**DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD**

Can we test geoengineering?

(Article begins on next page)

Energy & Environmental Science

PAPER

Can we test geoengineering?

Douglas G. MacMynowski,[a] **David W. Keith,**[b] **Ken Caldeira**[c] **and Ho-Jeong Shin**[c]

Solar radiation management (SRM), a form of geoengineering, might be used to offset some fraction of the anthropogenic radiative forcing of climate as a means to reduce climate change, but the risks and effectiveness of SRM are uncertain. We examine the possibility of testing SRM through sub-scale deployment as a means to test models of climate response to SRM and explore risks prior to full-scale implementation. Contrary to some claims, this could provide meaningful tests of the climate's response to SRM within a decade. We use idealized simulations with the HadCM3L general circulation model (GCM) to estimate the response to SRM and signal-to-noise ratio for global-scale SRM forcing tests, and quantify the trade-offs between duration and intensity of the test and it's ability to make quantitative measurements of the climate's response to SRM forcing. The response at long time-scales would need to be extrapolated from results measured by a short-term test; this can help reduce the uncertainty associated with relatively rapid climate feedbacks, but uncertainties that only manifest at long time-scales can never be resolved by such a test. With this important caveat, the transient climate response may be bounded with 90% confidence to be no more than 1.5 °C higher than it's estimated value, in a single decade test that used roughly 1/10th the radiative forcing perturbation of a CO_2-doubling. However, tests could require several decades or longer to obtain accurate response estimates, particularly to understand the response of regional hydrological fields which are critical uncertainties. Some fields, like precipitation over land, have as large a response to short period forcing as to slowly-varying changes. This implies that the ratio of the hydrological to the temperature response that results from a sustained SRM deployment will differ from that of either a short-duration test or that which has been observed to result from large volcanic eruptions.

[a]*Control and Dynamical Systems, California Institute of Technology, USA*
[b]*School of Engineering and Applied Sciences, and Kennedy School of Government, Harvard University, USA*
[c]*Dept of Global Ecology, Carnegie Institution, USA*

Introduction

While not new,[1,2] the idea of deliberate solar radiation management (SRM) has attracted renewed interest. Suggested approaches include increasing the amount of light-scattering stratospheric aerosols[3] or increasing the reflectivity of low-altitude marine clouds.[4] There are enormous uncertainties about the risks and effectiveness of SRM. Many uncertainties could be

Environmental impact

Solar radiation management (SRM), a form of geoengineering, might reduce the risks of climate change, but its effectiveness and risks are uncertain. While it is clearly premature to consider any testing at a scale large enough to measure climate response, it is not premature to understand what we can learn from such tests. Indeed, understanding how testing might be linked to gradual deployment is crucial to understanding our ability to manage SRM's deep uncertainties. The possibility of tests has been the subject of a conference at Asilomar and several opinion pieces, but while discussions of ethics and governance have begun, no analysis yet published has attempted to quantify what tests could learn. We show that by modulating the forcing, it would be possible to learn something about the climate's response from a subscale test, but that these tests could take decades. This is critical for understanding the timetable required if we want to have this option available. We also show that neither a short-duration test nor a volcanic eruption would have the same effect as a sustained implementation; this matters because volcanic eruptions are often pointed to as proxies for geoengineering, yet the precipitation response can be quite different. Both of these conclusions are essential to inform the current discussions on geoengineering.

reduced through a systematic program of theory and modeling. Important uncertainties arise from poorly understood atmospheric processes operating at small scales, such as the transport of sea salt particles from the ocean surface to cloud base, or aspects of aerosol formation and coagulation in the stratosphere. The first step beyond laboratory studies might be open-atmosphere experiments aimed at resolving these uncertainties. Because these processes act at small scales it is possible to test them – albeit imperfectly – with experiments that are at a scale that is far too small to have any measurable climatic impact.

If there was ever a serious intent to deploy SRM, then some initial phase of testing at a reduced amplitude could first be used to reduce (not eliminate) uncertainty about the effectiveness and risks of SRM, by improving our understanding of the climate response to SRM forcing. (The response per W m^{-2} SRM forcing will not be the same as that to CO_2 with the same radiative forcing.) The utility of tests prior to any full-scale implementation has been raised before.[5–8] Of course, testing at any amplitude large enough to be detectable at a global scale presents substantial ethical and governance challenges.[7–9] At this scale, such activities could accurately be described as sub-scale deployment, nevertheless we use the word "tests" to describe these activities; indeed an appropriate test signal could be superimposed on a gradual ramp up of SRM.

In this paper we focus on understanding the limits of what such a test could tell us; and in particular, we examine the trade-off between the duration and magnitude of the test and its ability to provide useful measures of the climate response in the presence of noise. This question has received little attention, yet is crucial to an understanding of our ability to manage SRM's deep uncertainties. (We do not address other important questions regarding the ability to conduct such a test, including the social and political implications of such testing, the technology required to produce a desired radiative forcing, or other impacts (e.g. ozone loss) of producing the radiative forcing that are specific to the forcing method.) While it is clearly premature to begin any large spatial-scale test, it is not premature to consider the implications. For SRM approaches to be available as an option by, for example, 2050, as an insurance policy against either high climate sensitivity or insufficient emissions reductions, then we would either need to begin tests decades earlier, or face the prospect of decisions about a full-scale implementation without sufficient information to ensure that we understand the effects.

Estimating the climate response to forcing is primarily a question of identifying a signal in the presence of the background "noise" of natural climate variability. The time required to detect a small signal may be significant, and thus there is a trade-off between the amplitude of the introduced perturbation in radiative forcing during a test, the length of time, and the uncertainty in estimating the effect on any relevant climate variable. Here our goal is to provide a quantitative test of assertions that "…geoengineering cannot be tested without full-scale implementation".[9] The response of climate models to radiative forcing is remarkably linear at the global scale, as illustrated both here (we quantify below the linearity of both temperature and precipitation) and elsewhere.[10,11] Thus while we agree with many of the points made in [9], our results demonstrate that useful knowledge can be obtained without full-scale

implementation. Note that there may also be nonlinearities involved in creating a desired radiative forcing (e.g., in the aerosol size distribution if SRM is implemented via stratospheric aerosols[12,13]); however, that issue is distinct from understanding the climate response to an applied radiative forcing.

