PROCEEDINGS A

rspa.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
Climate engineering, closed-loop control strategies

Keywords:
solar radiation management,climate engineering, adaptive control, PI control

[1]

Author for correspondence:
F. Bonetti
e-mail:
F.Bonetti.1@research.gla.ac.uk

Multiple input control strategies for robust and adaptive climate engineering in a low order 3-box model

F. Bonetti[1], C. McInnes[1]

[1] School of Engineering, University of Glasgow, Glasgow, G128QQ

A low order three-box energy balance model for the climate system is employed with a multivariable control scheme for the evaluation of new robust and adaptive climate engineering strategies using solar radiation management. The climate engineering measures are deployed in 3 boxes thus representing northern, southern and central bands. It is shown that, through heat transport between the boxes, it is possible to effect a degree of latitudinal control through the reduction of insolation. The approach employed consists of a closed-loop system with an adaptive controller, where the required control intervention is estimated under the $RCP4.5$ radiative scenario. Through the on-line estimation of the controller parameters, adaptive control can overcome key-issues related to uncertainties of the climate model, the external radiative forcing and the dynamics of the actuator used. In fact, the use of adaptive control offers a robust means of dealing with unforeseeable abrupt perturbations, as well as the parametrisation of the model considered, to counteract the $RCP4.5$ scenario, while still providing bounds on stability and control performance. Moreover, applying multivariable control theory also allows the formal controllability and observability of the system to be investigated in order to identify all feasible control strategies.

1. Introduction

It is clear that current concentrations of atmospheric CO_2 exceed measured historical levels in modern times [1], largely attributed to anthropogenic forcing since the industrial revolution. For this reason, efforts have focused on the long-term reduction of global greenhouse gas (GHGs) emissions through mitigation. However, the necessary decline in emissions rates at a global scale has never been achieved [2], leading to recent interest in climate engineering for future risk-mitigation strategies.

Climate engineering [3], also known as geoengineering, is the intentional manipulation of the climate system that aims to offset human-driven climate change. This involves techniques developed both to reduce the concentration of atmospheric carbon dioxide (Carbon Dioxide Removal (CDR)) and to counteract the radiative forcing that it generates (Solar Radiation Management (SRM) [4]).

SRM regards methods such as injecting scattering aerosols in the stratosphere [5] or deploying vast thin-film space mirrors to reduce direct solar insolation [6] and so reduce radiative forcing. In particular, regarding the injection of aerosol particles, in [7,8] a coupled atmosphere-ocean general circulation model with fully interactive stratospheric chemistry is employed to simulate aerosol injections at multiple locations to meet multiple simultaneous surface temperature objectives. Whereas in [9], multiple injection strategies are considered to achieve a desired radiative forcing profile using a two-dimensional chemistry-transport-aerosol model.

Moreover, CDR methods require long timescales to deliver a significant reduction in carbon dioxide concentration, while Solar Radiation Management (SRM) can be considered a relatively fast-acting method, although it does not directly affect the carbon cycle. This paper focuses on an evaluation of the effects of SRM methods on the climate system. In particular, a generic control function, representing the reduction of insolation, is initially considered until Sec. 5(d), where simple dynamics is considered for sulphur aerosols.

Prior work, where SRM methods are considered using a single control variable, can only influence global dynamics. Indeed, one of the criticisms of climate engineering is that regional impacts are not addressed. For this reason other recent work [7,8,10] has investigated Multi-Input-Multi-Output (MIMO) systems and control strategies to assess latitudinal disparities of SRM.

Therefore, a 3-box model for the climate system is considered in this paper and, as in previous work [11–13], the problem is considered in the frame of modern control theory. However, with respect to other work [11–14], a new closed-loop strategy involving an adaptive controller is considered for climate engineering. Performance and robustness of this strategy is then compared with a proportional-integral (PI) controller in feedback. Importantly, it is demonstrated that adaptive control can compensate for large uncertainties in the climate model, as well as abrupt perturbations and in the dynamics of the model considered. It therefore offers a robust strategy for closed-loop deployment of SRM. In particular, the methods is demonstrated to be of critical importance in the case of unknown perturbations, such as lack of information on the key parameters of the climate model or temperature measurements. Thus, despite the limited applicability of the simple 3-box model, the use of multiple control inputs allow such issues to begin to be addressed, albeit at coarse length-scales.

Moreover, the use of a 3-box model (rather than a single transfer function) allows latitudinal controllability and observability to begin to be investigated. Indeed, the formal controllability of the problem is assessed using the 3-box model. In particular, in this context, the asymmetry of the northern and southern bands are found to play an important role.

In the first section the simple 3-box model is described. The Earth is divided into three latitudinal bands to account for northern and southern zones and the equator, with heat transfer between the boxes to capture the poleward transport of energy from the equator. It is demonstrated that the model, with its simplicity, enables different control strategies to be implemented and evaluated: comparisons are made between the implementation of adaptive control and PI control. The simple 3-box model provides an initial model to begin to assess the performance of these strategies preceding more detailed studies. Validation and limitations of the model are found in Sec. 2.(a).

Section 3 then describes the study of controllability and observability of the 3-box model. In Sec. 4 a Model-Reference Adaptive Controller (MRAC) is investigated and in Sec. 5(a) its performance is investigated for three different control strategies. This type of controller overcomes issues related to the large uncertainties of the 3-box-model providing effective control, despite that the model of the plant is not well known. The stability of this control method is also investigated using Lyapunov methods. This is demonstrated through introducing significant changes in the model parameters in Sec. 5(b) and comparing results with a PI controller in feedback. Moreover, the robustness of the adaptive controller is tested in Sec. 5(c) where a scenario involving an abrupt perturbation is considered and in Sec. 5(d) where the dynamics of stratospheric aerosols is considered.

Finally, in Sec. 6, the 3-box model is expanded to 5 boxes to provide higher resolution of the polar bands. This modification of the model is employed to investigate effects of a collapse of the Arctic ice sheet and the required insolation reduction to counteract the resulting change of albedo. The analysis also demonstrates the utility of low order models to quickly investigate new climate engineering feedback control strategies.

2. Three-box model of the climate system

In this section a simple 3-box climate model is developed in order to investigate the use of multi-variable control and provide clarity to assess the performance of these control strategies.

In this paper, a new robust control strategy will be developed to minimize the largest latitudinal disparities from climate engineering deployment using multiple control inputs. For this task a low order model of the climate system with three latitudinal bands is used for illustration. The Earth's surface is divided in three bands: southern and northern bands (latitude bands in the ranges $(-65^\circ, -90^\circ)$, $(65^\circ, 90^\circ)$) and a central band $(-65^\circ, 65^\circ)$. In this way, coarse latitudinal dynamics are taken into account.

This subdivision can be represented through a 3-box model defined by Eq. (2.1-2.3). It is important to note that the model is not considered to be substitute for high fidelity General Circulation Models (GCM), and its use should not be extended to real-world applications. However, it can be used to assess the performance of adaptive control strategies relative to PI control and to allow an investigation of formal controllability properties which are key to multivariable control. It is also envisaged implementation of adaptive control using a GCM as the next step, but is beyond the scope of the paper.

Moreover, as it will be shown later, adaptive control is robust to uncertainties in the climate model itself; therefore the importance of the model employed can be de-emphasized.

The subdivision of the system into three latitudinal bands can now be developed to consider the use of three separate control processes. As noted earlier, this is motivated by the need to begin to investigate how to overcome issues associated with the largest latitudinal disparities of the impacts of SRM technologies.

In Eqs. (2.1-2.3) $T_i(t)$ is the surface temperature and $i = 1, 2, 3$ represents the northern band, the central band and the southern band, respectively. A coupled Energy Balance Model (EBM) is used with a diffusive term to describe heat transport between latitude bands, a term F_{ext} considers external forcing due to anthropogenic GHGs emissions and a function U_i (i=1,2,3) represents the generic reduction of the incoming solar radiation. In Sec. 5(d), the dynamics of stratospheric aerosols is considered for the function U_i (i=1,2,3) in order to demonstrate the robustness of adaptive control to the choice of the actuator dynamics.

The model is defined as 3 coupled linear equations which can be written as:

$$C_1 \frac{dT_1(t)}{dt} = S_1(1 - \alpha_1) - (a_1 + b_1 T_1) - k_1(T_1 - T_2) + F_{ext} + U_1 \tag{2.1}$$

$$C_2\frac{dT_2(t)}{dt} = S_2(1-\alpha_2) - (a_2 + b_2T_2) - \frac{1}{2}k_2(T_2 - T_1) - \frac{1}{2}k_2(T_2 - T_3) + F_{ext} + U_2 \qquad (2.2)$$

$$C_3\frac{dT_3(t)}{dt} = S_3(1-\alpha_3) - (a_3 + b_3T_3) - k_3(T_3 - T_2) + F_{ext} + U_3 \qquad (2.3)$$

where, S_i and α_i (i=1,2,3) are the mean annual insolation and the planetary albedo in each latitudinal band, respectively. These are assumed to be fixed, although time-dependent seasonal variation could in principle be included. The outgoing infra-red radiation for the EBM is well approximated by the expression $a + bT(t)$ [15] [16], where a and b are empirical constants selected to account for the effect of clouds, water vapour and CO_2. In particular, an infra-red parametrization for the northern and southern hemispheres can be used [17] for the northern band ($a_1,\ b_1$) and southern band ($a_3,\ b_3$), whereas, as found in [15], $a_2 = 203.3\ W/m^2,\ b_2 = 2.09\ W/m^2/°C$ are employed for the central band. According to [17] the out-going infra-red radiation can be represented as:

$$a + b\,T(t) = A_1 + A_2A_c + (B_1 + B_2A_c)T(t) \qquad (2.4)$$

where $A_1 = 257\ W/m^2,\ A_2 = -91\ W/m^2,\ B_1 = 1.63\ W/m^2/°C,\ B_2 = -0.11\ W/m^2/°C$ for the northern band and $A_1 = 262\ W/m^2,\ A_2 = -81\ W/m^2,\ B_1 = 1.64\ W/m^2/°C,\ B_2 = -0.09\ W/m^2/°C$ for the southern band, whereas A_c is the cloud cover fraction set to 0.5 [15]. Through climatological records of zonal surface temperature and satellite observations this fit has been proven to be quite accurate [18]. Moreveor, C_i (i=1,2,3) is the effective heat capacity for each latitudinal band, which is largely determined by the different hemispherical distributions of land and water. The heat capacity over land is approximately $1/30$ of the capacity over the ocean mixed layer [15], therefore, since a larger fraction of water is found in the southern hemisphere a larger heat capacity is expected. Considering the fraction of water and land in each hemisphere (oceans cover 61% of the northern hemisphere and the 82% of the southern hemisphere) the heat capacity, in terms of b_1 and b_3, is 2.88 b_1 years for the northern hemisphere and 3.79 b_3 years for the southern hemisphere[1]. As for the infra-red radiation, these values for the heat capacity are employed for northern band (C_1) and southern band (C_3) and their average is used for the central band (C_2).

