THE GEOENGINEERING MODEL INTERCOMPARISON PROJECT (GEOMIP): A CONTROL PERSPECTIVE

Andrew Jarvis and David Leedal,

Lancaster Environment Centre, Lancaster University, Lancaster, UK, LA1 4YQ.

Abstract

The Geoengineering Model Intercomparison Project (GeoMIP) has been designed as a method to compare a set of benchmark geoengineering interventions across modelling groups within the World Climate Research Program (WCRP) Coupled Model Intercomparison Project (CMIP). While we agree with the objectives of GeoMIP, this paper describes how the present experimental design could be extended by adding a simple control component. Using a Model Predictive Control framework we show this control provides an automated solution for the problem of balancing radiative forcings within a climate model as required by the G1, G2 and G3 GeoMIP scenarios. By automating this process, the control removes the need for expensive trial-and-error model run iterations as suggested by the present guidelines. In addition, the control allows for some further stardardisation of the experimental conditions, potentially making inference of the side effects of geoengineering more straightforward. Finally, the control provides an interesting analogue for geoengineering deployment governed by a policy agent acting under conditions of uncertainty over the effectiveness of the technology.

Keywords

Geoengineering, GeoMIP, climate, control, Model Pedictive Control.

Introduction

Concerns over the inability of society to engage with meaningful global-scale greenhouse gas emissions reduction is fuelling interest in direct interventions in the climate system designed to counteract the radiative effects of greenhouse gases (Lenton and Vaughan 2009; Vaughan and Lenton 2011; Rasch, Crutzen, and Coleman 2008). Investments in research are already being made to inform decisions on the effectiveness and side effects of these 'geoengineering' interventions, research that is almost exclusively conducted within global climate models (Virgoe 2009). The IPCC Fifth Assessment Report will consider geoengineering explicitly (IPCC 2011) and the Geoengineering Model Intercomparison Project (GeoMIP) has been established to co-ordinate the modelling efforts in support of this.

A proposal has emerged within GeoMIP for a suite of four benchmark scenarios to be used by the global climate modelling community (Kravitz et al. 2011a; Kravitz et al. 2011b), hereafter referred to jointly as K2011. Three of the four GeoMIP scenarios (G1, G2, and G3) call for the climate modeller to specify a pattern of Solar Radiation Management (SRM) that balances a prescribed pattern of forcing arising from a stylised greenhouse gas emission scenario. The fourth scenario, G4, simply investigates a constant SO_2 injection rate. The four scenarios are shown in Figure 1. These scenarios have largely been proposed to evaluate the side effects of geoengineering such as weakening of the African and Asian summer monsoon, ozone depletion, and untoward interference in localized patterns of temperature and precipitation (Robock et al. 2009). However, scenarios G1, G2 and G3 require the modeller to control the net radiative balance of the model atmosphere, thereby regulating changes in the global energy balance. Not surprisingly therefore, these tests have very strong links to the discipline of systems and control. In this paper we reflect on the proposed GeoMIP test cases from this standpoint and offer some additional experiments to complement those proposed by K2011.

To illustrate the parallels between the GeoMIP experiments set out in K2011 and control we offer a particularly simple control design for regulating the Top of Atmosphere (TOA) radiative forcings as could be used in the G1, G2 and G3 scenarios. We then show the performance results for this design using a simple stochastic climate model. The aim of the following sections is to not only make the control approach transparent, we have also explicitly set out to design a scheme that can be easily added to existing climate model code in the expectation of take-up within GeoMIP. Finally, we reflect on some as yet unexplored control issues associated with SRM geoengineering that are flagged by the development of the control design, particularly around the handling of uncertainty.

Model Predictive Control of global radiative balance

To illustrate the parallels between the K2011 GeoMIP scenarios and control we will develop a very simple Model Predictive Control (MPC) framework designed to achieve the objective of a neutral TOA radiative balance given a known greenhouse gas forcing disturbance. MPC is an appropriate approach here because it allows us to exploit a model that is very familiar to the climate modelling community, thereby aiding communication. However, it must be stressed that there are a broad range of alternative linear and nonlinear control methods that could and should be explored for applications such as this. The predictive model we exploit is the following simple TOA radiative balance,

$$N(t) = Q(t) - R(t) - f\left\{\Delta T(t)\right\} \tag{1}$$

where $N(t)$ (W m^{-2}) is the sum net TOA radiative flux; $Q(t)$ (W m^{-2}) is the net TOA radiative flux attributed to changes in anthropogenic greenhouse gas concentrations; $R(t)$ (W m^{-2}) is the net TOA radiative flux attributed to geoengineering actions and $f\left\{\Delta T(t)\right\}$ (W m^{-2}) the net TOA radiative flux attributed to disturbances in surface temperature, $\Delta T(t)$ (K). All terms are annual global means. Because $\Delta T(t)$ arises from imbalances in $N(t)$, $f\left\{\Delta T(t)\right\}$ are by definition feedback perturbations on the TOA radiative balance. The decadal timescales for the proposed GeoMIP runs and the desire to ignore seasonal variations in $N(t)$ means we will assume an annual sample interval, Δt, for the sampling of all variables and the specification of the control.