It will be difficult to distinguish between the effects of a small-amplitude forcing that is constant or slowly varying, and gradual changes due to other anthropogenic forcings or unforced climate variability. More information can thus be obtained using periodic (e.g., on/off/on/off…) or pseudo-random forcing, possibly superimposed on a gradual ramp-up of initial SRM deployment, and estimating the correlated climate response signal. Since the climate responds differently to forcing at different time-scales, the response measured in such a test would need to be extrapolated from the response at a relatively short period (a few years) to estimate the response on longer (century scale) time-scales most relevant to climate policy.

There are clear limitations in the ability of such a test to estimate the effects resulting from long time-scale feedbacks. In a model with a single time-constant, the feedback affects only the equilibration time and not the response to perturbations that vary much faster than this time-constant.[14,15] The case with multiple time-constants can be illustrated by a two-box model: after a few years, the "fast" dynamics (associated with the atmosphere and ocean mixed-layer) have equilibrated, and the two-box system exhibits a quasi-equilibrium response roughly equivalent to the transient climate response or TCR.[16] In general, a short-term time-varying test will yield information about these "fast" dynamics, with time constants faster than the test signal period. Note that many of the feedbacks that contribute most to the uncertainty in predictions of century-scale climate change, e.g., cloud, snow/ice-albedo, lapse-rate, and water vapor,[17,18] act sufficiently rapidly so that their effects on climate response would be apparent in a test that used short-period modulation, while uncertainties that manifest only at longer time-scales (e.g., due to uncertain ocean circulation changes) would not be resolved by such a test. This is a fundamental limitation of any short-term test, but does not mean that useful knowledge could not be gained.

It is also worth noting that because the land temperature responds to a radiative forcing perturbation more rapidly than the ocean,[19] and land-sea temperature contrast influences monsoonal precipitation,[20] then the relative precipitation response from a dynamic SRM test is much larger than what would be expected from a more slowly time-varying SRM implementation, particularly over areas such as the Indian sub-continent. Similarly, one should expect that the ratio of precipitation changes to temperature changes resulting from short-duration volcanic events[21] would be greater than those induced by SRM implementation with the same radiative forcing.

We explore these issues using HadCM3L GCM simulations with periodic forcing between 2- and 64-year periods. Since our goal is to estimate the detectability of atmospheric response, and not to choose a particular SRM scheme, we varied the solar forcing directly as a proxy for any SRM scheme that would produce global scale reduction of radiative forcing, as in [22–24]. While these and other studies have explored the effects of deploying SRM,[25–27] this is the first study to focus on the global-

scale testing phase. We focus on the changes in the temperature and precipitation that are correlated with the forcing signal, at the global scale, and over the Indian subcontinent as an example of regional scale.

In the presence of natural climate variability, then accurate estimates of the response (*e.g.*, 25% uncertainty) will require several decades, even with a test that introduced a 1 W m^{-2} maximum perturbation; this is a significant fraction of the amplitude required to offset a CO_2-doubling. The signal-to-noise ratio (SNR) is smaller for precipitation than for global temperature, and smaller still for regional-scale effects, requiring either higher amplitude testing or a longer test. Indeed, even a "full-scale" implementation would take time to accurately assess effects. However, to usefully bound the response, much less time and/or smaller radiative forcing perturbations are required.

Simulations

The HadCM3L fully-coupled atmosphere-ocean GCM from the UK Met Office is used here to estimate the climate response to SRM forcing and the SNR for multiple fields at different spatial scales as a function of the forcing period. HadCM3L has reduced ocean resolution compared to the more extensively used HadCM3. The model resolution is 3.75° in longitude by 2.5° latitude in both the atmosphere and ocean, with 19 vertical levels in the atmosphere and 20 in the ocean.[23,28,29] The version used here avoids the use of flux adjustment by removing Iceland,[28] and has climate sensitivity of 3 °C, similar to that of HadCM3; the transient climate response (temperature change averaged over 60–80 years due to a 1% annual increase in CO_2, see ref. 30 (p. 629)) is 2.2 °C. In addition to a reasonable model representation of climate sensitivity, the key characteristics we rely on here are a reasonable representation of precipitation and of climate variability. At the spatial scales considered herein, the climate variability of HadCM3L is quite similar to the real climate variability, as shown below in Fig. 4; it also captures ENSO.[31] This model has also been used in a slightly different configuration for exploring regional effects of SRM.[27] HadCM3, which has the same atmospheric model but higher ocean resolution, has been shown to have a reasonably realistic monsoon precipitation.[32]

Any desired time-varying perturbation can be expressed as a sum of its frequency components, so the frequency response provides a useful way to explore the behavior of the system to time-varying inputs.[33] Twelve 500-year simulations were conducted, each with sinusoidally varying solar constant at periods from 2- to 64-years and at 0.5%, 1% and 2% maximum variation. Additional simulations were run with no change in forcing to obtain variability statistics, and an ensemble of three 100-year simulations with a ramp decrease in solar constant that gives radiative forcing of 3.7W m^{-2}, equivalent to $2 \times CO_2$, at 70 years. We use the average from 60–80 years from these last simulations to be representative of the longer time-scale response relevant for SRM implementation; this gives the *transient climate response* for SRM forcing. All simulations used fixed pre-industrial greenhouse gas concentrations, and all of the response information herein is normalized by the perturbation in absorbed solar radiation, roughly 70% of the variation in incident solar radiation.

The response of each model climate field to the periodic SRM forcing was obtained by discarding the first 50 years of simulation to avoid initialization transients, and computing the Fourier transform of monthly averaged fields over an integer number of forcing cycles. This gives the magnitude and phase of the response that is correlated with the time-varying forcing; 450 years of output are sufficient to give less than a few percent error at the global scale. All of the response information in Fig. 1–4 plot this correlated component, per W m^{-2} forcing. At a 2-year forcing period, the results depend strongly on the phase of the forcing relative to the seasonal cycle (consistent with ref. 34); at 4-year and longer periods the correlated component evaluated here on monthly model output is nearly the same as evaluating the change in the annual mean climate that is correlated with the forcing.