The third term on the right of Eqs. (2.1-2.3) is the latitudinal heat transport rate that is considered proportional to the temperature difference between two contiguous latitudinal zones, which provides coupling between the boxes. In accordance with the 2^{nd} law of thermodynamics, it is the transport of heat from warmer tropical to colder polar regions that leads to a downgrading of energy and an increase of the Earth's global entropy [19]. The poleward heat transport can be approximated by a transport coefficient k given by $k_1 = 0.549\ W/m^2/K$ [15] for the northern band and $k_3 = 0.649\ W/m^2/K$ for the southern band. The values of k_1 and k_3 are selected so that the mean annual temperature at the equator represents the current climate ($T \simeq 30\ °C$ [15]). With regard to the central band, the transport coefficient needs to be larger at the equator than the higher latitudes [15], therefore a value of $0.73\ W/m^2/K$ is considered for k_2.

Considering variations of temperature around the equilibrium state of each latitudinal band T_{eq_i} ($i = 1,2,3$), the following transformation can be used for each band:

$$\zeta_i = \frac{T_i - T_{eq_i}}{T_{eq_i}} \qquad i = 1,2,3 \qquad (2.5)$$

Moreover, since Eqs. (2.1-2.3) form a linear system of differential equations, they can be written in the form $\frac{d\boldsymbol{\zeta}}{dt} = A\boldsymbol{\zeta} + \mathbf{F} + \mathbf{U}$ where $\boldsymbol{\zeta}$ is the 3x1 state vector defining the temperature anomalies

[1]The values of the heat capacities are given in years as in [15] to show the combination of the timescales of land and oceans for the southern and northern hemisphere. From Eq. (2.4) $b_1 = 1.575\ W/m^2/°C$ and $b_3 = 1.595\ W/m^2/°C$, therefore $C_1 = 4.542\ Wyr/m^2/°C$ and $C_3 = 6.048\ Wyr/m^2/°C$.

and where A is the system matrix and $\mathbf{F}$ is a forcing vector given by:

$$A = \begin{pmatrix} j_{11} & j_{12} & 0 \\ j_{21} & j_{22} & j_{23} \\ 0 & j_{32} & j_{33} \end{pmatrix} \tag{2.6}$$

$$\mathbf{F}(t) = \begin{pmatrix} S_1(1-\alpha_1) - (a_1 + b_1 T_1) - k_1\left(T_{eq_1} - T_{eq_2}\right) + F_{ext}(t) \\ S_2(1-\alpha_2) - (a_2 + b_2 T_2) - k_2\left(T_{eq_2} - \frac{1}{2}T_{eq_1} - \frac{1}{2}T_{eq_3}\right) + F_{ext}(t) \\ S_3(1-\alpha_3) - (a_3 + b_3 T_3) - k_3\left(T_{eq_3} - T_{eq_2}\right) + F_{ext}(t) \end{pmatrix} \tag{2.7}$$

with $j_{ii} = -\frac{(b_i + k_i)}{C_i}$ $(i = 1, 2, 3)$, $j_{12} = \frac{k_1}{C_1} h_{21}$, $j_{21} = \frac{1}{2}\frac{k_2}{C_2} h_{12}$, $j_{23} = \frac{1}{2}\frac{k_2}{C_2} h_{32}$, $j_{32} = \frac{k_3}{C_3} h_{23}$ where $h_{ij} = \frac{T_{eq_i}}{T_{eq_j}}$ $(i, j = 1, 2, 3)$. The equilibrium temperatures in the three zones can be computed considering the equilibrium state of the system in Eqs. (2.1-2.3). External forcing is then ignored so that:

$$\mathbf{F}(0) = \begin{pmatrix} S_1(1-\alpha_1) - (a_1 + b_1 T_{eq_1}) - k_1\left(T_{eq_1} - T_{eq_2}\right) \\ S_2(1-\alpha_2) - (a_2 + b_2 T_{eq_2}) - k_2\left(T_{eq_2} - \frac{1}{2}T_{eq_1} - \frac{1}{2}T_{eq_3}\right) \\ S_3(1-\alpha_3) - (a_3 + b_3 T_{eq_3}) - k_3\left(T_{eq_3} - T_{eq_2}\right) \end{pmatrix} \tag{2.8}$$

The system defined by Eq. (2.8) can then be solved to obtain the equilibrium temperatures of the three bands given by $(T_{eq_1}, T_{eq_2}, T_{eq_3}) = (-28.9°C, 14.7°C, -34.5°C)$ [15]. The terms S_i $(i = 1, 2, 3)$ can be written as $S_i = S_0 f_i$ with $S_0 = S/4$ $(S = 1370\ W/m^2$ [16]) and f_i are constants describing the latitudinal dependence of solar insolation. The data used for f_i and for the Earth's albedo α_i are reported in Table (1) and are the average of values reported by Warren [20]. The terms f_i are weight functions that determine the quantity of incoming solar radiation in each latitudinal band and are not necessarily bounded between 0 and 1 since S_0 is only the average value of the incoming solar radiation, not the maximum value.

Following the scheme in Eqs. (2.1-2.3), it would be possible to increase the number of boxes used in the model, indeed asymptotically achieve a continuous latitudinal model. In particular, this continuous problem has been investigated in [21] where a PDE model for the climate system is developed for closed-loop climate engineering control.

The external forcing F_{ext} is defined by the radiative forcing due to carbon dioxide in the atmosphere according to the Representative Concentration Pathway 4.5 ($RCP4.5$) [22] and is given by the following expression [23]:

$$F_{ext}(t) = F_{CO_2}(t) = 5.35 log\left(\frac{CO_2(t)}{CO_{2,0}}\right) \tag{2.9}$$

where $CO_{2,0}$ is the pre-industrial level of CO_2 in the atmosphere and $CO_2(t)$ is the concentration of CO_2 according to the $RCP4.5$ scenario, which is one of the intermediate stabilisation pathways in which radiative forcing is stabilised at approximately $4.5\ W/m^2$ after 2100 [24].

(a) Validation and limitations of the model

The response of the model described above to the four Representative Concentration Pathways (RCP) scenarios is shown in Fig. (1) where the average temperature between the three latitudinal bands is reported for the four cases. This result is obtained considering $U_i = 0$ $(i = 1, 2, 3)$ in Eqs. (2.1-2.3) and is comparable with other simulations in the literature, for example in [25] where the behaviour of the CMIP5 model under the RCP scenarios is reported. In particular, in Fig. (1), the uncertainty range of the temperature anomaly at t=2100 is reported as found in figure 1 from Ref. [25] for each radiative scenario and in every case the response of the model is within the relevant uncertainty range.

This approach is considered as the verification of the general correctness (limited to the application for which it has been considered in this paper) and usefulness of the model developed. Also, considering the step response to a doubling of CO_2, the climate sensitivity of the 3-box model is estimated to be $2.3°C$, which is within the acceptable range of values found for other climate models [26].

Moreover, in Fig. (1) as in all the other simulations, a Gaussian noise (w) with a normal distribution (0 mean and 25% standard deviation, considering a 50% uncertainty range for the temperature anomaly as in [27,28]) is added to the signal later to simulate climate variability.

This section demonstrates that the 3-box model can be employed to quickly evaluate climate engineering strategies. However, it is important to highlight that the model has been developed exclusively to demonstrate the usefulness of adaptive control for climate engineering and it does not aim to substitute high fidelity climate models. Indeed, in this section, only the global mean temperature is investigated and compared with RCP scenarios. Moreover, as seen in the previous section, the climate parameters for the three bands are chosen from the literature and/or to match results from observations; however, oversimplified expressions are employed for the infra-red radiation and for the heat diffusion between boxes and the mean circulation of the ocean is completely neglected. These simplifications allowed the rapid evaluation of qualitative results for SRM under many different scenarios, but the model is not considered suitable to plan real-world deployments and its use should be limited to preliminary climate engineering assessments prior to more detailed analysis.

Now that the 3-box model has been presented and validated, the development of the adaptive control strategies can proceed. In the following sections the investigation of the stability of the system, the controllability and the observability of the strategies considered (see Sec. refsec.controllability) and the description of the Model Reference Adaptive Control (MRAC) to assess adaptive control strategies for highly-uncertain systems is presented (see Sec. 4).

