate variability, making the signals $u(t)$ less "noisy", but would also reduce the phase margin.

We now provide a more detailed recipe for determining control gains for a particular application. Let $M_\mathrm{gc} = ||G(i\omega_\mathrm{gc})||$ and $\Phi_\mathrm{gc} = \phi(G(i\omega_\mathrm{gc}))$ be the magnitude and phase of the system at frequency ω_gc rad yr^{-1} (note that $\Phi_\mathrm{gc} < 0$, as the output lags the input). The additional phase added by proportional-integral control at frequency ω_gc is

$$\Phi_p = \tan^{-1}\left(\frac{\mathrm{Im}(K(i\omega_\mathrm{gc}))}{\mathrm{Re}(K(i\omega_\mathrm{gc}))}\right) = \tan^{-1}\left(-\frac{k_\mathrm{i}}{\omega_\mathrm{gc} k_\mathrm{p}}\right) = -\frac{\pi}{2} + \tan^{-1}\left(\frac{\omega_\mathrm{gc} k_\mathrm{p}}{k_\mathrm{i}}\right). \tag{14}$$

For pure integral control ($k_\mathrm{p} = 0$), $\Phi_p = -\pi/2$, which is the previously discussed addition of 90° of phase lag from the integral term. Addition of a non-zero proportional gain adds phase lead to this term. Let Φ_pm be the desired phase margin (a choice). Then by definition, $\Phi_\mathrm{pm} = \Phi_\mathrm{gc} - \pi/2 + \tan^{-1}(\omega_\mathrm{gc} k_\mathrm{p}/k_\mathrm{i}) + \pi$. Then

$$k_\mathrm{p} = \frac{k_\mathrm{i}}{\omega_\mathrm{gc}} \tan\left(\Phi_\mathrm{pm} - \frac{\pi}{2} - \Phi_\mathrm{gc}\right). \tag{15}$$

We choose k_i such that the loop transfer function gain is unity at ω_gc, i.e., $1 = ||G(i\omega_\mathrm{gc})K(i\omega_\mathrm{gc})|| = M_\mathrm{gc}\sqrt{k_\mathrm{p}^2 + (k_\mathrm{i}/\omega_\mathrm{gc})^2}$. Solving, the desired value of k_i is then

$$k_\mathrm{i} = \frac{\omega_\mathrm{gc}}{M_\mathrm{gc}} \Bigg/ \sqrt{1 + \tan^2\left(\Phi_\mathrm{pm} - \frac{\pi}{2} - \Phi_\mathrm{gc}\right)}. \tag{16}$$

Then by Eq. (15), k_p is also determined.

Note that Eqs. (15) and (16) only require information about the magnitude and phase at the loop crossover frequency ω_gc. This means that we can design "system identification" simulations (Sect. 3.5) in our design model using a sinusoidal input signal at the desired crossover frequency to estimate the magnitude and phase of the input/output response at

just that single frequency. This is also the reason why the first-order linear model is a sufficient description of system dynamics for designing the feedback algorithm, as no assumption needs to be made regarding the dynamics at frequencies away from ω_{gc}. A semi-infinite diffusion model (for example) that has the same magnitude and phase at ω_{gc} will require the same feedback gains to achieve the same bandwidth and phase margin. As such, knowing the model form is not essential for designing the feedback algorithm.

However, the model form does influence characteristics such as amplification of natural variability at frequencies away from ω_{gc} and convergence behavior. Figure 6 shows the sensitivity functions of the first order linear and semi-infinite diffusion models, defined as

$$S(s) = \frac{1}{1 + G(s)K(s)}. \tag{17}$$

From Eq. (11), this is the ratio of the system response to disturbances with and without the feedback, and applies both to slow variations in anthropogenic radiative forcing for which geoengineering is intended to compensate, as well as natural variability as discussed by MacMartin et al. (2014b). At low frequencies, $||S(i\omega)|| \leq 1$, consistent with the feedback algorithm maintaining the desired climate outcome independent of slow changes in greenhouse-gas concentrations. At high frequencies, $G(i\omega)$ is small so $||S(i\omega)|| \approx 1$. In between, there will be frequencies where $||S(i\omega)|| > 1$, meaning natural variability is amplified at those frequencies. While understanding the system response at a single frequency is sufficient to design a feedback algorithm, knowledge of the system response across a wider range of frequencies would be requried to fully understand how the system would react to natural variability. The time-domain convergence characteristics can be obtained from the inverse Laplace transform of $S(s)$. The predicted response is shown in Fig. 7 for both the first-order and semi-infinite diffusion models, and is relatively similar, with e-folding convergence rate in both cases of roughly $1/\omega_{\text{gc}}$. Knowledge of the system response at this one frequency is thus sufficient for understanding the time-domain convergence in response to differences between the desired and actual climate outcomes. This figure illustrates how little information is actually needed about the system to enable design of a feedback strategy that converges. Of course, in any actual deployment, it would be preferable to estimate

the full frequency-dependent input/output response from climate models in order to fully characterize the expected behavior. Understanding both climate system natural variability and how the imposed geoengineering affects different modes of variability is particularly important for detection and attribution of the climate effects of geoengineering.

Tradeoffs between convergence timescale and amplification of natural variability are choices in designing a feedback algorithm. Higher bandwidth leads to faster convergence and tighter management of the specified climate objectives. However, at higher frequencies, the system response has greater phase lag (see Fig. 5), and thus a higher bandwidth makes it more difficult to achieve a desired phase margin. Typically, in engineering applications, a phase margin of 60° is considered sufficient to avoid excessive amplification of natural variability (this gives $||S(i\omega_{\mathsf{gc}})|| = 1$, though $||S(i\omega)||$ may exceed unity at other frequencies). The other reason for ensuring adequate phase margin is that the estimated dynamics of the input/output response in the design model may not match the actual dynamics (or here, the dynamics of the evaluation model). We also have additional error here in estimating the design model response because of the influence of natural variability for the relatively short simulations used. For these reasons, phase margins larger than 60° are useful. A third consideration not noted earlier is that of the response of the input signal u to natural variability:

$$y(s) = S(s)d(s)u(s) = -K(s)y(s) \quad \Rightarrow \quad u(s) = -K(s)S(s)d(s).$$

Noting that $S(i\omega)$ is always near unity at high frequencies, the response of the input signal to natural variability is determined by $K(s)$ at high frequencies. Using a proportional-integral controller, then at high frequencies, $K(s) \approx k_{\mathsf{p}}$. Hence, increasing k_{p} to improve phase margin comes at a cost of the resulting input signal responding to high-frequency natural variability, that is, a "noisy" year-to-year variation in the amount of geoengineering. We do not claim that our choices herein give the best trade-off between these various factors, although we have endeavored to choose reasonable values.

3.4 Multi-input, multi-output (MIMO) feedback design

The discussion thus far has focused on a single-input, single-output, or SISO feedback algorithm case. However, both of our design examples (Sect. 2) are multivariate, that is, $\boldsymbol{y}$ and $\boldsymbol{u}$ are vectors related by a matrix-valued transfer function $\mathbf{G}(s)$. In both examples, the dimension of $\boldsymbol{y}$ and $\boldsymbol{u}$ are the same so that $\mathbf{G}(s)$ is a square matrix. It is also essential that $\mathbf{G}(s)$ be of full-rank, as otherwise there would be no choice of input $\boldsymbol{u}$ that could simultaneously drive every output in $\boldsymbol{y}$ to its desired value.

If $\mathbf{G}(s)$ is diagonal, then the SISO feedback design approach above can be applied directly. In this case, the multivariate goal simply corresponds to a set of decoupled SISO problems where each input variable only influences a single output variable. If $\mathbf{G}(s)$ is diagonally-dominant, so that each input *mostly* influences only one corresponding output variable, then stability is still guaranteed even if the off-diagonal coupling is ignored. A third, more general case that is relevant to both of our design examples is where $\mathbf{G}(s)$ is approximately triangular. For example, while high-latitude Northern Hemisphere insolation reduction influences both Arctic temperature and the precipitation centroid, high-latitude Southern Hemisphere insolation reduction only has a significant influence on the precipitation centroid, and hence this input/output system is roughly triangular. Note that while all complex-valued square matrices are triangularizable (e.g., by Gaussian elimination), the transformation for arbitrary $\mathbf{G}(s)$ will in general be frequency-dependent, and the ability to triangularize $\mathbf{G}(s)$ may not necessarily be useful. If $\mathbf{G}(s)$ is not nearly triangular nor readily triangularizable, then more complicated feedback design approaches will be required than are described herein.

As an illustrative example of how to design a multivariate feedback algorithm, we consider the following 3×3 system in which the influence matrix $\mathbf{M}$ is triangular:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} m_{11} & 0 & 0 \\ m_{21} & m_{22} & 0 \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}. \tag{18}$$

Although notation is omitted, all entries in Eq. (18) are frequency-dependent. From this representation, we note that $y_1 = m_{11}u_1$, i.e., y_1 is only influenced by changes in a single input. Therefore, designing a feedback strategy to converge to a desired value for y_1 only requires adjustments of a single input, using observations of a single output (Sect. 3.3). Next, $y_2 = m_{21}u_1 + m_{22}u_2$. However, u_1 is already determined by the previous relationship. One could adjust u_2 only in response to changes in y_2, but this neglects the known information about the effect of u_1 on y_2. A better strategy is to choose ~~$u_2 = -(m_{21}/m_{22})u_1 + k(y_2)$~~ $u_2 = -(m_{21}/m_{22})u_1 - k(y_2)$ where the feedback function $k(y)$ is again a SISO relationship: u_2 both responds to observed changes in y_2 and "corrects" for anticipated changes in y_2 that are caused by u_1. Similarly, $y_3 = m_{31}u_1 + m_{32}u_2 + m_{33}u_3$, where u_1 and u_2 have already been determined. Therefore, the problem of Multiple-Input, Multiple-Output (MIMO) feedback can be reduced to a set of SISO algorithms. This procedure is called *sequential loop closure*.

3.5 System identification

As was mentioned in Sect. 3.1, the goal of *system identification* is to estimate the transfer function matrix $\mathbf{G}(s)$ that describes the linear frequency–dependent relationship between the vector of inputs $\boldsymbol{u}$ and vector of outputs $\boldsymbol{y}$ sufficiently well to design a feedback algorithm and characterize its expected behavior. As noted before, the form $\boldsymbol{y}(s) = \mathbf{G}(s)\boldsymbol{u}(s)$ assumes that any nonlinearities (higher-order terms in the Taylor expansion) are sufficiently small that they do not present significant difficulties for feedback convergence. The matrix $\mathbf{G}(s)$ is estimated in the design model by introducing a signal $\boldsymbol{u}(t)$ and observing $\boldsymbol{y}(t)$. This can be done separately for each input signal, and the estimated responses of the outputs

to those inputs is then determined. There are several possible choices of input signal that could be useful in characterizing the dynamic behaviour of the system.

A step input perturbation is quite common in climate science, e.g., the abrupt4xCO2 simulation in CMIP5 (Taylor et al., 2012) in which the CO_2 concentration is abruptly quadrupled from its preindustrial value. While these have been used to estimate system dynamics (e.g., Caldeira and Myhrvold, 2013), one limitation is that the input signal is heavily-weighted towards low-frequencies: the Laplace transform of a step input is $H(s) = 1/s$. While this input contains information about all frequencies, the signal-to-noise ratio can be poor at higher frequencies and may require averaging multiple ensemble members.

An alternative is to use single-frequency sinusoidal input signals, as was done by MacMynowski et al. (2011). Evaluating the response to a broad range of input frequencies can be computationally expensive. However, as was discussed in Sect. 3.3, a feedback algorithm can be designed with a characterization of the system response at a single frequency. If the system is approximately linear, then after transient system behavior subsides, the output $y(t)$ will also be sinusoidal at the same frequency. This typically requires two full periods to be simulated; using the example of $\omega_{\mathsf{gc}} = 0.2\,\mathsf{rad\,yr}^{-1}$ as in Sect. 3.3 would require a $(2)(2\pi/0.2) \approx 63\,\mathsf{year}$ simulation to characterize the system response at that single frequency. Then choosing gains k_{p} and k_{i} so that the desired loop crossover frequency and phase margin are obtained requires two pieces of information: the magnitude of y relative to the input u and the phase-shift between y and u. If a first-order response is assumed, the parameters β and τ in Eq. (7) can be determined, although this is not strictly necessary to design a feedback algorithm.

Another alternative is to input a band-limited signal, which is useful for characterizing system behavior over a small range of frequencies; this can be helpful if the different input–output relationships have different timescales of response. This method has an advantage over step-response simulations, in that the input signal is not heavily weighted toward some frequencies at the expense of others. This has a disadvantage as compared to sinusoidal inputs, in that the input signal is more distributed, resulting in lower signal-to-noise ratios.

If the loop crossover frequency falls within the quasi-static response of the system, then a sinusoidal input and a band-limited input will yield similar information.

Natural climate variability limits the accuracy of estimating the transfer function in simulations. Errors can be estimated from the frequency-dependent signal-to-noise ratio (SNR) denoted ~~$S(\omega)$, where S^2~~ $Z(\omega)$, where Z^2 is the variance due to the input divided by the variance that would have occurred without the input. The error in estimating transfer function magnitude and phase can be related to the SNR as

$$\begin{aligned} \sigma_M(i\omega) &\approx [1/Z(\omega)]\,||G(i\omega)|| \\ \sigma_\phi(i\omega) &\approx \tan^{-1}[1/Z(\omega)]\,. \end{aligned} \tag{19}$$

The SNR can be estimated from a control run with no input, or for a sufficiently long time-series can be estimated as in MacMartin and Tziperman (2014) from the coherence ~~$\gamma^2 = S^2/(1+S^2)$~~ $\gamma^2 = Z^2/(1+Z^2)$ (the fraction of the total output variance that is associated with the input). For single-sinusoid input signals used below, we estimate the SNR at the frequency of the input signal from the output variance averaged over nearby frequencies. With two full periods at ω_{gc} simulated, projecting the output time-series onto sinusoids at $\omega_{gc}/2$ and $3\omega_{gc}/2$ gives estimates for how large the output signal might have been in the absence of the input signal.

3.6 2×2 design example

Our characterization of the 2×2 system begins with a series of step response simulations and is followed by a set of sinusoidal response simulations. This is sufficient for us to design a control algorithm that sufficiently meets the prescribed climate objectives.

For the step response simulations, beginning from a stable preindustrial control run, insolation over the Arctic or Antarctic was abruptly reduced by 2, 4, 8, and 12 %; the results from these simulations are summarized in Fig. 8. The choice of 12 % was informed by simulations performed by MacCracken et al. (2013), and lower magnitudes were chosen to test linearity of the climate response and the noise threshold. Linearity is illustrated in Fig. 9, showing that the precipitation centroid (Eq. 1) is a robust metric for our purposes.

These simulations can already inform the influence matrix for this particular case. Reductions in Arctic insolation reduce Arctic temperature and shift tropical precipitation southward. Reductions in Antarctic insolation shift tropical precipitation northward but do not discernibly affect Arctic temperatures. Therefore, using notation to suppress any potential time or frequency dependence, we can write the influence matrix as

$$\begin{bmatrix} T_{\text{Arctic}} \\ \chi \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ -\xi & \eta \end{bmatrix} \begin{bmatrix} S_{\text{Arctic}} \\ S_{\text{Antarctic}} \end{bmatrix} \tag{20}$$

where T_{Arctic} denotes Arctic temperature change, χ denotes shifts in the meridional-centroid of zonal-mean precipitation (Eq. 1; positive northward), S_{Arctic} denotes adjustments in Arctic insolation, $S_{\text{Antarctic}}$ denotes adjustments in Antarctic insolation, and λ, ξ, and η are positive functions of, as of yet, undetermined form. As noted above, this particular system is inherently triangular.

The step-response results can be fit to the functional form in Eq. (6). However, the exponential fits in Fig. 8 are relatively poor and can only be done for the highest amplitude of step response, partly due to a large amount of high frequency variability. This is an inherent problem with step response simulations, as the input signal contains nonzero content at all frequencies. Nevertheless, they are useful in that they provide confidence that a first order linear model captures the system behavior sufficiently to design a control algorithm.

To circumvent these shortcomings, it is useful to complement the step response information with sinusoidal input signals. Beginning from a preindustrial control run, insolation over the Arctic or Antarctic was reduced according to the function $u(t) = 0.12 \sin(2\pi/\omega t)$. The amplitude 0.12 was chosen simply because the step response with a 12 % amplitude appeared to give a good signal, and the step response did not show any evidence of substantial nonlinearity.

Figure 10 shows the sinusoidal response of the system for an input signal with a period of 10π years, corresponding to $\omega = 0.2\,\text{rad}\,\text{yr}^{-1}$. From these simulations, the influence matrix can be computed via the amplitude ratio (the gain ratio of the output signal to the input signal) and the phase shift (the difference in phase lag between the output and input

signals). If $\mathbf{G}(s)$ is the 2×2 transfer function representing the input–output relationships in the system, then the results from the sinusoidal input simulations give

$$\|\mathbf{G}(0.2i)\| = \begin{bmatrix} 0.447\,^\circ\mathrm{C}/\% & 0.047\,^\circ\mathrm{C}/\% \\ 0.030\,^\circ/\% & 0.028\,^\circ/\% \end{bmatrix} \qquad \phi(\mathbf{G}(0.2i)) = \begin{bmatrix} 27^\circ & -30^\circ \\ 49^\circ & 38^\circ \end{bmatrix}. \tag{21}$$

Consistent with physical understanding of the system, the top-right entry of the magnitude matrix is small and is ignored for the purpose of designing a feedback algorithm. Assuming the first-order linear model described by Eq. (7), also note that part of the phase lag is due to the inherent dynamics of the system, and part is due to the half-year time delay introduced by annual averaging. At this frequency, $D = 0.5$ introduces a phase lag of approximately 6°, which is incorporated into the estimates of phase lag in Eq. (21). While the system identification simulation varies the input continuously and only introduces $D = 0.5$, our feedback implementations update the input and hold it constant for the following year, introducing another half-year delay and an extra 6° phase lag at $\omega = 0.2\,\mathrm{rad\,yr}^{-1}$ on top of the estimate in Eq. (21).