The surface air temperature and precipitation responses at a 2-, 4- (representative of an SRM test), and a 64-year forcing period (more representative of SRM implementation) are shown in Fig. 1. There is some delay between the maximum of the
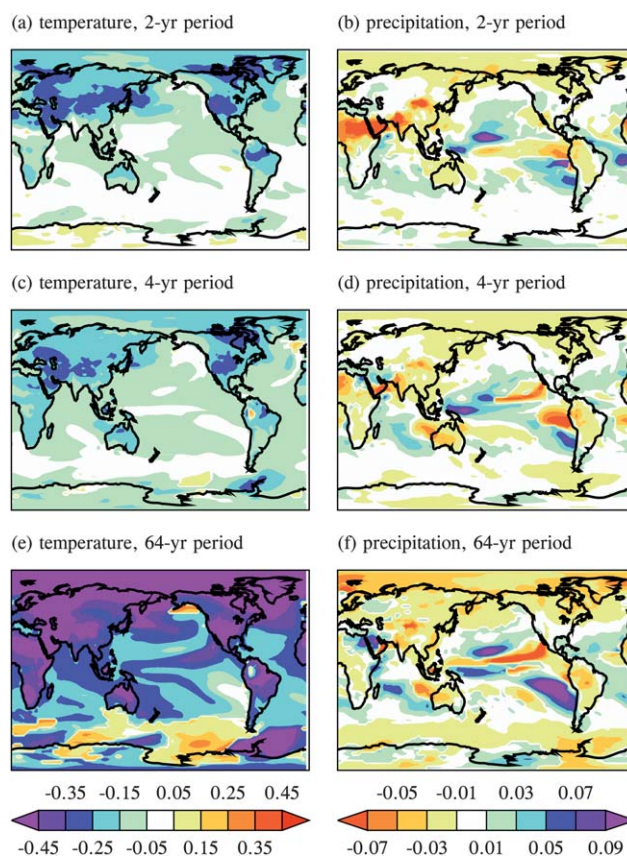

Fig. 1 Frequency dependence of the climate response that is correlated with the forcing: surface temperature (°C per W m^{-2}, left) and precipitation (relative change compared to baseline, per W m^{-2}, right), due to forcing at 2 yr (top), 4 yr (middle) and 64 yr (bottom row) periods. The middle and bottom cases are representative of what might be expected from an SRM test and SRM implementation respectively. The sign is estimated from the phase of the response and plotted as negative (decreased temperature or precipitation when solar radiation is decreased) if the phase is between a 45° lead and a 135° lag; blue thus indicates cooling or increased precipitation due to an SRM perturbation.

sinusoidal radiative forcing and the maximum response. At short periods, the dominant temperature response is over land. For longer periods, the average response increases. Both of these are expected from considerations of thermal inertia. In addition, the response pattern changes, due in part to spatially varying thermal inertia; this raises issues regarding the extrapolation of estimated responses from short-period tests (or volcanic eruptions) to longer time-scales.

The frequency response of several spatially-averaged fields is shown in Fig. 2. Of particular concern for SRM approaches is the impact on rainfall.[24,26] The global land-averaged precipitation has a different dependence on perturbation period than the global mean temperature, and the relative change in precipitation for a given change in temperature is thus much larger for short-period perturbations than for long. Precipitation changes over India have also been noted for volcanic events;[21] while such changes can be expected from short-period SRM tests, the effect for implementation should be much smaller. In this model, and depending on the the phase of the forcing relative to the seasonal cycle, the ratio of Indian precipitation to global temperature response at a 2-year forcing period can be ∼7 times larger than the ratio at long time-scales (given by the transient climate response defined earlier). Similar behavior also exists for other regions, see Fig. 1.
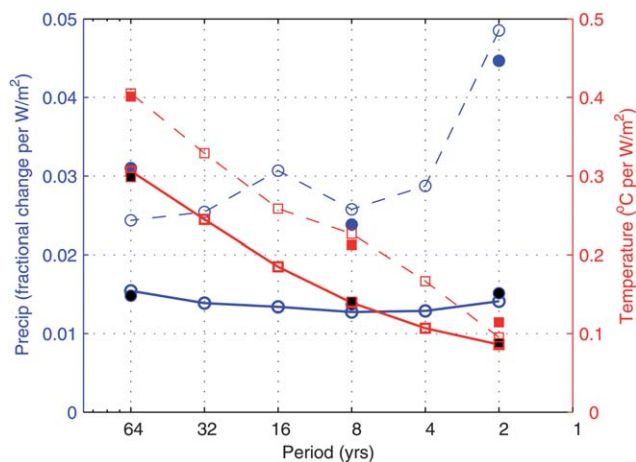
Fig. 2 Response to solar radiative forcing at different periods: comparison of temperature response (red '□', right-axis) and land-average fractional change in precipitation (blue '○', left-axis). Solid lines are global averages, dashed lines show the response for the Indian subcontinent, defined here as the land-average over 8.75–33.75°N and 73.125–91.875°E (chosen to include full grid-boxes and capture the dominant monsoonal precipitation). The degree of linearity is also illustrated; the response per W m^{-2} obtained with forcing at half the amplitude (0.5% instead of 1%) is shown for the 2, 8, and 64-year periods with solid symbols (black for global response, red/blue for India). Short period forcing results in significantly larger precipitation response per unit temperature response than long period forcing. For the Indian subcontinent, this ratio of responses is a factor of 6 larger at a 2-year period than it is at a 64-year period. The phase of the 2-year data point here corresponds to maximum/minimum forcing in Northern Hemisphere (NH) summer; this is appropriate for maximizing the signal in the Arctic temperature, but also leads to a more significant Indian monsoon response than for a phase chosen to peak in NH winter.
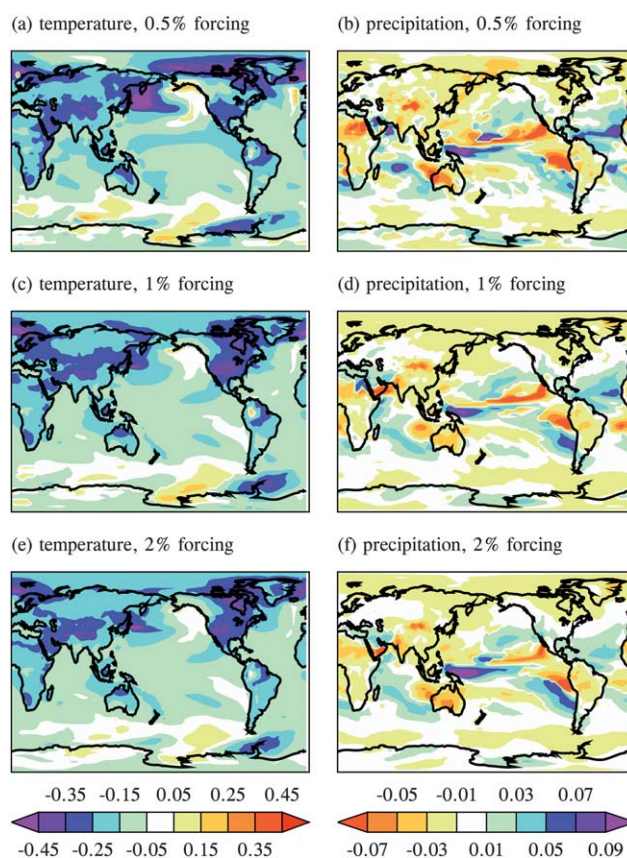
Fig. 3 Linearity of response to forcing: surface air temperature response (°C per W m^{-2}, left) and precipitation response (change relative to baseline, per W m^{-2}, right), estimated for solar perturbations of 0.5%, 1%, and 2% forcing amplitude at an 8-year period. The sign is computed as in Fig. 1. The difference in response per W m^{-2} SRM between tests conducted at 1% (2.4 W m^{-2}) and 2% (4.8 W m^{-2}) change in radiative forcing is relatively small, but the difference is higher in a few regions between 1% and 0.5% changes. While the global mean response is reasonably linear, local nonlinearities can be significant, particularly at small amplitudes.