Assuming $R(t)$ is directly proportional to the level of SRM geoengineering, $U(t)$, without any lag because of the annual timescales we will consider (Kaufman, Tanré, and Boucher 2002), i.e., $R(t) = \varepsilon U(t)$ where ε is the effectiveness of the geoengineering at causing changes in $R(t)$ (Lenton and Vaughan 2009), then to impose the condition $N(t) = 0$ in Equation (1) gives the inverse MPC law,

$$U(t) = \varepsilon^{-1}\left[Q(t) - f\left\{\Delta T(t)\right\}\right] \tag{2}$$

We will assume that there are no political, socio-economic or engineering constraints on the specification of $U(t)$ in line with the GeoMIP experiments (although see later discussion on this). In addition to assuming $N(t)$ is observed (e.g., by satellite), albeit with measurement error, we will assume that $Q(t)$ is predictable so we can specify the one year ahead values $Q(t+\Delta t)$ and hence $U(t+\Delta t)$ as appears to be assumed in the GeoMIP experiments. The inertia in both the global socio-economic system driving emissions and the carbon cycle accepting these emissions would lead one to conclude this is realistic. That said, we will make provision for uncertainty when specifying $Q(t+\Delta t)$ (see below) as this appears more realistic.

Within GeoMIP the carbon cycle-climate system will invariably be represented by complex climate models. The degree of internal variability generated by these models is such that the observation of the terms required to solve Equation (2) will be uncertain (Deser et al. 2010). For example, the magnitude of climate feedback, $f\left\{\Delta T(t)\right\}$, is effectively impossible to calculate a priori. Likewise, the effectiveness of a given geoengineering technology, ε, is also subject to significant uncertainty for similar reasons (Heckendorn et al. 2009). Because of this, any approach to regulate $N(t)$ based on Equation (2) alone is likely to fail when, for all cases of practical interest, the terms in Equation (2) are either not known with sufficient accuracy or the observations used to infer these values are subject to significant natural variability and/or measurement noise. The effect this uncertainty has on the deployment of any geoengineering measure could be an important issue for GeoMIP to address because this would be one of the defining features of any proposed real world deployment.

One approach to overcoming these uncertainties is to recognise that the terms in Equation (1) can be estimated from time-series data describing aggregate properties of climate models (e.g., Gregory et al. 2004). At $t=0$ (the initiation of the geoengineering) we have some prior knowledge of what ε should be (e.g., for G1 and G2, $\varepsilon=(1-\alpha)$ where α is the global albedo of the climate model). Because of uncertainty over the value of the global albedo we start with an initial estimate $\hat{\varepsilon}(0)$. However, rather than use $\hat{\varepsilon}(0)$ to do a trial run followed by a series of re-runs with supervised iterative updates of $\hat{\varepsilon}(0)$ as, in effect, advocated by K2011, we propose using Recursive Least Squares (RLS) estimation to provide annual updates of $\hat{\varepsilon}(t)$ within Equation (1). We then use this recursive update of $\hat{\varepsilon}(t)$ to update $U(t+\Delta t)$ in Equation (2) in a single automated simulation. Providing $\hat{\varepsilon}(t)$ converges on ε within an appropriate timeframe, then $N(t)\rightarrow 0$ as required.

While $N(t) \neq 0$ the climate system accumulates heat leading to transients in $\Delta T(t)$ and hence the feedbacks leading to the perturbation $f\left\{\Delta T(t)\right\}$. As a result, if we want to ensure $N(t) \to 0$ we also need to account for these feedback effects when adjusting $U(t)$. Because we are uncertain over the nature of the feedbacks driving $f\left\{\Delta T(t)\right\}$ (Bony et al. 2006) we assume $f\left\{\Delta T(t)\right\} \approx \hat{\lambda}(t)\Delta T(t)$ (Gregory et al. 2004; Forster and Taylor 2006) and again use RLS to update the estimate of the feedback parameter $\hat{\lambda}(t)$ (W m^{-2} K^{-1}) starting from $\hat{\lambda}(0)$.

Building on Equation (1) the regression equation used here is given by,

$$N(t) - Q(t) = \hat{\varepsilon}(t)U(t) - \hat{\lambda}(t)\Delta T(t) + \hat{\omega}(t) + \xi(t) \tag{3}$$

with the additional offset, $\hat{\omega}(t)$ (W m^{-2}), accommodating any structural uncertainty in $f\left\{\Delta T(t)\right\} \approx \hat{\lambda}(t)\Delta T(t)$ in addition to lack of closure in the projections of $Q(t)$ relative to the 'observations' of $N(t)$. $\xi(t)$ (W m^{-2}) are the regression model residuals. Here we use the RLS algorithm given in Young (2011, Ch. 3). RLS estimation will be inefficient if the statistical properties of $\xi(t)$ deviate too far from the ideal of an independent, white noise series. In extreme cases this may produce parameter estimates which generate an unstable control scheme. For the simple stochastic climate model used in this paper this is not an issue and inspection of the properties of $\xi(t)$ for complex climate models such as Atmosphere-Ocean General Circulation Models (A-OGCMs) suggests the same is likely to be true for these also (Forster and Taylor 2006). In the event this was an issue constraints could be applied to the parameter estimates to prevent such instabilities, or more sophisticated recursive estimation procedures could be exploited (Kalman 1960; Young 2011).