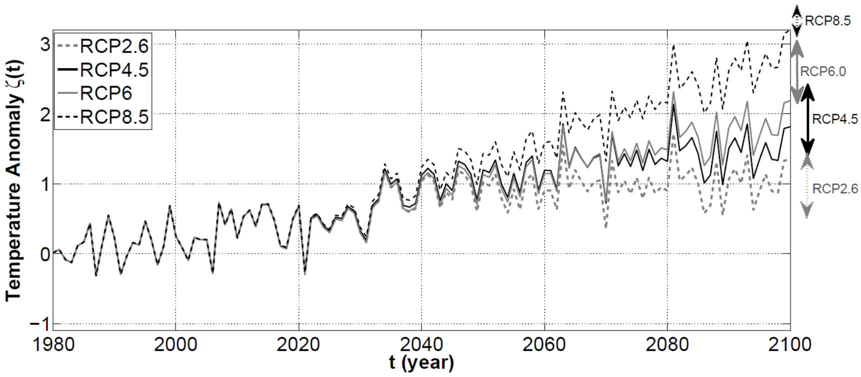

Figure 1: Response of the 3-box model described in Eqs. (2.1-2.3) to RCP scenarios. Uncertainty ranges for RCP scenarios found in [25] are reported on the right-hand side.

Table 1: Values of f_i and the Earth's albedo for the three latitudinal bands [20].

	Northern Band	Central Band	Southern Band
f_i	0.60625	1.0882	0.545625
α_i	0.498	0.281	0.498

3. Controllability and Observability of the system

In order to investigate the stability of the 3-box model defined by Eqs. (2.1-2.3), the solutions of the system $(\lambda I - A)$ are evaluated, where A is the state matrix of the system reported in Eq. (2.6). The eigenvalues are real and negative as expected demonstrating that the system is asymptotically stable.

From the point of view of the control system,low order models with more than one input and one output, such as the 3-box model, are of particular interest since it allows for the investigation of the formal controllability and observability of the system by exploring the control architectures.

In this paper, four cases are investigated: reduction of the insolation (a) in the northern band, (b) in the northern and southern band, (c) in all the three zones and (d) in the central band only. The four strategies are summarized in Table (2). The system in Sec. 2 now becomes $\frac{d\boldsymbol{\zeta}}{dt} = A\boldsymbol{\zeta} + \mathbf{F} +$

Table 2: Summary of the four control strategies investigated. The area of deployment provides the latitudinal band in which the deployment of SRM takes place. The number of controllers (or actuators) indicates the number of bands in which SRM is deployed. The objective represents the latitudinal band where the temperature anomaly is driven to zero.

Case	Area of deployment	# Controllers	Objective to minimize
a	Northern band	1	Northern band
b	Northern and Southern bands	2	Northern and Southern bands
c	Northern band, Southern band and Central band	3	all 3 bands
d	Central band	1	all 3 bands

$B\mathbf{U}$ where B is the control distribution matrix, with as many columns as the number of controllers and 3 rows equal to the dimension of the system. The number of controllers (or actuators) are different for each strategy (a-d) as reported in Table (2).

For each strategy, the controllability of the system can be verified. Since A is a non-singular 3×3 matrix, the controllability matrix [29] associated with the system in Eqs. (2.1-2.3) is given by $\Sigma = [B\ AB\ A^2B]$. Matrix Σ is used to evaluate when $rank(\Sigma) = rank(A)$. In that case the system is fully controllable and so it is in principle possible to drive the three internal states of the system from any initial state to any other final state in a finite time interval.

The matrix Σ is determined for cases (*a*-*d*) and the system is found to be always controllable for the strategies (*a*-*c*) and for strategy (*d*) only if the asymmetries of the poles are taken into account in the model. The actuator matrices, B_a, B_b and B_c, used to compute the controllability matrices above, are given by:

$$B_a = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad B_b = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \quad B_c = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.1}$$

With regard to strategy (*d*), it has been noted that if the asymmetries of the southern band with respect to the northern band are neglected and the same climate parameters are used for both latitudinal bands ($C_1 = C_3$, $b_1 = b_3$, $k_1 = k_3$), it can be demonstrated that the system is uncontrollable. In this case, if only the central band is controlled and so the matrix B is given by the vector $(0,1,0)^T$ then the controllability matrix of the system is:

$$\Sigma_d = \begin{pmatrix} 0 & j_{21} & j_{11}j_{21} + j_{21}j_{22} \\ 1 & j_{22} & j_{12}j_{21} + j_{22}^2 + j_{23}j_{32} \\ 0 & j_{23} & j_{22}j_{23} + j_{23}j_{33} \end{pmatrix} = \begin{pmatrix} 0 & 0.1185 & -0.0941 \\ 1 & -0.3973 & 0.1678 \\ 0 & 0.1203 & -0.0955 \end{pmatrix} \tag{3.2}$$

The numerical values for the components of Σ_d in Eq. (3.2) are obtained considering $C = 7.3\ W\ yr/m^2/K$, $b = 2.17\ W/m^2/^\circ C$ and $k = 0.73\ W/m^2/K$ for both poles [15].

In this case, the poles are not significantly different to each other, therefore the associated rows $\left(1^{st} \text{ and } 3^{rd}\right)$ are not independent. For this reason, $rank(\Sigma) < rank(A)$ and the system cannot be controlled.

However, when the asymmetry of the poles (mainly due to the different fraction of land and water) are taken into account and the climate parameters in Sec. 2 are considered, the controllability matrix for strategy (*d*) can be written as:

$$\Sigma_d = \begin{pmatrix} 0 & 0.1693 & -0.1492 \\ 1 & -0.3920 & 0.1628 \\ 0 & 0.0661 & -0.0470 \end{pmatrix} \tag{3.3}$$

The system is now controllable and this means that a temperature anomaly can in principle be driven to zero in all the three latitudinal bands, even if SRM is deployed only in the central band. This case has been considered to investigate the limits of controllability of the system, however it is not further investigated in this paper.

In a similar way it is possible to investigate the observability of the system in several cases. This feature is also useful to understand the number of observable states when not all the measurements are available. This could in principle occur, for example, if some measurements cannot be considered reliable or if a geopolitical disagreement causes a disruption. The observability matrix of a system is given by $\mathcal{O} = [\Psi \quad \Psi A \quad \Psi A^2]$, where Ψ is the matrix of measurements with as many rows as the number of sensors and 3 columns equal to the dimension of the system. The model is completely observable if the matrix $\mathcal{O}$ has full rank 3. Considering the cases in Tab (2), as expected, the results confirm the outcomes from the analysis of controllability: the system is fully observable in cases (*a*-*c*) and in case (*d*) if asymmetries of the poles are taken into account. In particular, this last case is investigated in Sec. 5(c), where an abrupt lack of information on the temperature anomalies in the northern and southern bands is simulated to show the response of the adaptive controller with a sudden major perturbation.

4. Adaptive control strategy

In this section, adaptive control strategies which able to deal with large uncertainties are described in detail. The adaptive controller investigated in this paper is the direct model reference adaptive

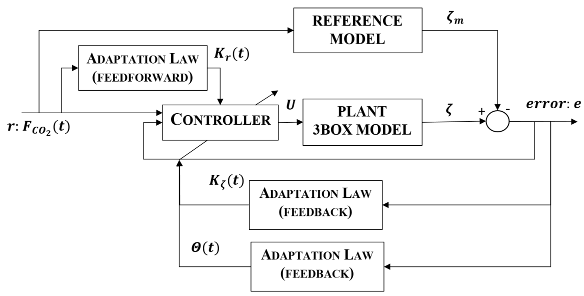

Figure 2: Model Reference Adaptive Control (MRAC) scheme.

control (MRAC) [30,31]. As it is shown in Fig. (2), the main elements of this controller are: (1) a reference model which specifies the desired response to external commands; (2) a plant model whose general structure is known but its parameters are uncertain (the 3-box model); (3) a controller that provides tracking; (4) adaptation laws to adjust the parameters of the control law during the process (feedforward and feedback).

The goal of the MRAC is to create a closed loop controller with parameters that can be updated to change the response of the uncertain system to that of an ideal model.

Consider a system (given by the reference model in Fig. (2)) where the system matrix A_m and the output $\boldsymbol{\zeta}_m$ (i.e. vector of temperature anomalies) are unknown and a stable system (given by the 3-box model) whose system matrix A and output $\boldsymbol{\zeta}$ are provided. The goal is to find a control law $\boldsymbol{U}$ such that the error $\mathbf{e}$ in Fig. (2) between the output of the 3-box model ($\boldsymbol{\zeta}$) and the reference model ($\boldsymbol{\zeta}_m$) vanishes when $t \to \infty$. The system in Eqs. (2.1-2.3) is an uncertain system and can be written as:

$$\dot{\boldsymbol{\zeta}} = A\boldsymbol{\zeta} + B(\mathbf{U} + \Theta^T\boldsymbol{\zeta}) + \mathbf{w} \tag{4.1}$$

where the term $\Theta^T\boldsymbol{\zeta}$ is now used to match the uncertainties of the system, where Θ is an unknown parameter matrix that will be part of the control law (see Eq. (4.2)) and $\mathbf{w}$ is a bounded disturbance (white noise with zero mean and 0.01 standard deviation).

The adaptive control law is parametrized as follows:

$$\mathbf{U} = K_\zeta^T\boldsymbol{\zeta} + K_r^T\mathbf{r} - \Theta^T\boldsymbol{\zeta} \tag{4.2}$$

where K_ζ, K_r and Θ are the dynamical gain matrices whose parameters are estimated at each iteration and $\mathbf{r}$ is the external forcing given by $\mathbf{F}_{CO_2}(t)$, as reported in Fig. (2). According to [30], it can be shown that the error dynamics can be written as follow:

$$\dot{\mathbf{e}} = \dot{\boldsymbol{\zeta}}(t) - \dot{\boldsymbol{\zeta}}_m(t) = A_m\mathbf{e} + B\left(\Delta K_\zeta^T\mathbf{e} + \Delta K_r^T\mathbf{r} + \Delta\Theta^T\mathbf{e}\right) \tag{4.3}$$

where $\boldsymbol{\zeta}_m$ and A_m are the unknown state vector and the unknown system matrix of the reference model (see Fig. 2). Also, ΔK_ζ, ΔK_r and $\Delta\Theta$ are given by the difference between the estimated and the ideal gain matrices.