The global mean response to radiative forcing is quite linear in this model at the magnitudes and durations under consideration here (see additional data points in Fig. 2). Nonlinearity could be a more significant issue in some regions of the world; see Fig. 3. However, in this model, the grid-scale temperature response per W m^{-2} SRM estimated for a 0.5% periodic change in solar radiation (1.2 W m^{-2}) is within 25% of the value resulting from a forcing that is 4 times larger, for more than 68% of the Earth surface area, and nearly 80% of the land surface area. The precipitation response is less linear, with the grid-scale response per W m^{-2} SRM forcing for the 0.5% case falling within 25% of the value for the 2% forcing case over 1/3 of the Earth surface, and within 33% for half the surface area. There are also nonlinearities in the relationship between radiative forcing and aerosol injection rate,[12,13] for example, but these do not affect the ability to estimate the climate response to radiative forcing perturbations at an amplitude smaller than expected for SRM implementation.

Thus, a short-term small-amplitude SRM test will not give the same response per W m^{-2} SRM forcing as full-scale

implementation, and models would still be required to extrapolate to full-scale behavior. Because the model system response is mostly linear, the amplitude mostly affects the ability to estimate the signal in the presence of noise. In contrast, the frequency dependence strongly affects characteristics of the signal. This is particularly important in comparing the relative reductions in precipitation and temperature, especially on a regional scale.

Detection

The information in Fig. 2 is useful in understanding the trade-off between forcing amplitude (u in W m^{-2}), length of test (N in years), and the uncertainty in estimating the response of any particular variable at the forcing frequency (σ, dimensionless ratio of standard deviation of an estimate to its value). This trade-off depends on the SNR. The response of some variable of interest due to forcing (the signal) is measured in response units, *e.g.* °C, per W m^{-2}. The broadband natural variability (the noise) has amplitude spectrum measured in response units per square root of frequency, *e.g.* °C per (1/year)$^{1/2}$. We define the *normalized* SNR s as the ratio of these two values, in (1/year)$^{1/2}$ per W m^{-2}. Averaging reduces the variance of the estimated response inversely with the length of the test, thus after N years the standard deviation of the error in estimating the response, normalized by the response itself, is

$$\sigma = \frac{1}{su\sqrt{N}} \qquad (1)$$

This is just the *un-normalized* SNR of the test, that is, the ratio of the standard deviation to the mean response. Also note that the variability statistics are nearly Gaussian (using a Lilliefors test on the annual-mean variables considered here then the differences in the distributions from Gaussian are not statistically significant for either the model or the detrended data sources described below). Hence with no prior being used here, the uncertainty distribution on the estimated response is also Gaussian.

