Check for updates

ADVANCING EARTH AND SPACE SCIENCES

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2025JD044319

Solar Geoengineering Strategies Based on Reinforcement Learning

Heng Quan[1,2], Daniel D. B. Koll[1], Nicholas Lutsko[3], and Janni Yuval[4]

[1]Department of Atmospheric and Oceanic Sciences, Peking University, Beijing, China, [2]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA, [3]Scripps Institution of Oceanography, La Jolla, CA, USA, [4]Department of Earth, Atmospheric, and Planetary Sciences, MIT, Cambridge, MA, USA

Key Points:

- We let a reinforcement learning (RL) algorithm control the stratospheric aerosol in a global climate model (GCM)
- RL learns to produce stable and plausible stratospheric aerosol injection strategies within several dozen GCM simulations
- RL shows that the optimal geoengineering strategy depends on the time when geoengineering is initiated

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

D. D. B. Koll,
dkoll@pku.edu.cn

Citation:

Quan, H., Koll, D. D. B., Lutsko, N., & Yuval, J. (2025). Solar geoengineering strategies based on Reinforcement Learning. *Journal of Geophysical Research: Atmospheres*, *130*, e2025JD044319. https://doi.org/10.1029/2025JD044319

Received 10 MAY 2025
Accepted 21 NOV 2025

Abstract Solar geoengineering via stratospheric aerosol injection (SAI) poses an optimization problem. How exactly should aerosol be deployed to maximize its benefits while minimizing undesirable side-effects, such as shifts in rainfall patterns? Previous work explored this problem using feedback control based on linear algorithms. Here, we investigate an alternative approach, which also naturally incorporates feedback. We let a reinforcement learning (RL) algorithm control the distribution of stratospheric aerosol concentration in an idealized global climate model (GCM). Within several dozen GCM simulations, RL learns to produce stable and plausible strategies. RL also learns that the optimal geoengineering strategy depends on the time when geoengineering is initiated, which we further explain using a simple energy-balance model. Our results provide a first proof-of-concept that RL can identify promising SAI strategies.

Plain Language Summary Society might be able to temporarily mitigate the worst impacts of climate change by injecting reflective aerosols into the stratosphere. What is the best way to deploy aerosol without creating additional problems, such as disrupting monsoons and storm tracks? Previous research tackled this question using linear algorithms from the control literature. Our goal is to investigate the potential of an alternative algorithm class, namely an AI technique called reinforcement learning (RL). We train an RL algorithm to control the pattern of stratospheric aerosol concentration inside a global climate model. Initially, the algorithm produces random aerosol patterns. Over time, it learns how to best use aerosol to keep temperature and rainfall patterns in the model close to a desired target state. The algorithm also learns nonobvious strategies, such as how to vary the concentration of aerosol over time to better overcome Earth's thermal inertia and cool the climate faster. These results show RL is a feasible and promising technique for future geoengineering research. More work is needed to compare the strategies identified here to those produced by alternative algorithms.

1. Introduction

Solar geoengineering (SG) via stratospheric aerosol injection (SAI) is a proposal to counteract climate change by injecting reflective aerosol particles into the upper atmosphere (Crutzen, 2006; National Research Council, 2015; Smith & Wagner, 2018; D. G. MacMartin et al., 2019). SAI promises to cool the planet but it does not simply reverse climate change. Stratospheric aerosols impact the planet's surface and top-of-atmosphere energy budgets differently than greenhouse gases such as CO_2 (Bala et al., 2008; Seeley et al., 2021). Depending on how SAI is implemented, a deployment that would reduce global-mean temperatures may not fully reverse all aspects of climate change. SAI can result in residual polar amplification (Kravitz et al., 2013), disruptions to Asian and African monsoons (Bala et al., 2008; K. L. Ricke et al., 2010; K. Ricke et al., 2023), and a weakening or shift of the storm tracks (Gertler et al., 2020).

The exact nature and magnitude of these side-effects can be minimized by tailoring the SAI response to better offset the impacts of CO_2 forcing, posing a complex optimization problem (Ban-Weiss & Caldeira, 2010; D. G. MacMartin et al., 2018; D. G. MacMartin & Kravitz, 2019). For example, previous work showed that by preferentially deploying aerosol at high latitudes, SAI could counteract polar amplification in addition to global-mean warming (Ban-Weiss & Caldeira, 2010; Lutsko et al., 2020). By adjusting both the pattern and the timing of aerosol deployment SAI could also meet more complicated objectives, such as simultaneously minimizing temperature and precipitation changes (Bala et al., 2008; K. L. Ricke et al., 2010; D. G. MacMartin et al., 2013; Brody et al., 2025), limiting the rate of warming (MacMartin, Caldeira, & Keith, 2014), keeping global warming within the goals of the Paris Agreement (D. G. MacMartin et al., 2018; Tilmes et al., 2020), or minimizing Arctic sea ice changes (Jackson et al., 2015; Visioni et al., 2020; W. R. Lee et al., 2023; Zhang et al., 2024). The degrees of freedom available to optimize the climate's response, also known as SAI's design space, are the amount of aerosol deployed plus the spatial and temporal pattern with which it is deployed.

How can one use these degrees of freedom to design a successful SAI strategy? One feature that will likely be crucial is feedback or feedback control (Kravitz et al., 2014; MacMartin, Kravitz, et al., 2014; Kravitz et al., 2017; D. G. MacMartin & Kravitz, 2019). Feedback means that, as SAI commences, the climate's response is monitored by an algorithm. The algorithm then continuously adjusts the amount and pattern of aerosol to better achieve the desired target state. This allows one to achieve robust outcomes even with imperfect knowledge, for example, due to model error (the climate's response to SAI is different than what one thought it would be) or internal variability (the climate varies not just in response to SAI).

How is feedback control implemented in practice? The state-of-the-art in SAI research is linear control algorithms, such as proportional-integral controllers with feedforward (Kravitz et al., 2017; Mills et al., 2017; Richter et al., 2017; D. G. MacMartin et al., 2017; Tilmes et al., 2017; D. G. MacMartin & Kravitz, 2019). These algorithms typically use an initial guess (the feedforward) as to how the climate will respond to SAI forcing, then use linear feedback to compensate for deviations from the initial guess. Among the advantages of this approach are that the resulting SAI strategies are usually stable, and that the algorithms only have few tuning parameters, so computationally the design process is relatively cheap. In recent years, the specific linear feedback control algorithm developed by Kravitz et al. (2017) has been widely used in several major SAI model intercomparisons and ensemble experiments (Tilmes et al., 2018; Richter et al., 2022; D. MacMartin et al., 2022; Visioni et al., 2021, 2024).