Through the method of Lyapunov [30,32] it is possible to choose adaptive laws, i.e. control laws to suitably update the gain matrices at each iteration. These adaptive laws are chosen such that the time-derivative of a Lyapunov function decreases along the error dynamics trajectory.

The Lyapunov function candidate for the design of an MRAC system of 3^{th} order is given by:

$$V(\mathbf{e},\Delta K_\zeta,\Delta K_r,\Delta\Theta)=\mathbf{e}^T P\mathbf{e}+Tr(\Delta K_\zeta^T\Gamma_\zeta^{-1}\Delta K_\zeta)+Tr(\Delta K_r^T\Gamma_r^{-1}\Delta K_r)+Tr(\Delta\Theta^T\Gamma_\theta^{-1}\Delta\Theta) \tag{4.4}$$

where Γ_ζ, Γ_r and Γ_θ are symmetric positive definite matrices and P is a unique symmetric positive definite solution of the algebraic Lyapunov equation ($PA+A^TP=-M$ with M a symmetric positive definite matrix). Also, Tr is the trace of the matrix. If the adaptive control laws are chosen as follows:

$$\dot{K}_\zeta=-\Gamma_\zeta\mathbf{e}\mathbf{e}^TPB \qquad \dot{K}_r=-\Gamma_r\mathbf{r}\mathbf{e}^TPB \qquad \dot{\Theta}=-\Gamma_\theta\mathbf{e}\mathbf{e}^TPB \tag{4.5}$$

the time derivative of the Lyapunov function becomes semi-negative definite:

$$\dot{V}(\mathbf{e}(t),\Delta K_\zeta(t),\Delta K_r(t),\Delta\Theta(t))=-\mathbf{e}^T(t)Q\mathbf{e}(t) \tag{4.6}$$

The invariant set theorems of La Salle and Barbalat's Lemma extend the concept of the Lyapunov function providing asymptotic stability analysis tools for autonomous and non-autonomous systems with a negative semi-definite time-derivative of a Lyapunov function [30]. Therefore, when $t\to\infty$, $\dot{V}(\mathbf{e},t)=0$ and it follows from Eq. (4.6) that $||\mathbf{e}(t)||=0$.

The standard MRAC is usually known to become unstable in the presence of time delay. However, in this specific problem, since the external forcing ($\mathbf{r}(t)$) is persistently exciting the system ($RCP4.5$ scenario [24]), it is demonstrated in [30] that MRAC systems are robust despite uncertainties. Comparisons of the adaptive strategy to the conventional PI control is provided later in Secs. (5(b),5(c),5(d)).

5. Results and discussion

(a) Adaptive control applied to cases a,b,c

Control strategies with PI control in feedback have been employed for climate engineering [10,11,13]. Such control strategies can handle certain classes of parametric and dynamic uncertainties. However, adaptive control can tolerate much larger uncertainties because of the on-line estimation of the control law gains. Adaptive control therefore represents the natural solution for problems where only a nominal model of the real-world plant is available for control design and the plant parameters can vary [33], and thus appears of significant benefit to climate engineering. Moreover, it is demonstrated that adaptive control is also able to deal with unforeseen major perturbations, that can likely occur in the decades when SRM is deployed, and also the modelling of the actuator dynamics.

The method is based on the dynamical estimation of the parameters of the gain matrices and it has been proved that, despite the uncertainties on the parameters of the plant (the climate model), a suitable control strategy can always be developed. Even if the actual gains deviate from the nominal control gains, adaptive control guarantees that the values for the control gain matrices are always included in the admissible domain that would not result in loss of system stability.

The adaptive control strategy is now applied in all the three cases reported in Tab. (2). As noted in Sec. 3, the dimensions of B depend on the configuration of the control strategy (i.e. on the number of actuators) and for this reason a different controllability matrix is obtained in each case. Expressions in Eq. (4.5) are employed to adjust the control matrices K_x,K_r and Θ at each iteration.

The symmetric matrices Γ_ζ, Γ_r, Γ_Θ and M are chosen so that all the system's variables have the same order of magnitude and are comparable during the estimation of the adaptive control laws. It is important to note that these matrices are set before the beginning of the process and it is not necessary to modify them despite changes in external inputs because of the on-line update of the control laws. The chosen values for these matrices for cases (a-c) (Tab. (2)) are reported below.

$$\Gamma_{\zeta_A} = \begin{pmatrix} 3^9 & 10^8 & 2\times 10^8 \\ 10^8 & 1 & 1 \\ 2\times 10^8 & 1 & 3\times 10^9 \end{pmatrix} \quad \Gamma_{r_A} = \begin{pmatrix} 10^9 & 10^5 & 10^5 \\ 10^5 & 1 & 1 \\ 10^5 & 1 & 10^9 \end{pmatrix} \tag{5.1}$$

$$\Gamma_{\zeta_B} = \begin{pmatrix} 10^9 & 2\times 10^8 & 10^9 \\ 2\times 10^8 & 10^{-10} & 1 \\ 10^9 & 1 & 1 \end{pmatrix} \quad \Gamma_{r_B} = \begin{pmatrix} 10^8 & 10^8 & 10^8 \\ 10^8 & 1 & 1 \\ 10^8 & 1 & 1 \end{pmatrix} \tag{5.2}$$

$$\Gamma_{\zeta_C} = \begin{pmatrix} 10 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 10 \end{pmatrix} \quad \Gamma_{r_C} = \begin{pmatrix} 100 & 10 & 100 \\ 10 & 10 & 10 \\ 100 & 10 & 100 \end{pmatrix} \tag{5.3}$$

$$\Gamma_{\Theta_A} = \begin{pmatrix} 3\times 10^3 & 10^5 & 10^5 \\ 10^5 & 1 & 1 \\ 10^5 & 1 & 3\times 10^3 \end{pmatrix} \Gamma_{\Theta_B} = \begin{pmatrix} 100 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 100 \end{pmatrix} \Gamma_{\Theta_C} = \begin{pmatrix} 10 & 10 & 10 \\ 10 & 20 & 10 \\ 10 & 10 & 10 \end{pmatrix} \tag{5.4}$$

Moreover, the matrices M and P, given by the solution of the Lyapunov equation (see Sec. 4) are given below for cases (a-c).

$$M_A = 10^{-11} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.9 \end{pmatrix} \quad P_A = 10^{-10} \begin{pmatrix} 0.1092 & 0.0092 & 0.0028 \\ 0.0092 & 0.0120 & 0.0111 \\ 0.0028 & 0.0111 & 0.1250 \end{pmatrix} \tag{5.5}$$

$$M_B = 10^{-9} \begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 10^{-5} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad P_B = 10^{-8} \begin{pmatrix} 0.0011 & 0.0004 & 0.0016 \\ 0.0004 & 0.0014 & 0.0108 \\ 0.0016 & 0.0108 & 0.1379 \end{pmatrix} \tag{5.6}$$

$$M_C = \begin{pmatrix} 10^{-4} & 0 & 0 \\ 0 & 10^{-4} & 0 \\ 0 & 0 & 2\times 10^{-3} \end{pmatrix} \quad P_C = 10^{-4} \begin{pmatrix} 1.1136 & 0.2564 & 0.3621 \\ 0.2564 & 1.2678 & 2.2847 \\ 0.3621 & 2.2847 & 27.6137 \end{pmatrix} \tag{5.7}$$

The matrices M, Γ_ζ, Γ_r and Γ_Θ depend on the control strategy and their values are chosen in order to control the temperature anomaly in the northern band in case (a), in both northern and southern bands in case (b) and in all the latitudinal bands in case (c).

The temperature anomaly, defined in Eq. (2.5), and the required insolation reduction are shown in Figs. (3.a-c) for each of the three cases using the adaptive control strategy. In the first case (Fig. (3.a)) the insolation reduction is considered in the northern band. In this case, the temperature anomaly in the northern band is reduced in approximately 15 years from the deployment of SRM and a maximum 1% insolation reduction is required.

It can be noted that in the second case (Fig. (3.b)) two controllers (or actuators) are considered and SRM is deployed in both northern and southern bands. The temperature in the central band is influenced by the heat transport between the models and it levels up to $1.2^\circ C$. Again, the control law is able to minimize the temperature anomaly in both bands by 2050-2060 with a maximum insolation reduction of 1% in the northern band and 0.7% in the southern band. The third case (Fig. (3.c)) regards the full system where SRM is deployed in all bands. It is important to note that the main objective of case (c) is to minimize the temperature in the central band, although effort has been made during the control design so that the temperature anomaly in the other two bands do not become negative. In particular, in this case, the temperature anomaly in all the latitudinal bands is minimized within approximately 10 years. As can be seen, the minimization in the central band is also connected with the decline of the temperature in the other two bands. A maximum insolation reduction of approximately 0.8% and 0.6% is required in the northern and southern bands, respectively, and 0.2% in the central band.

As can be seen, adaptive control provides the necessary radiative forcing in the 3-box model to counteract human-driven climate change under the $RCP4.5$ scenario with a 50% uncertainty range for the temperature anomaly due to climate variability [27,28] (see Sec. 2 for more details). Moreover, the overall solar reduction required in case c is approximately 1.6% which is broadly comparable with literature [28].