The SNR obtained for varying incoming solar forcing in HadCM3L is shown in Fig. 4 for surface air temperature and land-average fractional precipitation changes, for both the globe and the Indian sub-continent. Each plot shows the response as a function of the forcing period, and the natural variability amplitude spectrum estimated from both the model and monthly anomaly data. The actual climate variability is obtained from NOAA National Climatic Data Center (reconstructed data from 1880–2009) for the global mean surface air temperature, and from NCEP/NCAR reanalysis[35] from 1948–2009 for the remaining fields. The shorter data record results in a more uncertain spectrum, and thus the deviations between model and actual spectra should not be over-interpreted. Since there is good agreement between the model and the actual spectrum of natural
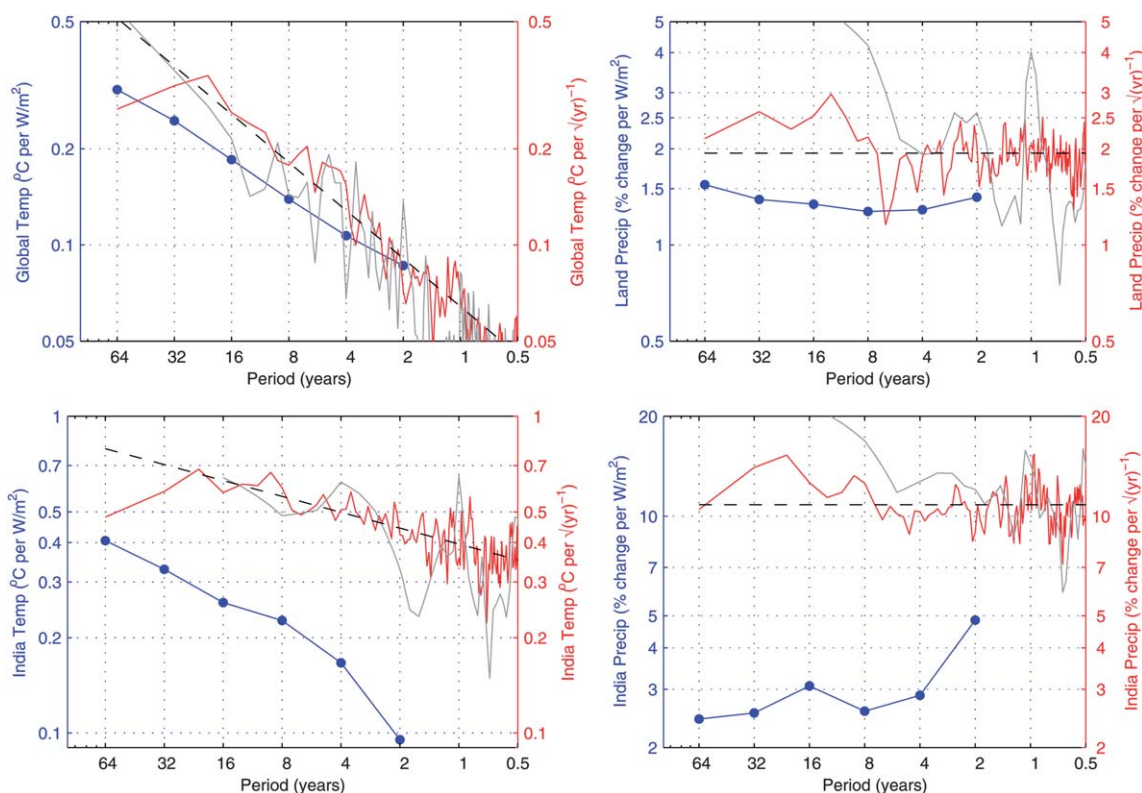
Fig. 4 Signal-to-noise ratio (SNR) for climate forcing at different periods and for different fields. Upper plots are for global mean temperature (left) and global-land-averaged precipitation (right), while lower plots are averaged only over the Indian subcontinent. Each plot includes the response to 1 W m^{-2} forcing (blue circles, left axis) and the amplitude spectrum of natural variability per $\sqrt{\text{yr}^{-1}}$ (red, right axis) obtained from the model, and with a fit (dashed line). The plots are thus normalized to show the effective SNR per year of test at a 1 W m^{-2} amplitude. The actual background spectrum (gray) is estimated from monthly anomalies using NOAA NCDC data from 1880–2009 for the global mean temperature and using NCEP/NCAR reanalysis from 1948–2009 for the remaining fields.

climate variability, but the former is better known, we use the model spectrum for SNR calculations; uncertainty in the amplitude of variability would not affect the estimated response to SRM, but would affect the estimated confidence in the response-estimates. The background variability of global mean temperature is known to have a $1/f$ power spectrum dependence on frequency f; the best fit to this is shown. The precipitation variability is white noise for these time-scales at both the global and regional scale (best fits shown), while the temperature variability over India appears to have a power spectrum dependence of $f^{-2/3}$, and this is used in the fit shown. Some variability can be attributed to known factors (*e.g.*, volcanoes, El Niño), and thus our SNR estimates are slightly conservative.

The SNR is not strongly dependent on frequency for any of the fields, although slightly higher for global mean temperature at shorter periods, and the 2-year period case has greater SNR for Indian precipitation due to the larger response. This two-year response is not representative of the long-term response to sustained forcing, and also depends on the phase relative to the seasonal cycle. It would thus be preferable to avoid these very short periods, but otherwise the frequency or frequency content of forcing is not critical (it may also be worth using periods long enough to avoid excitation of or attribution issues associated with El Niño, which has a period ranging from 2–7 years). The forcing signal may be chosen with multiple frequencies to understand climate response at different time-scales, but this decision does not have a significant influence on SNR and hence on the required length of a test. The SNR depends on the root-mean-square amplitude of the time-varying forcing, so an on/off/on/off pattern yields the same SNR as a sinusoidal variation of the same peak magnitude, used in constructing Fig. 4.

We estimate the SNR based on a 4-year forcing period, and normalize by a 1 W m^{-2} forcing perturbation in absorbed solar radiation and per $(year)^{1/2}$ of test-time (this approximation is valid only for tests longer than the forcing period). This gives, for this model:

Global mean surface air temperature:	$s \simeq 0.83$
Global land-average precipitation change:	$s \simeq 0.66$
Indian sub-continent surface air temperature:	$s \simeq 0.33$
Indian sub-continent land precipitation:	$s \simeq 0.27$

The global mean temperature SNR is consistent with the estimate in ref. 5. A factor of two in SNR requires either double the forcing amplitude or four times the length of the test to achieve comparable accuracy, so the effect of the reduced SNR at regional scales is quite significant. Using these SNRs, the implications of eqn (1) are illustrated in Fig. 5 for a 1 W m^{-2} case. At half this amplitude, comparable accuracy requires four times as much time. Note that the climate variability limits the absolute estimation error; using the SNR means that the relative accuracies in Fig. 5 are normalized by the response per W m^{-2} SRM forcing of this particular model.