However, current linear feedback algorithms also have disadvantages. Both feedforward and feedback have to be designed through a manual process called system identification. This process is complex and requires manual inputs at multiple stages, especially if the goal is to control multiple (potentially interacting) climate objectives (Kravitz et al., 2014; MacMartin, Kravitz, et al., 2014; Kravitz et al., 2017; D. G. MacMartin & Kravitz, 2019). As a result, no algorithm has yet been used to control more than 3 climate objectives simultaneously. Moreover, the climate objectives considered so far have typically been simple metrics. For example, the widely used algorithm of Kravitz et al. (2017); Tilmes et al. (2018); Richter et al. (2022) controlled the global-mean temperature, interhemispheric temperature gradient, and equator-to-pole temperature gradient; W. Lee et al. (2020) controlled the ITCZ position; Jackson et al. (2015); Visioni et al. (2020); W. R. Lee et al. (2023); Zhang et al. (2024) controlled the Arctic sea ice. Similarly, linear feedback control algorithms have been found to do a reasonably good job of keeping surface temperature stable down to regional scales but not precipitation (Richter et al., 2022; Visioni et al., 2021) due to a trade-off between temperature and precipitation (Tilmes et al., 2013; W. Lee et al., 2020). Very limited work has tried to simultaneously minimize temperature and precipitation changes (Brody et al., 2025). It is thus still an open question how to design appropriate linear algorithms in scenarios in which one would want to control the spatial distribution of many variables simultaneously (e.g., precipitation plus temperature patterns in many regions), or in which a high degree of control might be important in some regions but not in others (e.g., controlling precipitation patterns over land vs. ocean).

What alternatives exist to linear control algorithms? One algorithm class which, to our knowledge, has not yet been explored in geoengineering research is reinforcement learning (RL). RL is a machine learning optimization technique in which an algorithm repeatedly interacts with a system to develop an optimal response or strategy. This approach has already proven itself in a wide range of domains including game playing, autonomous navigation, and robotics (Annaswamy, 2023; Bellemare et al., 2020; Fawzi et al., 2022; Silver et al., 2016). Because RL continuously interacts with a system, it naturally incorporates feedback. One advantage of RL over linear control algorithms is that it easily generalizes to complex scenarios with many interacting climate objectives that can have arbitrarily complex spatial distributions. Similarly, RL training explores a large parameter space and can formulate potentially nonlinear responses, which means RL might be able to identify novel SAI strategies that are inaccessible to linear algorithms.

RL's attractive features come at a cost. RL training requires large amounts of data. In practice, this might mean RL is too computationally expensive for geoengineering research, particularly research based on complex global climate models (GCMs). Moreover, there is no guarantee that the policies developed by RL will be stable (Khan et al., 2012). Finally, the outcome of an RL algorithm is a trained policy function typically encoded by a neural network (see below). This means any strategies identified by RL might be difficult to understand intuitively.

Our goal in this paper is thus to explore whether RL is a potentially feasible approach for geoengineering research. We let an RL algorithm control the distribution of stratospheric aerosol concentration in a GCM and consider three idealized warming scenarios in which CO_2 is abruptly doubled. RL is then tasked with maintaining surface temperature and precipitation minus evaporation (P-E) patterns simultaneously. Starting from random, the RL algorithm trains itself to manipulate temperature and P-E within several dozen GCM simulations. The resulting RL strategies are plausible and match considerations discussed in previous work; in particular, RL recovers the "kicking the can down the road effect" (Brody et al., 2024; Pflüger et al., 2024) and a preference for employing more aerosol in the polar regions (Ban-Weiss & Caldeira, 2010; Lutsko et al., 2020). Overall, our results provide a first proof-of-concept for RL in geoengineering research. They also set the stage for future work to more directly compare RL strategies against those identified by linear feedback control and other algorithms.

2. Methods

A priori, it is unclear whether the RL algorithm (described in Text S1 in Supporting Information S1) will be stable once coupled to a climate model. To test its stability and set the algorithm's hyperparameters, we first use a 1D energy-balance model (EBM), a computationally cheap model of Earth's climate in which one can analytically solve for the optimum aerosol forcing pattern. The RL algorithm remains stable across a 21-member ensemble of EBMs and recovers the true optimal aerosol forcing pattern in this model (see Text S2 in Supporting Information S1).

Next, we couple the RL algorithm to a more complex climate model in which the optimum aerosol pattern is unknown. Because it is unclear how quickly the RL algorithm will converge in this model, we use several idealizations to keep computational costs low. In doing so we closely follow Ban-Weiss and Caldeira (2010), who developed one of the first proofs-of-concept for solar geoengineering (SG) optimization methods using an idealized GCM. We use a fully 3D but older-generation GCM, namely the Community Atmosphere Model 3.0 (Collins et al., 2004) coupled to the Community Land Model 3.0 (Oleson et al., 2004) and a slab ocean. Horizontal resolution is kept modest (32 × 64 grid points) and the model's aerosol is noninteractive. That means aerosol exerts a radiative effect, but its particle size and spatial distribution do not change due to winds or microphysics.

Consistent with Ban-Weiss and Caldeira (2010), we perform two GCM reference simulations without geoengineering: a "present-day" control simulation with 390 ppm of CO_2 (1°C global-mean warming over the model's preindustrial state) and a 2x CO_2 simulation with 780 ppm of CO_2. The control simulation equilibrates within 20 years and then is run for another 30 years. The 2x CO_2 simulation starts from the equilibrated control simulation. To remain agnostic about the future emission pathway, CO_2 is then doubled instantaneously and the simulation is run for 30 years, ultimately resulting in 2.67°C of warming above present-day.

We modify the (noninteractive) sulfate aerosol concentration in the GCM's top vertical layer, which corresponds to the stratosphere at $p \approx 3$ mbar and $z \approx 40$ km. Following previous proof-of-concept studies, particle size is assumed to be log-normally distributed with a dry median radius of 0.05 μm and geometric standard deviation of 2.0 (Ban-Weiss & Caldeira, 2010; Rasch et al., 2008). Realistically, the size of aerosol particles would not be fixed and an increase in stratospheric aerosol burden would tend to increase the average particle size, which limits the maximum forcing. To mimic these effects without interactive aerosol, we impose an upper limit on the aerosol contration (see Text S1 in Supporting Information S1).

The geoengineering strategy that we seek to optimize is the stratospheric aerosol concentration as a function of space and time. To ensure that the aerosol distribution is reasonably realistic and to reduce the computational cost, we fix the aerosol concentration to be zonally uniform and limit its latitudinal resolution to 7 latitude bins, which is similar to Zhang et al. (2022) who injected aerosol at 6–8 latitude bins. The seven-point latitudinal profile of aerosol concentration is then linearly interpolated to the GCM's resolution. Overall, our model setup and aerosol treatments are thus identical to Ban-Weiss and Caldeira (2010) but to optimize the aerosol concentration profile we use RL as follows.