From visual inspection of Figs. 10 and 11, it is clear that climate system noise can result in errors in the sinusoidal fits, which can introduce errors into estimates of the transfer functions. MacMartin and Tziperman (2014) discuss how to calculate the estimation error in the transfer function; we repeat the salient equations here.

Based on calculations of SNR (Eq. 19), the standard deviations of the estimation error in G at the loop crossover frequency $\omega_{\mathrm{gc}} = 0.2\,\mathrm{rad\,yr}^{-1}$ are

$$\boldsymbol{\sigma}_M(0.2i) = \begin{bmatrix} 0.029 & 0.027 \\ 0.009 & 0.007 \end{bmatrix} \qquad \boldsymbol{\sigma}_\phi(G(0.2i)) = \begin{bmatrix} 4^\circ & 30^\circ \\ 17^\circ & 13^\circ \end{bmatrix}. \tag{22}$$

One standard deviation of the errors in magnitude are between 7 and 31 % of the values given in Eq. (21) (excluding the top right entry), and one standard deviation of the errors in phase are between 13 and 25 % (with the exception of the top right entry). Kravitz et al. (2014) showed that the feedback algorithm considered there was robust to at least 50 % error in magnitude, and we choose a sufficient phase margin below to accommodate the

maximum phase error of $\sim 17°$. Therefore, we conclude that any potential errors in the fits are unlikely to substantially affect the feedback design.

From Eq. (21), the time-constant β and the timescale τ in Eq. (7) can be computed. For example, the relationship between S_{Arctic} and T_{Arctic} can be described by $\beta = 0.447$ and $\tau = 1.946$. These values differ somewhat from the best-fit values obtained from the step response simulations, potentially due to difficulties with the step-response system identification introduced by high frequency variability, or due to the first-order linear model not adequately describing the dynamics. In principle, additional simulations could improve estimates of these parameters, but as we showed in Sect. 3.3, this is not necessary for designing a successful feedback algorithm.

Following the procedure described in the previous sections, we first choose SISO feedback gains to adjust high-latitude Northern Hemisphere forcing in response to deviation in Arctic temperature from the desired value. Choosing $k_{\text{p}}/k_{\text{i}} = 2.5$ adds 27° phase lead from the proportional term ($\tan^{-1}(2.5 \times 0.2)$; Eq. 16). ($k_{\text{p}}/k_{\text{i}} = 2.5$ is approximately the same value that one would obtain Equation 15, but it is not identical. Changing k_{p} changes the phase margin; to some extent, these are design choices.) Then by Eq. (15), choosing $k_{\text{i}} = 0.4$ gives the desired crossover frequency of $\omega_{\text{gc}} = 0.2\,\text{rad}\,\text{yr}^{-1}$. These choices result in a phase margin of 84°.

As described in Sect. 3.4, the high-latitude Southern-hemisphere forcing $S_{\text{Antarctic}}$ can be adjusted both in response to changes in the precipitation centroid χ and in response to the expected change in χ due to S_{Arctic}. We again choose $k_{\text{p}}/k_{\text{i}} = 2.5$, although because the system itself has slightly greater phase lag than the Arctic temperature response to Arctic insolation changes, this choice will lead to lower phase margin. The value of k_{i} that yields $\omega_{\text{gc}} \approx 0.2\,\text{rad}\,\text{yr}^{-1}$ is 6.4, which we round to $k_{\text{i}} = 6$; then $k_{\text{p}} = 15$. Although better performance would be achieved by also adjusting $S_{\text{Antarctic}}$ in direct response to changes in S_{Arctic}, we neglect this here, as the adjustment only in response to changes in χ results in acceptable performance.

Thus in summary, we have

$$S_{\text{Arctic}} = 0.4 \int_0^t (T_{\text{Arctic,ref}} - T_{\text{Arctic}})\,\mathrm{d}t + (T_{\text{Arctic,ref}} - T_{\text{Arctic}}) \tag{23}$$

$$S_{\text{Antarctic}} = 6 \int_0^t (\chi_{\text{ref}} - \chi)\,\mathrm{d}t + 15(\chi_{\text{ref}} - \chi). \tag{24}$$

(Although it is abuse of notation, integrals are used instead of sums for clarity.)

3.7 3×3 design example

We now consider system identification and feedback design for the 3-input, 3-output design example described in Sect. 2, where the inputs are L_0, L_1, and L_2 patterns of solar reduction (Fig. 2b), and the outputs are the corresponding projections of zonal-mean temperature: global-mean (T_0), a linear dependence on sine of latitude that captures interhemispheric asymmetry (T_1), and a quadratic dependence on sine of latitude that captures equator-to-pole temperature gradients (T_2). The first (SISO) entry in this 3×3 problem is the same input/output system for which feedback was designed in earlier work (MacMartin et al., 2014b; Kravitz et al., 2014), based on an extensive frequency-domain system identification of the HadCM3L general circulation model (MacMynowski et al., 2011).

Similarly to Eq. 20, we can write the influence matrix for the 3×3 problem as

$$\begin{bmatrix} T_0 \\ T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} m_{00} & 0 & 0 \\ m_{10} & m_{11} & 0 \\ m_{20} & m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} L_0 \\ L_1 \\ L_2 \end{bmatrix} \tag{25}$$

This system is also inherently triangular.

We characterize the system response solely through sinusoidal input signals, as shown in Fig. 11. As in the 2×2 design example, the system is naturally nearly-triangular. A globally uniform solar reduction leads to changes in all three output measures – global mean

temperature, as well as the linear and quadratic meridional dependence. A (zero-mean) solar reduction that is linear with the sine of latitude has minimal influence on global mean temperature, but influences both the linear and quadratic terms in the zonal-mean temperature. As in the 2×2 example in Sect. 3.6, there is an implicit 6-month time-delay in the plotted simulation output from the annual-averaging, corresponding to a 6° phase delay at this frequency.

The best estimate of the magnitude and phase of the input/output response at $\omega = 0.2\,\mathrm{rad\,yr}^{-1}$ is

$$\|\mathbf{G}(0.2i)\| = \begin{bmatrix} 0.66 & - & - \\ 0.13 & 0.15 & - \\ 0.07 & 0.03 & 0.03 \end{bmatrix} \qquad \phi(\mathbf{G}(0.2i)) = \begin{bmatrix} 38 & - & - \\ 22 & 20 & - \\ 60 & -12 & 10 \end{bmatrix}. \tag{26}$$

where dashes in the matrices indicate that the estimate is indistinguishable from error (see Eq. 27 below) and does not have a strong physical connection.

The phase estimates include a half year of time delay due to annual averaging. Climate variability clearly introduces uncertainty in these estimates, particularly for the small elements (Fig. 11); as such, the bottom-middle entry of the phase lag matrix is taken to be zero. The upper triangular entries of the transfer function matrix are indistinguishable from zero, consistent with physical understanding of the system, and are left blank. Note from Fig. 1 in MacMartin et al. (2014b) that at this frequency, the corresponding $(1,1)$ entry for the HadCM3L general circulation model had a gain of 0.6 and a phase of 30° (36° if including a half-year time delay), which is very similar to the results here.

Performing the same error calculations as in Sect. 3.6 yields Eq. (27):

$$\boldsymbol{\sigma}_M(0.2i) = \begin{bmatrix} 0.030 & - & - \\ 0.039 & 0.022 & - \\ 0.016 & 0.013 & 0.018 \end{bmatrix} \qquad \boldsymbol{\sigma}_\phi(0.2i) = \begin{bmatrix} 3 & - & - \\ 17 & 8 & - \\ 13 & 24 & 30 \end{bmatrix}. \tag{27}$$

The errors in magnitude (1σ) for the lower triangle are $\sim 4\,\%$ for the 1×1 sub-case, up to 30 % for the 2×2 sub-case, and up to 59 % for the full 3×3 case. As in the 2×2 case, no

error in Eq. (27) is going to substantially impact the performance of the feedback design, although the large phase uncertainty in the final entry leads us to choose a larger phase margin than we might otherwise.

We first choose feedback gains to adjust globally-uniform solar reduction to maintain global-mean temperature, corresponding to the $(1,1)$ entry of the system dynamics matrix. Again, note that there is a one-year time-delay introduced by averaging over the previous year before making a decision and holding that decision fixed for an entire year; at a frequency of $0.2\,\mathrm{rad\,yr}^{-1}$, this yields a phase lag of approximately $11.4°$, only half of which is included in Eq. (26). Thus with zero proportional gain, choosing $k_\mathrm{i} = 0.2/0.7$ (after rounding) would give a loop crossover frequency of $0.2\,\mathrm{rad\,yr}^{-1}$ and a 48° phase margin. This low phase margin would yield significant amplification of natural variability at high frequencies and could lead to additional problems if the evaluation model dynamics are not the same as those of the design model. A proportional gain $k_\mathrm{p} = k_\mathrm{i}$ adds 11.3° degrees phase at frequency $\omega = 0.2$ ($\tan^{-1}(0.2)$), for a total phase margin of roughly 60°. (The phase margin, and hence k_p, is a design choice; as noted earlier, 60° is a reasonable choice.) Decreasing k_i by the factor $\sqrt{1+(1\times 0.2)^2}$ compensates for the increase in gain at $0.2\,\mathrm{rad\,yr}^{-1}$ due to the proportional gain, and thus maintains the desired loop crossover frequency of $0.2\,\mathrm{rad\,yr}^{-1}$. Thus, $k_\mathrm{i} = 0.2/0.7/\sqrt{1+(1\times 0.2)^2} \approx 0.28$.

If the system were diagonal, the additional degrees of freedom could be similarly adjusted with just a rescaling of both k_i and k_p by the inverse of the diagonal elements of $\|\mathbf{G}(0.2i)\|$; this would maintain the same loop cross-over frequency for each degree of freedom, and the expected phase margin would be slightly higher for the remaining degrees of freedom. While this approach would converge, better performance can be achieved by using the knowledge of the coupling described by the off-diagonal elements, as described in Sect. 3.4. (Note that the sign of the effect of a uniform solar reduction on all three degrees of freedom is confident from physical principles; the influence of L_1 on T_2 is less obvious.) Thus, the Multiple-Input,

Multiple-Output (MIMO) feedback design for this problem can be summarized as:

$$\Delta L_0 = 0.28 \int_0^t (T_0 - T_{0,\text{ref}})\,\mathrm{d}t + 0.28\,(T_0 - T_{0,\text{ref}}) \tag{28}$$

$$\Delta L_1 = -\Delta L_0 + 1.3 \int_0^t (T_1 - T_{1,\text{ref}})\,\mathrm{d}t + 1.3\,(T_1 - T_{1,\text{ref}}) \tag{29}$$

$$\Delta L_2 = -0.6\Delta L_1 - 1.4\Delta L_0 + 3.9 \int_0^t (T_2 - T_{2,\text{ref}})\,\mathrm{d}t + 3.9\,(T_2 - T_{2,\text{ref}}). \tag{30}$$

(As in Eqs. 23 and 24, integrals are used instead of sums for clarity.)

4 Results from the 2×2 case

We now proceed with an evaluation of the effectiveness of our designed feedback algorithms.

Figure 12 shows results for the 2×2 case in the design model (CESM) with no feedback (1pctCO2; black lines), only adjusting Arctic insolation to modify Arctic temperature (abbreviated "Arctic Only"; blue lines), and the full 2×2 case (red lines). The feedback algorithm does an excellent job of meeting the specified climate objectives, with total root-mean-square (RMS) differences from the objectives given in Table 1. Because the feedback algorithm adjusts every year, this strategy is not designed to remove one-year timescale deviations from the objectives. Arctic insolation reductions in both the Arctic Only case and the full 2×2 case are approximately linear with CO_2 forcing and reach approximately 14 % by the end of the 70 year simulation, a similar magnitude to that used in the system identification simulations. In the Arctic Only case, the precipitation centroid χ shifts southward relative to the 1pctCO2 simulation, as expected, but does not return to the baseline value.

It is not obvious a priori whether the amount of Arctic insolation reduction that returns Arctic temperature would over- or under-compensate the CO_2-induced shift in the precipitation centroid, and indeed the two models used here show different behavior in this respect.

Because of the net northward shift with only Arctic insolation reductions, bringing the precipitation centroid southward actually requires an increase in Antarctic insolation in this model. (The feedback algorithm was not given any information regarding feasibility of the applied radiative forcing; there is no known method of modifying shortwave radiation between 60°N and 90°N, let alone how to increase downward radiative flux in this region.) As might be expected from the results in Fig. 10, the magnitude of increase in Antarctic insolation in any particular year is on average greater than the magnitude of decrease in Arctic insolation. This is clearly not representative of choices that would be made in an actual geoengineering implementation, but serves as a useful demonstration of multivariable feedback in part because this behavior is model-specific. The GISS results below show that the effectiveness of the feedback algorithm in this case is not dependent on whether Arctic-only insolation over- or under-compensates the CO_2-shift in precipitation.

Figure 12 illustrates that (in the design model) the feedback algorithm works as designed, meeting the objectives as specified. However, it is valuable to explore the resulting climate in more detail, as it informs the complexity of defining objectives for geonengineering. Figure 13 provides more spatial detail for the results in Fig. 12. The 1pctCO2 simulation results in widespread warming, with temperature amplification at high latitudes and an increase in global precipitation. In the Arctic Only simulation, most of the land mass in the Arctic remains slightly warmer than in the preindustrial control run, and the ocean regions are cooled, resulting in no average warming over the Arctic region. Tropical precipitation is shifted southward as compared to the 1pctCO2 case. (See Appendix A for mechanistic explanations of tropical precipitation shifts.) In the full 2×2 simulation, the Arctic is cooled, again with a land-ocean contrast, and the Antarctic is warmed more; these results are consistent with those of Fig. 12. Tropical precipitation is shifted farther South than in the Arctic Only simulation and is slightly strengthened, but there is substantial drying north of the equator relative to baseline. Overall, although the feedback algorithm is effective at meeting

the specified objectives, there are residual changes in precipitation (Fig. 13) for which the design does not account. Additional degrees of freedom would be required to offset these local changes as well, assuming there exist such degrees of freedom.

Figure 14 shows changes in the seasonal cycle of precipitation and the centroid χ for the design model. All simulations show an increase in tropical precipitation, which is consistent with increased CO_2 concentration (e.g., Held and Soden, 2006). Precipitation patterns in the 1pctCO2 simulation show an increase in Northern Hemisphere precipitation in months commonly associated with the summer monsoon, consistent with understood mechanisms governing monsoon changes (e.g., May, 2004). The centroid χ is shifted northward as compared to the preindustrial climatology in nearly all months (with the exception of boreal late spring), especially in the boreal winter when the ITCZ is at its most southward position. In the Arctic Only simulation, the position of χ is restored quite well except in the boreal winter, which may be expected, as there is essentially no change in Arctic insolation during the polar winter. There is also a reduction in boreal winter/spring precipitation in the Northern Hemisphere subtropics due to a decrease in precipitable water (not shown). The full 2×2 simulation shows that even though the mean position of χ is restored, the seasonal cycle is not, with precipitation slightly too far north in boreal winter and too far South in boreal summer. Other than the increase in tropical precipitation, the only large anomaly in precipitation is in the Southern Hemisphere subtropics during austral autumn, consistent with a warmer Southern Hemisphere and enhanced Australian monsoon precipitation (Lau and Wu, 1999). According to these results, restoration of the mean position of tropical precipitation would not require the same feedback design as restoration of the seasonal cycle of tropical precipitation.

Given sufficient simulation time, the results above with the design model could have been achieved without feedback simply by estimating the model sensitivities to forcing from CO_2 and both patterns of solar reduction, computing the amount of each pattern that would achieve the objectives, and conducting multiple simulations if necessary to get the correct answer. However, while that approach would demonstrate what might be achievable with perfect knowledge, it would not demonstrate a viable implementable strategy. The true

power of using feedback to adjust the input degrees of freedom is demonstrated by using the exact same algorithm to also achieve acceptable performance in the evaluation model which, up to this point, has not been exercised at all in developing the strategy. Figure 15 shows the results from implementation of the feedback design in GISS ModelE2, the "evaluation" model. The magnitude of change of Arctic temperature and shift in χ are notably smaller for 1pctCO2 than in the design model, consistent with previous evaluations of the differences in general behavior between the two models (Kravitz et al., 2014). Regardless, the feedback algorithm still does an excellent job at meeting the objectives; RMS differences are listed in Table 1 next to the results from CESM. The Arctic-Only simulation shows a more southward value of χ than in the 1pctCO2 simulation, similar to CESM; the full 2×2 case results in better performance on this objective than the Arctic Only case.