Accurate estimates at a local scale would require greater time or larger forcing. The normalized uncertainty $\sigma(N)$ in (1) can be computed from the SNR at each location; this assumes that the variance of the (Gaussian) natural variability is known, rather than simultaneously estimating this variance during the test. The time required to detect a change with 95% confidence is the time at which $1/\sigma(N) < 1.65$ (the 95th percentile of the Gaussian
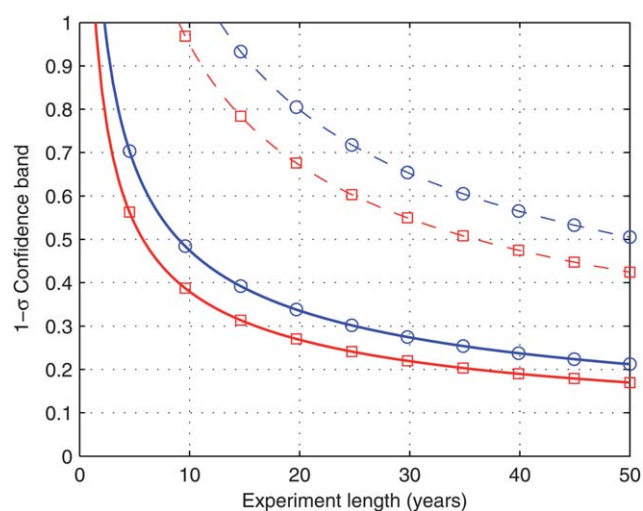

Fig. 5 Testing time required to achieve a specified accuracy in estimating the response to SRM forcing (standard deviation, normalized by the estimate) for 1 W m^{-2} forcing; the required time scales inversely with the square of the forcing amplitude. Temperature (red '□') and land-average precipitation (blue '○'). Fields averaged only over India (dashed) require significantly longer than global averages (solid) to obtain comparable accuracy.

distribution). This will only give an estimate for the detection time, since the variability statistics themselves are likely to also be varying during the test due to climate change. At the grid scale of this model (3.75° by 2.5°), and using a 1 W m^{-2} test, the time required to detect a local-scale temperature change with 95% confidence exceeds 50 years for most of the planet, longer for detecting local-scale precipitation changes (see Fig. 6), and longer still to provide accurate response estimates. This implies that it will be difficult to confidently attribute local changes to any SRM test. This inability to detect such changes locally means both that it is difficult to establish the spatial pattern of change resulting from a small-scale test and that the local temperature and precipitation changes are likely to be undetectably small.

The discussion above relates to estimating the response to SRM forcing of any particular field to some specified accuracy; next we discuss the potential of tests to rule out the possibility of high sensitivity. After N years of a test with peak amplitude of u W m^{-2} and background noise of n response-units per $(1/year)^{1/2}$, there is a 90% confidence that the actual response per W m^{-2} SRM is less than its measured value plus $1.28n/(u\sqrt{N})$.
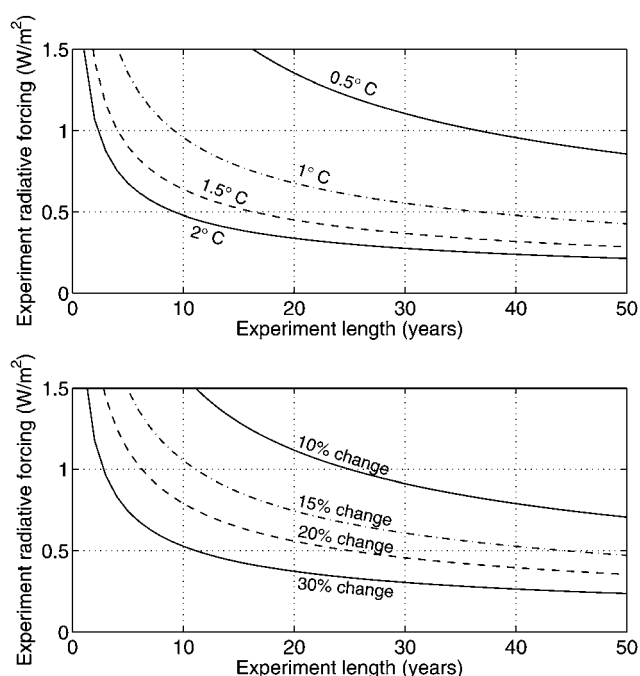
As noted earlier, effects on temperature and precipitation would need to be extrapolated from the values measured at the period of the test to estimate the response on longer time-scales relevant for full-scale deployment. For precipitation, the response of the HadCM3L model is roughly independent of frequency for forcing slower than a 4-year period. For temperature changes, we use the ratio of the transient climate response (defined earlier) to the response at the forcing period in the HadCM3L model (roughly a factor of 3.5), and this could be done for other models. While this will give an estimate of the transient climate response, uncertain climate feedbacks that change the response at long time-scales will also change the system time-constants,[14,15] and so this ratio of responses can itself depend on the climate sensitivity.
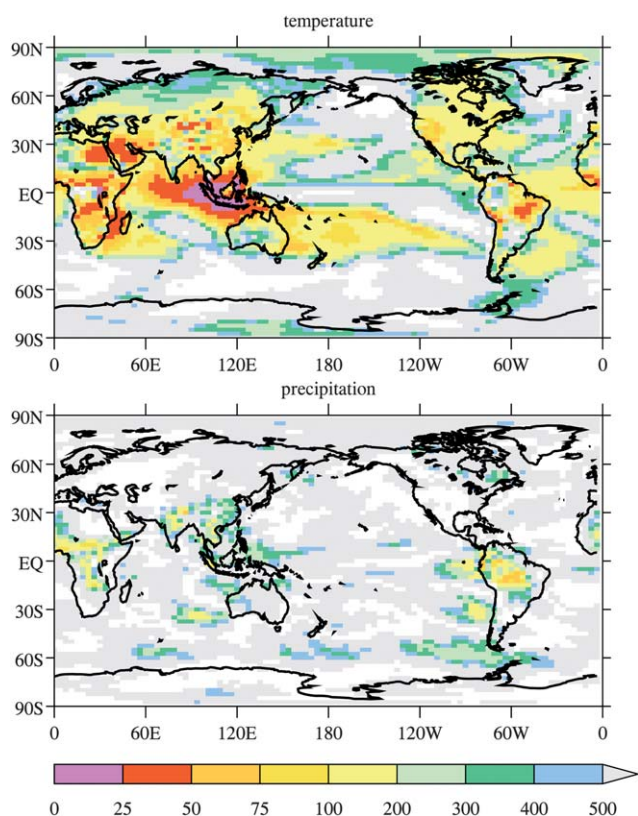
Fig. 6 Time required (years) to detect a change with 95% confidence on a grid-box scale (3.75° by 2.5°) for a test with forcing level of 1 Wm^{-2} for temperature (top) and precipitation (bottom). Confidently attributing specific local effects to an SRM test will be challenging.