The goal for RL is to find the stratospheric aerosol distribution that maximizes a reward function R. Intuitively, R measures how much and how long the climate deviates from a desired target climate, analogous to a *negative* loss function. For the target climate, we use the present-day control simulation. For the reward function, we exclude discounting and define

$$R = \sum_t r(t), \tag{1}$$

where $r(t)$ is the instantaneous reward, t is time, and the default time-resolution is monthly data. Two important variables for climate impacts are changes in temperature and precipitation, but the two variables respond differently to aerosol forcing (Bala et al., 2008; K. L. Ricke et al., 2010; D. G. MacMartin et al., 2013) while changes in precipitation are moreover constrained by the hydrological cycle. We therefore define r to reflect both changes in surface temperature, ΔT, and precipitation minus evaporation, $\Delta(P - E)$. For changes in surface temperature r is

$$r[\Delta T(t)] = -\langle|\Delta\overline{T(t,\phi)}|\rangle = -\langle|\overline{T}(t,\phi) - \overline{T}_{control}(t,\phi)|\rangle \tag{2}$$

where ϕ is latitude, overlines denote a zonal mean and five-year running mean in time (discussed below), vertical lines denote absolute value, and angle brackets denote an area-weighted latitudinal mean. Notice the negative sign —a larger reward means a smaller deviation. For changes in precipitation minus evaporation r is

$$r[\Delta P(t) - \Delta E(t)] = -\langle|\Delta\overline{P(t,\phi)} - \Delta\overline{E(t,\phi)}|\rangle. \tag{3}$$

The units of $r[\Delta T]$ and $r[\Delta(P - E)]$ are not equal so to simultaneously optimize for both quantities the reward functions are weighed relative to their changes in the 2x CO_2 scenario without geoengineering,

$$r = -\sqrt{\left(\frac{r[\Delta T]}{r_{2\times CO_2}[\Delta T]}\right)^2 + \left(\frac{r[\Delta(P-E)]}{r_{2\times CO_2}[\Delta(P-E)]}\right)^2}. \tag{4}$$

We define r in terms of a five-year running mean to account for internal variability. One potential issue is that the RL algorithm might learn to artificially suppress the model's internal variability, because doing so would maximize R even further. By using a five-year running mean, we find that RL avoids the problem of large changes in variability compared to the control simulation (see below).

To maximize the reward given by Equations 1 and 4, we implement an actor-critic algorithm (Mnih et al., 2016). Briefly, we train a policy neural network whose input is the current climate state in the model (5-year running means of ΔT and $\Delta(P - E)$ as a function of latitude and longitude) and whose output is the stratospheric aerosol concentration as a function of time and latitude (see Figure 1a). The neural network's input and output are updated monthly. Just as in linear feedback algorithms, this means the aerosol is continually adjusted to reflect the most up-to-date information about the climate's state.

At the beginning of training, the neural network produces random outputs. The network parameters are then updated every month, reinforcing aerosol profiles, which lead to a more favorable reward R. Each training episode lasts 30 model years (30 × 12 updating steps). At the end of each episode, the model is reset, then run another 30 years from the same initial conditions. This means the model repeatedly trains on the same CO_2 emissions scenario. Because the neural network is stochastic (see Text S1 in Supporting Information S1), the exact climate state is different each time. Training is stopped once R no longer increases. To evaluate a trained neural network, we fix the network parameters and run the model for a final episode (i.e., 30 years).

We consider three geoengineering tasks:

1. *Stabilize climate change (Stabilize CC)*: Starting from the "present-day" control climate with 390 ppm of CO_2 (1°C above preindustrial), CO_2 is doubled, which sets the climate on a warming course of 3.67°C above preindustrial. Simultaneously, geoengineering commences. RL is used to keep zonal-mean surface temperature $T(\phi)$ and precipitation minus evaporation $(P - E)(\phi)$ patterns stable at present-day values via Equation 4, thus preventing any further climate change.
2. *Revert climate change (Revert CC)*: CO_2 is doubled, but geoengineering commences only after global-mean warming exceeds 2.5°C above preindustrial (1.5°C above present-day, approximately 5 years after doubling CO_2 concentration). In this task, RL is again used to achieve present-day values of $T(\phi)$ and $(P - E)(\phi)$.

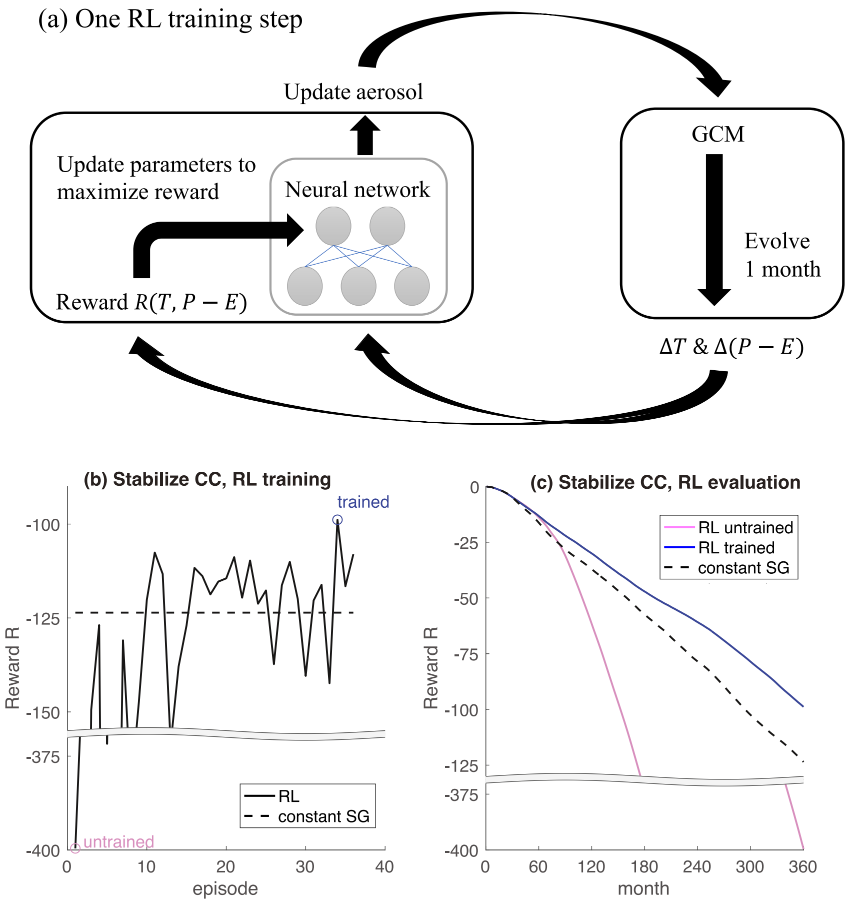

Figure 1. Reinforcement learning (RL) with a global climate model (GCM). (a) Schematic showing how RL is applied to a GCM. A neural network takes in climate variables from the GCM and outputs the GCM's aerosol distribution. As the GCM steps forward in time, an RL algorithm updates the neural network's parameters, and thus the aerosol distribution to achieve a higher cumulative reward function R. (b),(c) RL is used to solve the task Stabilize Climate Change (Stabilize CC) described in the text. (b) The cumulative reward R increases from an initially untrained policy (pink circle) to a trained policy (blue circle). RL eventually outperforms an idealized reference policy in which geoengineering is constant in space and time (dashed line). (c) Evaluation of R for untrained and trained RL policies (pink and blue) as well as the idealized reference policy (dashed). Note the breaks in y-axes for (b) and (c).