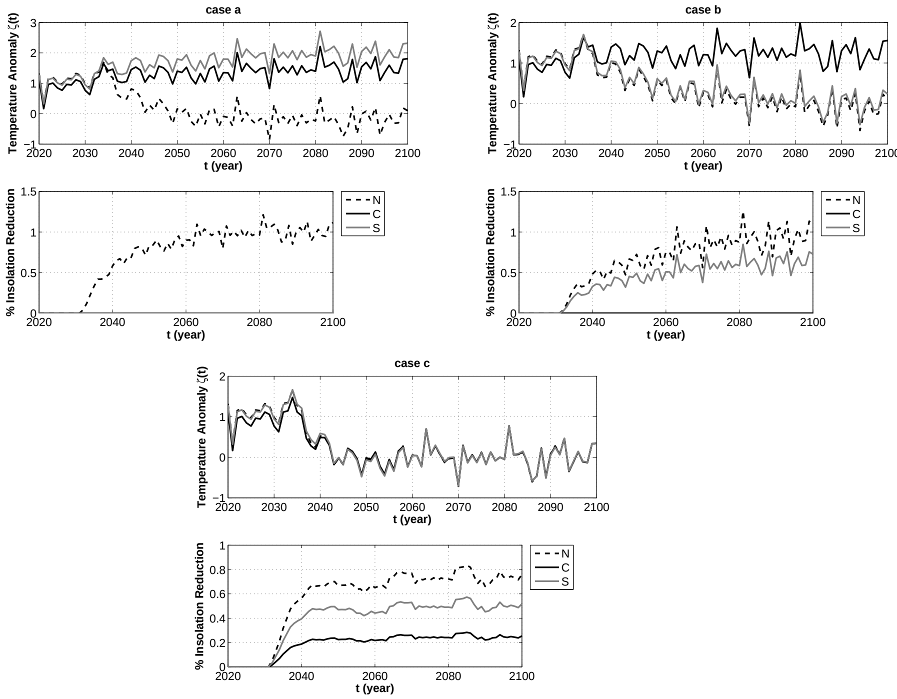

Figure 3: Temperature perturbation $\zeta(°C)$ in the 3 latitudinal bands and required insolation reduction under the $RCP4.5$ radiative scenario with SRM (adaptive control strategy) deployed in 2030 in the northern band for case (a) (1 actuator), in the northern and southern bands for case (b) (2 actuators), in all the bands for case (c) (3 actuators).

In this section, it is demonstrated that adaptive control, as with PI control (see [28]), deals well with uncertainties due to climate variability. However, in the next section, the robustness of adaptive control to large uncertainties in the climate model, unforeseen perturbations and to the choice of the actuator is demonstrated. Comparisons of the results from these investigations with the implementation of PI control are reported for each case and key differences between the two approaches are noted.

(b) Performance of adaptive and PI control with variation of the model parameters

Next simulations are used to illustrate a comparison between the implementation of PI and adaptive control in case (3).

In particular, an ideal PI controller is designed and tuned in order to minimize the temperature anomaly in the three latitudinal bands for case (3). The employed control law can be written as:

$$\mathbf{U} = -K_P \boldsymbol{\zeta} - K_I \int_0^t \boldsymbol{\zeta} dt \tag{5.8}$$

where K_P(3x3) and K_I(3x3) are proportional and integral gain matrices, respectively, whose components are obtained through linear quadratic regulator (LQR) optimization [34]. Because of the coupling between the three boxes in the climate model both the matrices are completely full. Their expressions are:

$$K_P = \begin{pmatrix} 3.214 & 0.084 & -0.00084 \\ 0.3191 & 0.4606 & 0.151 \\ -0.00082 & 0.0399 & 0.0399 \end{pmatrix} \quad K_I = \begin{pmatrix} 0.053 & 0.0025 & -0.00027 \\ -0.01494 & 0.0312 & -0.0071 \\ -0.0003 & 0.0012 & 0.0533 \end{pmatrix} \tag{5.9}$$

As discussed in Sec. 4, the adaptive control strategy is able to tolerate larger uncertainties in the parameters of the plant model with respect to a PI controller, providing good performance in all circumstances. Thus, simulations are performed where the three main parameters of the 3-box model (the heat capacity C_i $(i = 1, 2, 3)$, the transport coefficient k_i $(i = 1, 2, 3)$ and the infrared parameter b_i $(i = 1, 2, 3)$) are modified and the system's response is investigated in the cases when the adaptive and the PI controller are employed. In particular, three sets of climate parameters are reported in Tab. (3) for the northern and southern band and the central band. As

Table 3: Sets of climate parameters considered for the three latitudinal bands to compare adaptive and PI control.

Case	C_1	k_1	b_1	C_3	k_3	b_3	C_2	k_2	b_2
I	3.2	0.649	1.675	7.04	0.2325	1.195	5.12	0.44	1.8
II	6	1.049	2.675	10.04	0.5325	1.995	8.02	0.79	2.9
III	10.54	1.249	2.175	12.04	0.6325	2.045	11.29	0.94	2.5

will be seen, the PI control shows deteriorated performance when the model parameters drift from their nominal values. In particular, an issue is found in the minimization of the temperature anomaly in the northern and southern bands: plots in the bottom of Fig. (4) show that control with the PI controller in feedback is not able to correctly track the dynamics of the state and drive the anomalies to zero. Although the temperature anomaly for the central band is more or less minimized, the control strategy is not considered successful because it is not effective in the other latitudinal bands. Moreover, it is noticeable that with a larger drift of the model parameters it is more unlikely that the control is effective (see bottom plot of case I Fig. (4)), therefore PI control is considered unreliable with major uncertainties in the climate model. Whereas, plots in the top of Fig. (4) indicate that the adaptive control provides good performance in all the cases considered and no marked variations are reported with respect to the nominal case.

Therefore, as expected, the simulations with PI control demonstrate that the controller is rather sensitive to the parameterisation of the climate model. The results do not provide acceptable solutions and the control methods cannot be considered as robust as adaptive control techniques in dealing with the inevitable large (and unknown) uncertainties of the climate system. This outcome seems to be in contrast with results in [28], where PI control is found to be robust to parametric uncertainties. However, in [28] only the global mean climate state is investigated to test robustness of PI control to uncertainties; whereas, in this paper, it is shown the distinctive response of the northern and southern latitudinal bands and the central band provide more insight into the effect of feedback architectures on large latitudinal disparities. In fact, the PI control demonstrated good performances minimizing the central band, but unacceptable results are found for the reduction of temperature anomalies in the northern and southern bands.

A key point to note that, since adaptive control demonstrates robustness to large uncertainties in the climate model employed, the fidelity of the model used is of less importance if this control strategy is considered.

(c) Performance of adaptive and PI control with unforeseen major perturbations

In the previous section it was demonstrated that adaptive control is able to deal with large uncertainties in the model by adjusting the control parameters as required to minimize the temperature anomalies. A further interesting analysis regards the possible occurrence of unforeseen major perturbations, such as a sudden partial failure of the climate engineering system. In principle this event could be caused, for example, by the deterioration or failure of one or more actuators (of whatever type) during the course of the implementation of the climate engineering strategy.

In particular, in this section, employing the control law in Eq. (5.8) with the gain matrices in Eq. (5.9), it is assumed that in 2065 the actuators in the northern and southern latitudinal bands become only 50 % efficient (actuator effectiveness is only 50% of the commanded effort).

In these circumstances, the performance of adaptive control and PI control are considered and the results found in top and bottom of Fig. (5), respectively. In the simulations, in both cases, a large perturbation occurs at 2065, where the control effectiveness of the temperature anomaly in the northern and southern bands is reduced by 50%. In the case of the adaptive control, the

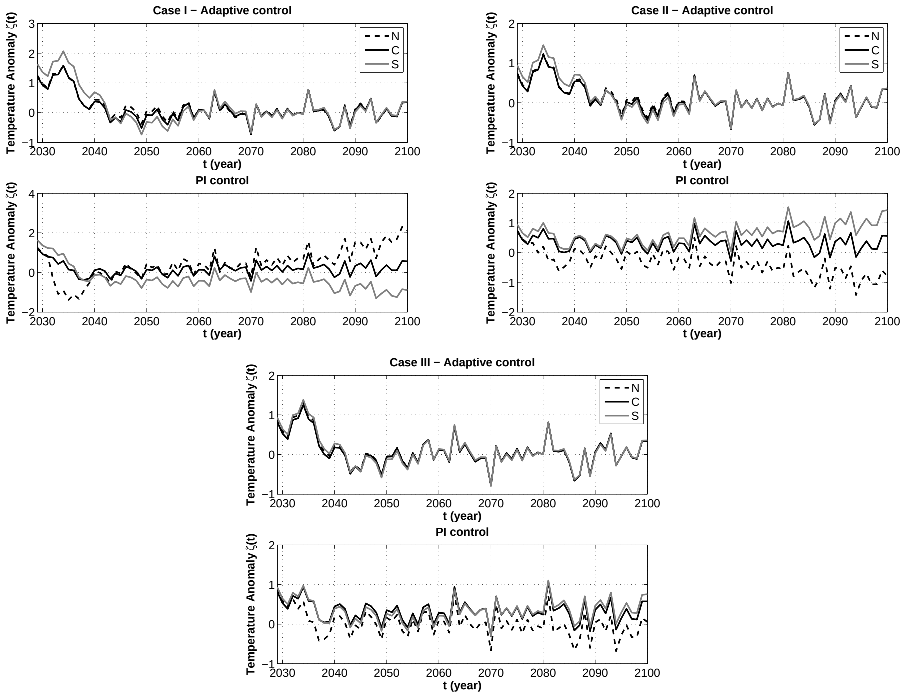

Figure 4: Temperature perturbation for the first (top left), second (top right) and third sets (bottom) of climate parameters when adaptive control and PI control are applied to the three latitudinal bands (see Tab. (3) for details).

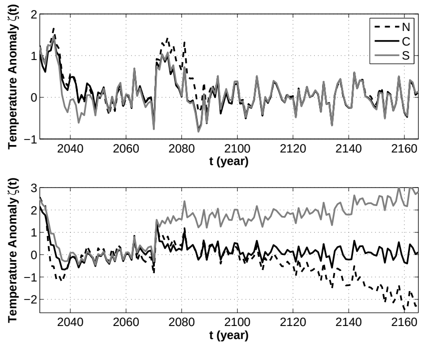

Figure 5: Temperature anomalies for the three latitudinal bands with a sudden disruption of SRM occurring at 2065 with the application of adaptive control (top) and PI control (bottom).

controller parameters are automatically adjusted in order to counteract the perturbation in these new circumstances. Specifically, the adaptive control registers the change, increases the control effort in the central band and so provides the required control in the northern and southern bands through the coupling between the boxes via heat diffusion. Whereas, when the PI control is applied (bottom of Fig. (5)), the temperature anomaly in the central band is minimized but in the other bands diverges reaching approximately $-2.1°C$ in the northern band and $3°C$ in the southern band.