From Equation (2) the one step ahead control law using the recursive estimates $\hat{\varepsilon}(t)$, $\hat{\lambda}(t)$ and $\hat{\omega}(t)$ is given by,

$$U(t + \Delta t) = \hat{\varepsilon}(t)^{-1}\left[Q(t + \Delta t) - \hat{\lambda}(t)\Delta T(t) + \hat{\omega}(t)\right] \tag{4}$$

In equation (3) the fact that there are non-anthropogenic components to both $N(t)$ and $\Delta T(t)$ is incidental given the RLS acts as a low pass filter when returning estimates of $\hat{\varepsilon}(t)$,

$\hat{\lambda}(t)$ and $\hat{\omega}(t)$. However, in equation (4) the natural variability in $\Delta T(t)$ could be amplified in the specification of $U(t+\Delta t)$, something we may wish to avoid. The ideal solution to this is to have a robust detection/attribution system that is able to partition $\Delta T(t)$ into its natural and anthropogenic components (Mann and Lees 1996). A simple proxy for this is to derive recursive estimates of the variations in $\Delta T(t)$ that filter out the 'unwanted' inter-annual variations i.e.,

$$\Delta \bar{T}(t) = \Delta \bar{T}(t-1) + \beta(\Delta T(t) - \Delta \bar{T}(t-1)) \quad (5)$$

where $\Delta \bar{T}(t)$ is the filtered global mean surface temperature perturbations and β defines the cut-off timescale of the filter, which is taken as 5 years in this illustration (i.e., $\beta = 1 - e^{-\Delta t/5}$).

A simple climate model example

Equations (3), (4) and (5), in conjunction with the RLS algorithm, form the MPC for the candidate GeoMIP experiments. Here we illustrate solving the G1 and G2 scenarios shown in Figure 1 using this MPC design. The simple climate model structure we use for the experiments is shown in Figure 2. This model has the core global scale feedback dynamics of A-OGCMs (Jarvis 2011) and is expressed in such a way that $f\left\{\Delta T(t)\right\}$ can be computed directly from, but is not a static function of, $\Delta T(t)$. A 10^3 member Monte Carlo evaluation of the proposed MPC was performed for both the G1 and G2 scenarios where variability was introduced into the ensemble as follows:

1. The fast, intermediate and slow feedbacks in the simple climate model were varied in strength by sampling the transfer function amplitude parameter (labelled b_{fast}, b_{int} and b_{slow} in Figure 2) from a trivariate normal distribution with means -2.3, 1.0 and 0.5 W m^{-2} K^{-1} and variances 0.080, 0.014 and 0.004 respectively. The trivariate distribution also included covariance between the feedback amplitudes to reflect interactions between the feedback processes they attempt to represent, where cov(b_{fast}, b_{int}) = 0.0173, cov(b_{fast}, b_{slow}) = 0.0091, and cov(b_{int}, b_{slow}) = 0.0037. This increased the 5th and 95th percentiles for the 2 x CO_2 equilibrium range for ΔT from 2.33 - 6.64 to 2.18 - 7.62 K.

2. $N(t)$ had two zero mean white noise components added to it to reflect both annual stochastic variations and measurement error. The annual stochastic variations, which had a standard deviation of 0.5 W m^{-2}, were allowed to pass into the climate model where they are translated into a significant red noise component on $\Delta T(t)$ by the climate model dynamics, reminiscent of the effects of large scale stochastic climate phenomena such as volcanic eruptions or El and La Nina events (Roe 2009; see Figure 3). The measurement error on $N(t)$ had a standard deviation of 0.25 W m^{-2}.

3. The albedo of the climate model was assumed to be 0.3 and hence ε = 0.7. $Q_{2\mathrm{xCO2}}$ was assumed to be 3.7 W m^{-2} (IPCC 2001a ; Ch. 6). For the initial parameter estimates $\hat{\varepsilon}(0)$, $\hat{\lambda}(0)$ and $\hat{\omega}(0)$ we assumed prior distributions with means of 0.8, 2.0 W m^{-2} K^{-1} and 0 W m^{-2} and standard deviations of 0.1, 0.5 W m^{-2} K^{-1} and 0.5 W m^{-2} respectively (note the deliberate mismatches here). A value of $\hat{\lambda}(0)$ = 2.0 W m^{-2} K^{-1} was adopted because only the faster feedbacks were assumed to be important because of the relatively short timeframe of the GeoMIP experiments.

From equation (1) the condition $N(t) \to 0$ arising from the 'correct' cancellation of $Q(t)$ by $R(t)$ (the stated aim of the G1, G2 and G3 scenarios in K2011) implies $f\left\{\Delta T(t)\right\} \to 0$. Therefore, although not stated explicitly in K2011, $\Delta T(t) \to 0$ is a co-requisite of $N(t) \to 0$ and any failure to achieve the dual condition $N(t) \to 0$; $\Delta T(t) \to 0$ must imply the incorrect specification of $R(t)$. Whilst $N(t) \neq 0$, especially following the initial application of the disturbance in $Q(t)$, the climate system accumulates heat resulting in a perturbation in $\Delta T(t)$. The transient for this perturbation will take many years to play out. For example, the simple climate model illustrated in Figure 2 has a reference system time constant of $\sim$5 years and feedback time constants of 0, 50 and 500 years. Given these dynamics will result in perturbations lasting significantly longer than the GeoMIP experiment runs of 50 years, a prudent step to take would be to implement an additional set of control feedbacks that extinguish the perturbation in $\Delta T(t)$ well within the 50 year GeoMIP timeframe. Not only would this lead to a much more standardised set of A-OGCM experiments free from the confounding factors introduced by $\Delta T(t) \neq 0$, through ultimately removing the feedback effect $f\left\{\Delta T(t)\right\}$ this will ensure the MPC becomes more robust, ultimately depending only on the recursive estimation of $\hat{\varepsilon}(t)$. This does not mean that the pattern of regional changes in temperature as a result of the geoengineering are abolished, they are simply constrained to occur within the context of no change to the global average i.e., they are standardised.