3. *Revert tropical climate change (Revert tropical CC)*: Similar to "Revert CC", CO_2 is doubled and geoengineering commences once global-mean warming exceeds 2.5°C above preindustrial (approximately 5 years after doubling CO_2 concentration). However, in this task, the reward function for RL only includes surface temperature Equation 2 and only within the tropics (equatorward of 30°). This task corresponds to a scenario in which geoengineering is used to cool down the tropics, for example, because some tropical regions are exceeding the human heat stress limit (Sherwood & Huber, 2010).

To compare RL against a simpler reference strategy, we use a manual optimization method similar to that in early geoengineering optimization papers (K. L. Ricke et al., 2010; Ban-Weiss & Caldeira, 2010; D. G. MacMartin et al., 2013). We simulate a finite set of stratospheric aerosol concentration profiles that are constant in both space and time. Of these, we choose the profile, which maximizes the reward in Equation 4. The resulting best "constant SG" strategy has a global-mean aerosol concentration of 3.0×10^{-5} kg m^{-2} (see Text S4 in Supporting Information S1). Note, this constant strategy is only provided as a simple reference. Future work is needed to compare RL against the more sophisticated algorithms used in current SAI research.

Figure 1 shows the RL training process for the first task, Stabilize CC. Starting from an initially low reward of $R = -400$, the neural network learns to produce more favorable aerosol profiles with the best value $R = -99$ achieved after 34 training episodes; further episodes do not lead to further improvement. The trained RL strategy

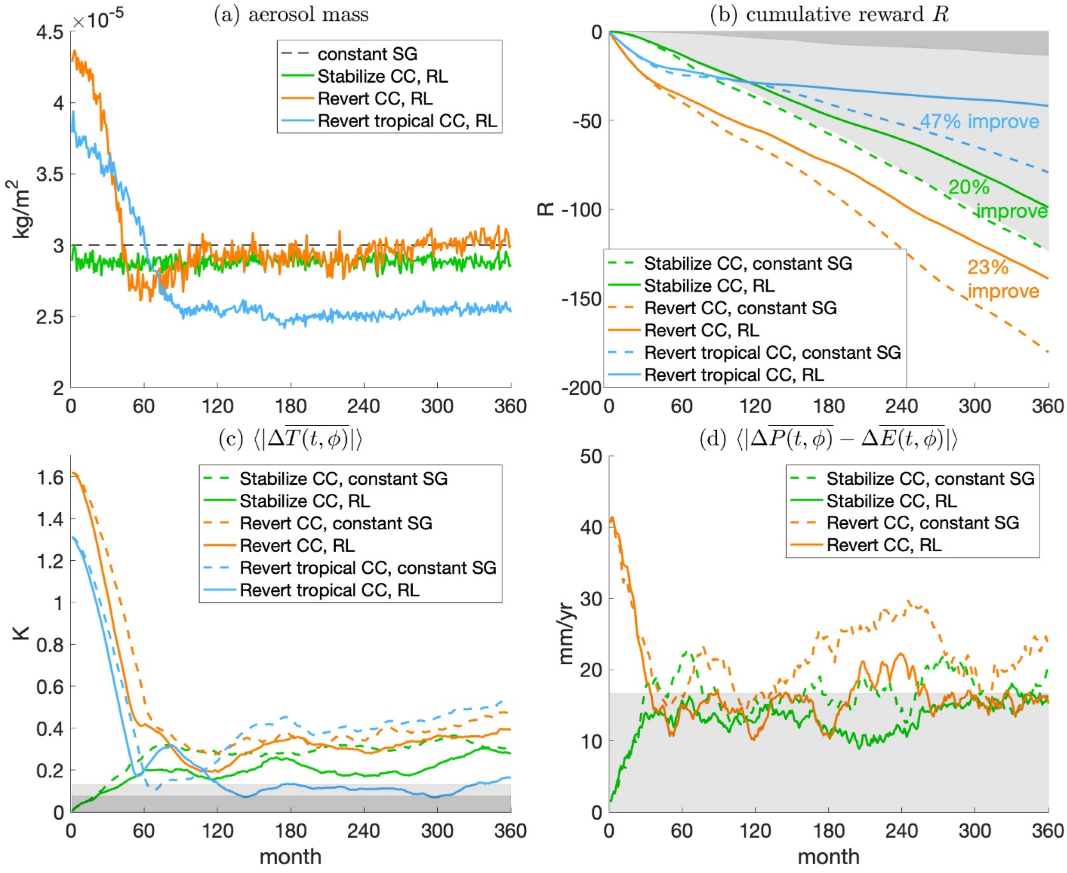

Figure 2. Global-mean climate metrics for trained reinforcement learning policies (solid lines). For reference, we also show metrics for an idealized Solar geoengineering policy with constant aerosol concentration (dashed lines). All curves start from the time when geoengineering commences. (a) Global-mean aerosol concentration. (b) Cumulative reward R. (c) Global-mean deviation of surface temperature, ΔT with respect to the desired reference climate state. (d) Same, but for deviation of precipitation minus evaporation, $\Delta(P - E)$. Gray shadings in (c) and (d) are estimates of internal variability for ΔT and $\Delta(P - E)$ in an extended control run, while the gray shadings in (b) show the impact of internal variability on the cumulative reward R. Light shadings are global means relevant for the tasks Stabilize Climate Change and Revert Climate Change, dark shadings are tropics-only means relevant for the task Revert Tropical Climate Change.

outperforms the simpler constant SG strategy, which only achieves $R = -124$. Results are similar for the other two RL tasks in which RL training converges in less than 20–40 episodes and the trained RL strategy again outperforms the constant SG strategy (see Text in Supporting Information S1). Our results demonstrate it is computationally feasible to use a GCM to train an RL algorithm.

3. Results

To outperform the simpler constant SG strategy, RL leverages time and spatial variations simultaneously. First, we show that RL improves the time variation of aerosol concentration.