The required reduction in Arctic insolation to achieve the Arctic temperature goal is approximately 7 %, or about half of the required value for CESM. Unlike the design model, achieving the goal for χ requires a reduction in Antarctic insolation (Arctic-only reductions very slightly over-compensate rather than under-compensate the centroid shift due to CO_2 alone). This result indicates that as long as the sign of the response is understood (i.e., that insolation reduction in one hemisphere will tend to shift tropical precipitation away from that hemisphere), the feedback algorithm is robust to substantial uncertainties in the details of the response, which is indeed the entire point of using feedback. However, as the results in Fig. 16 show, the residual climate effects may differ, depending upon the different model-dependent spatial patterns of response to forcing. Accounting for the residuals would require modifying additional degrees of freedom that are known to modify the temperature and precipitation patterns in Fig. 16. This may or may not be possible; there is likely a practical limit (which has not yet been discovered) to what is achievable by geoengineering.

Like the results for CESM (Fig. 13), the 1pctCO2 simulation in GISS results in widespread warming, with Arctic amplification and an acceleration of the hydrological cycle. The Arctic Only simulation is similarly effective at reducing Arctic temperature change, although with many of the changes over land as well as ocean. In the full 2×2 case, both poles are cooled,

consistent with the changes in insolation in Fig. 15. Tropical precipitation is enhanced in all three simulations, consistent with an increase in CO_2.

Figure 15 indicates that the change in the precipitation centroid χ is fairly well compensated in the Arctic Only case, suggesting that the reductions in Antarctic insolation are not strictly necessary to achieve the specified objectives. Moreover, Antarctic insolation reduction reaches 16 % by the end of the simulations, whereas Arctic insolation reduction reaches only 7 %. (Explanations of these results are provided in Appendix A.)

Overall, we have demonstrated the ability to successfully design a 2×2 feedback algorithm for the case we have investigated here. In doing so, we met all four criteria outlined in the introduction, including a multi-model assessment of the feedback algorithm, demonstrating that the designed algorithm is robust to inter-model differences. In the next section, we follow the same investigations for the 3×3 design case.

5 Results from the 3×3 case

Figure 17 shows results in the design model for the 3×3 case, where 1×1 (black lines) indicate only modifications of L_0 to offset changes in T_0, 2×2 (blue lines) indicate modifications of L_0 and L_1 to offset changes in T_0 and T_1, and 3×3 (red lines) indicates the full 3×3 case as described in Sect. 2. Implementation of feedback in CESM shows excellent performance for the objectives being managed in each of these cases, and relatively poorer performance for any objectives not being managed in a particular case (e.g., ΔT_1 for the 1×1 case). RMS values of departures from the specified goals are given in Table 2.

Reductions in L_0 (a uniform insolation reduction) increase in magnitude approximately linearly with CO_2 concentration, which is consistent with previous results (the 1×1 case is effectively the same as GeoMIP experiment G2; Kravitz et al., 2011; Jones et al., 2013). The gradual reduction in L_0 keeps global mean temperature roughly constant, but cooling tends to be greater in the Northern Hemisphere than in the Southern Hemisphere. Maintaining the interhemispheric temperature gradient in the 2×2 simulation requires an increase in L_1 in this model, which increases Northern Hemisphere insolation and decreases South-

ern Hemisphere insolation relative to applying L_0 alone (i.e., less reduction in Northern Hemisphere insolation and more in the south; see Fig. 18). Even with these two patterns of change, the poles remain warmer than the equator, requiring an insolation reduction in L_2 (Fig. 19).

In Fig. 17, the 2×2 sub-case and the full 3×3 case have an initial increase in L_1, followed by a slow asymptote toward no net change in L_1. This indicates that changes in L_0 and L_1 primarily affect processes on two different timescales. Initially, the interhemispheric temperature gradient is driven by processes associated with a land-sea contrast, in large part because the Northern Hemisphere has more land than the Southern Hemisphere. After a few years, the land-ocean temperature contrast remains relatively constant (Lambert et al., 2011), and a large driver of interhemispheric temperature gradient is Arctic amplification (Holland and Bitz, 2003). These different timescales are reflected in the results of Fig. 17. The value of ΔL_1 reaches a maximum after 9–10 years; the associated e-folding timescale is therefore 2–3 years, which is consistent with known timescales of land surface feedbacks (Andrews et al., 2009). After this time period, interhemispheric temperature differences are largely due to greater Arctic temperature increases than Antarctic increases. These differences are effectively suppressed by reductions in L_0 and L_2, so smaller changes in L_1 are needed to restore T_1 to its objective. Reductions in L_0 and L_2 would be effective at offsetting changes in T_1 early in the simulation as well, but such modifications would also result in departures of T_0 and T_2 from their respective objectives.

Figure 19 shows the effectiveness of the different patterns of solar reduction on the pattern of temperature changes. As has been discussed previously (e.g., Kravitz et al., 2013a), an increase in CO_2 causes warming everywhere with polar amplification, and a decrease in L_0 will result in "overcooling" of the tropics and "undercooling" of the poles, with more residual temperature change in the Arctic than the Antarctic. The 2×2 case still has overcooling of the tropics and undercooling of the poles, but the temperature residuals at the poles have smaller interhemispheric disparity. In the full 3×3 case, these residuals are reduced substantially on average, although the results in Fig. 19 show differential effects over land and ocean. In principle, additional degrees of freedom might be included to correct

for additional residuals (again, assuming such degrees of freedom exist), but using three degrees of freedom is quite effective at removing most large temperature changes due to CO_2 increases.

Figure 18 shows results for CESM that are consistent with the above descriptions. For the 1×1 sub-case, because the CO_2 concentration gradually increases over the course of the simulation, insolation reduction must also increase to offset global mean temperature changes.

Early in the 2×2 sub-case, CO_2 warming calls for a reduction in L_0, which results in negative values of T_1 in CESM, i.e., Northern Hemisphere cooling is greater than Southern Hemisphere cooling, whereas countervailing Arctic amplification from increased CO_2 is not yet large enough to offset this cooling pattern. The net effect is a change in the inter-hemispheric temperature gradient, resulting in a negative value of T_1, so L_1 must increase to compensate. Thus, early in the 2×2 sub-case, there is a net reduction in insolation in the Southern Hemisphere, and net changes are small in the Northern Hemisphere. Later in the simulation, both CO_2 increases and L_0 reductions are larger in magnitude. As has been seen in previous simulations of geoengineering (e.g., Kravitz et al., 2013a), this combination of forcing results in greater net warming in the Northern Hemisphere than in the Southern Hemisphere, so the previously seen increase in L_1 is less than at the beginning of the simulation.

For the full 3×3 case, the net polar warming and tropical cooling that occurs in the 2×2 sub-case leads to changes in T_2, which are compensated by reductions in L_2. An increase in L_1 results in Arctic warming and Antarctic cooling, but the Arctic warming is amplified by the mechanisms involved in Arctic amplification, resulting in a net increase in T_2, requiring a decrease in L_2 to compensate. As can be seen in the results for GISS ModelE2, some of the net effects are model-dependent, i.e., it is not obvious whether a CO_2 increase and an L_0 decrease will result in positive or negative T_1.

Because the objectives of the 3×3 design case were framed solely in terms of temperature, it might be expected that changes due to CO_2 in other fields would not be compensated as well by this particular design. Figure 20 shows these residuals for precipitation changes.

In all three geoengineering cases, precipitation residuals are strongly reduced as compared to the 1pctCO2 simulation, consistent with previous results (e.g., Tilmes et al., 2013). Table 3 shows changes in global mean precipitation and χ as compared to the preindustrial control simulation. All feedback sub-cases show a decrease in global mean precipitation, overcompensating the increase caused by CO_2, as well as a southward shift in χ relative to 1pctCO2 that is not enough to compensate for the northward shift due to increased CO_2.

All of the GISS results (Fig. 21) show that the T_0 goal is met quite well by all three sub-cases. In GISS ModelE2, reductions in L_0 also result in small net changes in T_1, so in the 2×2 sub-case and the full 3×3 case, the feedback algorithm does not call for large changes in L_1. As can be seen in Fig. 22, by the end of the simulation, there is a slight equator-to-pole temperature difference, for which the feedback algorithm compensates by calling for reductions in L_2. Figure 22 shows that the full 3×3 case is quite effective at offsetting many temperature changes throughout the globe. The residual precipitation changes (Fig. 23) look qualitatively similar to those of the CESM simulations. Global mean changes in precipitation are of the same sign and similar magnitude to the CESM results. The northward shift in χ due to CO_2 is smaller than in CESM, and the feedback here slightly overcompensates rather than undercompensates this shift.

6 Discussion and conclusions

Geoengineering is not a binary decision of “on” or “off”. Rather, if it is ever deployed, multiple separate degrees of freedom could be adjusted to simultaneously meet multiple objectives. Climate models can be used to predict the response of multiple “output” variables in response to multiple “input” variables, but the actual climate response will not be identical. For this reason, the radiative forcing introduced by geoengineering would need to be adjusted in response to the observed climate outcomes; this feedback process compensates for uncertainty between models and reality. Here we have demonstrated this design process, and in particular the ability to simultaneously adjust multiple patterns of radiative forcing in response to multiple observed climate variables. Using a two-model approach

with separate design and evaluation models is essential for demonstrating that the feedback process results in a strategy that is not overly dependent on the specific details of an individual model but is instead robust across models.

We reiterate two key points. First, attempts to generically characterize the climate effects of solar geoengineering are ill-posed, because these effects depend both upon the specific technology used and the objectives. There is a broad range of potentially achievable climates, each with its associated impacts on society (such as effects on water scarcity or agriculture). Second, by demonstrating a multivariable feedback strategy to adjust multiple distinct spatial patterns of radiative forcing, and demonstrating that a strategy designed in one model can meet defined objectives in a separate evaluation model, this work reinforces previous research, suggesting that an accurate climate model is not necessarily required to implement solar geoengineering (Kravitz et al., 2014), even when balancing multiple climate objectives.

As we stated in Sect. 1, determining the objectives of the solar geoengineering efforts is an important first step. In our examples, we chose straightforward, unambiguous objectives, such as returning some aspects of climate back to a preindustrial baseline state. As was noted in Sections 4 and 5, each case had residuals for which the feedback algorithm did not control (e.g., the seasonal position of the precipitation centroid in the 2×2 case or any precipitation pattern in the 3×3 case). These effects are somewhat independent of the objectives for which we did control, so modifying them would require additional degrees of freedom, assuming such degrees of freedom could be found. Furthermore, if we had performed the 3×3 design case against (for example) an RCP8.5 scenario in which we were attempting to prevent global mean temperature change from exceeding 2 °C above its preindustrial value, there is flexibility as to what the goals of the other two design criteria (L_1 and L_2) ought to be. One potential goal would be to maintain whatever temperature pattern there was in 2020. Another would be to cut the warming rate in half in the Arctic. There are numerous other potential specifications, each with potentially different feedback algorithm designs; carefully specifying the problem to be solved is crucial.

There are two obvious directions for future research.

First, what are the limits to such a strategy? We have intentionally chosen a small number of objectives and chosen corresponding input variables where the physical relationship between inputs and outputs is well understood, so the input/output response is likely to be similar between different models, as well as between models and reality. Increasing the number of adjusted patterns of radiative forcing and the number of different climate objectives is likely at some point to be limited by uncertainty. Put more bluntly, one cannot necessarily control 100 different climate fields in 100 regions just because a model says it's possible. While feedback provides robustness, some knowledge is required about the input/output dynamics; if not even the sign of the relationship is known, for example, then it is challenging to design an algorithm that converges. This is where the role of clear physical mechanisms becomes crucial: in the absence of mechanisms, it is not known whether any discovered input/output relationships are robust on the timescales of interest, or if a mechanism is known to have highly nonlinear behavior, linear feedback may not be effective even with large expenditure of effort on feedback design. Furthermore, even if some complicated strategy converged to a slightly better solution than a simpler one, natural variability may limit the ability to detect that difference on societally-relevant timescales, let alone attribute those changes to geoengineering. Understanding the boundaries of what is achievable, as well as what robust conclusions can be obtained about any particular strategy, are open questions that require further research.

Second, we have demonstrated the ability to simultaneously manage multiple climate criteria using the common approach of changing solar irradiance, here as a function of latitude. Accomplishing the objectives with physically achievable mechanisms, such as with stratospheric aerosols or marine cloud brightening, introduces additional complications~~.~~, even beyond the example shown in Fig. 12, where meeting the objectives required an increase in Antarctic insolation (i.e., it may not be possible to achieve all objectives due to physical constraints). For example, in the case of stratospheric aerosols, one could choose both the latitude and altitude of injection. However, (i) this does not give arbitrary ability to influence the resulting latitudinal dependence of aerosol optical depth or radiative forcing, (ii) the resulting radiative forcing patterns cannot be adjusted instantaneously, and (iii) the rela-

tionship between injection parameters and spatial patterns of radiative forcing introduces additional uncertainty, in no small part due to model-dependent results and insufficient validation of models as compared to reality. Using our methodology with stratospheric aerosols requires two distinct steps. One is to characterize the relationship between injection parameters (e.g., altitude, latitude, season) and distributions of aerosol radiative forcing. The second is determining the relationship between that radiative forcing and climate effects. Each of these steps has substantial uncertainties, and overcoming these uncertainties to meet climate objectives by using stratospheric aerosols (again, assuming those objectives are even achievable, independent of the ability of feedback to meet those objectives) would require a separate feedback process for each step. Marine cloud brightening would introduce further challenges and opportunities from the spatial heterogeneity of radiative forcing in both latitude and longitude, as well as the potentially rapid temporal response. This is intimately tied to the above mentioned area of research: feedback is essential for managing some of these uncertainties, but there are limits to what feedback can achieve.

Appendix A

The position of the ITCZ, a large determining factor in the position of χ (Eq. 1), is effectively determined by the magnitude of cross-equatorial atmospheric energy transport (Kang et al., 2008). Following the discussion of Donohoe et al. (2013), cross-equatorial atmospheric energy transport (AT) is

$$\mathrm{AT} = -[B] - [A] \tag{A1}$$

where B is the total flux of energy into the atmospheric column (including all net top-of-atmosphere radiative components, surface radiative components, and turbulent components), A is the net storage of energy in the atmosphere, and brackets indicate a spatial integral over the Northern Hemisphere. A is the time derivative of column-integrated moist

static energy, i.e.,

$$A = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{g}\int_0^{p_\mathrm{s}} c_p T + L_\mathrm{v} q \,\mathrm{d}P\right] \tag{A2}$$

where p_s denotes surface pressure, P is pressure (Pa), $g = 9.81\,\mathrm{m\,s^{-2}}$ is acceleration due to gravity, $c_p = 1004\,\mathrm{J\,kg^{-1}\,K^{-1}}$ is the specific heat at constant pressure, T is temperature (K), $L_\mathrm{v} = 2.5\times10^6\,\mathrm{J\,kg^{-1}}$ is the latent heat of vaporization of water, and q is specific humidity ($\mathrm{kg\,kg^{-1}}$). (Note that the term zg normally present in the definition of moist static energy has been removed, as the gravitational potential of the atmosphere as a whole is assumed to not change.) Positive values of AT indicate northward transport of energy by the atmosphere across the equator.

Figure 24 shows annually-averaged timeseries of AT for the 2×2 design case. The values in this figure are consistent with the results in Fig. 12: the full 2×2 simulation is effective at restoring the position of the precipitation centroid because it approximately equilibrates the cross-equatorial energy transport. The Arctic Only simulation shows promise in stabilizing cross-equatorial energy transport at a new steady-state value, whereas the 1pctCO2 simulation has a continuing negative trend, representing further northward shifts of tropical precipitation, accompanied by more southward energy transport to compensate for the hemispheric energy imbalance.

Fig. 25 shows cross-equatorial heat transport for the GISS ModelE2 simulations. The annually averaged timeseries of change in AT were quite noisy, so both the raw timeseries and ordinary least-squares regression on those timeseries are shown. Because any trends in the timeseries are small, R^2 values are also predictably small, and none of the regressions is statistically robust; nevertheless, these results can give an indication of physical mechanisms explaining system behavior. The 1pctCO2 simulation shows heat transport into the Southern Hemisphere that steadily increases in magnitude throughout the simulation, consistent with the results from CESM and with expectations. The Arctic Only simulation shows overcompensation, in that the increasingly large Arctic cooling to compensate

for CO_2 warming actually results in net heat transport into the Northern Hemisphere. To offset this change in cross-equatorial heat transport, the full 2×2 case calls for Antarctic cooling, which returns cross-equatorial heat transport to the "correct" direction. Due to the poor regression fits, it is difficult to comment on the relative magnitudes of cross-equatorial heat transport and whether the full 2×2 case actually returns AT to preindustrial values.

Acknowledgements. The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830. CESM simulations were performed using PNNL institutional computing resources. GISS ModelE2 simulations were supported by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.

References

Andrews, T., Forster, P. M., and Gregory, J. M.: A surface energy perspective on climate change, J. Climate, 22, 2557–2570, doi:10.1175/2008JCLI2759.1, 2009.

Åström, K. J. and Murray, R. M.: Analysis and Design of Feedback Systems, Princeton, New Jersey, USA, 2008.

Ban-Weiss, G. A. and Caldeira, K.: Geoengineering as an optimization problem, Environ. Res. Lett., 5, 034009, doi:10.1088/1748-9326/5/3/034009, 2010.

Bintanja, R. and Selten, F. M.: Future increases in Arctic precipitation linked to local evaporation and sea-ice retreat, Nature, 509, 479–482, doi:10.1038/nature13259, 2014.