Given the surface ocean mixed layer dominates the dynamics of $\Delta T(t)$ on timescales < 10 years (Watterson 2000) the simplest control law for $\Delta T(t)$ that offers full control of these first order mixed layer dynamics is given by

$$U_c(t+\Delta t) = U_c(t) - g(1)(\Delta \bar{T}(t) - \Delta \bar{T}(t-\Delta t)) - g(2)\Delta T(t) \quad (6)$$

where $U_c(t)$ are the adjustments to $U(t)$ required to extinguish the perturbation in $\Delta \bar{T}(t)$ and the feedback control gains, g = [4.9288   0.6906], yield a closed loop response that again only responds on timescales of ~5 years, i.e., it will act to abolish any $\Delta \bar{T}(t)$ perturbation well within the GeoMIP 50 year run but will overlook the interannual variability.

Results

Figures 3 and 4 summarise the G1 and G2 MPC experiments. As can be seen, the TOA energy balance is achieved within the observable limits for all but the first few years following the introduction of the G1 forcing disturbances when the estimate of ε differs significantly from its 'true' value and the forcing disturbance is large enough to expose this. Not surprisingly, the imposition of the large change in $Q(t)$ at $t = 0$ in G1 yields enough information on ε for the RLS to rapidly converge on the 'true' value of 0.7. Note that the corrective actions are always lagged one year behind unforeseen disturbances because, in a truly sequential setting like this, the unforeseen cannot be pre-empted.

Thereafter, convergence of $\hat{\varepsilon}(t)$ on its true value is more gradual and associated with increases in the confidence of the RLS estimates as noise effects are filtered out and the effects of $f\left\{\Delta T(t)\right\}$ is eliminated by the control action in equation (6). For G2 the information content of the time series data is greatly reduced because of the more gradual disturbance being imposed. This makes estimation of ε more problematic, but at the same time the consequences of any mismatch between the estimated and 'true' value of effectiveness greatly reduced. As a result, a point of note for the G2 scenario is that the effectiveness of geoengineering is much harder to evaluate but less important for the control of the TOA forcings.

The estimates of $\hat{\lambda}(t)$ remain rather uncertain throughout both simulations due to the very small perturbation in $\Delta T(t)$ experienced as a result of the control (see Figure 3 and 4). However, the perturbations in $\Delta T(t)$ are large enough to initiate a feedback response $f\left\{\Delta T(t)\right\}$ and it is the weakening of this feedback over time (mimicking the loss of ocean heat uptake; Jarvis 2011) that is causing the estimates of $\hat{\lambda}(t)$ to drift up. Again, the transient in $f\left\{\Delta T(t)\right\}$ is ultimately extinguished by the control action driving $\Delta T(t)$ back to zero. Not only does this help fulfil the TOA radiative balance, it ensures $\hat{\varepsilon}(t)$ converges on its 'true' value by ultimately removing all effects of $f\left\{\Delta T(t)\right\}$ from the RLS. For both scenarios $\hat{\omega}(t)$ remains near zero because, in this simple case, the TOA model is structurally accurate.

The success of the MPC is borne out in the second perturbation in the G1 scenario when $Q(t)$ is reduced back to zero. As can be seen, the RLS estimates of $\hat{\varepsilon}(t)$ and $\hat{\lambda}(t)$ are sufficiently good when allied to the MPC design for there to be no net TOA disturbance at this point. However, as with all simulation exercises, there is always a tendency to over-represent prior knowledge because of the difficulty of divorcing the modelling from the modeller. Even in the simple example presented here the priors for $\hat{\varepsilon}(t)$ and $\hat{\lambda}(t)$ were probably unrealistically close to their 'true' values if we wanted to represent a realistic scenario when faced with a complex model. As a result, it would be valuable to execute such simulations as blind trials where a GeoMIP-agreed set of priors for $\hat{\varepsilon}(t)$ and $\hat{\lambda}(t)$ were applied to all models in the GeoMIP ensemble as if there were some initial consensus on effectiveness and climate sensitivity informing the early stage deployment of geoengineering.

Discussion

The results from the G1 and G2 examples show that it appears possible to successfully automate the specification of the global level of SRM geoengineering within GeoMIP using a relatively simple MPC approach. This approach also conveniently estimates the effectiveness of SRM within the climate model in question. The most obvious advantage of the MPC approach advocated here over the iterative learning advocated in K2011 is the significant time saving that could be gained. Not only does the MPC reduce the evaluation to a single run, it also removes the need for repeat interventions by the modeller. This would help increase the size of ensembles that could be considered within the GeoMIP evaluation. Furthermore, through removing significant global mean surface temperature perturbations this should help standardise model output making the intercomparison of the side effects of geoengineering more transparent.