In this section it is demonstrated that adaptive control is of critical importance in effectively compensating unforeseeable perturbations and failures in the climate engineering system, whereas the PI controller is not able to deal with these abrupt changes.

(d) Performance of adaptive and PI control with aerosol dynamics

Hitherto, a generic control function (U) representing the reduction of the incoming solar radiation has been considered in the closed loop control system and the performance of adaptive and PI control have been considered in several circumstances (variation of the system's parameters in Sec. 5(b), and abrupt disruption of SRM in Sec. 5(c)).

In this section, the robustness of adaptive and PI control methods is investigated in the case when a specific actuator is chosen. Among SRM methods, the emission of sulphate aerosols in the stratosphere will now be considered in the closed-loop control system.

Here, the model employed considers the decay process of the aerosol particles from the stratosphere in each latitudinal band and the poleward diffusion of aerosols injected in the central band. More detailed models describing the aerosols dynamics can be found in [7–9].

Thus, considering a latitude-average aerosol radiative forcing as in [35], the first-order dynamics of the control inputs can be written as a 3-components vector as follow:

$$\frac{dU_i}{dt} = \frac{S_i}{2}\mathcal{T}_a^2(1-A_c)\omega_i\beta_i(1-R_s)^2\tau_{0_i}(t)\, e^{-t/\Gamma_i} \qquad i=1,2,3 \tag{5.10}$$

where U_i are the components of the control vectors also found in Eqs. (2.1-2.3) and Γ_i (i=1,2,3) denotes the time constant associated with the rate of removal of aerosols, which is of order $1\ year$ for stratospheric aerosols in the central band (Γ_2) and $3\ months$ for aerosols in the northern and southern bands ($\Gamma_1,\ \Gamma_3$) [36]. Moreover, in Eq. (5.10) the incident radiation properties (S_i is the incoming solar radiation in the i^{th} band, as discussed in Sec. 2), the optical properties of the atmosphere ($\mathcal{T}_a = 0.76$ is the atmospheric transmission, $A_c = 0.6$ is the fractional cloud cover and $R_s = 0.15$ is the surface reflectance [35]) and the optical properties of aerosols (single scattering albedo $\omega_1 = \omega_3 = 0.9$, $\omega_2 = 0.5$ [37], aerosol optical depth $\tau_i(t)$ $(i=1,2,3)$, average upscatter fraction $\beta_1 = \beta_3 = 0.21$, $\beta_2 = 0.27$ [38]) are defined.

Considering Eq. (5.10), the dynamics associated with the removal of the aerosol particles can be expressed through a first order transfer function for each latitudinal band that links the actuated and the commanded control, U_{ai} and U_{ci} (i=1,2,3) respectively. A simple proportional term is considered for the diffusion from the central band to the northern and southern bands. These relationships are summarized for each latitudinal band as follow:

$$U_{a_1}(s) = \frac{1}{\Gamma_1 s+1}U_{c_1}(s) + k_d U_{a_2}(s) \tag{5.11}$$

$$U_{a_2}(s) = \frac{1}{\Gamma_2 s+1}U_{c_2}(s) - 2k_d U_{a_2}(s) \tag{5.12}$$

$$U_{a_3}(s) = \frac{1}{\Gamma_3 s+1}U_{c_3}(s) + k_d U_{a_2}(s) \tag{5.13}$$

where k_d is a diffusion coefficient set to 0.05 as in [39], where the latitudinal diffusion of the aerosol particles ejected after the eruption of El Chicon is considered.

These expressions are used directly in the closed-loop system for the three latitudinal bands in order to take into account the decay process of aerosol particles and their poleward diffusion while the control laws are estimated.

Finally, in order to make comparisons with results from the literature, the emission rate of the aerosol particles ($Tg/year$) is given as $E_{a_i} = dB_{a_i}(t)/dt\ \ (i=1,2,3)$ [41], where $B_{a_i}(t)$ is the mass (Tg) of aerosol particles (or sulphur burden [35]) in the i^{th} latitudinal band which depends on the area covered and the extinction parameter which can be estimated through Mie Theory (usually $3.5\ m^2/g$ for aerosol particles [38]). In this paper particles of radius between $0.1\ \mu m$ and $1\ \mu m$ [40] are considered and the altitude of injection is $25\ km$ for the central band and $20\ km$ for the northern and southern bands [36].

The area covered is given by 13.9% for the southern and northern band (from $65^\circ N$ to $90^\circ N$ and from $65^\circ S$ to $90^\circ S$) and 72.2% (from $65^\circ S$ to $65^\circ N$) for the central band.

In order to compare the results with data in the literature, the mass of aerosol particles is converted into units of sulphur according to [40], where the following equivalence is reported: $1\ Tg\ S = 4\ Tg$ of aerosol particles.

Thus, applying adaptive control and PI control to case c of Table (2) and considering the aerosol dynamics described above in the closed-loop control system plots in Figs. (6.a-b) are obtained.

In particular, Fig. (6.a) shows the trend of the temperature anomalies and the time history of the required sulphur burden for each band when adaptive control is applied. Whereas, Fig. (6.b) shows results when the PI control is applied.

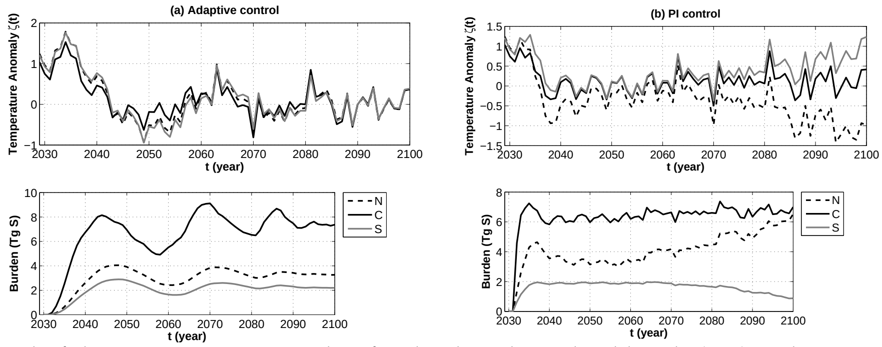

Figure 6: Trend of the temperature anomalies for the three latitudinal bands (top) and required sulphur burden ($Tg\ S$) (bottom) when adaptive control (left) and PI control (right) are applied to the 3-box model.

It is important to note that the control parameters employed for the two control methods are the same used to minimize the temperature anomalies with a generic SRM control function as in Sec. (a) and Sec. (b). Specifically, the control parameters reported in Eqs. (5.3,5.4.c,5.7) are employed for the adaptive control and those found in Eq. (5.9) for the PI control.

The analysis aims to demonstrate that the control parameters for the adaptive control do not depend on the SRM strategy considered showing that this method is robust to the choice of the actuator employed. This result is shown in Fig. (6.a) where the temperature anomalies are minimized in all latitudinal bands when adaptive control is employed. Whereas, for the PI control, it can be seen in Fig. (6.b), that the temperature anomalies largely diverge, specifically in the northern and southern bands, where they reaches $\approx -1°C$ and $\approx 1.3°C$, respectively. Adaptive control is able to deal with aerosol diffusion because the control gains are updated during every iteration in order to minimize the temperature anomaly while also considering the actuator dynamics. For this reason, it is expected also that in case the aerosol dynamics is different from that assumed in this paper, or another actuator is considered, an adaptive controller would be able to minimize the temperature anomaly and to estimate the required SRM effort. In the case of PI control, the control gains need to be selected for every specific case because once chosen they are not automatically up-dated; therefore, when different actuator dynamics are considered, the PI control does not deliver the same performance.

With regards to the sulphur burden, in order to offset the radiative forcing from the $RCP4.5$ scenario (see Fig. (1)), the time-decay of the aerosol particles and their poleward diffusion from the central band, the aerosol mass needs to increase with the time. In particular, at the bottom of Fig. (6.a), the aerosol mass shows a periodic trend with a steady increase up to $\approx 8\ Tg\ S$ in the central band, whereas the concentration in the northern and southern bands rises up to only $2-3\ Tg\ S$. Otherwise, in Fig. (6.b), the trend of the aerosol burden (bottom figure) reaches a peak of $6\ Tg\ S$ for the central band, providing enough insolation reduction to minimize the temperature anomaly. Whereas, with PI control an incorrect estimation of the required control is found for the northern and southern bands due to the fact that the control parameters are not taking the aerosol dynamics into account.

Values of the sulphur burden in Fig. (6.a) are in accordance with values found in [42] (see Tab.2 in [42]) for a given value of the optical depth for aerosol deployed at approximately 25 km ($30hP$).

Case (c) of Tab. (2) is the most critical because it involves the control of the temperature anomaly in the central band, which is higher than the other two bands. However, the required emission rates are still within an acceptable range of values: around $0.6-0.4\ Tg\ S$ per year in the northern and southern bands, respectively, and around $2\ Tg\ S$ per year in the central band. Although comparisons with the literature are complicated because the experiments differ in size, area of injection and environment, in [43] a similar experiment with the $RCP4.5$ scenario has been performed and an emission rate of $6\ Tg\ S/yr$ found (in an open-loop simulation) to counteract radiative forcing [42].

Moreover, GeoMIP (Geoengineering Model Intercomparison Project) experiment G3 considers an emission rate of $5\ Tg\ SO_2$ per year ($2.5\ Tg\ S/yr$) on the equator to balance the $RCP4.5$ scenario with stratospheric aerosols for the period between 2020 and 2070 [44].