Broccoli, A. J., Dahl, K. A., and Stouffer, R. J.: Response of the ITCZ to Northern Hemisphere cooling, Geophys. Res. Lett., 33, L01702, doi:10.1029/2005GL024546, 2006.

Caldeira, K. and Myhrvold, N.: Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration, Environ. Res. Lett., 8, 034039, doi:10.1088/1748-9326/8/3/034039, 2013.

Caldeira, K. and Wood, L.: Global and Arctic climate engineering: numerical model studies, Philos. T. R. Soc. A, 366, 4039–4056, doi:10.1098/rsta.2008.0132, 2008.

Crook, J. A., Jackson, L. S., Osprey, S. M., and Forster, P. M.: A comparison of temperature and precipitation responses to different Earth radiation management geoengineering schemes, J. Geophys. Res., 120, 9352–9373, doi:10.1002/2015JD023269, 2015.

Crutzen, P. J.: Albedo enhancement by stratospheric sulfur injections: a contribution to resolve a policy dilemma?, Climatic Change, 77, 211–220, doi:10.1007/s10584-006-9101-y, 2006.

Donohoe, A., Marshall, J., Ferreira, D., and McGee, D.: The Relationship between ITCZ Location and cross-equatorial atmospheric heat transport: from the seasonal cycle to the Last Glacial Maximum, J. Climate, 26, 3597–3618, doi:10.1175/JCLI-D-12-00467.1, 2013.

Ferraro, A. J., Highwood, E. J., and Charlton-Perez, A. J.: Weakened tropical circulation and reduced precipitation in response to geoengineering, Environ. Res. Lett., 9, 014001, doi:10.1088/1748-9326/9/1/014001, 2014.

Govindasamy, B. and Caldeira, K.: Geoengineering Earth's radiation balance to mitigate CO_2-induced climate change, Geophys. Res. Lett., 27, 2141–2144, doi:10.1029/1999GL006086, 2000.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yaho, M., and Zhang, S.: Efficacy of climate forcings, J. Geophys. Res., 110, D18104, doi:10.1029/2005JD005776, 2005.

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

Held, I. M. and Soden, B. J.: Robust responses of the hydrological cycle to global warming, J. Climate, 19, 5686–5699, doi:10.1175/JCLI3990.1, 2006.

Holland, M. M. and Bitz, C. M.: Polar amplification of climate change in coupled models, Clim. Dynam., 21, 221–232, doi:10.1007/s00382-003-0332-6, 2003.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: a framework for collaborative research, B. Am. Meteorol. Soc., 94, 1339–1360, doi:10.1175/BAMS-D-12-00121.1, 2013.

Jackson, L. S., Crook, J. A., Jarvis, A., Leedal, D., Ridgwell, A., Vaughan, N., and Forster, P. M.: Assessing the controllability of Arctic sea ice extent by sulfate aerosol geoengineering, Geophys. Res. Lett., 42, 1223–1231, doi:10.1002/2014GL062240, 2015.

Jarvis, A. and Leedal, D.: The Geoengineering Model Intercomparison Project (GeoMIP): a control perspective, Atmos. Sci. Lett., 13, 157–163, doi:10.1002/asl.387, 2012.

Jones, A., Haywood, J. M., Alterskjær, K., Boucher, O., Cole, J. N. S., Curry, C. L., Irvine, P. J., Ji, D., Kravitz, B., Kristjánsson, J. E., Moore, J. C., Niemeier, U., Robock, A., Schmidt, H., Singh, B., Tilmes, S., Watanabe, S., and Yoon, J.-H.: The impact of abrupt suspension of solar radiation management (termination effect) in experiment G2 of the Geoengineering Model Intercomparison Project (GeoMIP), J. Geophys. Res., 118, 9743–9752, doi:10.1002/jgrd.50762, 2013.

Kalidindi, S., Bala, G., Modak, A., and Caldeira, K.: Modeling of solar radiation management: a comparison of simulations using reduced solar constant and stratospheric sulfate aerosols, Clim. Dynam., 44, 2909–2925, doi:10.1007/s00382-014-2240-3, 2014.

Kang, S. M., Held, I. M., Frierson, D. W., and Zhao, M.: The Response of the ITCZ to extratropical thermal forcing: idealized slab-ocean experiments with a GCM, J. Climate, 21, 3521–3532, doi:10.1175/2007JCLI2146.1, 2008.

Kang, S. M., Seager, R., Frierson, D. M. W., and Liu, X.: Croll revisited: why is the Northern Hemisphere warmer than the Southern Hemisphere?, Clim. Dynam., 44, 1457–1472, doi:10.1007/s00382-014-2147-z, 2015.

Kravitz, B., Robock, A., Boucher, O., Schmidt, H., Taylor, K. E., Stenchikov, G., and Schulz, M.: The Geoengineering Model Intercomparison Project (GeoMIP), Atmos. Sci. Lett., 12, 162–167, doi:10.1002/asl.316, 2011.

Kravitz, B., Caldeira, K., Boucher, O., Robock, A., Rasch, P. J., Alterskjær, K., Karam, D. B., Cole, J. N. S., Curry, C. L., Haywood, J. M., Irvine, P. J., Ji, D., Jones, A., Kristjánsson, J. E., Lunt, D. J., Moore, J., Niemeier, U., Schmidt, H., Schulz, M., Singh, B., Tilmes, S., Watanabe, S., Yang, S., and Yoon, J.-H.: Climate model response from the Geoengineering Model Intercomparison Project (GeoMIP), J. Geophys. Res., 118, 8320–8332, doi:10.1002/jgrd.50646, 2013a.

Kravitz, B., Forster, P. M., Jones, A., Robock, A., Alterskjær, K., Boucher, O., Jenkins, A. K. L., Korhonen, H., Kristjánsson, J. E., Muri, H., Niemeier, U., Partanen, A.-I., Rasch, P. J., Wang, H., and Watanabe, S.: Sea spray geoengineering experiments in the Geoengineering Model Intercomparison Project (GeoMIP): experimental design and preliminary results, J. Geophys. Res., 118, 11175–11186, doi:10.1002/jgrd.50856, 2013b.

Kravitz, B., MacMartin, D. G., Leedal, D. T., Rasch, P. J., and Jarvis, A. J.: Explicit feedback and the management of uncertainty in meeting climate objectives with solar geoengineering, Environ. Res. Lett., 9, 044006, doi:10.1088/1748-9326/9/4/044006, 2014.

Kravitz, B., Robock, A., Tilmes, S., Boucher, O., English, J. M., Irvine, P. J., Jones, A., Lawrence, M. G., MacCracken, M., Muri, H., Moore, J. C., Niemeier, U., Phipps, S. J., Sillmann, J., Storelvmo, T., Wang, H., and Watanabe, S.: The Geoengineering Model Intercomparison Project Phase 6 (GeoMIP6): simulation design and preliminary results, Geosci. Model Dev., 8, 3379–3392, doi:10.5194/gmd-8-3379-2015, 2015a.

Kravitz, B., MacMartin, D. G., Rasch, P. J., and Jarvis, A.: A new method of comparing forcing agents in climate models, J. Climate, 28, 8203–8218, doi:10.1175/JCLI-D-14-00663.1, 2015b.

Lambert, F. H., Webb, M. J., and Joshi M. M.: The relationship between land-ocean surface temperature contrast and radiative forcing, J. Climate, 24, 3239–3256, doi:10.1175/2011JCLI3893.1, 2011.

Latham, J.: Control of global warming?, Nature, 347, 339–340, doi:10.1038/347339b0, 1990.

Lau, K. M. and Wu, H.-T.: Assessment of the impacts of the 1997–98 El Niño on the Asian–Australia monsoon, Geophys. Res. Lett., 26, 1747–1750, doi:10.1029/1999GL900307, 1999.

MacCracken, M. C., Shin, H.-J., Caldeira, K., and Ban-Weiss, G. A.: Climate response to imposed solar radiation reductions in high latitudes, Earth Syst. Dynam., 4, 301–315, doi:10.5194/esd-4-301-2013, 2013.

MacMartin, D. G. and Tziperman, E.: Using transfer functions to quantify El Niño Southern Oscillation dynamics in data and models, Philos. T. Roy. Soc. A, 40, 20140272, doi:10.1098/rspa.2014.0272, 2014.

MacMartin, D. G., Keith, D. W., Kravitz, B., and Caldeira, K.: Management of trade-offs in geoengineering through optimal choice of non-uniform radiative forcing, Nat. Clim. Change, 3, 365–368, doi:10.1038/nclimate1722, 2013.

MacMartin, D. G., Caldeira, K., and Keith, D. W.: Solar geoengineering to limit the rate of temperature change, Philos. T. Roy. Soc. A, 372, 20140134, doi:10.1098/rsta.2014.0134, 2014a.

MacMartin, D. G., Kravitz, B., Keith, D. W., and Jarvis, A.: Dynamics of the coupled human-climate system resulting from closed-loop control of solar geoengineering, Clim. Dynam., 43, 243–258, doi:10.1007/s00382-013-1822-9, 2014b.

MacMynowski, D. G., Shin, H.-J., and Caldeira, K.: The frequency response of temperature and precipitation in a climate model, Geophys. Res. Lett., 38, L16711, doi:10.1029/2011GL048623, 2011.

Marshall, J., Donohoe, A., Ferreira, D., and McGee, D.: The ocean's role in setting the mean position of the Inter-Tropical Convergence Zone, Clim. Dynam., 42, 1967–1979, doi:10.1007/s00382-013-1767-z, 2014.

May, W.: Potential future changes in the Indian summer monsoon due to greenhouse warming: analysis of mechanisms in a global time-slice experiment, Clim. Dynam., 22, 389–414, doi:10.1007/s00382-003-0389-2, 2004.

NAS: Climate Intervention: Reflecting Sunlight to Cool Earth, Tech. rep., National Research Council, available at: http://www.nap.edu/catalog/18988/climate-intervention-reflecting-sunlight-to-cool-earth, last access: 7 May 2015.

Niemeier, U., Schmidt, H., Alterskjær, K., and Kristjánsson, J. E.: Solar irradiance reduction via climate engineering: Impact of different techniques on the energy balance and the hydrological cycle, J. Geophys. Res., 118, 11905–11917, doi:10.1002/2013JD020445, 2013.

Robock, A., Oman, L., and Stenchikov, G.: Regional climate responses to geoengineering with tropical and Arctic SO_2 injections, J. Geophys. Res., 113, D16101, doi:10.1029/2008JD010050, 2008.

Schaefer, K., Zhang, T., Bruhwiler, L., and Barrett, A. P.: Amount and timing of permafrost carbon release in response to climate warming, Tellus B, 63, 165–180, doi:10.1111/j.1600-0889.2011.00527.x, 2011.

Serreze, M. C., Holland, M. M., and Stroeve, J.: Perspectives on the Arctic's shrinking sea-ice cover, Science, 315, 1533–1536, doi:10.1126/science.1139426, 2007.

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., Bréon, F.-M., Church, J. A., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory, J. M., Hartmann, D. L., Jansen, E., Kirtman, B., Knutti, R., Krishna, K., Kumar, S., Lemke, P., Marotzke, J., Masson-Delmotte, V., Meehl, G. A., Mokhov, I. I., Piao, S., Ramaswamy, V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L. D., Vaughan, D. G., and Xie, S.-P.: Technical Summary, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013.

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

Taylor, P. C., Ellingson, R. G., and Cai, M.: Seasonal variations of climate feedbacks in the NCAR CCSM3, J. Climate, 24, 3433–3444, doi:10.1175/2011JCLI3862.1, 2011.

Tilmes, S., Fasullo, J., Lamarque, J.-F., Marsh, D. R., Mills, M., Alterskjær, K., Muri, H., Kristjánsson, J. E., Boucher, O., Schulz, M., Cole, J. N. S., Curry, C. L., Jones, A., Haywood, J., Irvine, P. J., Ji, D., Moore, J. C., Karam, D. B., Kravitz, B., Rasch, P. J., Singh, B., Yoon, J.-H., Niemeier, U.,

Schmidt, H., Robock, A., Yang, S., and Watanabe, S.: The hydrological impact of geoengineering in the Geoengineering Model Intercomparison Project (GeoMIP), J. Geophys. Res., 118, 11036–11058, doi:10.1002/jgrd.50868, 2013.

Tilmes, S., Jahn, A., Kay, J. E., Holland, M., and Lamarque, J.-F.: Can regional climate engineering save the summer Arctic sea ice?, Geophys. Res. Lett., 41, 880–885, doi:10.1002/2013GL058731, 2014.

Table 1. Root mean square (RMS) differences in Arctic temperature (°C) and χ (Eq. 1; degrees latitude) from the temperature and latitude objectives. Values are calculated over the entire 70 year simulation as the RMS of interannual deviations from the ~~long-term~~ preindustrial control (piControl) mean.

		Arctic temperature	χ
CESM	piControl	0.60	0.17
	1pctCO2	3.37	0.59
	Arctic Only	0.70	0.38
	2×2	0.70	0.17
GISS	piControl	0.41	0.11
	1pctCO2	1.77	0.17
	Arctic Only	0.43	0.09
	2×2	0.40	0.09

Table 2. Root mean square (RMS) differences in T_0, T_1, and T_2 (see Sect. 2 for definitions) for all of the simulations in the 3×3 design case. All units are in °C. Values are calculated over the entire 70 year simulation as the RMS of interannual deviations from the ~~long-term~~ preindustrial control (piControl) mean.

		T_0	T_1	T_2
CESM	piControl	0.14	0.04	0.04
	1pctCO2	1.34	0.16	0.24
	1×1	0.19	0.19	0.14
	2×2	0.18	0.08	0.16
	3×3	0.20	0.08	0.05
GISS	piControl	0.08	0.05	0.03
	1pctCO2	1.46	0.28	0.17
	1×1	0.13	0.06	0.06
	2×2	0.13	0.05	0.05
	3×3	0.13	0.06	0.04

Table 3. Residual changes in global mean precipitation ($\overline{P}$) and the precipitation centroid (χ; Eq. 1) for all of the simulations in the 3×3 design case. Changes are compared to the preindustrial control simulation. Units for $\overline{P}$ are in $\mathrm{mm\,yr^{-1}}$, and units for χ are in degrees latitude. Reported values are averages over years 61–70 of simulation.

		$\overline{P}$	χ
CESM	1pctCO2	29	0.81
	1×1	-22	0.57
	2×2	-23	0.66
	3×3	-16	0.73
GISS	1pctCO2	21	0.25
	1×1	-20	-0.03
	2×2	-19	-0.14
	3×3	-17	-0.08

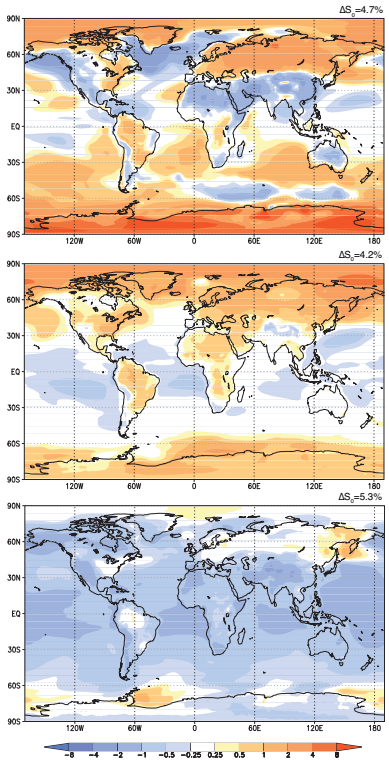

Figure 1. These three panels show that the "canonical" temperature response to offsetting global mean temperature increases from CO_2 (an abrupt quadrupling of the CO_2 concentration from its preindustrial value) with total solar irradiance reduction (top panel; e.g., Govindasamy and Caldeira, 2000; Kravitz et al., 2013a) still involves a degree of freedom, in that the resulting temperature pattern depends upon the amount of solar reduction. The middle panel shows the temperature response if total solar irradiance is reduced such that the mean tropical temperature does not represent an overcooling. The bottom panel shows the temperature response if total solar irradiance is reduced such that the mean polar temperature does not represent an undercooling. Values above each panel indicate the percentage reduction in total solar irradiance. All simulations were conducted with CESM and represent an average over years 11–20 of simulation.

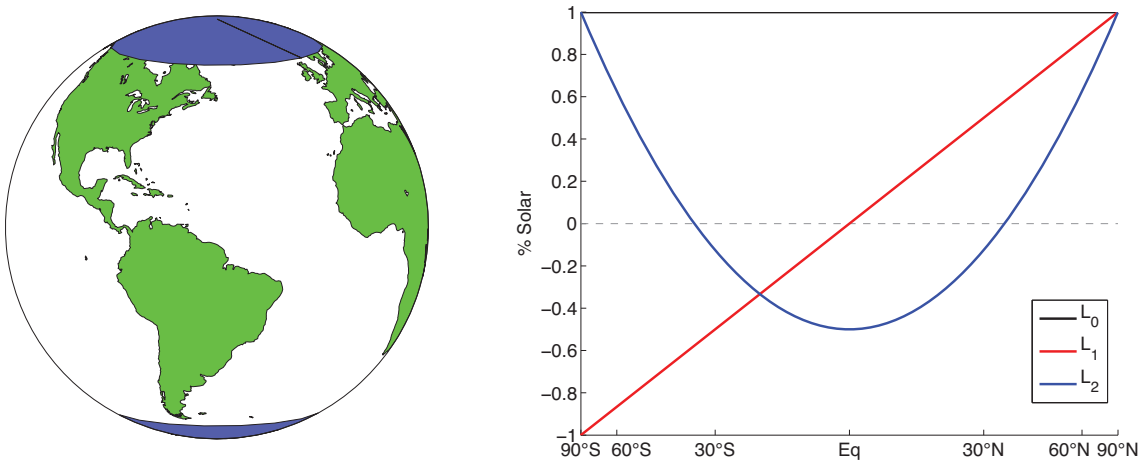

Figure 2. The degrees of freedom that were modified in the two cases considered here, referred to as 2×2 (left panel) and 3×3 (right panel). In the 2×2 case, Arctic and Antarctic insolation (shaded regions) are modified to minimize changes in Arctic temperature and the latitude of the precipitation centroid due to increasing CO_2 (see Sect. 2 for details). In the 3×3 case, the three patterns of insolation reduction illustrated here are modified to minimize changes in global mean temperature, the inter-hemispheric temperature gradient, and the equator-to-pole temperature gradient caused by increasing CO_2 (see Sect. 2).