The trade-offs are two-fold: Firstly, the control requires the online calculation of aggregate metrics such as $N(t)$ and $\Delta T(t)$. This is not currently standard practice within A-OGCM code given it is usually performed offline in higher level software such as IDL or Matlab. However, given all the necessary disaggregated states (both past and present) should be available within an A-OGCM run, this simply requires a small piece of additional code to perform the online spatial averaging. Secondly, the RLS structure also needs embedding within the A-OGCM along with the associated control law(s). Again, this requires very little additional code (see Young 2011, ch 3). If these two logistic barriers were insurmountable, e.g., if a more sophisticated control architecture was required, an alternative approach would be to design a simple client/server middleware scheme that allowed for an online data exchange between the A-OGCM and, for example, IDL or Matlab, where the annually revised control could be specified, returning the sequentially updated level of SRM back to the A-OGCM. Experience of applying equation (3) to A-OGCM data (Forster and Taylor 2006) would indicate the signal to noise ratio in $N(t)$ and $\Delta T(t)$ are such that the likelihood is that the RLS should converge, although finding circumstances where this broke down is likely to yield important insights into some of the risks presented by SRM deployment.

Beyond these important practical issues the MPC approach offers a very different perspective on geoengineering evaluation. Unlike the model world, any proposed real world deployment of geoengineering would not have the benefit of reruns of reality in order to fully identify the effectiveness and side effects of technologies or portfolios of technologies prior to deployment. When this is allied to the conditions of deep uncertainty surrounding the climate science and associated models (Allen et al. 2000; Reilly et al. 2001; Stott and Kettleborough 2001) it is clear that research on the sequential decision making and adaptive learning processes for handling this uncertainty is as important an aspect of geoengineering research as any attempt to predict the pattern of side effects geoengineering deployments might generate.

Control approaches such as MPC can contribute significantly to the way we might address the potential sequential nature of geoengineering. Learning protocols such as RLS can act as useful proxies for data assimilation and model updating strategies, whilst the control these models inform would act as a proxy for a 'rational' decision making process. In this context the design criteria of simply maintaining a neutral TOA radiative balance without any constraints is clearly too narrow. Traditionally, control strategies such as MPC would consider a more nuanced approach which might include trading off objectives, including costs relating to the environmental side effects, whilst handling constraints describing the capacity of the human system to monitor and respond to dangerous climate change. Similarly, model independent control strategies would act as useful proxies for robust observation-led policies for geoengineering deployment. Using this, one could explore the controllability of climate states using geoengineering. For example, the G2 scenario shown in Figure 4 raises an interesting question about SRM deployment where the effectiveness remains undetected despite the control objective being met. Under these circumstances the pressure to abandon SRM would be obvious, particularly if deleterious side effects were perceived to be being incurred.

Although the GeoMIP experiment is not intended to be interpreted in this way, it is possible that the neutral TOA generated by a modelling effort after using multiple model iterations and adjustments could be wrongly interpreted as a measure of success. The single iteration approach is clearly more realistic in this context. However, again the apparent sucess of the scenarios shown in Figure 3 and 4 must not be interpreted as a demonstration of the controllability of climate states using geoengineering. Such evaluations need to be performed under much more realistic conditions where constraints are applied to the control and the evaluation encompasses more realistic conditions of uncertainty and observability. The spatially explicit case where there are point sources for SRM attempting to regulate local anomalies will no doubt prove to be extremely challenging to control. Whether climate models can act as accurate proxies for the real world in such an evaluation remains an open question. They could, however, provide useful virtual environments where different adaptive strategies could be explored and lessons learnt.

Acknowledgement

This work was supported under UK-EPSRC research contract: Integrated Assessment of Geoengineering Proposals - EPSRC EP/I014721/1.

References

Allen MR, Stott PA, Mitchell JFB, Schnur R, Delworth TL. 2000. Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* 407 (6804):617-620.

Bony S, Colman R, Kattsov VM, Allan RP, Bretherton CS, Dufresne J, Hall A, Hallegatte S, Holland MM, Ingram W, Randall DA, Soden BJ, Tselioudis G, Webb MJ. 2006. How Well Do We Understand and Evaluate Climate Change Feedback Processes? *Journal of Climate* 19 (15):3445-3482.

Deser C, Phillips A, Bourdette V, Teng H. 2010. Uncertainty in climate change projections: the role of internal variability. *Climate Dynamic*s. DOI: 10.1007/s00382-010-0977-x.

Evensen G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics* 53 (4):343-367.

Forster PM, Taylor KE. 2006 Climate Forcings and Climate Sensitivities Diagnosed from Coupled Climate Model Integrations. *Journal of Climate*, 19: 6181 – 6194

Gregory JM, Ingram WJ, Palmer MA, Jones GS, Stott PA, Thorpe RB, Lowe JA, Johns TC, Williams KD. 2004. A new method for diagnosing radiative forcing and climate sensitivity. *Geophysics Research Lett*er 31:L03205.

Heckendorn P, Weisenstein D, Fueglistaler S, Luo BP, Rozanov E, Schraner M, Thomason LW, Peter T. 2009. The impact of geoengineering aerosols on stratospheric temperature and ozone. *Environmental Research Letters* 4:045108.

IPCC. 2011. Agreed Reference Material for the IPCC Fifth Assessment Report. Available online at http://www.ipcc.ch/pdf/ar5/ar5-outline-compilation.pdf. [accessed November 2011]

Jarvis A. 2011. The magnitude-timescale relationship of surface temperature feedbacks in climate models. *Earth System Dynamics* 2, 213-221.

Kalman, RE. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82 (1):35-45.

Kaufman YJ, Tanré D, Boucher O. 2002. A satellite view of aerosols in the climate system. *Nature* 419 (6903):215.