We find that the optimal aerosol time series generated by RL strongly depends on the underlying task. Figure 2 shows the evolution of global-mean ΔT, $\Delta(P - E)$, R, and sulfate aerosol concentration for all three tasks after training. In the first task, "Stabilize CC", RL keeps the global-mean aerosol concentration essentially constant at about 3×10^{-5} kg m^{-2} (green line). By contrast, in the other two tasks, RL has learned that it is advantageous to initially produce large amounts of aerosol before switching to a steady-state with moderate aerosol. For "Revert CC", RL initially produces almost 4.5×10^{-5} kg m^{-2} of aerosol before ramping down to about 3×10^{-5} kg m^{-2} within the first 120 months, after which the aerosol concentration is again constant (orange line). For "Revert Tropical CC" RL identifies essentially the same strategy even though the total aerosol concentration in this case is lower (blue line). These results match recent work, which found that, if SAI's goal is to actively reverse warming, it is advantageous to first use large amounts of aerosol before later switching to moderate amounts of aerosol (Brody et al., 2024; Pflüger et al., 2024).

To further elucidate this "kicking the can down the road" effect (Brody et al., 2024), we show in Text S6 in Supporting Information S1 that the same incentive structure also exists in a 0D EBM, which represents the global-mean climate,

$$C\frac{dT}{dt} = S(1 - \alpha(t)) - (a + bT) + F(t). \tag{5}$$

Here, C is the climate system's thermal inertia, t is time, S is insolation, α is albedo, $a + bT$ is the outgoing longwave radiation, and F is the CO_2 radiative forcing. In the EBM, SG counteracts F by increasing the albedo α; both are functions of time.

We find the same qualitative incentive structure in the EBM as that reported in previous work and which RL has identified in the GCM. If geoengineering starts at the same time as F increases, one can always pick a suitable geoengineering intensity to prevent future warming. In contrast, if geoengineering only starts after T has already warmed, the thermal inertia C makes it advantageous to first use strong geoengineering before switching to a steady-state strategy with moderate geoengineering (see Figure S4 in Supporting Information S1). Even though the RL algorithm has no explicit knowledge of the climate system's thermal inertia, our results show it must have implicitly deduced this knowledge.

Although we cannot prove that the RL strategies found in the GCM are globally optimal, we can measure how much the RL strategies can be further improved by comparing them to the effect of internal variability. Internal variability causes ΔT and $\Delta(P - E)$ to deviate from zero at any instant in time, which implies a nonzero reward function R. We estimate ΔT, $\Delta(P - E)$ and R due to internal variability using our extended 30-year control simulation (gray shadings in Figure 2). In the "Stablize CC" task, RL produces a reward R, which is slightly larger than the reward for the internal variability, showing that RL slightly suppresses internal variability (the light gray shading in Figure 2b). In the "Revert CC" task, the reward of RL is smaller than, but close to, that of internal variability. Therefore, in these two tasks, there is limited room to improve beyond the RL strategies found here. By contrast, in the "Revert Tropical CC" task, the reward of RL is much smaller than the internal variability (the dark gray shading in Figure 2b). This suggests that there is still room to improve beyond the RL strategy, although it already significantly outperforms the constant SG strategy.

Next, we show that RL also improves the spatial variation of aerosol concentration. Figure 3 shows the zonal-mean profiles of ΔT, $\Delta(P - E)$ and aerosol concentration for the three trained RL strategies and their constant SG counterparts. To understand the aerosol concentration profiles, we consider ΔT and $\Delta(P - E)$ induced by the 2x CO_2 forcing (red curves in Figure 3). ΔT is polar-amplified, whereas $\Delta(P - E)$ is hemispherically asymmetric larger in the tropics and middle to high latitudes and smaller in the subtropics. Theoretically, a higher aerosol concentration at high latitudes can minimize ΔT (Ban-Weiss & Caldeira, 2010; Lutsko et al., 2020), while a higher aerosol concentration in the tropics can minimize $\Delta(P - E)$ because the baseline $(P - E)$ is large there; the two are linearly related, $\Delta(P - E) \approx \left(0.07\,\mathrm{K}^{-1}\right)\Delta T(P - E)$ (Held & Soden, 2006). For the "Stabilize CC" task and the "Revert tropical CC" task, the aerosol concentration profiles generated by RL favor higher aerosol concentration in the inner tropics and at high latitudes consistent with the theoretical expectations above. For the "Revert CC" task, RL favors a tropically amplified aerosol concentration in the beginning to minimize $\Delta(P - E)$ and a nearly uniform aerosol concentration in the end.

Comparing RL strategies with the constant SG strategy, we find that RL is better at controlling both surface temperature and P-E (see Figure 3 and inset text). Ban-Weiss and Caldeira (2010) studied the "Stabilize CC" task using the same model and same aerosol treatments, but they controlled surface temperature and P-E separately. They found that a polar-amplified aerosol concentration profile controls the surface temperature better but P-E worse than the constant SG strategy, and they did not find an aerosol concentration profile that significantly outperforms the constant SG strategy in controlling P-E. With RL, we find an aerosol concentration profile that outperforms the constant SG strategy in controlling surface temperature and P-E simultaneously (the first column in Figure 3), which is an improvement upon previous results.