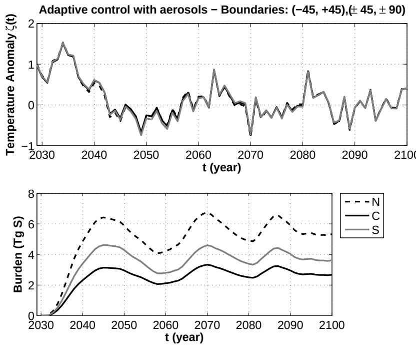

Figure 7: Trend of the temperature anomalies for the three latitudinal bands (top) and required sulphur burden ($Tg\ S$) (bottom) when adaptive control is applied to the 3-box model with new regional boundaries: $(-45^{\circ}, +45^{\circ})$ for the central band and $(\pm 45^{\circ}, \pm 90^{\circ})$] for northern and southern band.

Moreover, the 3-box model can be employed to evaluate the influence of the boundaries chosen for the three latitudinal bands. Thus, assuming the new boundaries, given by $(-45^{\circ}, 45^{\circ})$ for the central band and $(\pm 45^{\circ}$ to $\pm\, 90^{\circ})$ for northern and southern bands, the analysis of adaptive control with sulphur aerosols dynamics is performed again. Results are reported in Fig. (7) and it is found that, choosing boundaries closer to mid-latitudes caused the behaviour of each box to become similar to each other, loosing some information on the most important differences between the polar and central regions. This effect is particularly visible for the northern and southern bands in Fig. (7).

Also, as expected, considering the new boundaries, it is found that the distribution of aerosols required is different in each band with respect to the previous case investigated but the overall quantity of aerosols is estimated to be the same in both cases (see Fig. (6) for comparison). This result highlights the independence of the aerosol injection strategy from the boundaries of the 3-box model.

6. Performance of adaptive control with collapsing Arctic ice-sheet in a 5-box climate model

In this section, the 3-box model is modified in order to take into account the climate conditions near the North Pole and South Pole. In fact, between $\pm 70^{\circ}$ and $\pm 90^{\circ}$ ice sheets provide a considerably different albedo with respect to any other region on Earth with the Arctic and Antarctic albedo as high as 0.6-0.7 [45]. Several analytical climate models in the literature consider a step function albedo in order to model the insolation of these regions [15], but these are difficult to manage and can easily generate mathematical artefacts [46].

Otherwise, the 3-box model can be easily expanded to n-boxes, following the structure in Eqs. (2.1-2.3), and so provides a useful tool to quickly investigate the effect of climate engineering over the polar regions and their interaction with the other latitudinal bands.

Thus, the 3-box model becomes a 5-box model considering the following subdivision: southern polar band $(-90^{\circ}, 70^{\circ})$, southern band $(-70^{\circ}, -50^{\circ})$, central band $(-50^{\circ}, 50^{\circ})$, northern band $(50^{\circ}, 70^{\circ})$, northern polar band $(70^{\circ}, 90^{\circ})$. In particular, for the northern polar band the model parameters are $C_n = 4.2\ \ W\ \ yr/m^2/^{\circ}$, $S_n = 176.56\ \ W/m^2$, $\alpha_n = 0.6665$, $b_n = 1.45\ W/m^2/^{\circ}$, $k_n = 0.52\ W/m^2/K$; and for the southern polar band $C_s = 6.5\ W\ yr/m^2/^{\circ}$, $S_s = 194.21\ W/m^2$, $\alpha_s = 0.7095$, $b_s = 1.47\ W/m^2/^{\circ}$, $k_s = 0.76\ W/m^2/K$ [20,45].

As before the $RCP4.5$ radiative scenario is considered as uniformly distributed external disturbance. However, in this case, for example, it is now assumed that in 2060 a major collapse of the ice sheet in the Arctic (northern polar band) occurs. Consequently, a rapid reduction of the Arctic albedo (1% per year) takes place. In this simulation, SRM is deployed in 2030 with an adaptive controller. Thus, it is expected that the controller adjusts the control gains to counteract the radiative forcing due to increasing CO_2 as well as providing the necessary insolation reduction in the north polar band when the change in the albedo occurs in 2060. Results

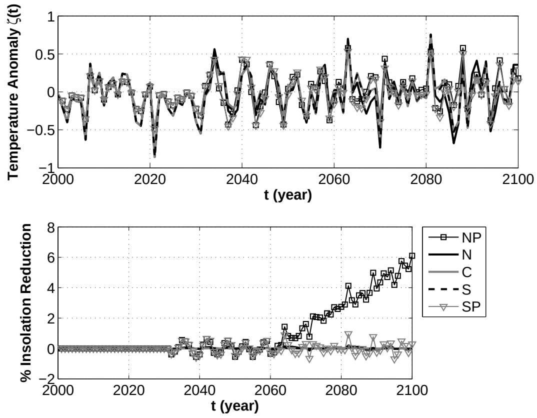

Figure 8: Temperature anomalies (top) due to the RCP4.5 scenario and insolation reduction (bottom) for the five latitudinal bands with adaptive control deployed in all the latitudinal bands in 2030 and a collapse of the ice sheet in the northern polar band occurring in 2060.

of the simulation can be found in Fig. (8), where it can be seen that adaptive control is able to deal with the sudden change in albedo due to the collapsing ice sheets in the Arctic and rapidly provide the required insolation reduction to counteract all external disturbances.

7. Conclusion

A 3-box model for the climate system has been employed: the surface is divided into the northern, southern and central latitude bands to account for temperature disparities between mid and high latitudes. Assuming independent climate engineering interventions in each band, the model provides a multiple control input system to explore strategies to mitigate latitudinal climate warming and provide clarity to assess the performance of adaptive and PI control strategies. A new control strategy involving an adaptive controller is considered for the first time for climate engineering to counteract human-driven climate change.

The 3-box model does not aim to be a substitute for high fidelity General Circulation Models (GCM) models for the description of the climate system and its use should be limited to initial investigations preceding more complete analysis. However, for the purposes of this paper (i.e. to establish a control strategy able to overcome issues related to large uncertainties), it is a useful tool to demonstrate the performance of adaptive control strategies.

The multi-variable approach with the 3-box model also allows for an investigation of the formal controllability and observability of the system for the first time. The controllability analysis demonstrates that the three of the strategies considered are always controllable with the 3-box model. The fourth strategy involving the control of the three latitudinal bands through the deployment of SRM in the central band only shows that the system can be controlled only if the asymmetries of the northern and southern bands are taken into account. Considering the controllability matrix, it is demonstrated that if the same climate parameters are considered for these two bands, the matrix has two linearly-dependent rows and it is therefore uncontrollable. Otherwise, considering the largely different fraction of land and water in the two hemispheres the system then becomes controllable.

Performance and robustness of the adaptive control has been tested and compared with a PI controller. Considering variations of the three main model parameters of the 3-box model, adaptive and PI control methods were compared in order to investigate the susceptibility of the control strategies to uncertainties in the model. Results show that adaptive control is robust to large uncertainties in the climate model itself, de-emphasizing therefore the importance of the model employed. Whereas, as expected, the PI control shows poor performance when large variations from the nominal model parameters are considered and, therefore, it does not provide satisfactory results in any of the cases investigated.

Moreover, adaptive control shows excellent performance in case of unforeseen perturbations, such as a sudden partial failure in the climate engineering system in the northern and southern bands. In fact, the controller parameters are automatically adjusted in order to counteract the perturbation in these new circumstances. In this case results from the implementation of the PI control show poor performance.

Finally, it has been demonstrated that adaptive control is also robust to the choice of the method employed to deploy SRM. When the dynamics of stratospheric sulphur aerosols is considered for the actuator in the closed loop system, without modifying the adaptive control parameters, again, the controller is able to respond properly in order to minimize the temperature anomalies in all the latitudinal bands. Also, no significant changes are found with respect to the case where a generic control function is used. Results from this simulation indicate that the temperature anomaly under the $RCP4.5$ scenario could be offset in all the three latitudinal bands injecting 0.6-0.4 $Tg\ S/yr$ in the northern and southern bands, respectively, and 2 $Tg\ S/yr$ in the central band. The values estimated for the emission rates are within acceptable bounds and are broadly comparable with results from the literature. Also, as expected, it is found that the overall quantity of aerosols required does not depend on the boundaries chosen for the 3-box model and only their distribution within the boxes is affected.

Applying a PI controller in the same case it is noted that the estimated sulphur burden required to minimize the temperature anomalies is incorrectly estimated in the northern and southern bands leading to poor performance.

Thus, since adaptive control has shown superior performance in several scenarios, it has been chosen as controller for the last simulation, involving a collapse of the ice-sheets in the Arctic. In particular, in this case extreme climatic conditions of northern and southern polar bands are considered by adding two additional latitudinal bands to the model. Again, adaptive control provided the necessary insolation reduction to counteract radiative forcing due to carbon dioxide as well as a rapid change in albedo caused by the melting ice in the Arctic region.

Therefore, as expected, these simulations demonstrate that the adaptive control works well with large uncertainties in the climate model, with unforeseeable perturbations and does not depend on the method chosen to deploy SRM. Since the control gain matrices are updated during every iteration, adaptive control guarantees the convergence of the strategy.

Ethics. No ethics statement is required for this work.

Data Accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material.

Competing Interests. We declare we have no competing interests.

Authors' Contributions. FB conceived the mathematical model, implemented and performed the simulations, interpreted the computational results in consultation with CM and wrote the paper. Furthermore, CM provided guidance and supervision for the entire process. All authors gave final approval for publication.

Funding. This work was supported by a University of Glasgow PhD scholarship (FB) and a Royal Society Wolfson Research Merit Award (CM).

Acknowledgements. Federica Bonetti acknowledges a University of Glasgow PhD scholarship and Colin McInnes acknowledges support from a Royal Society Wolfson Research Merit Award.