Kravitz B, Robock A, Boucher O, Schmidt H, Taylor K, Stenchikov G, Schulz M. 2010. Specifications for GeoMIP experimentsG1 through G4. Available online at http://climate.envsci.rutgers.edu/GeoMIP/docs/specificationsG1 G4.pdf [accessed November 2011]

Kravitz B, Robock A, Boucher O, Schmidt H, Taylor K, Stenchikov GL, Schulz M. 2011b. The geoengineering model intercomparison project (GeoMIP). *Atmospheric Science Letters*. DOI: 10.1002/as1.316

Lenton TM, Vaughan NE. 2009. The radiative forcing potential of different climate geoengineering options. *Atmospheric Chemistry and Physics* 9 (15):5539-5561.

Mann ME, and Lees JM. 1996. Robust estimation of background noise and signal detection in climatic time series. *Climatic Change* 33 (3):409-445.

Rasch PJ, Crutzen PJ, Coleman DB. 2008. Exploring the geoengineering of climate using stratospheric sulfate aerosols: The role of particle size. *Geophysical Research Letters* 35 (2):L02809.

Reilly J, Stone PH, Forest CE, Webster MD, Jacoby HD, Prinn RG. 2001. Uncertainty and climate change assessments. *Science* 293 (5529):430.

Robock A, Marquardt A, Kravitz B, Stenchikov G. 2009. Benefits, risks, and costs of stratospheric geoengineering. *Geophysical Research Letters* 36 (19):L19703.

Roe GH. 2009. Feedbacks, timescales and seeing red. *Annual Review of Earth and Planetary Sciences* 37: 93-115.

Stott PA, Kettleborough JA. 2001. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Journal of Climate* 14:1055-1068.

Vaughan NE, Lenton TM. 2011. A review of climate geoengineering proposals. *Climatic Change*: 109, 745-790.

Virgoe J. 2009. International governance of a possible geoengineering intervention to combat climate change. *Climatic Change* 95 (1):103-119.

Young PC. 2011. *Recursive Estimation and Time-Series Analysis*. 2nd ed. Springer.

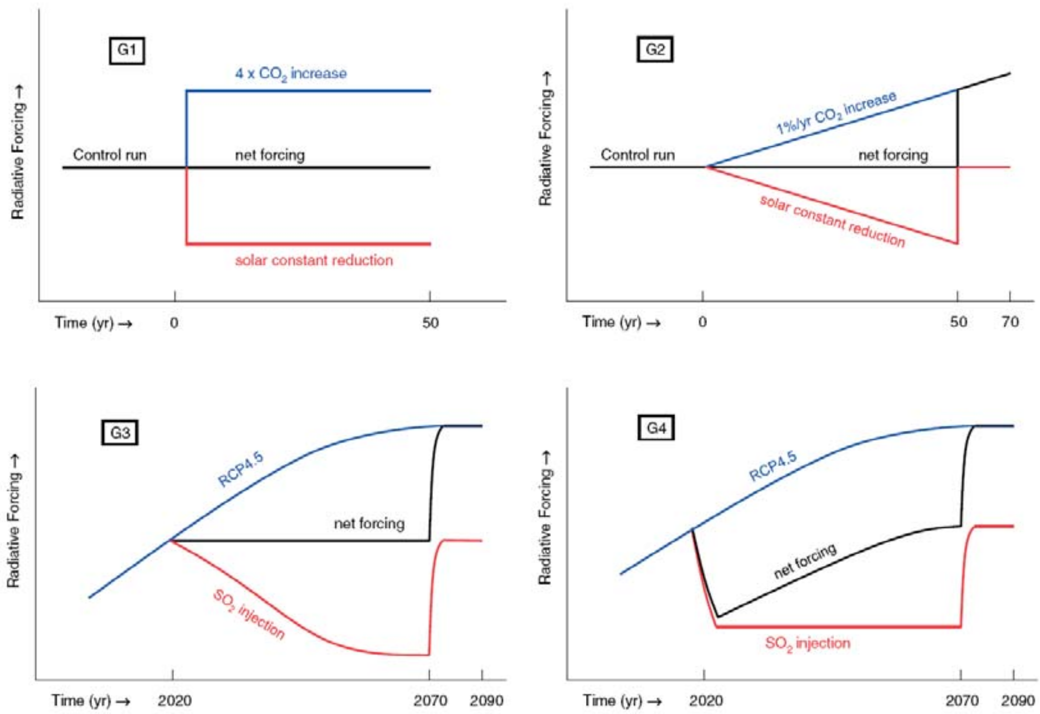

Figure 1. The proposed GeoMIP standard test scenarios as advocated by Kravitz et al. (2011a,b).