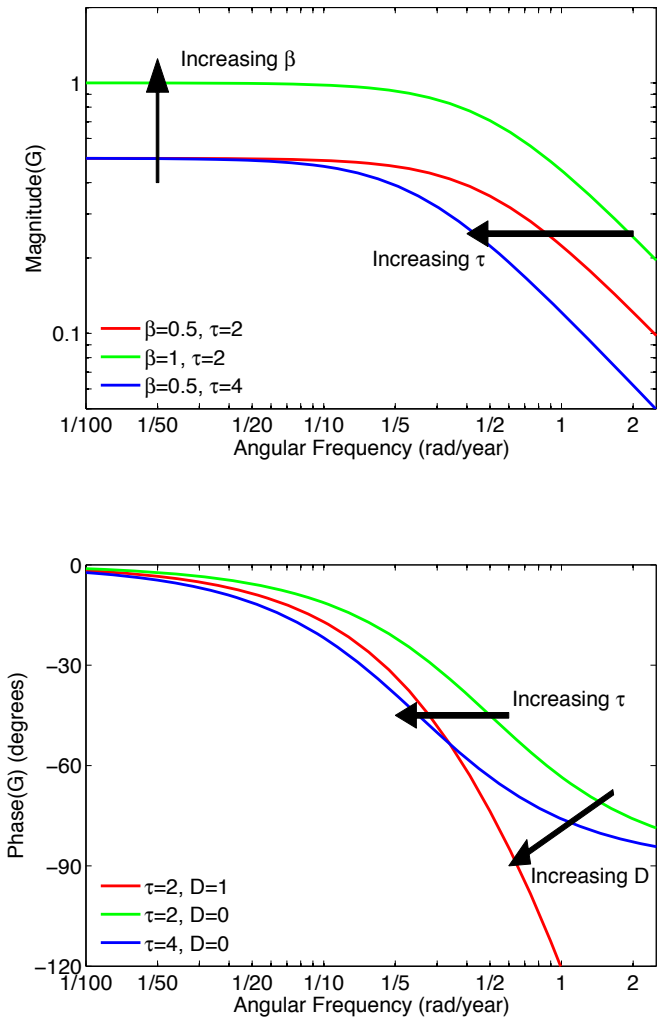

Figure 3. Bode plot showing the frequency response of the transfer function $G(s) = e^{-sD}\beta/(1+s\tau)$ for various values of β, τ, and D. The top panel shows the magnitude of the frequency response $||G(s)||$, and the bottom panel shows the phase $\phi = \tan^{-1}(\text{Im}(G(s))/\text{Re}(G(s)))$. β only affects magnitude, D only affects phase, and τ affects both. The red lines approximately correspond to the estimated response of Arctic temperature to Arctic insolation reduction in the 2×2 design example.

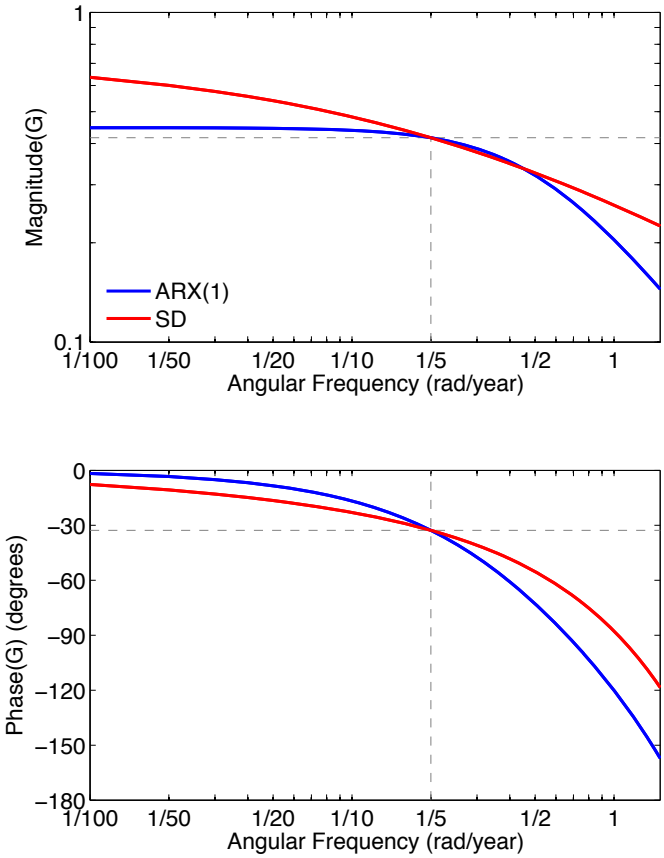

Figure 4. Bode plot (as in Fig. 3) comparing the first order linear model (ARX(1); Eq. 7) and a semi-infinite diffusion model (SD; Eq. 8). Values for ARX(1) are $\beta = 0.447$, $\tau = 1.946$, and $D = 1.0$. Values for SD are $\beta_d = 0.732$, $\tau_d = 4.063$, and $D = 1.0$. Values are chosen so that the magnitude and phase are identical for both models at $\omega = 0.2\,\mathrm{rad\,yr^{-1}}$.

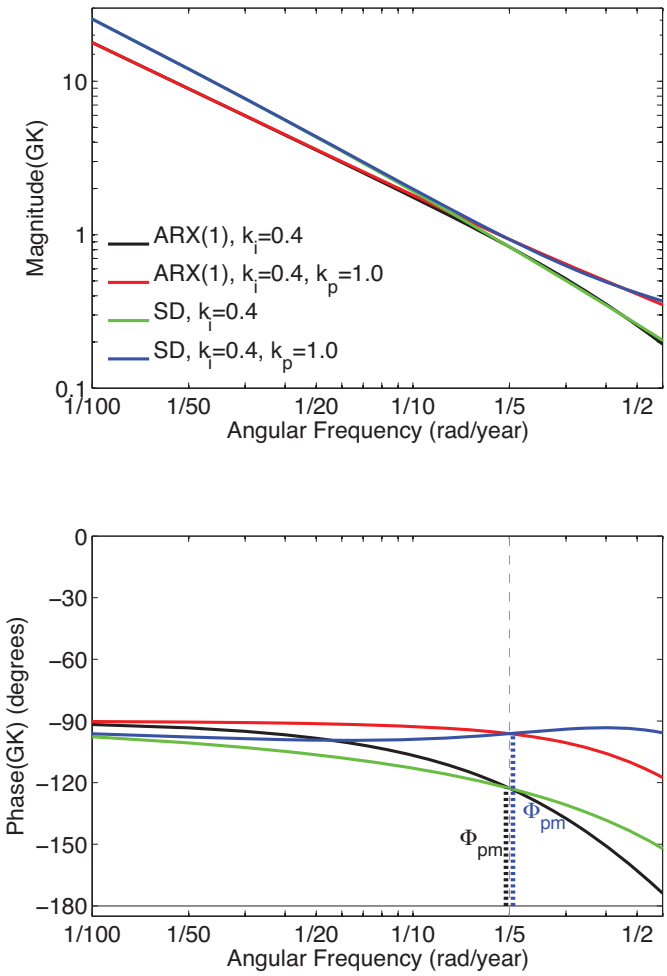

Figure 5. As in Fig. 4 but for the loop transfer function $G(s)K(s)$ for values of control gains k_i and k_p used in the 2×2 case. Plots are shown for the first order linear model (ARX(1)) and the semi-infinite diffusion model (SD). Grey dashed lines indicate a loop crossover frequency of $\omega_\mathrm{gc} = 0.2$, corresponding to the frequency where $||GK|| = 1$. Φ_pm denotes the phase margin (Sect. 3.3; the distance between the curves and a phase lag of 180°) for two cases with and without proportional gain k_p. Pure integral gain adds 90° of phase lag, which can be partially compensated by adding proportional gain.

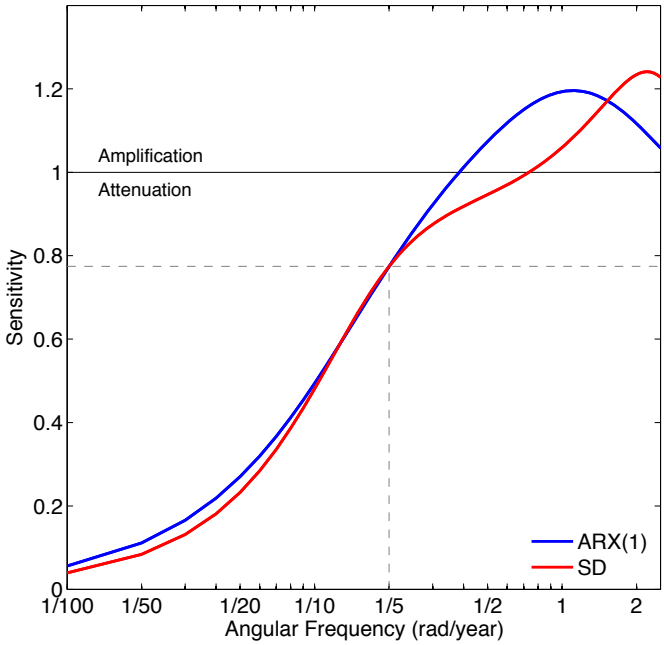

Figure 6. Magnitude of the sensitivity function $S(s) = (1 + G(s)K(s))^{-1}$ for the first order linear model (ARX(1)) and semi-infinite diffusion model (SD). Model parameters are the same as in Figs. 4 and 5, and $K(s) = 1.0 + 0.4/s$. Black horizontal line indicates $||S(s)||$=1; values above this line indicate amplification at that frequency, and values below this line indicate attenuation. Grey vertical line indicates the loop crossover frequency $\omega_{\text{gc}} = 0.2\,\text{rad}\,\text{yr}^{-1}$.

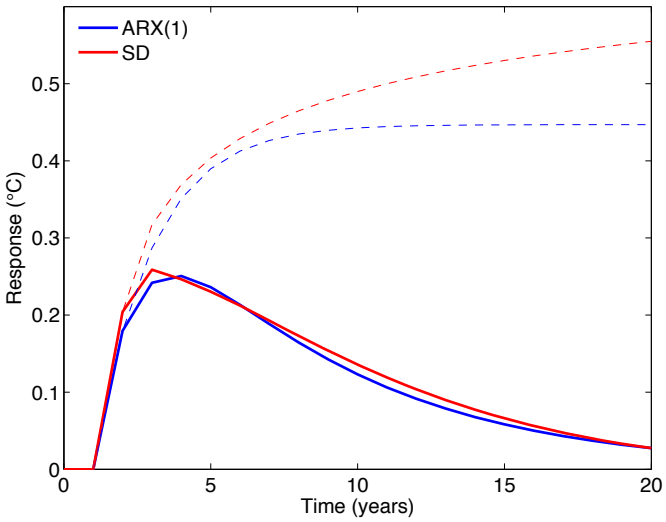

Figure 7. Time domain response of the first-order linear (ARX(1); Eq. 6) and the semi-infinite diffusion model (SD; inverse Laplace transform of Eq. 8) due to a step change in radiative forcing at $t = 1\,\text{year}$. Parameter values for the models are the same as in the text and the caption of Fig. 4. Dashed lines show the open-loop response, and solid lines show the closed-loop response with $k_\text{i} = 0.4$ and $k_\text{p} = 1.0$.

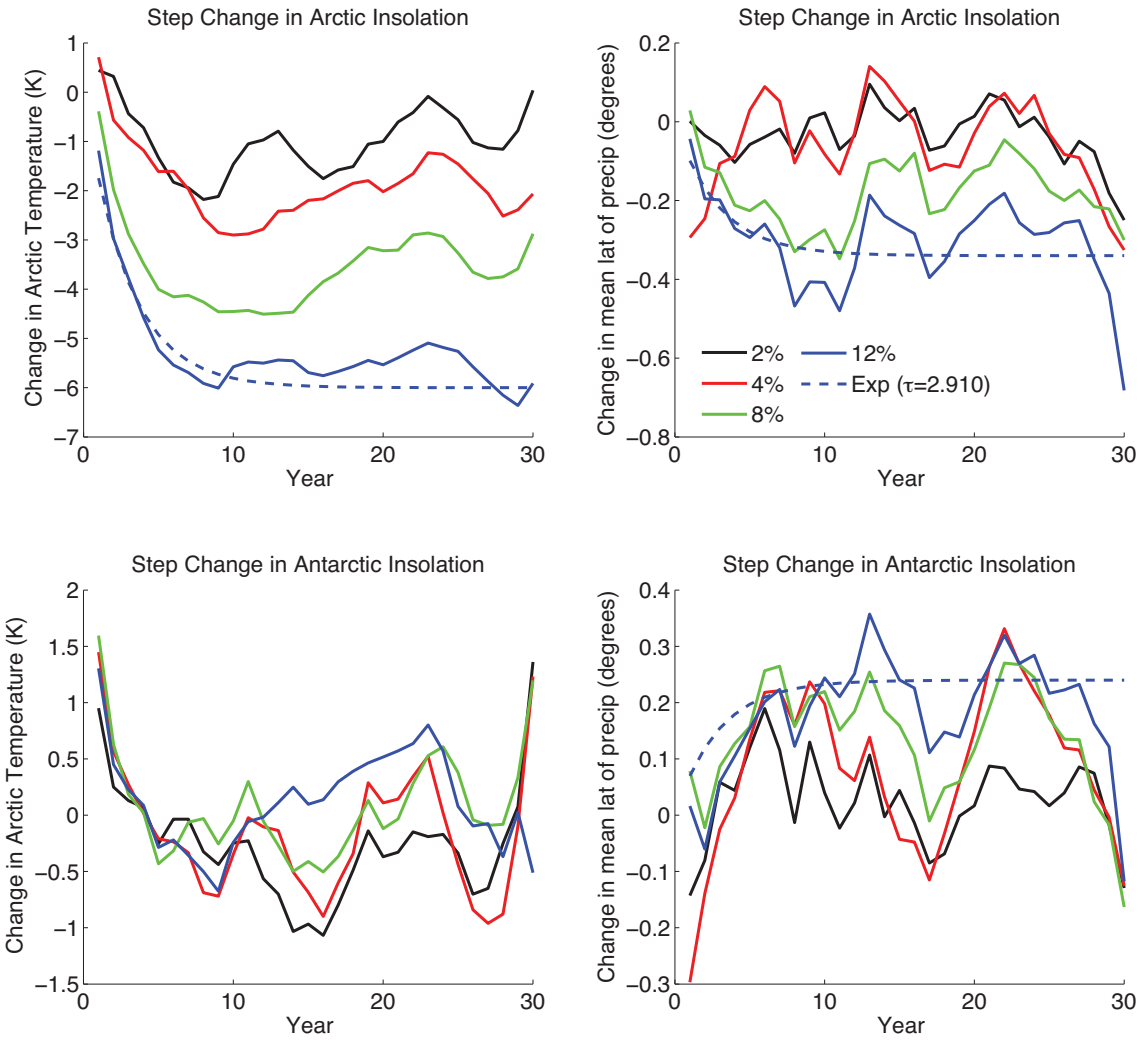

Figure 8. Step responses for the 2×2 design case. Step perturbations in the Arctic (top row panels) and Antarctic (bottom row panels) were 2, 4, 8, and 12 % of total solar irradiance in those regions. Performing a best fit with an exponential function (Eq. 6) to the 12 % step response results yields $\beta = 0.5$ and $\tau = 2.410$, plus a half year of time delay due to annual averaging.

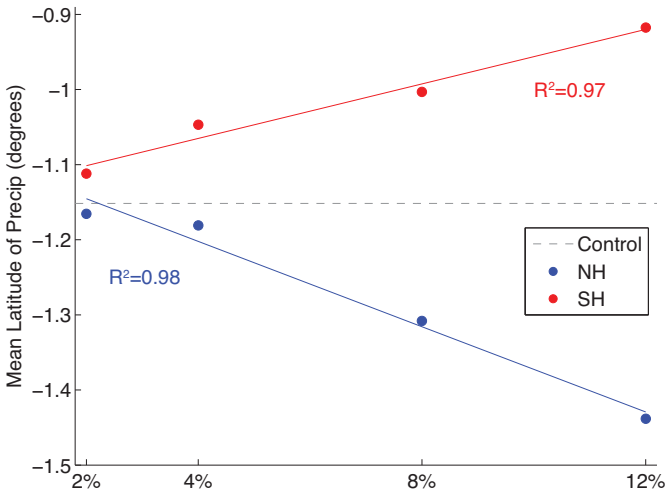

Figure 9. Linear regression over the precipitation centroid (χ; Eq. 1) results from all of the step response simulations. Regressions were performed over the average values of χ over years 11–30 of simulation. χ is approximately linear with perturbation amplitude.