Fig. 7 Required test time and forcing amplitude to achieve a 90% confidence that the actual climate response to a 3.7 W m^{-2} radiative forcing perturbation does not exceed the measured response by more than a given bound for global mean temperature (top) and regional Indian precipitation (bottom). The 4-year period response is extrapolated to the predicted error in transient climate response for this model.

Fig. 7 shows the combination of forcing and length of test that would be required to constrain the maximum error in estimating the response per W m^{-2} SRM forcing, for the global mean temperature and the Indian precipitation change. The response is scaled to the longer-time response of this model as described above, and scaled by 3.7 W m^{-2} to illustrate the error in constraining the transient climate response either due to $2 \times CO_2$ (assuming comparable sensitivity for the same radiative forcing) or to the SRM forcing required to fully counteract that.

Relatively short tests would thus significantly constrain our current uncertainty about high climate sensitivity. Current estimates of the climate sensitivity [ref. 30, p. 749] of 2–4.5 °C ("likely" range) have a normalized 1-σ confidence of roughly 0.4 (*i.e.* 3.2 ± 1.3), although the distribution is not Gaussian; the normalized uncertainty in the transient climate response ("very likely" range 1–3 °C) is about 0.3. A 20-year test at 1 W m^{-2} would thus improve our knowledge (1-σ confidence from Fig. 5 of 0.25), while a single-decade test with even 1/10th the radiative forcing from $2 \times CO_2$ could still provide a useful constraint on the chance that the transient climate response is above the high end of IPCC estimates. Given that climate impacts and thus climate policy are driven by the possibility that climate sensitivity is at the high-impact low-probability tail of the distribution, such a subscale test might make an important contribution to the assessment of climate risks, which may, in turn, improve our ability to manage those risks.

Finally, all of the estimates here are upper bounds on the required time in the sense that they are based purely on signal detection and do not take into account either any priors nor any understanding of physics. Fingerprint analysis has been quite successful in global warming attribution.[30] Similarly, multiple fields could be considered in order to test between different hypotheses predicted by different models regarding the impact of SRM on some particular response such as the Indian monsoon. Detection at the global scale could also be improved by taking advantage of the optimal spatial pattern for discriminating the signal, as in analysis of the response to the solar cycle.[36] Note that while this would improve detection time, it results in an estimate of the response of a particular spatial pattern, not the global mean.

Summary

Simulations with the HadCM3L GCM illustrate trade-offs associated with small-amplitude SRM testing that might precede any possible future deployment. The relative accuracies quoted herein depend on the sensitivity of the model used, and will therefore vary for different models, while the absolute accuracy is a function of only the background variability and not the model.

Time-dependent forcing can yield more information, however, the climate responds differently at different time-scales, and in particular, there can be a much larger precipitation response for short-period forcing than for gradual changes. In this model, the change in Indian precipitation for a given change in global mean temperature can be a factor of 7 higher for a two-year forcing period than it is for long time-scale changes (here we use the

transient climate response, defined as the response due to a steady increase in radiative forcing equivalent to $2 \times CO_2$ at 70 years). It would be valuable to understand these relative frequency response characteristics in different models. The time-scale of forcing is also important in using volcanic events to understand SRM; volcanoes are not direct analogues for understanding expected precipitation changes, although the ability to successfully reproduce the consequences of volcanic aerosols would increase confidence in predictions for a sustained aerosol layer. A related issue with short-duration tests is that they will yield estimates of the response to short-period forcing, which need to be extrapolated to estimate the response on longer time-scales. While some use of models is thus still required, the information gained would nonetheless reduce SRM risks and help validate models.

The trade-off between forcing amplitude, length of test, and confidence in estimating climate response can be estimated from our simulations and the spectrum of natural climate variability (which is quite similar to the model variability at the spatial scales considered here). A 1 W m^{-2} test would require about 20 years to estimate the global mean temperature response to SRM to ~25% accuracy, but ~50 years to estimate the precipitation response over India to 50% accuracy. However, high accuracy is not required to constrain the probability of outliers. *E.g.*, if the transient climate response to SRM is x °C, then a single decade with 0.4 W m^{-2} periodic forcing is sufficient to constrain the upper bound on the error in estimating the transient climate response to be less than $x + 1.5$ °C with 90% confidence. Any test will only probe Earth system feedbacks that respond detectably on the time scales of the test, and thus these estimates do not include the uncertainty in extrapolating from short time-scale response to long. Nonetheless, an initial test (or sub-scale deployment phase) could provide important tests of the climate's response to geoengineering within a decade, although accurate estimates could require several decades or longer. Testing cannot eliminate uncertainty about the risks posed by geoengineering, but testing by modulation could improve understanding of risks of geoengineering and might also constrain our estimate of the climates sensitivity to CO_2.