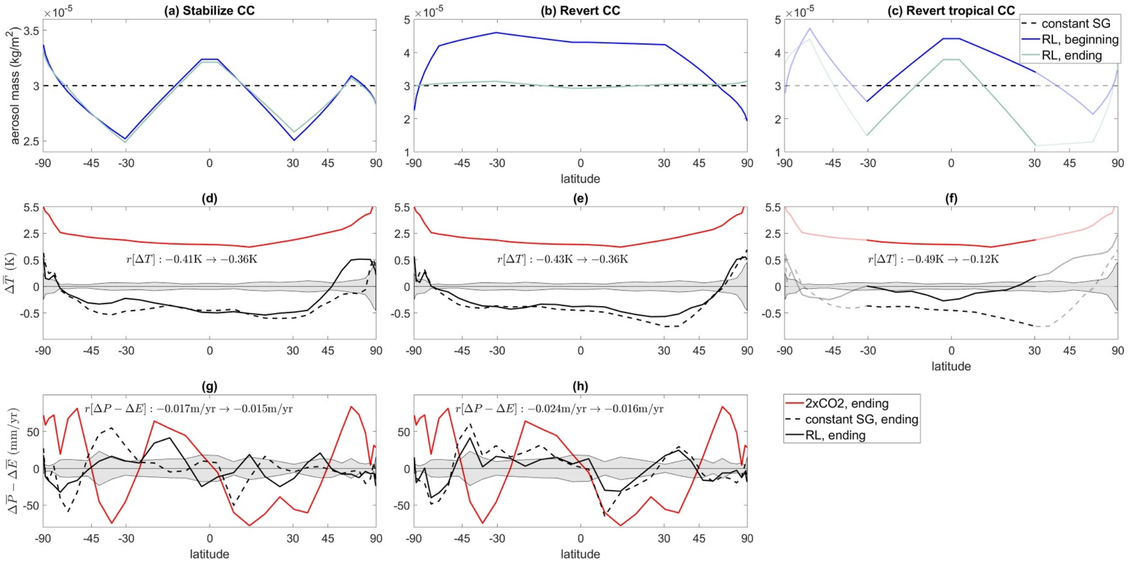

Figure 3. Comparison of latitudinal profiles between the constant Solar geoengineering strategy (dashed) and trained reinforcement learning (RL) strategies (solid). Top row: Zonal-mean aerosol concentration. Dark versus light blue lines indicate aerosol concentration profiles at the beginning (averages in the first 5 years) and end (averages in the last 5 years) during the 30-year evaluation simulation of the trained RL strategies. Middle row: Zonal-mean change in surface temperature ΔT with respect to the desired reference climate state at the end (averages in the last 5 years) of each 30-year evaluation simulation. The red curves show corresponding climate states with double CO_2 concentration but no geoengineering. Bottom row: Same, but for precipitation minus evaporation, $P - E$. In-figure texts show T-only and $P - E$ only rewards for the average profiles in the last 5 years (Equations 2 and 3). Shadings show estimated internal variability for ΔT and $\Delta(P - E)$ based on an extended control run (see Methods).

RL is particularly effective at restoring the pattern of P-E; the residual $\Delta(P - E)$ patterns for RL are largely consistent with internal variability, whereas the $\Delta(P - E)$ residuals for constant SG are not. In terms of ΔT, both RL and constant SG lead to overcooling in the tropics and residual warming at the poles so neither strategy fully compensates for polar amplification. Uncompensated polar amplification is a common outcome in SG simulations presumably due to the spatial mismatch between CO_2 versus insolation forcing (Duffey et al., 2023; Lutsko et al., 2020). Here, we use a reward function, which is sensitive to the magnitude but not the sign of ΔT, so it is plausible that RL finds it advantageous to accept some residual polar amplification. Meanwhile, some recent studies using linear feedback control and state-of-the-art fully coupled GCMs found that subtropical aerosol injection can more effectively reduce the global mean surface temperature (Brody et al., 2025; Kravitz et al., 2019; Zhang et al., 2024). Therefore, future work could explore alternative reward functions and use more realistic climate models to further minimize ΔT.

4. Discussion and Conclusion

In this work, we provide a first proof-of-concept for RL in geoengineering research. We couple an RL algorithm to an idealized GCM, and let the RL algorithm optimize geoengineering strategies to solve three different geoengineering tasks. The resulting RL strategies are physically reasonable and outperform a reference geoengineering strategy in which the aerosol concentration is constant in both space and time.

Our results show that the optimal geoengineering strategy depends on when geoengineering is initiated. We find that Earth's thermal inertia creates a risk asymmetry between SAI proposals, which aim to moderate or prevent future climate change (Irvine et al., 2019; Keith & MacMartin, 2015) versus SAI proposals which aim to revert an already-existing "climate emergency" or which invoke looming tipping points (Battisti et al., 2009; Smith et al., 2024). The latter have an incentive to initially emit large amounts of aerosol and only switch to moderate emissions later. This is problematic since it will be difficult, perhaps impossible, to ramp up the stratospheric aerosol concentration quickly while also testing the SAI technology and its impacts before any large-scale deployment (National Research Council, 2015). Geoengineering proposals should thus be evaluated not only based on technical merit but also on their underlying risk incentives.

Our study does not yet directly compare RL against linear control algorithms; future work is needed to perform such comparisons. RL's main promise over linear control algorithms is that it can be more easily generalized to more complex problems, such as geoengineering scenarios involving the full spatial distributions of many interactive climate objectives and multiple climate interventions. As a potentially promising sign in this direction, recent geoengineering studies using linear control algorithms documented a trade-off between temperature and precipitation (Tilmes et al., 2013; W. Lee et al., 2020). Although our work does not attempt to optimize for precipitation, RL shows promise at learning how to control the meridional profiles of surface temperature and P-E simultaneously.

Given that it is only a proof-of-concept, our work relies on multiple simplifications. Our idealized model setup includes a coarse GCM resolution, slab ocean, an instantaneous CO_2-doubling scenario, and no interactive aerosol. We further optimize the stratospheric aerosol concentration as a function of latitude and time, whereas in reality the actual control parameters would be the aerosol injection rate at certain locations. A next step for RL would be to repeat our work with a higher-resolution model that includes interactive aerosol and explore whether, under more realistic conditions, RL can control surface temperature and P-E down to regional scales, or whether RL can simultaneously control more than two climate variables of interest. Repeating the same exercise with a more sophisticated model will increase its computational cost but might potentially still be feasible. As an upper limit, we estimate that if one applied the RL algorithm to a state-of-the-art fully-coupled GCM, a training process consisting of ∼3,000 simulation years (∼30 episodes multiplied by ∼100 years in each episode) might take about ∼2 months using ∼2,000 CPUs. In the near future, more idealized models such as intermediate-complexity Earth system models might thus be an easier testing ground for RL than state-of-the-art GCMs.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The source code of CAM3, the training scripts, and the plotting scripts are stored in https://zenodo.org/records/15009398 (Quan et al., 2025). We use the Reinforcement Learning Toolbox 3.0 in MATLAB R2021a, which can be downloaded in MATLAB via "APPS" and "get more apps".

Acknowledgments

This research is supported by High-performance Computing Platform of Peking University. We thank two reviewers for their helpful suggestions.

References

Annaswamy, A. M. (2023). Adaptive control and intersections with reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, *6*(1), 65–93. https://doi.org/10.1146/annurev-control-062922-090153

Bala, G., Duffy, P., & Taylor, K. (2008). Impact of geoengineering schemes on the global hydrological cycle. *Proceedings of the National Academy of Sciences*, *105*(22), 7664–7669. https://doi.org/10.1073/pnas.0711648105

Ban-Weiss, G. A., & Caldeira, K. (2010). Geoengineering as an optimization problem. *Environmental Research Letters*, *5*(3), 034009. https://doi.org/10.1088/1748-9326/5/3/034009

Battisti, D., Blackstock, J. J., Caldeira, K., Eardley, D. E., Katz, J. I., Keith, D. W., et al. (2009). Climate engineering responses to climate emergencies. *IOP Conference Series: Earth and Environmental Science*, *6*(45), 452015. https://doi.org/10.1088/1755-1307/6/45/452015

Bellemare, M. G., Candido, S., Castro, P. S., Gong, J., Machado, M. C., Moitra, S., et al. (2020). Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature*, *588*(7836), 77–82. https://doi.org/10.1038/s41586-020-2939-8

Brody, E., Visioni, D., Bednarz, E. M., Kravitz, B., MacMartin, D. G., Richter, J. H., & Tye, M. R. (2024). Kicking the can Down the road: Understanding the effects of delaying the deployment of stratospheric aerosol injection. *Environmental Research: Climate*, *3*(3), 035011. https://doi.org/10.1088/2752-5295/ad53f3

Brody, E., Zhang, Y., MacMartin, D. G., Visioni, D., Kravitz, B., & Bednarz, E. M. (2025). Using optimization tools to explore stratospheric aerosol injection strategies. *Earth System Dynamics*, *16*(4), 1325–1341. https://doi.org/10.5194/esd-16-1325-2025

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., et al. (2004). Description of the NCAR community atmosphere model (CAM 3.0). *NCAR Tech. Note NCAR/TN-464+ STR*, *226*, 1326–1334.