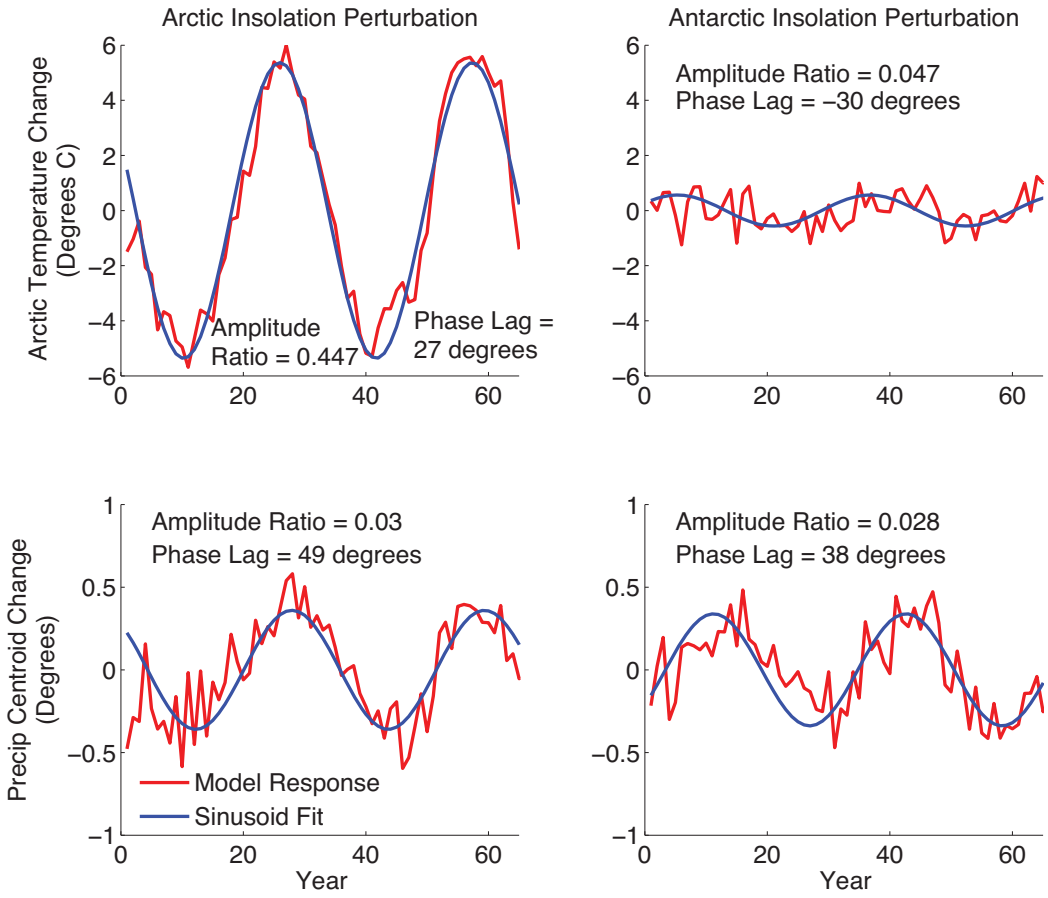

Figure 10. Results from the sinusoidal perturbations in the 2×2 design case. Input signal was $u(t) = 0.12 \sin(0.2t)$, where 0.12 corresponds to a maximum amplitude of 12 % reduction or increase in solar irradiance in the region, and $0.2\,\mathrm{rad\,yr^{-1}}$ is the chosen bandwidth (Sect. 3.3). Results are summarized in Eq. (21).

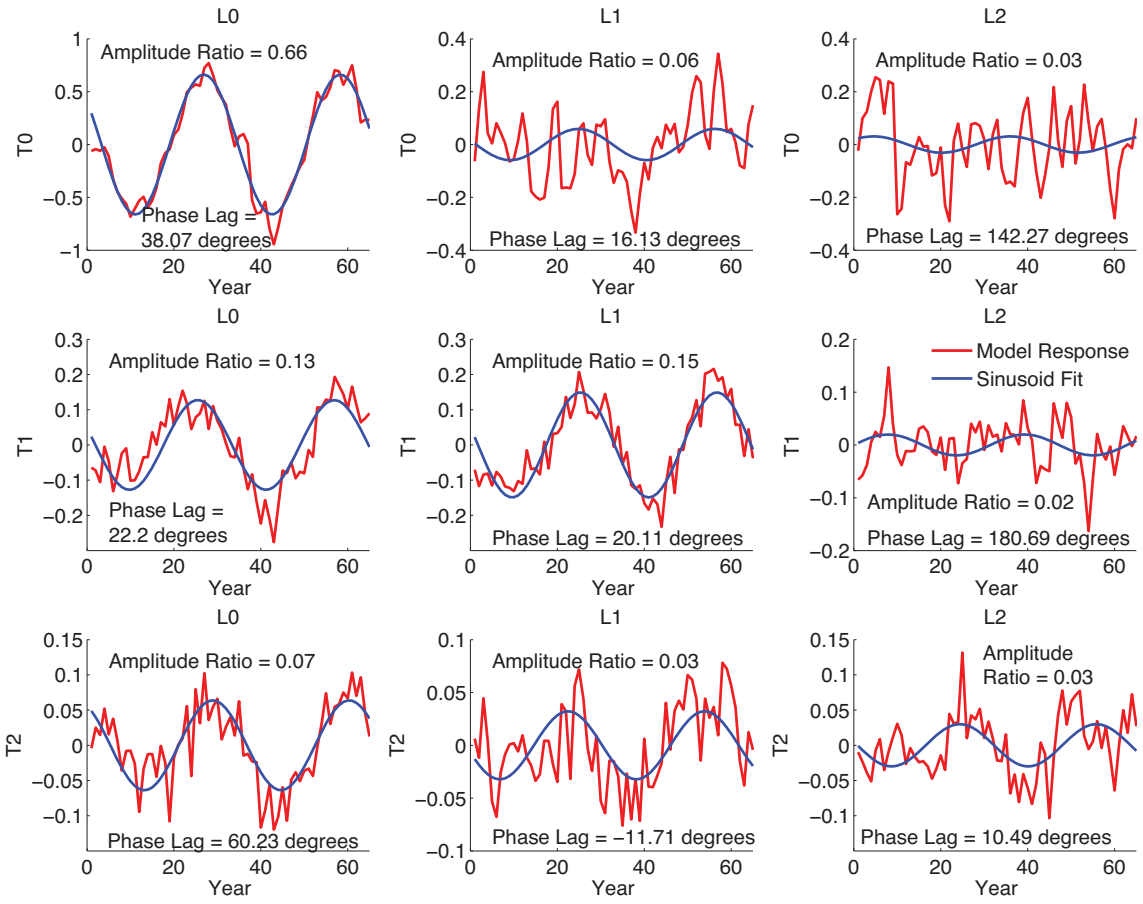

Figure 11. Results for the sinusoidal perturbations in the 3×3 design case. Inputs (L_0, L_1, L_2) and outputs (T_0, T_1, T_2) are described in Sect. 2. Input signal was $u(t) = 0.01 \sin(0.2\,t)$, where 0.01 corresponds to a maximum amplitude of 1 % reduction or increase in the pattern of insolation change (Eq. 3), and 0.2 $\mathrm{rad\,yr^{-1}}$ is the chosen bandwidth (Sect. 3.3). Results are summarized in Eq. (26).

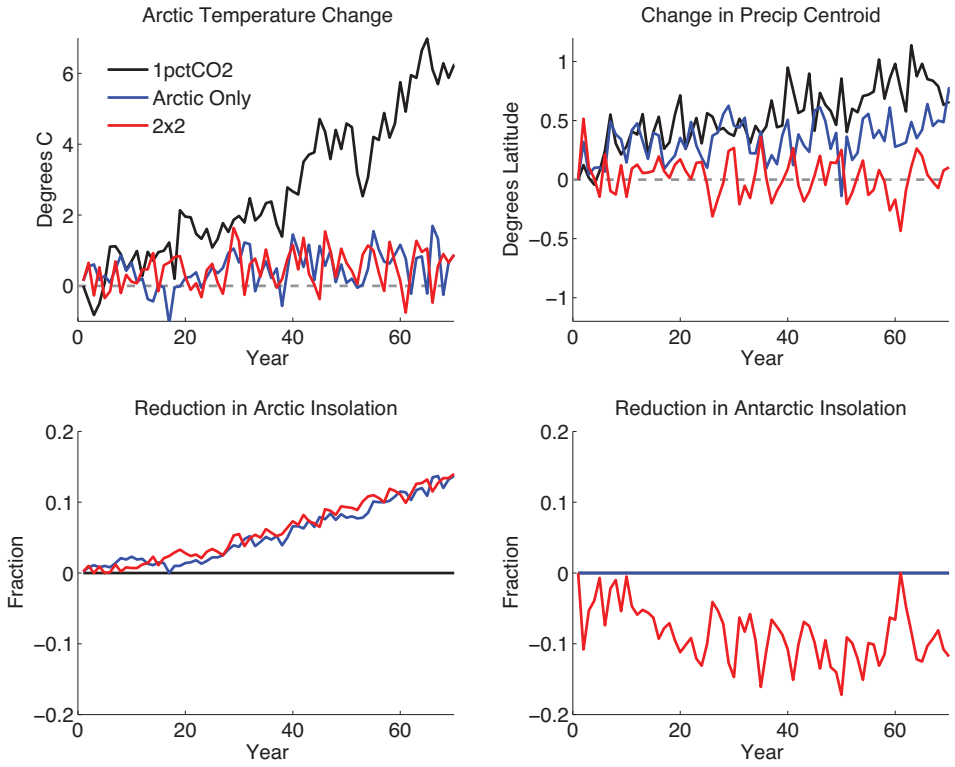

Figure 12. Results for the 2×2 case in the design model. Black lines indicate the 1pctCO2 simulation. The feedback algorithm adjusts Arctic and Antarctic insolation (bottom left and bottom right panels, respectively) to offset these changes, returning Arctic Temperature (top left panel) and (for the 2×2 case) the precipitation centroid (χ, Eq. 1); top right panel) to the dashed grey lines. Blue lines indicate simulations in which only Arctic insolation is adjusted. Red lines indicate the full 2×2 case.

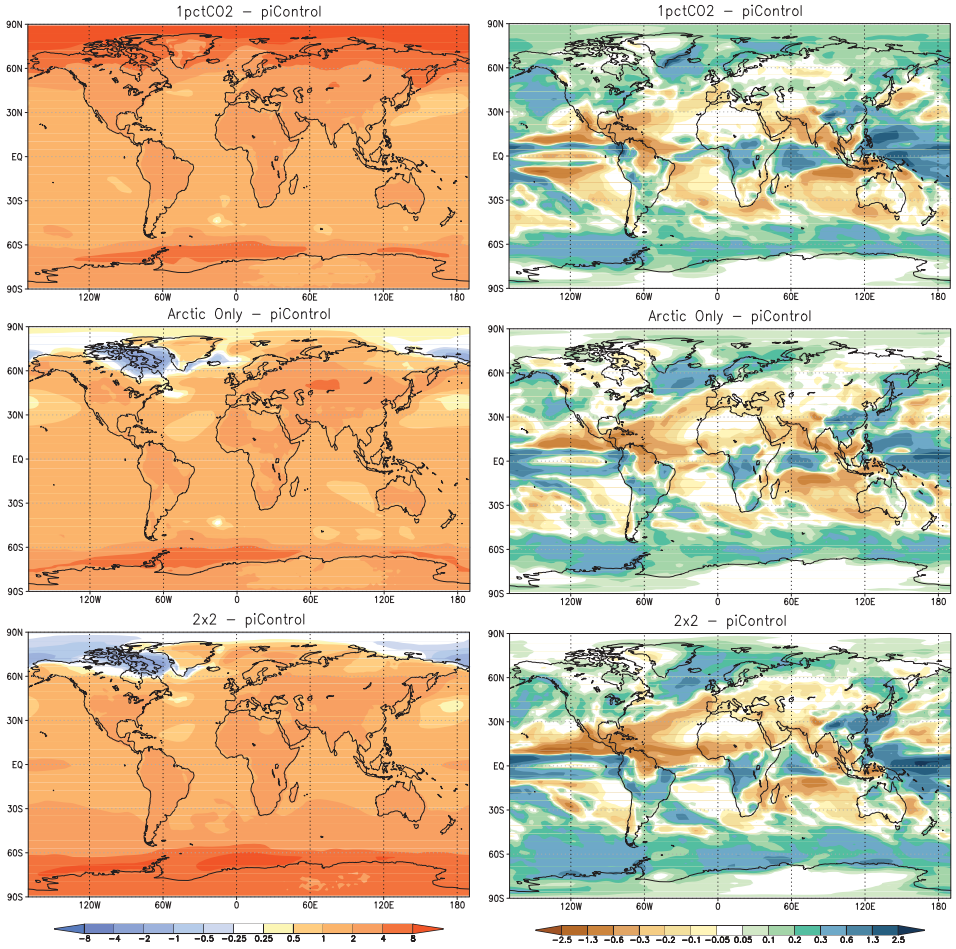

Figure 13. Maps of temperature change (left column; degrees C) and precipitation change (right column; mm day^{-1}) for the 2×2 case in the design model. Changes are calculated from the average of a preindustrial control simulation. Top panel corresponds to a 1pctCO2 simulation, middle row indicates simulations in which only Arctic insolation is adjusted, and bottom row indicates the full 2×2 case. All panels are averages over the last ten years of a 70 year simulation.

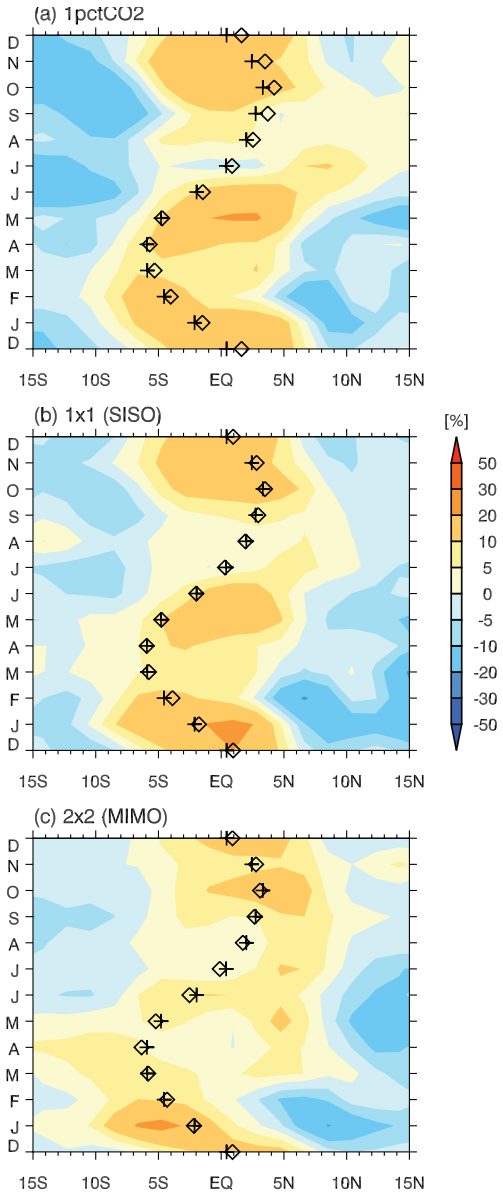
(a) 1pctCO2
(b) 1x1 (SISO)
(c) 2x2 (MIMO)
D
N
O
S
A
J
J
M
A
M
F
J
D
15S
10S
5S
EQ
5N
10N
15N
[%]
50
30
20
10
5
0
-5
-10
-20
-30
-50

Figure 14. Climatology of percent change (with respect to piControl) in total precipitation (shading; %) and shift in the precipitation centroid (χ, Eq. 1); symbols) for the 2×2 case in the design model. Top panel corresponds to 1pctCO2, middle panel corresponds to the Arctic Only simulation, and bottom panel corresponds to the full 2×2 simulation. Plus signs indicate piControl, and diamonds indicate the perturbed simulation. All values are averaged over years 59–68 of simulation and are linearly bias-corrected to account for small differences in background conditions between the perturbed simulations and piControl.