References

1. Flato G, Marotzke J, Abiodun B, Braconnot P, Chou SC, Collins WJ, Cox P, Driouech F, Emori S, Eyring V, Forest C. *Evaluation of Climate Models. 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*. Climate Change 2013. 2013;5:741-866.
2. Keller DP, Feng EY, Oschlies A. *Potential climate engineering effectiveness and side effects during a*

high carbon dioxide-emission scenario. Nature Communications. 2014 Feb 25;5.
3. Blackstock JJ, Battisti DS, Caldeira K, Eardley DM, Katz JI, Keith DW, Patrinos AA, Schrag DP, Socolow RH, Koonin SE. *Climate engineering responses to climate emergencies* (2009) Novim, Santa Barbara, CA, USA (29 July 2009)
4. Vaughan NE, Lenton TM. *A review of climate geoengineering proposals*. Climatic Change. 2011 Dec 1;109(3-4):745-90. DOI: 10.1007/s10584-011-0027-7.
5. Teller E, Wood L, Hyde R. *Global warming and ice ages: I. Prospects for physics-based modulation of global change*. Lawrence Livermore National Lab., CA (United States).
6. McInnes CR. *Space-based geoengineering: challenges and requirements*. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 2010 Mar 1;224(3):571-80.
7. Kravitz, B., MacMartin, D. G., Mills, M. J., Richter, J. H., Tilmes, S., Lamarque, J.-F., Tribbia, J. J.,Vitt, F. (2017). *First simulations of designing stratospheric sulfate aerosol geoengineering to meet multiple simultaneous climate objectives*. Journal of Geophysical Research: Atmospheres, $122, 12, 61612, 634$. https://doi.org/10.1002/2017JD026874.
8. MacMartin, D. G., Kravitz, B., Tilmes, S., Richter, J.H., Mills, M. J., Lamarque, J.-F., Tribbia, J. J., & Vitt, F. (2017). *The climate response to stratospheric aerosol geoengineering can be tailored using multiple injection locations*. Journal of Geophysical Research: Atmospheres, $122, 12, 57412, 590$.
9. Dai, Z., Weisenstein, D. K., & Keith, D. W. (2018). *Tailoring meridional and seasonal radiative forcing by sulfate aerosol solar geoengineering*. Geophysical Research Letters, 45, $1030-1039$.
10. Kravitz, B., MacMartin, D. G., Wang, H., and Rasch, P. J. *Geoengineering as a design problem*, Earth System Dynamics, 7, 469-497, DOI: https://doi.org/10.5194/esd-7-469-2016, 2016.
11. MacMartin DG, Kravitz B, Keith DW. *Geoengineering: The world's largest control problem*. In American Control Conference (ACC), 2014 2014 Jun 4 (pp. 2401-2406). IEEE.
12. Weller SR, Schulz BP. *Geoengineering via solar radiation management as a feedback control problem: Controller design for disturbance rejection*. In Control Conference (AUCC), 2014 4th Australian 2014 Nov 17 (pp. 101-106). IEEE. DOI:10.1109/AUCC.2014.7358684.
13. MacMartin DG, Kravitz B, Keith DW, Jarvis A. *Dynamics of the coupled human climate system resulting from closed-loop control of solar geoengineering*. Climate Dynamics. 2014 Jul 1;43(1-2):243-58.
14. Jarvis A, Leedal D, Taylor CJ, Young P. *Stabilizing global mean surface temperature: a feedback control perspective*. Environmental Modelling & Software. 2009 May 31;24(5):665-74.
15. North GR, Cahalan RF, Coakley JA. *Energy balance climate models*. Reviews of Geophysics. 1981 Feb 1;19(1):91-121. DOI: 10.1029/RG019i001p00091.
16. McGuffie K, Henderson-Sellers A. *A climate modelling primer*. John Wiley & Sons; 2005 Mar 11.
17. Budyko MI. *The effect of solar radiation variations on the climate of the Earth*. Tellus. 1969 Oct 1;21(5):611-9. DOI: 10.1111/j.2153-3490.1969.tb00466.x.
18. Cess RD. *Climate change: An appraisal of atmospheric feedback mechanisms employing zonal climatology*. Journal of the Atmospheric Sciences. 1976 Oct;33(10):1831-43.
19. Kleidon A., Lorenz R. () *1 Entropy Production by Earth System Processes*. In: Kleidon A., Lorenz R.D. (eds) Non-equilibrium Thermodynamics and the Production of Entropy. Understanding Complex Systems. Springer, Berlin, Heidelberg.
20. Warren SG, Schneider SH. *Seasonal simulation as a test for uncertainties in the parameterizations of a Budyko-Sellers zonal climate model*. Journal of the Atmospheric Sciences. 1979 Aug;36(8):1377-91.
21. Bonetti F., McInnes C., 2018, Climate Dynamics, under review.
22. Kelly DL, Kolstad CD. *Integrated assessment models for climate change control*. International Yearbook of Environmental and Resource Economics. 1999;2000:171-97.
23. Myhre G, Highwood EJ, Shine KP, Stordal F. *New estimates of radiative forcing due to well mixed greenhouse gases*. Geophysical Research Letters. 1998 Jul 15;25(14):2715-8.
24. Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, Van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA. *The next generation of scenarios for climate change research and assessment*. Nature. 2010 Feb 11;463(7282):747-56.
25. Knutti R, Sedláček J. *Robustness and uncertainties in the new CMIP5 climate model projections*. Nature Climate Change. 2013 Apr 1;3(4):369-73. DOI: 10.1038/NCLIMATE1716.
26. Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor (2012), *Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models*, Geophys. Res. Lett., 39, L09712.
27. Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver and M. Wehner, *2013:*

Long-term Climate Change: Projections, Commitments and Irreversibility. 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 [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

28. Kravitz, B., MacMartin, D. G., Leedal, D. T., Rasch, P. J., Jarvis, A. J. (2014). *Explicit feedback and the management of uncertainty in meeting climate objectives with solar geoengineering*. Environmental Research Letters, 9(4), 044006.
29. Moore B. *Principal component analysis in linear systems: Controllability, observability, and model reduction*. IEEE Transactions on Automatic Control. 1981 Feb;26(1):17-32.
30. Lavretsky E, Wise KA. *Robust Adaptive Control*. In Robust and Adaptive Control 2013 (pp. 317-353). Springer London.
31. Barkana I. *Simple adaptive control: a stable direct model reference adaptive control methodology–brief survey*. International Journal of Adaptive Control and Signal Processing. 2014 Jul 1;28(7-8):567-603.
32. Nguyen NT. *Adaptive Control for linear uncertain systems with unmodeled dynamics revisited via optimal control modification*. In AIAA Guidance, Navigation, and Control (GNC) Conference 2013 (p. 4988).
33. Seborg, D. E., Edgar, T. F. and Shah, S. L. (1986), *Adaptive control strategies for process control: A survey*. AIChE J., 32: 881-913. doi:10.1002/aic.690320602.
34. Anderson, Brian DO and Moore, John B *Optimal control: linear quadratic methods*. 2007. Courier Corporation.
35. Charlson, R.J., S.E. Schwartz, J.M. Hales, R.D. Cess, J.A. Coakley, Jr., J.E. Hansen, and D.J. Hoffman, 1992: *Climate forcing by anthropogenic aerosols*. Science, 255, 423-430,
36. Robock A, Oman L, Stenchikov GL. *Regional climate responses to geoengineering with tropical and Arctic SO_2 injections*. Journal of Geophysical Research: Atmospheres. 2008 Aug 27;113(D16).
37. Bergin, M. H., 2000: *Aerosol radiative properties and their impacts*. From Weather Forecasting to Exploring the Solar System, C. Boutron, Ed., EDP Sciences, 5165.
38. Penner JE, Andreae MO, Annegarn H, Barrie L, Feichter J, Hegg D, Jayaraman A, Leaitch R, Murphy D, Nganga J, Pitari G. *Aerosols, their direct and indirect effects*. In Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change 2001 (pp. 289-348). Cambridge University Press.
39. King, M. D., & Arking, A. (1984). *A model of the radiative properties of the El Chichon Stratospheric Aerosol layer*. Journal of Climate and Applied Meteorology, 23(7), 1121-1137.
40. Rasch PJ, Tilmes S, Turco RP, Robock A, Oman L, Chen CC, Stenchikov GL, Garcia RR. *An overview of geoengineering of climate using stratospheric sulphate aerosols*. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. 2008 Nov 13;366(1882):4007-37.
41. Brovkin V, Petoukhov V, Claussen M, Bauer E, Archer D, Jaeger C. *Geoengineering climate by stratospheric sulfur injections: Earth system vulnerability to technological failure*. Climatic Change. 2009 Feb 1;92(3):243-59.
42. Niemeier U, Timmreck C. *What is the limit of climate engineering by stratospheric injection of SO_2?*. Atmospheric Chemistry and Physics. 2015 Aug 18;15(16):9129-41.
43. Niemeier U, Schmidt H, Alterskjar K, Kristjánsson JE. *Solar irradiance reduction via climate engineering: Impact of different techniques on the energy balance and the hydrological cycle*. Journal of Geophysical Research: Atmospheres. 2013 Nov 16;118(21).
44. Tilmes S, Mills MJ, Niemeier U, Schmidt H, Robock A, Kravitz B, Lamarque JF, Pitari G, English JM. *A new Geoengineering Model Intercomparison Project (GeoMIP) experiment designed for climate and chemistry models*. Geoscientific Model Development. 2015 Jan 15;8(1):43-9.
45. Winton, M. (2013). *Sea Ice–Albedo Feedback and Nonlinear Arctic Climate Change. In Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications* (eds E. T. DeWeaver, C. M. Bitz and L. Tremblay). doi:10.1029/180GM09.
46. Griffel, D.H. and P.G. Drazin, 1981: *On Diffusive Climatological Models*. Journal of Atmospheric Science, 38, 2327–2332, https://doi.org/10.1175/1520-0469(1981)038<2327:ODCM>2.0.CO;2.