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

Duffey, A., Irvine, P., Tsamados, M., & Stroeve, J. (2023). Solar geoengineering in the polar regions: A review. *Earth's Future*, *11*(6), e2023EF003679. https://doi.org/10.1029/2023EF003679

Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatain, M., et al. (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, *610*(7930), 47–53. https://doi.org/10.1038/s41586-022-05172-4

Gertler, C. G., O'Gorman, P. A., Kravitz, B., Moore, J. C., Phipps, S. J., & Watanabe, S. (2020). Weakening of the extratropical storm tracks in solar geoengineering scenarios. *Geophysical Research Letters*, *47*(11), e2020GL087348. https://doi.org/10.1029/2020GL087348

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of Climate*, *19*(21), 5686–5699. https://doi.org/10.1175/jcli3990.1

Irvine, P., Emanuel, K., He, J., Horowitz, L. W., Vecchi, G., & Keith, D. (2019). Halving warming with idealized solar geoengineering moderates key climate hazards. *Nature Climate Change*, *9*(4), 295–299. https://doi.org/10.1038/s41558-019-0398-8

Jackson, L., Crook, J., Jarvis, A., Leedal, D., Ridgwell, A., Vaughan, N., & Forster, P. (2015). Assessing the controllability of Arctic sea ice extent by sulfate aerosol geoengineering. *Geophysical Research Letters*, *42*(4), 1223–1231. https://doi.org/10.1002/2014gl062240

Keith, D. W., & MacMartin, D. G. (2015). A temporary, moderate and responsive scenario for solar geoengineering. *Nature Climate Change*, *5*(3), 201–206. https://doi.org/10.1038/nclimate2493

Khan, S. G., Herrmann, G., Lewis, F. L., Pipe, T., & Melhuish, C. (2012). Reinforcement learning and optimal adaptive control: An overview and implementation examples. *Annual Reviews in Control*, *36*(1), 42–59. https://doi.org/10.1016/j.arcontrol.2012.03.004

Kravitz, B., Caldeira, K., Boucher, O., Robock, A., Rasch, P. J., Alterskjær, K., et al. (2013). Climate model response from the Geoengineering Model Intercomparison Project (GeoMIP). *Journal of Geophysical Research: Atmospheres*, *118*(15), 8320–8332. https://doi.org/10.1002/jgrd.50646

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. https://doi.org/10.1088/1748-9326/9/4/044006

Kravitz, B., MacMartin, D. G., Mills, M. J., Richter, J. H., Tilmes, S., Lamarque, J.-F., et al. (2017). First simulations of designing stratospheric sulfate aerosol geoengineering to meet multiple simultaneous climate objectives. *Journal of Geophysical Research: Atmospheres*, *122*(23), 12–616. https://doi.org/10.1002/2017jd026874

Kravitz, B., MacMartin, D. G., Tilmes, S., Richter, J. H., Mills, M. J., Cheng, W., et al. (2019). Comparing surface and stratospheric impacts of geoengineering with different SO2 injection strategies. *Journal of Geophysical Research: Atmospheres*, *124*(14), 7900–7918. https://doi.org/10.1029/2019JD030329

Lee, W., MacMartin, D., Visioni, D., & Kravitz, B. (2020). Expanding the design space of stratospheric aerosol geoengineering to include precipitation-based objectives and explore trade-offs. *Earth System Dynamics*, *11*(4), 1051–1072. https://doi.org/10.5194/esd-11-1051-2020

Lee, W. R., MacMartin, D. G., Visioni, D., Kravitz, B., Chen, Y., Moore, J. C., et al. (2023). High-latitude stratospheric aerosol injection to preserve the Arctic. *Earth's Future*, *11*(1), e2022EF003052. https://doi.org/10.1029/2022ef003052

Lutsko, N. J., Seeley, J. T., & Keith, D. W. (2020). Estimating impacts and trade-offs in solar geoengineering scenarios with a moist energy balance model. *Geophysical Research Letters*, *47*(9), e2020GL087290. https://doi.org/10.1029/2020gl087290

MacMartin, D., Visioni, D., Kravitz, B., Richter, J., Felgenhauer, T., Lee, W., et al. (2022). Scenarios for modeling solar radiation modification. *Proceedings of the National Academy of Sciences*, *119*(33), e2202230119. https://doi.org/10.1073/pnas.2202230119

MacMartin, D. G., Caldeira, K., & Keith, D. W. (2014). Solar geoengineering to limit the rate of temperature change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *372*(2031), 20140134. https://doi.org/10.1098/rsta.2014.0134

MacMartin, D. G., Irvine, P. J., Kravitz, B., & Horton, J. B. (2019). Technical characteristics of a solar geoengineering deployment and implications for governance. *Climate Policy*, *19*(10), 1325–1339. https://doi.org/10.1080/14693062.2019.1668347

MacMartin, D. G., Keith, D. W., Kravitz, B., & Caldeira, K. (2013). Management of trade-offs in geoengineering through optimal choice of non-uniform radiative forcing. *Nature Climate Change*, *3*(4), 365–368. https://doi.org/10.1038/nclimate1722

MacMartin, D. G., & Kravitz, B. (2019). The engineering of climate engineering. *Annual Review of Control, Robotics, and Autonomous Systems*, *2*(1), 445–467. https://doi.org/10.1146/annurev-control-053018-023725

MacMartin, D. G., Kravitz, B., Keith, D. W., & Jarvis, A. (2014). Dynamics of the coupled human–climate system resulting from closed-loop control of solar geoengineering. *Climate Dynamics*, *43*(1), 243–258. https://doi.org/10.1007/s00382-013-1822-9

MacMartin, D. G., Kravitz, B., Tilmes, S., Richter, J. H., Mills, M. J., Lamarque, J.-F., et al. (2017). The climate response to stratospheric aerosol geoengineering can be tailored using multiple injection locations. *Journal of Geophysical Research: Atmospheres*, *122*(23), 12–574. https://doi.org/10.1002/2017jd026868

MacMartin, D. G., Ricke, K. L., & Keith, D. W. (2018). Solar geoengineering as part of an overall strategy for meeting the 1.5°C paris target. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2119), 20160454. https://doi.org/10.1098/rsta.2016.0454

Mills, M. J., Richter, J. H., Tilmes, S., Kravitz, B., MacMartin, D. G., Glanville, A. A., et al. (2017). Radiative and chemical response to interactive stratospheric sulfate aerosols in fully coupled CESM1 (WACCM). *Journal of Geophysical Research: Atmospheres*, *122*(23), 13–061. https://doi.org/10.1002/2017jd027006

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., et al. (2016). In *International conference on machine learning* (pp. 1928–1937).