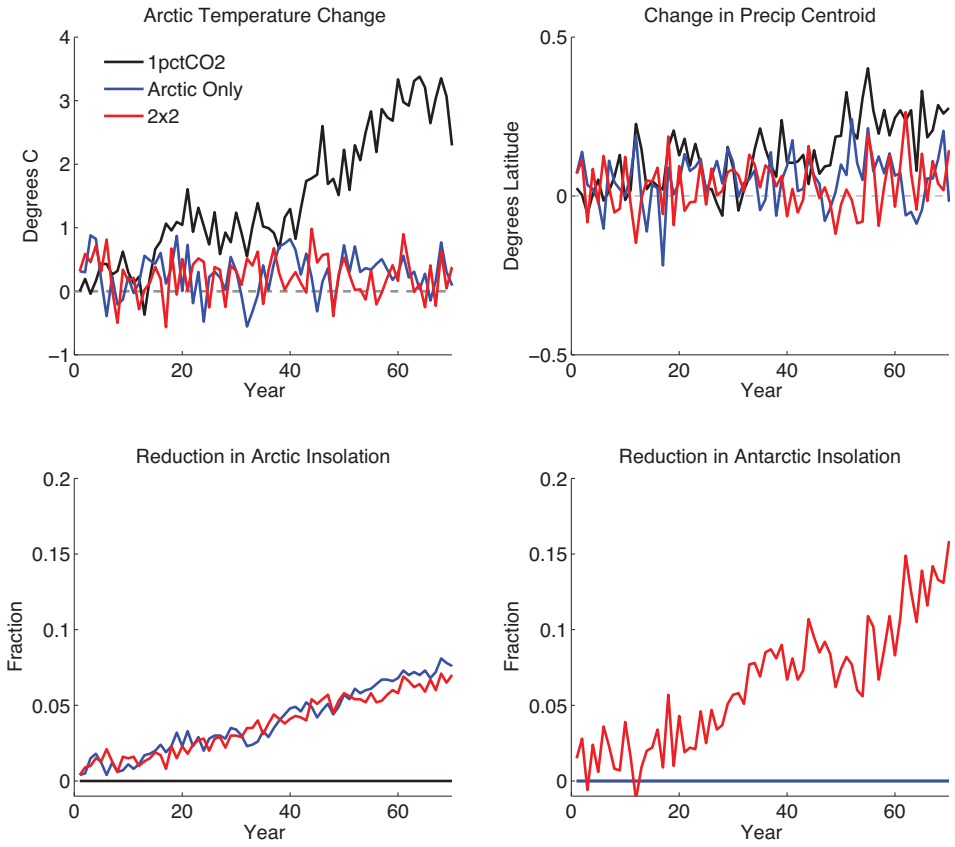

Figure 15. Same as Fig. 12 but for GISS ModelE2 (the evaluation model).

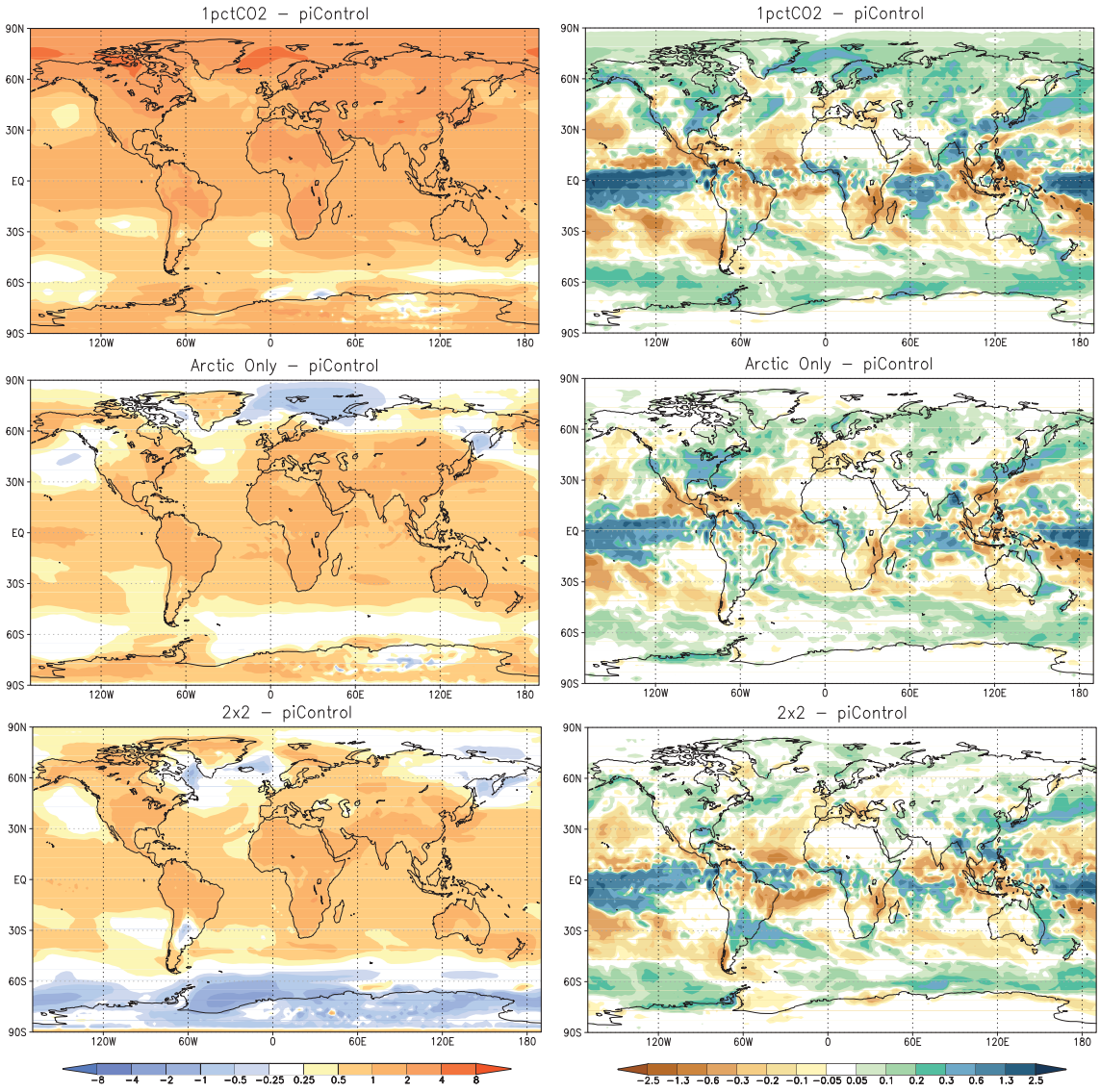

Figure 16. Same as Fig. 13 but for GISS ModelE2 (the evaluation model).

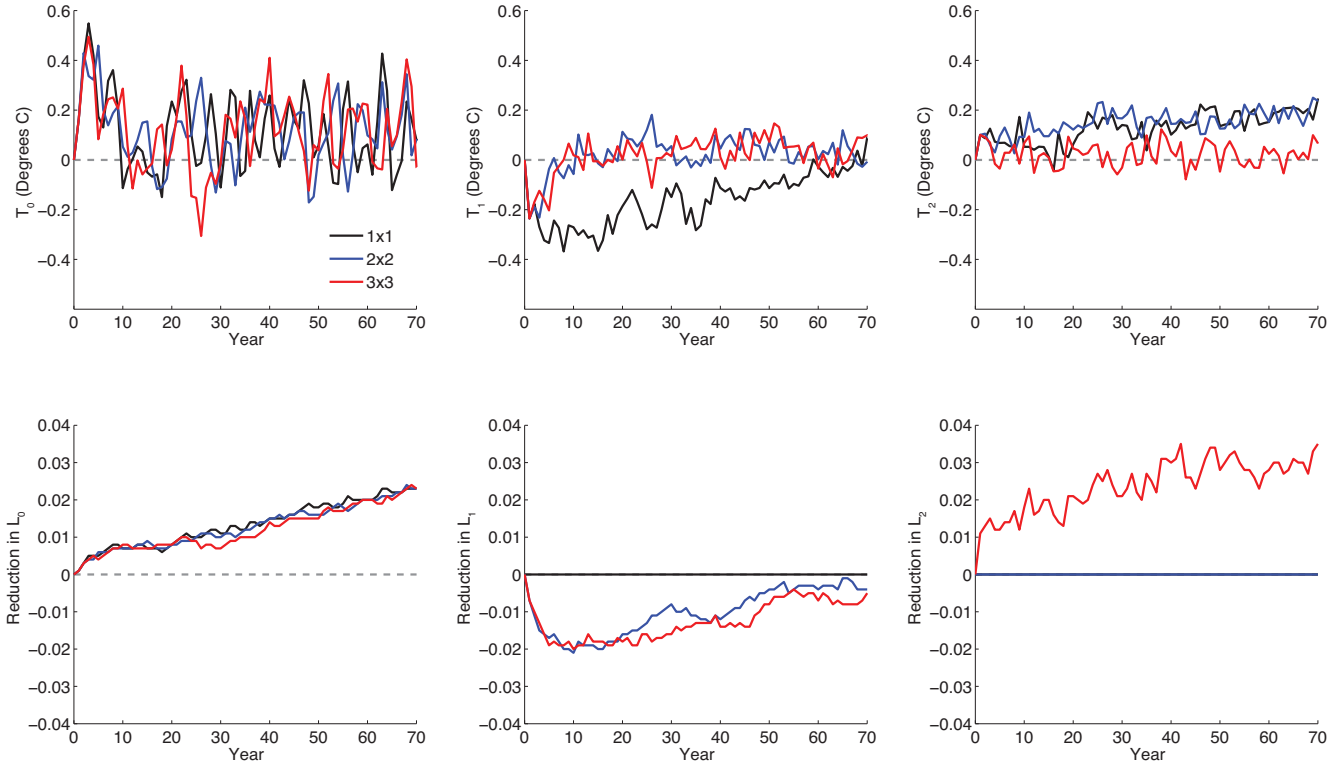

Figure 17. Results for the 3×3 case in the design model. Black lines indicate the 1×1 sub-case where L_0 is adjusted to offset changes in T_0 due to 1pctCO2. Blue lines indicate the 2×2 sub-case where L_0 and L_1 are adjusted to offset changes in T_0 and T_1. Red lines indicate the full 3×3 case.

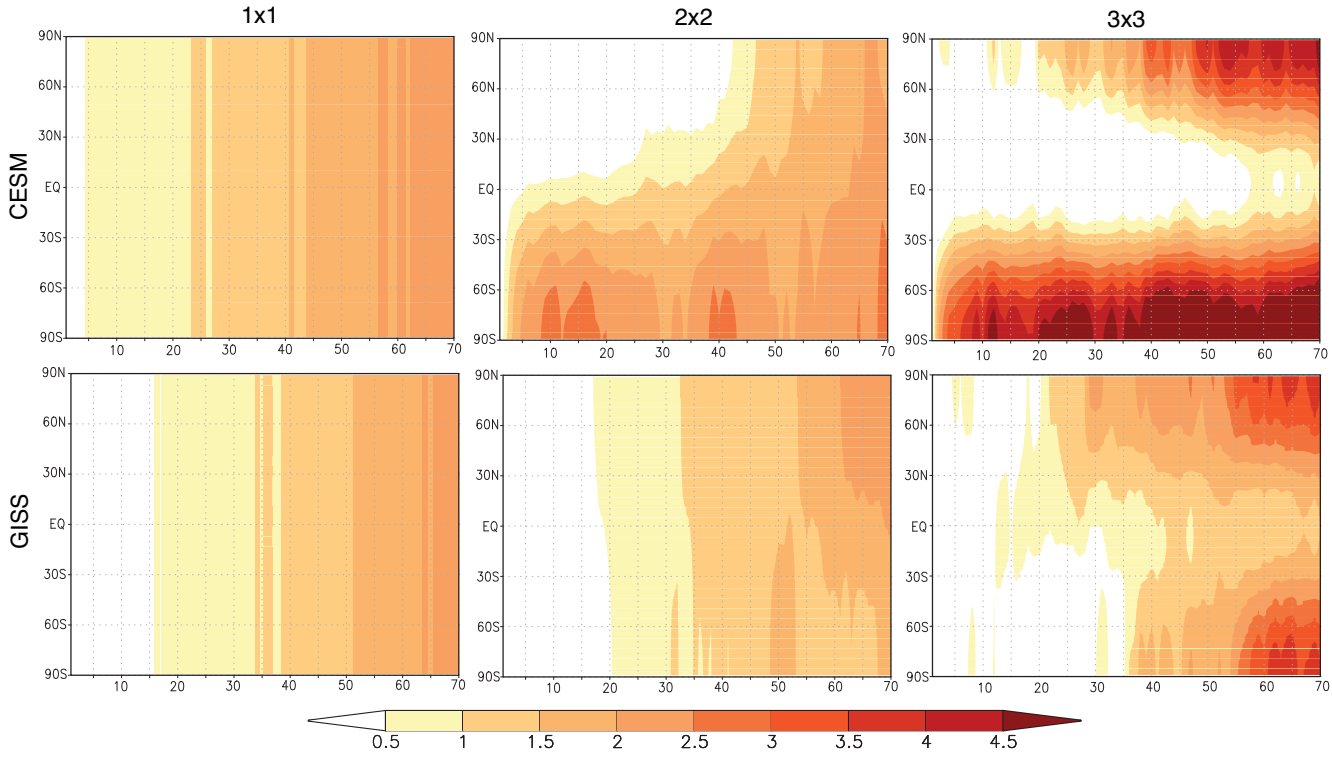

Figure 18. Percent reduction in insolation for the 3×3 design case. Left column is the 1×1 simulation (offsetting T_0 changes via L_0 changes), middle column is the 2×2 simulation (offsetting T_0 and T_1 changes via L_0 and L_1 changes), and right column is the full 3×3 simulation. Top row corresponds to the design model, and bottom row is the evaluation model. All results are zonally and annually averaged.

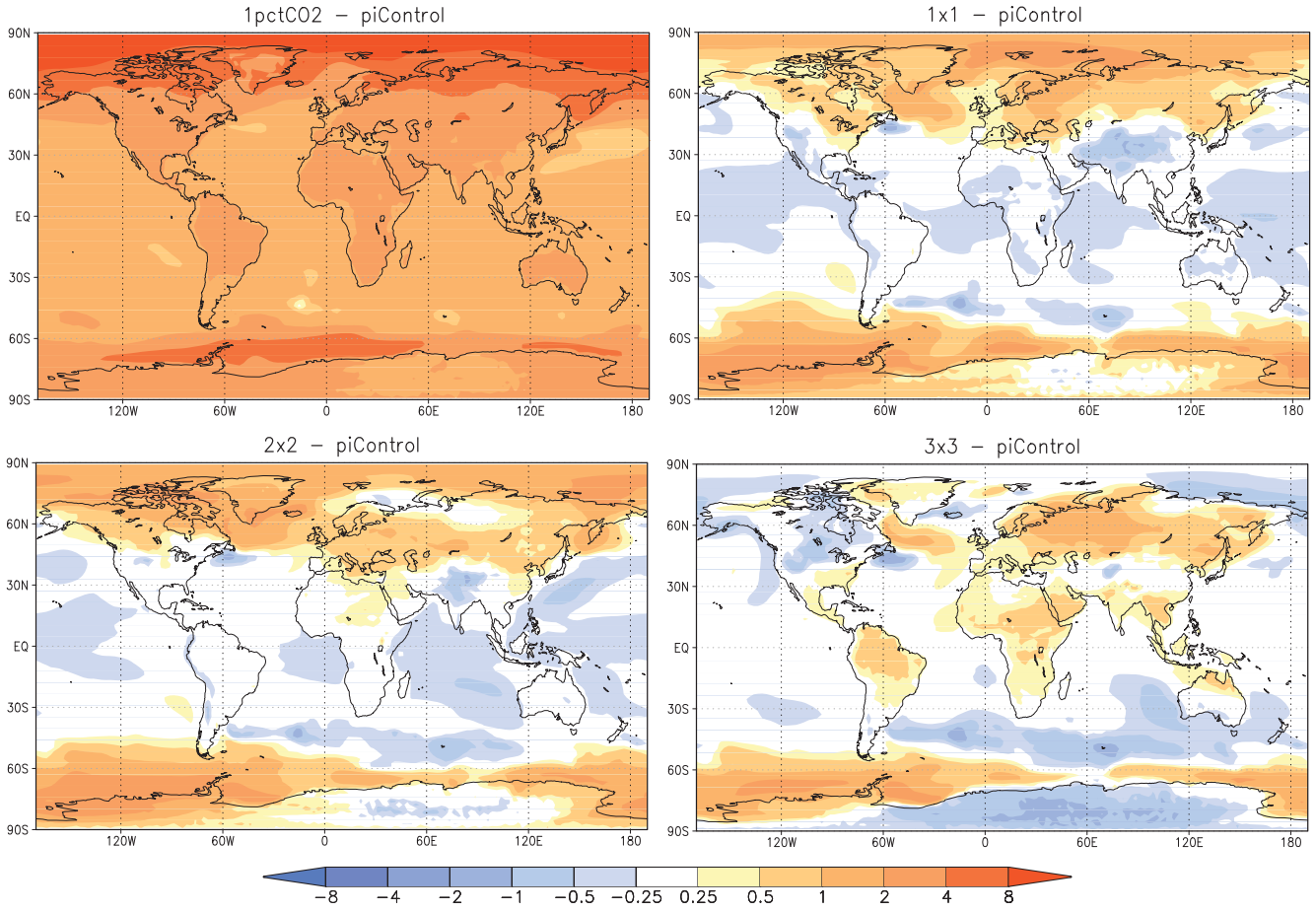

Figure 19. Maps of temperature change from the preindustrial control simulation (°C) for the 3×3 design case and its sub-cases in the design model. All panels are averages over the last ten years of a 70 year simulation.