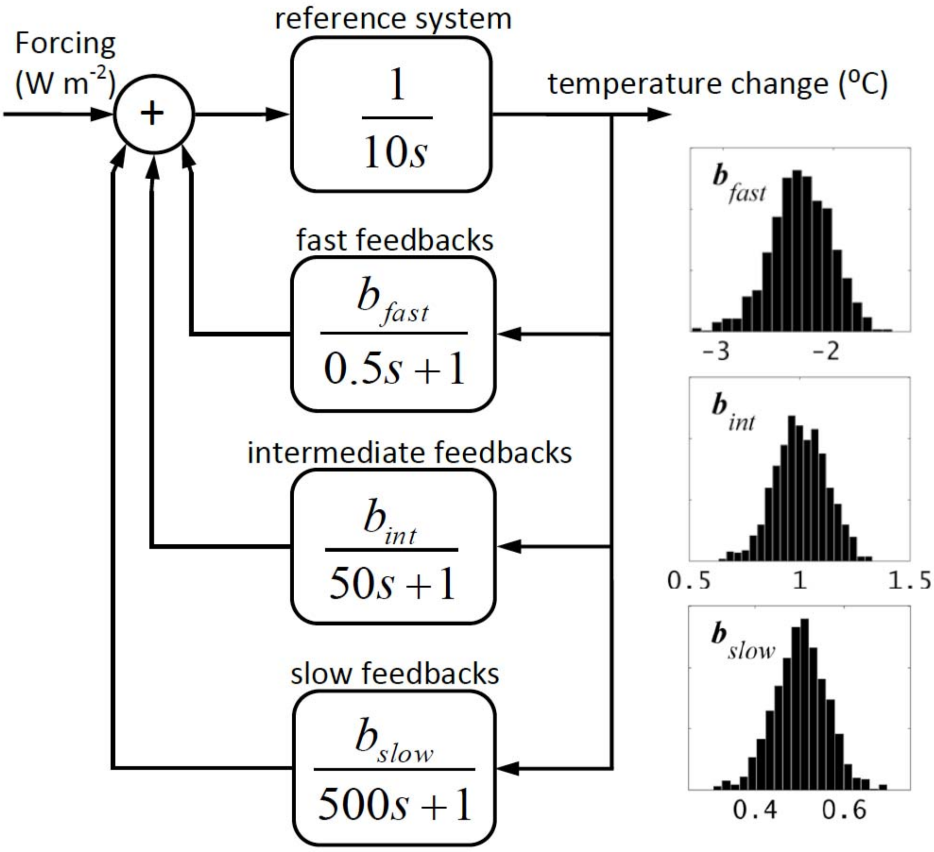

Figure 2. Feedback system block diagram of the simple climate model used in the MPC evaluation (adapted from Jarvis, 2011). The blocks represent continuous time transfer functions where s is the differential operator. The multiplier for s is the time constant for the feedback response in years. To represent uncertainty in the climate model a 10^3 member ensemble of climate models was generated where the fast, intermediate and slow feedback amplitude parameters (labelled b_{fast}, b_{int} and b_{slow}) were sample from the distributions shown. Covariance was added to represent reinforcement between the feedback components further increasing the dynamic range of the model ensemble (see text).