National Research Council. (2015). *Climate intervention: Reflecting sunlight to cool Earth*. The National Academies Press. https://doi.org/10.17226/18988

Oleson, K. W., Dai, Y., Bonan, G., Bosilovich, M., Dickinson, R., Dirmeyer, P., et al. (2004). *Technical description of the community land model (CLM)*. Tech. Note NCAR/TN-461+ STR.

Pflüger, D., Wieners, C. E., van Kampenhout, L., Wijngaard, R. R., & Dijkstra, H. A. (2024). Flawed emergency intervention: Slow Ocean response to abrupt stratospheric aerosol injection. *Geophysical Research Letters*, *51*(5), e2023GL106132. https://doi.org/10.1029/2023GL106132

Quan, H., Koll, D., Lutsko, N., & Yuval, J. (2025). Data and codes for "Solar Geoengineering Strategies Based on Reinforcement Learning" [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.15009398

Rasch, P. J., Crutzen, P. J., & Coleman, D. B. (2008). Exploring the geoengineering of climate using stratospheric sulfate aerosols: The role of particle size. *Geophysical Research Letters*, *35*(2). https://doi.org/10.1029/2007gl032179

Richter, J. H., Visioni, D., MacMartin, D. G., Bailey, D. A., Rosenbloom, N., Dobbins, B., et al. (2022). Assessing responses and impacts of solar climate intervention on the Earth system with stratospheric aerosol injection (ARISE-SAI): Protocol and initial results from the first simulations. *Geoscientific Model Development*, *15*(22), 8221–8243. https://doi.org/10.5194/gmd-15-8221-2022

Richter, J. H., Tilmes, S., Mills, M. J., Tribbia, J. J., Kravitz, B., MacMartin, D. G., et al. (2017). Stratospheric dynamical response and ozone feedbacks in the presence of so2 injections. *Journal of Geophysical Research: Atmospheres*, *122*(23), 12–557. https://doi.org/10.1002/2017jd026912

Ricke, K., Wan, J. S., Saenger, M., & Lutsko, N. J. (2023). Hydrological consequences of solar geoengineering. *Annual Review of Earth and Planetary Sciences*, *51*(1), 447–470. https://doi.org/10.1146/annurev-earth-031920-083456

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

Seeley, J. T., Lutsko, N. J., & Keith, D. W. (2021). Designing a radiative antidote to CO2. *Geophysical Research Letters*, *48*(1), e2020GL090876. https://doi.org/10.1029/2020GL090876

Sherwood, S. C., & Huber, M. (2010). An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, *107*(21), 9552–9555. https://doi.org/10.1073/pnas.0913352107

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Smith, W., Bartels, M. F., Boers, J. G., & Rice, C. V. (2024). On thin ice: Solar geoengineering to manage tipping element risks in the cryosphere by 2040. *Earth's Future*, *12*(8), e2024EF004797. https://doi.org/10.1029/2024EF004797

Smith, W., & Wagner, G. (2018). Stratospheric aerosol injection tactics and costs in the first 15 years of deployment. *Environmental Research Letters*, *13*(12), 124001. https://doi.org/10.1088/1748-9326/aae98d

Tilmes, S., Fasullo, J., Lamarque, J.-F., Marsh, D. R., Mills, M., Alterskjaer, K., et al. (2013). The hydrological impact of geoengineering in the geoengineering model intercomparison project (GeoMIP). *Journal of Geophysical Research: Atmospheres*, *118*(19), 11–036. https://doi.org/10.1002/jgrd.50868

Tilmes, S., MacMartin, D. G., Lenaerts, J., Van Kampenhout, L., Muntjewerf, L., Xia, L., et al. (2020). Reaching 1.5 and 2.0°C global surface temperature targets using stratospheric aerosol geoengineering. *Earth System Dynamics*, *11*(3), 579–601. https://doi.org/10.5194/esd-11-579-2020

Tilmes, S., Richter, J. H., Kravitz, B., MacMartin, D. G., Mills, M. J., Simpson, I. R., et al. (2018). CESM1 (WACCM) stratospheric aerosol geoengineering large ensemble project. *Bulletin of the American Meteorological Society*, *99*(11), 2361–2371. https://doi.org/10.1175/bams-d-17-0267.1

Tilmes, S., Richter, J. H., Mills, M. J., Kravitz, B., MacMartin, D. G., Vitt, F., et al. (2017). Sensitivity of aerosol distribution and climate response to stratospheric SO2 injection locations. *Journal of Geophysical Research: Atmospheres*, *122*(23), 12–591. https://doi.org/10.1002/2017jd026888

Visioni, D., MacMartin, D. G., Kravitz, B., Boucher, O., Jones, A., Lurton, T., et al. (2021). Identifying the sources of uncertainty in climate model simulations of solar radiation modification with the G6sulfur and G6solar geoengineering model intercomparison project (GeoMIP) simulations. *Atmospheric Chemistry and Physics*, *21*(13), 10039–10063. https://doi.org/10.5194/acp-21-10039-2021

Visioni, D., MacMartin, D. G., Kravitz, B., Richter, J. H., Tilmes, S., & Mills, M. J. (2020). Seasonally modulated stratospheric aerosol geoengineering alters the climate outcomes. *Geophysical Research Letters*, *47*(12), e2020GL088337. https://doi.org/10.1029/2020gl088337

Visioni, D., Robock, A., Haywood, J., Henry, M., Tilmes, S., MacMartin, D. G., et al. (2024). G6-1.5 K-SAI: A new geoengineering model intercomparison project (GeoMIP) experiment integrating recent advances in solar radiation modification studies. *Geoscientific Model Development*, *17*(7), 2583–2596. https://doi.org/10.5194/gmd-17-2583-2024

Zhang, Y., MacMartin, D. G., Visioni, D., Bednarz, E. M., & Kravitz, B. (2024). Hemispherically symmetric strategies for stratospheric aerosol injection. *Earth System Dynamics*, *15*(2), 191–213. https://doi.org/10.5194/esd-15-191-2024

Zhang, Y., MacMartin, D. G., Visioni, D., & Kravitz, B. (2022). How large is the design space for stratospheric aerosol geoengineering? *Earth System Dynamics*, *13*(1), 201–217. https://doi.org/10.5194/esd-13-201-2022