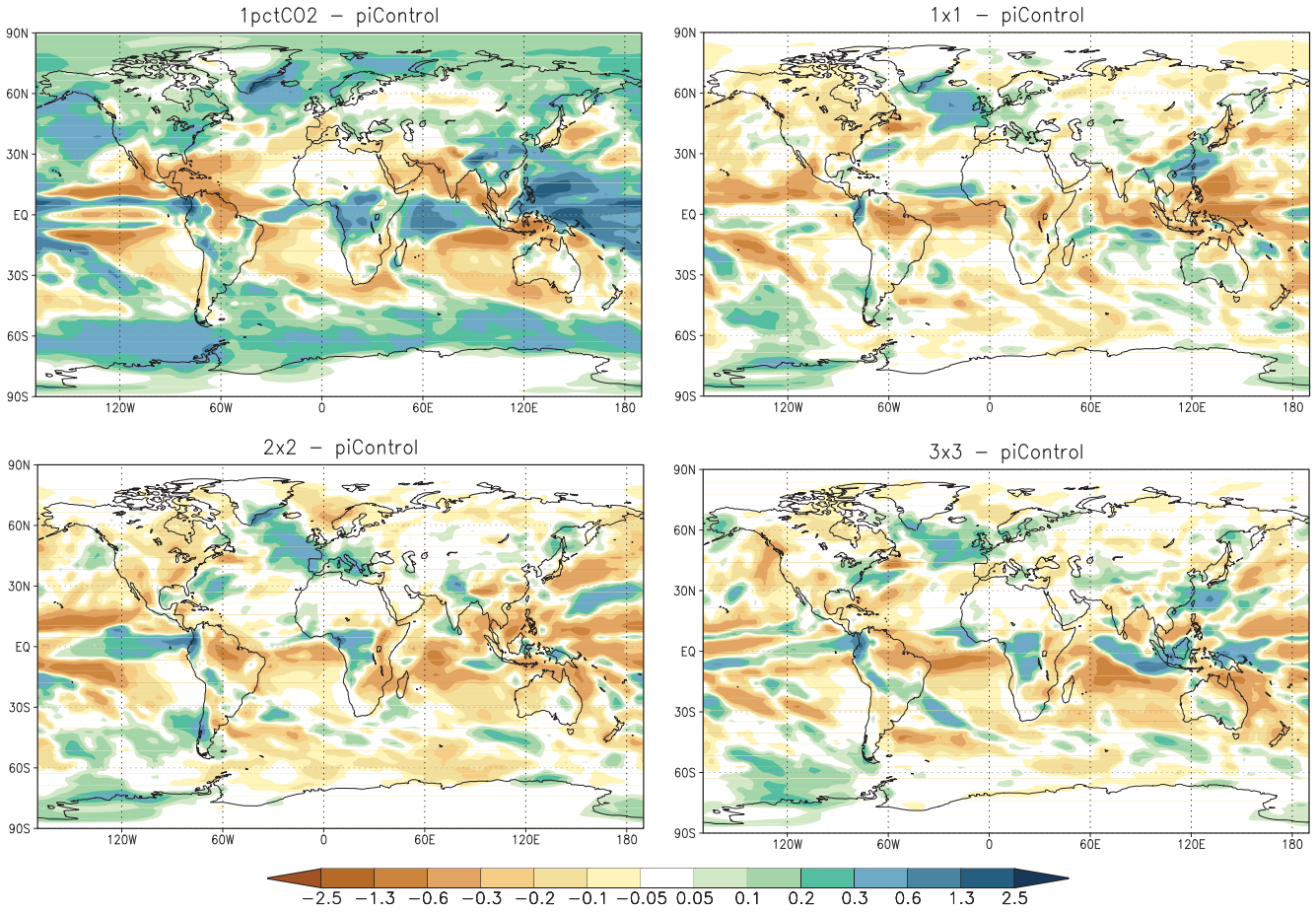

Figure 20. Same as Fig. 19 but for precipitation changes in the design model. Values are in mm day^{-1}.

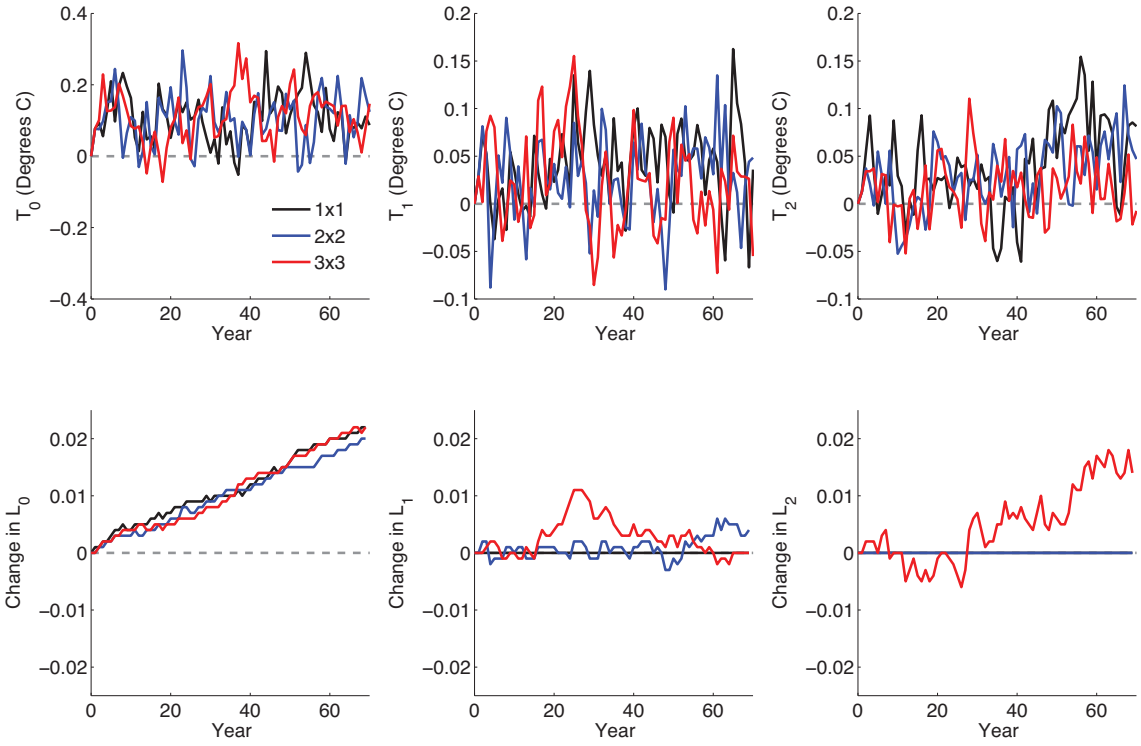

Figure 21. Same as Fig. 17 but for GISS ModelE2 (the evaluation model). Note different axis scaling in the top row.

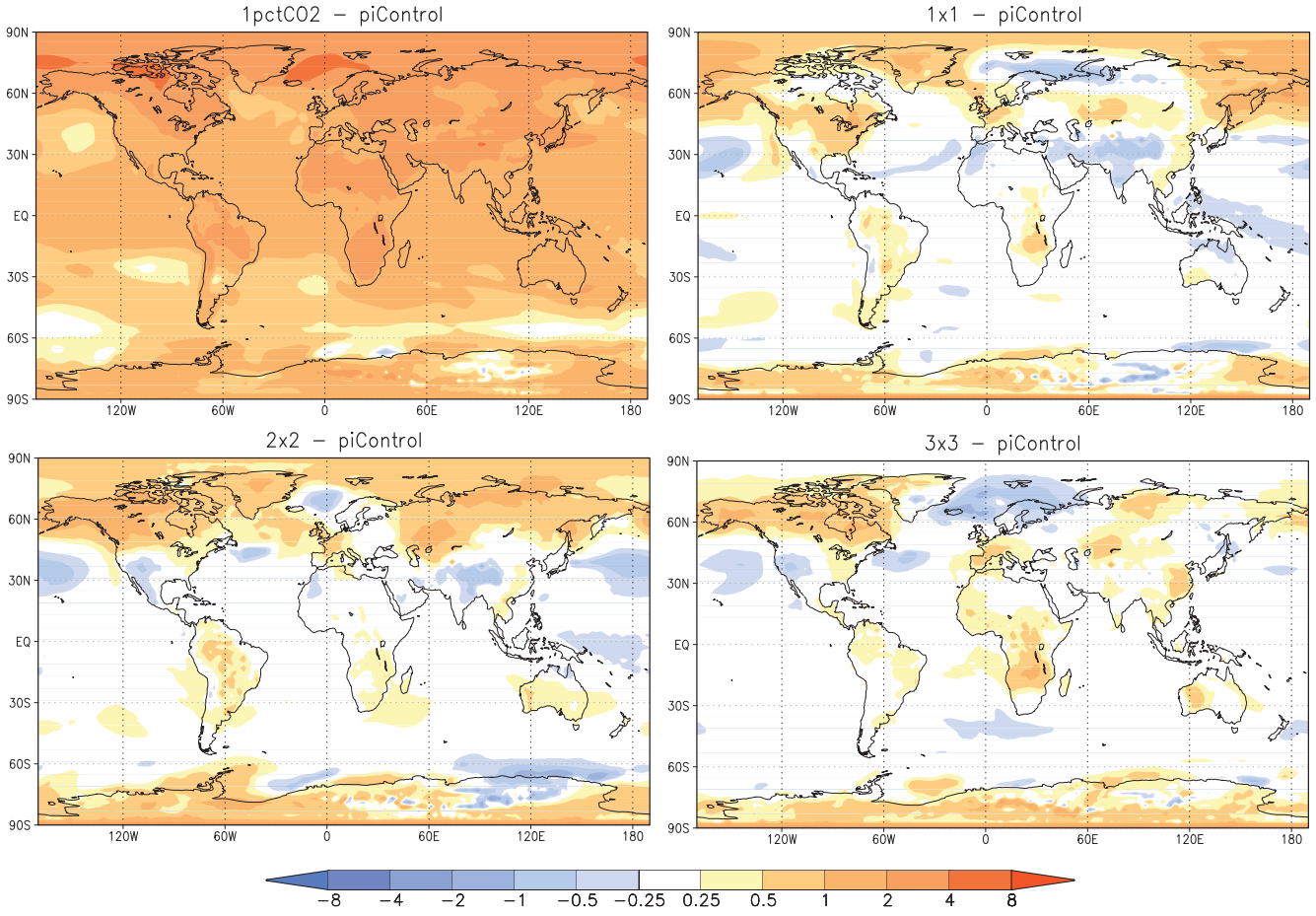

Figure 22. Same as Fig. 19 but for GISS ModelE2 (the evaluation model).

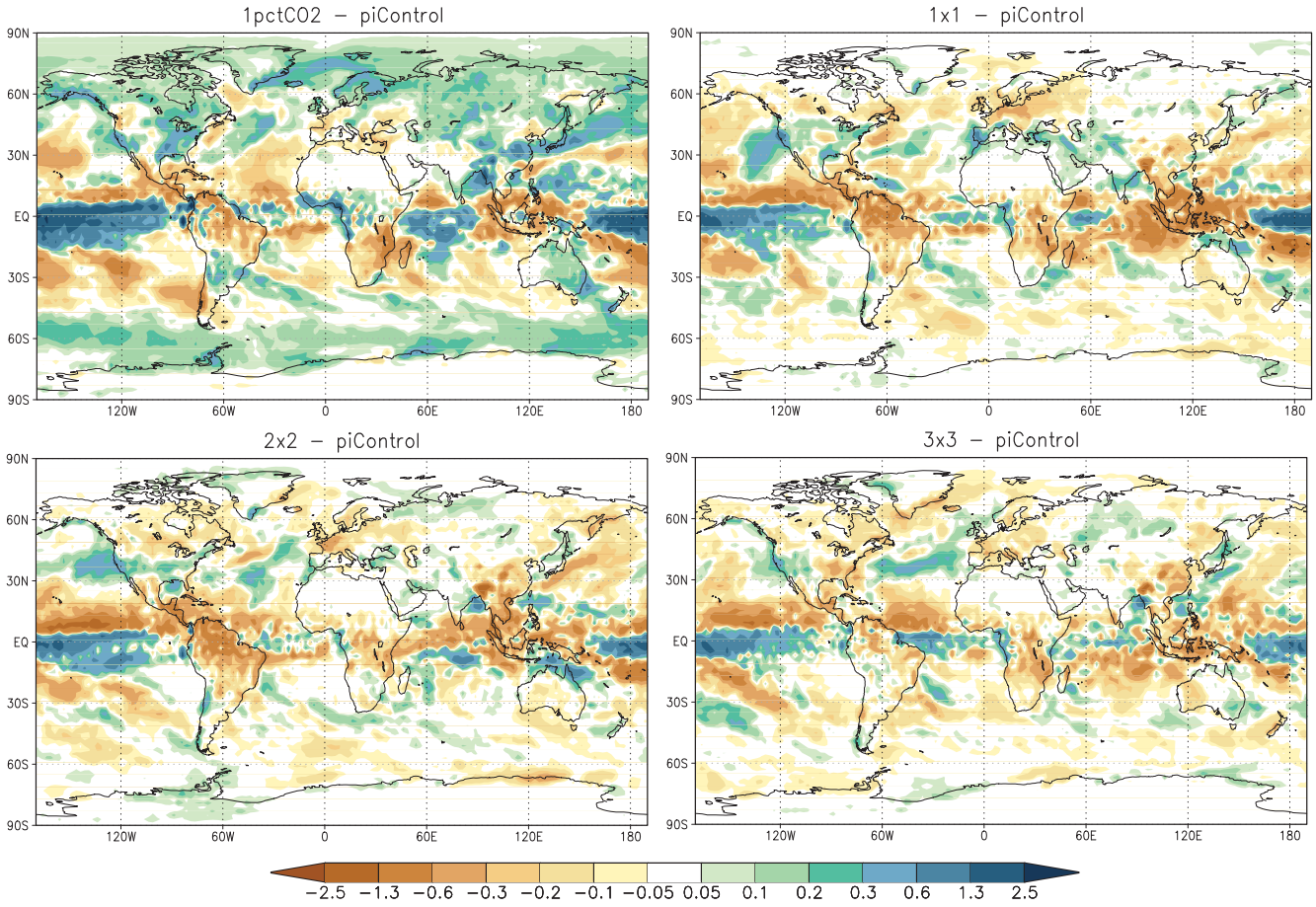

Figure 23. Same as Fig. 20 but for GISS ModelE2 (the evaluation model).

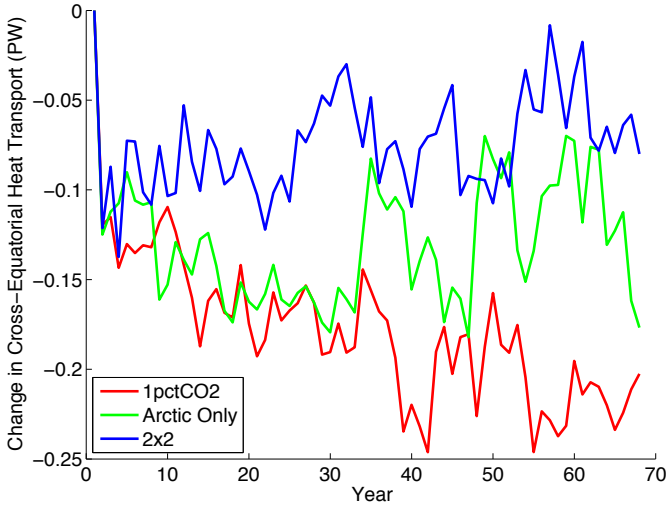

Figure 24. Change from the preindustrial control in cross-equatorial energy transport by the atmosphere (Eq. A1) for the 2×2 case in the design model. All values are annually averaged and expressed in PW. For clarity, all plotted values were annually averaged and then smoothed (5-point centered moving average).

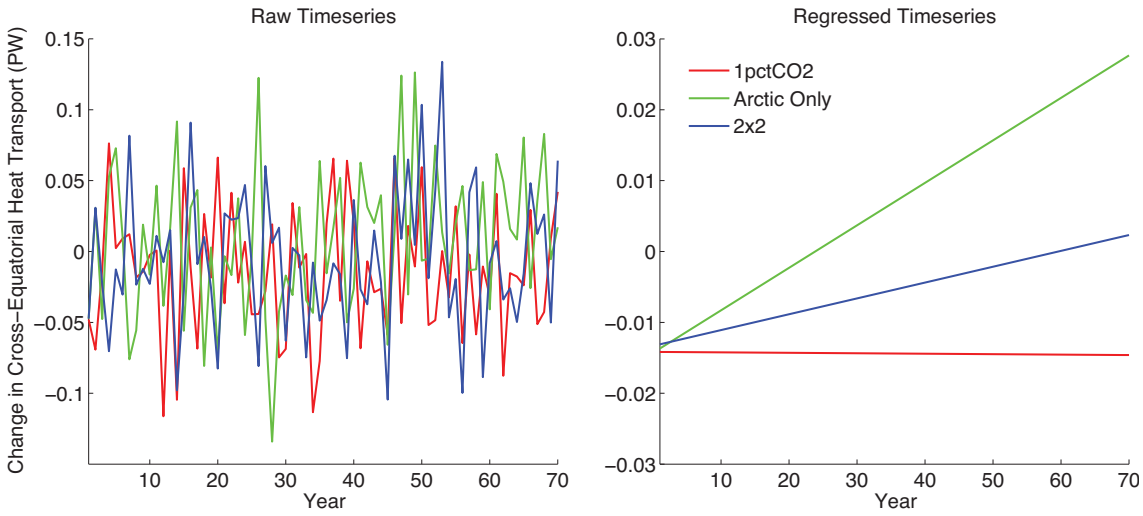

Figure 25. Similar to Fig. 24 but for GISS ModelE2 (the evaluation model). Left panel shows annually averaged change from the preindustrial control in cross-equatorial energy transport by the atmosphere (Eq. A1). Right panel shows ordinary least-squares linear regressions performed on those timeseries. All values are annually averaged and expressed in PW.