Acknowledgements

References

1 S. H. Schneider, Geoengineering: Could or Should we do it?, *Clim. Change*, 1996, **33**, 291–302.
2 D. Keith, Geoengineering the Climate: History and Prospect, *Annu. Rev. Energy Environ.*, 2000, **25**, 245–284.
3 J. I. Katz, Stratospheric albedo modification, *Energy Environ. Sci.*, 2010, **3**, 1634–1644.
4 J. Latham, P. Rasch, C.-C. J. Chen, L. Kettles, A. Gadian, A. Gettelman, H. Morrison, K. Bower and T. Choularton, Global temperature stabilization *via* controlled albedo enhancement of low-level maritime clouds, *Phil. Trans. R. Soc. A*, 2008, **366**(No. 1882), 3969–3987.
5 J. J. Blackstock, D. S. Battisti, K. Caldeira, D. M. Eardley, J. I. Katz, D. W. Keith, A. A. N. Patrinos, D. P. Schrag, R. H. Socolow, and S. E. Koonin, *Climate Engineering Responses to Climate*

Emergencies, Novim, 2009. Archived online at: http://arxiv.org/pdf/0907.5140.
6 K. Caldeira and D. W. Keith, The need for Climate Engineering Research, *Issues Sci. Technol.*, 2010, 57–62.
7 D. W. Keith, E. Parson and M. Granger Morgan, Research on global sun block needed now, *Nature*, 2010, **463**, 426–427.
8 J. J. Blackstock and J. C. S. Long, The Politics of Geoengineering, *Science*, 2010, **327**, 527.
9 A. Robock, M. Bunzl, B. Kravitz and G. L. Stenchikov, A test for Geoengineering?, *Science*, 2010, **327**, 530–531.
10 P. M. de F. Forster, M. Blackburn, R. Glover and K. P. Shine, An examination of climate sensitivity for idealised climate change experiments in an intermediate general circulation model, *Clim. Dyn.*, 2000, **16**, 833–849.
11 G. J. Boer and B. Yu, Climate sensitivity and response, *Clim. Dyn.*, 2003, **20**, 415–429.
12 P. Heckendorn, D. Weisenstein, S. Fueglistaler, B. P. Luo, E. Rozanov, M. Schraner, L. W. Thomason and T. Peter, The impact of geoengineering aerosols on stratospheric temperature and ozone, *Environ. Res. Lett.*, 2009, **4**, 045108.
13 J. R. Pierce, D. K. Weisenstein, P. Heckendorn, T. Peter and D. W. Keith, Efficient formation of stratospheric aerosol for climate engineering by emission of condensible vapor from aircraft, *Geophys. Res. Lett.*, 2010, **37**, L18805.
14 J. Hansen, G. Russel, A. Lacis, I. Fung and D. Rind, Climate response times: Dependence on climate sensitivity and ocean mixing, *Science*, 1985, **229**, 857–859.
15 G. Roe, Feedbacks, Timescales, and Seeing Red, *Annu. Rev. Earth Planet. Sci.*, 2009, **37**, 93–115.
16 I. M. Held, M. Winton, K. Takahashi, T. Delworth, F. Zeng and G. V. Vallis, Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, *J. Clim.*, 2010, **23**, 2418–2427.
17 S. Bony, *et al.*, How well do we understand and evaluate climate change feedback processes?, *J. Clim.*, 2006, **19**, 3445–3482.
18 K. Zickfeld, M. G. Morgan, D. J. Frame and D. W. Keith, Expert judgments about transient climate response to alternative future trajectories of radiative forcing, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 12451–12456.
19 B. Dong, J. M. Gregory and R. T. Sutton, Understanding land-sea warming contrast in response to increased greenhouse gases. Part I: Transient adjustment, *J. Clim.*, 2009, **22**, 3079–3097.
20 J. M. Wallace, and P. V. Hobbs, *Atmospheric Science: An Introductory Survey*, Academic Press, 1977.
21 K. E. Trenberth and A. Dai, Effects of Mount Pinatubo volcanic eruption on the hydrological cycle as an analog of geoengineering, *Geophys. Res. Lett.*, 2007, **34**, L15702.
22 B. Govindasamy and K. Caldeira, Geoengineering Earth's radiation balance to mitigate CO2-induced climate change, *Geophys. Res. Lett.*, 2000, **27**, 2141–2144.
23 D. J. Lunt, A. Ridgwell, P. J. Valdes and A. Seale, "Sunshade World": A fully coupled GCM evaluation of the climatic impacts of geoengineering, *Geophys. Res. Lett.*, 2008, **35**, L12710.
24 G. Bala, P. B. Duffy and K. E. Taylor, Impact of geoengineering schemes on the global hydrological cycle, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 7664–7669.
25 A. Jones, J. Haywood, O. Boucher, B. Kravitz and A. Robock, Geoengineering by stratospheric SO2 injection: results from the Met Office HadGEM2 climate model and comparison with the Goddard Institute for Space Studies ModelE, *Atmos. Chem. Phys.*, 2010, **10**, 5999–6006.
26 A. Robock, L. Oman and G. Stenchikov, Regional climate responses to geoengineering with tropical and Arctic SO2 injections, *J. Geophys. Res.*, 2008, **113**, D16101.
27 K. L. Ricke, M. Granger Morgan and M. R. Allen, Regional climate response to solar-radiation management, *Nat. Geosci.*, 2010, **3**, 537–541.
28 C. Jones, A fast ocean GCM without flux adjustments, *J. Atmos. Oceanic Technol.*, 2003, **20**, 1857–1868.
29 P. M. Cox, R. A. Betts, C. D. Jones, S. A. Spall and I. J. Totterdell, Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, *Nature*, 2000, **408**, 184–187.
30 IPCC, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 2007.

31 C. D. Jones, M. Collins, P. M. Cox and S. A. Spall, The carbon cycle response to ENSO: A coupled climate-carbon cycle model study, *J. Clim.*, 2001, **14**, 4113–4129.

32 H. Annamalai and K. Hamilton, The South Asian summer monsoon and its relationship with ENSO in the IPCC AR4 simulations, *J. Clim.*, 2007, **20**, 1071–1092.

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

34 B. Kravitz and A. Robock, Climate effects of high-latitude volcanic eruptions: Role of the time of year, *J. Geophys. Res.*, 2011, **116**, D01105.

35 Kalnay, *et al.*, The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 1996, **77**, 437–470.

36 K. K. Tung and C. D. Camp, Solar cycle warming at the Earth's surface in NCEP and ERA-40 data: A linear discriminant analysis, *J. Geophys. Res.*, 2008, **113**, D05114.