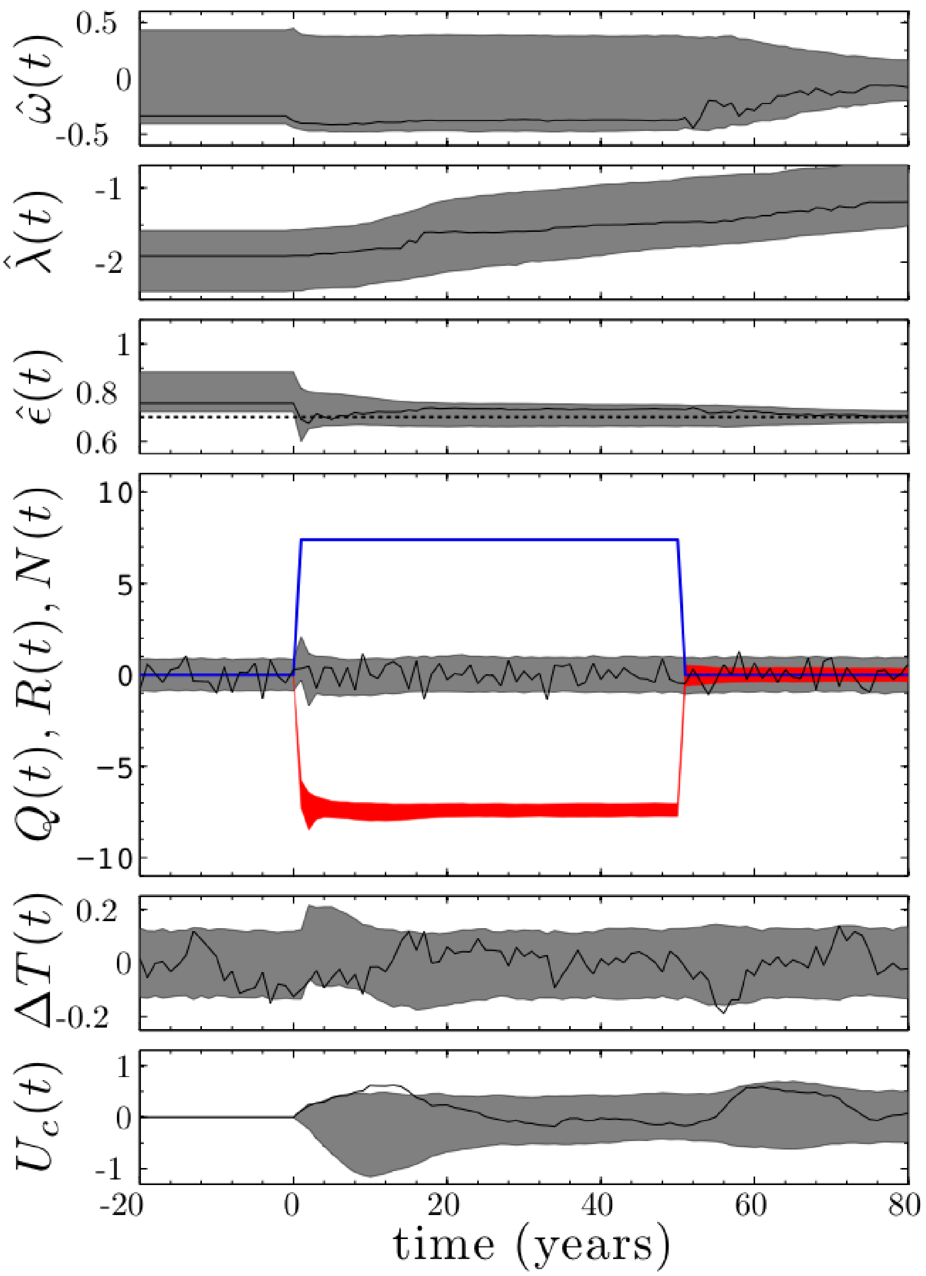

Figure 3. Results from the MPC implementation of the G1 scenario. The bands are 95 percent confidence intervals for the 10^3 member ensemble (see text for simulation conditions). Single solid black lines show one realisation. For the net TOA radiative fluxes the colour convention follows Figure 1.

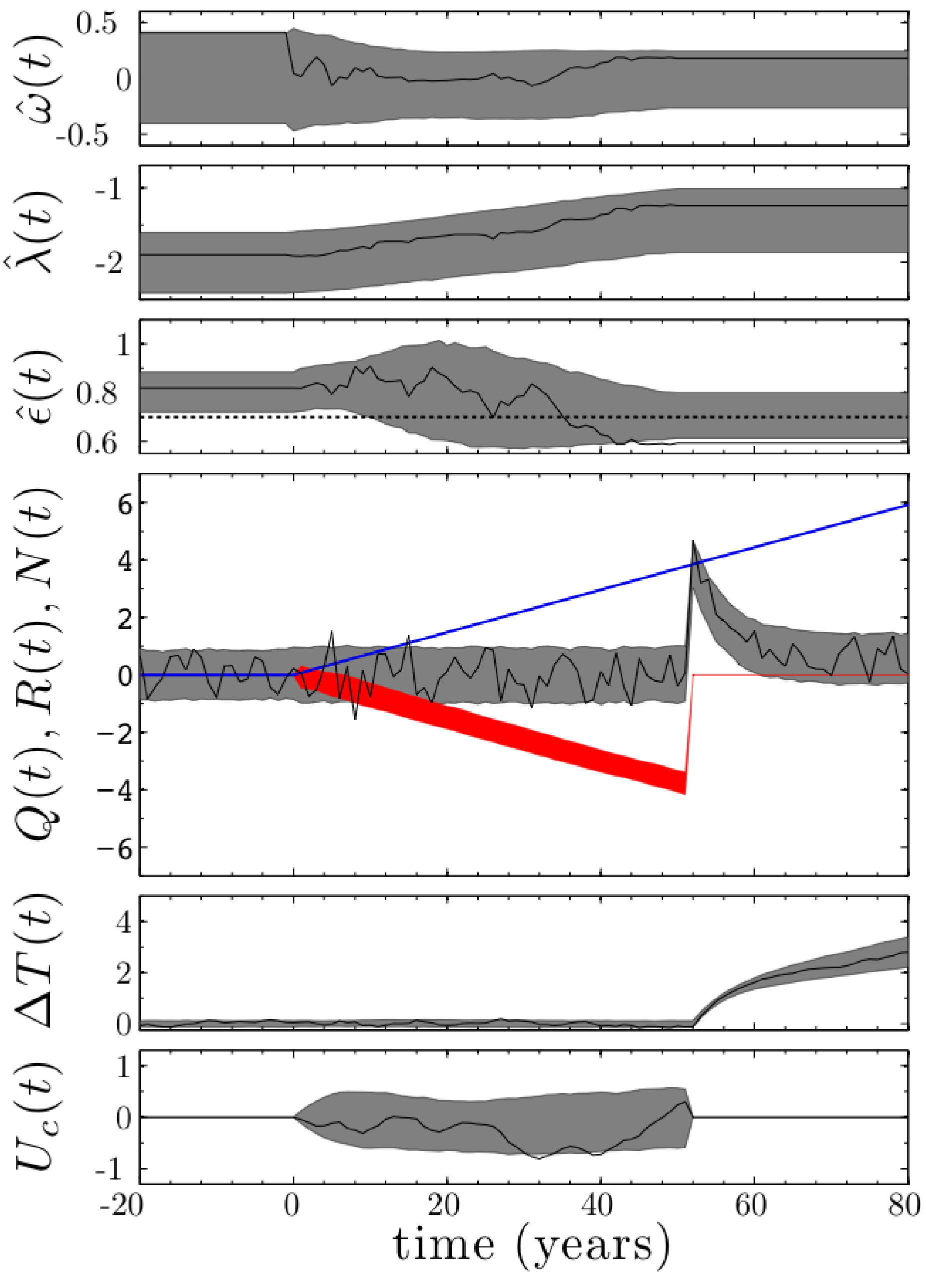

Figure 4. Results from the MPC implementation of the G2 scenario. The bands are 95 percent confidence intervals for the 10^3 member ensemble (see text for simulation conditions). Single solid black lines show one realisation. For the net TOA radiative fluxes the colour convention follows Figure 1.