Stratospheric Aerosol Injection as a Deep Reinforcement Learning Problem

Anonymous Authors[1]

1. Introduction

As global greenhouse gas emissions continue to rise, the use of geoengineering in order to artificially mitigate climate change effects is increasingly considered. Stratospheric aerosol injection (SAI), which reduces solar radiative forcing and thus can be used to offset excess radiative forcing due to the greenhouse effect, is both technically and economically feasible (Crutzen, 2006; MacMartin, 2014; Smith, 2018). However, naive deployment of SAI has been shown in simulation to produce highly adversarial regional climatic effects in regions such as India and West Africa (Ricke, 2010). Wealthy countries would most likely be able to trigger SAI unilaterally, i.e. China, Russia or US could decide to fix their own climates and, by collateral damage, drying India out by disrupting the monsoon or inducing termination effects with rapid warming (Jones, 2013). Understanding both how SAI can be optimised and how to best react to rogue injections is therefore of crucial geostrategic interest (et al, 2015b).

In this paper, we argue that optimal SAI control can be characterised as a high-dimensional Markov Decision Process (MDP) (Bellman, 1957). This motivates the use of deep reinforcement learning (DRL) (Mnih, 2015) in order to automatically discover non-trivial, and potentially time-varying, optimal injection policies or identify catastrophic ones. To overcome the inherent sample inefficiency of DRL, we propose to emulate a Global Circulation Model (GCM) using deep learning techniques. To our knowledge, this is the first proposed application of deep reinforcement learning to the climate sciences.

2. Related work

General Circulation Models (GCMs), which simulate the earth's climate on a global scale, are inherently computationally intensive. Simple statistical methods are routinely used in order to estimate climate responses to slow forcings (Castruccio, 2013). Recently, the advent of deep learning has led to a number of successful emulation attempts of both full GCMs used for weather prediction (Düben, 2018), as well as for sub-grid scale processes (Brenowitz, 2018; Rasp, 2018), including precipitation (O'Gorman, 2018). This suggests that the emulation of the response of regional variables, such as precipitation and surface temperature, to aerosol injection forcings may now be within reach.

Investigation of optimal SAI control within the climate community is currently constrained to low-dimensional injection pattern parametrisations (Ban-Weiss & Caldeira, 2010) or manual grid search over edge cases of interest (et al, 2015a). Even in simple settings, it has been shown that regional climate response is sensitive to the choice of SAI policy (MacMartin, 2013). In addition, super-regional impacts on El Nino/Southern Oscillation have been demonstrated (Gabriel, 2015). This suggests that climate response to SAI is sensitive enough to warrant a high-dimensional treatment.

Altering the injection altitude, latitude, season, or particle type - possibly even with the use of specially engineering photophoretic nanoparticles (Keith, 2010) - may provide the ability to "tailor" fine-grained SAI. But, presently, stratospheric aerosol models have substantially different responses to identical injection strategies (Pitari, 2014), suggesting directly simulating the implications of these strategies requires further model development.

3. GCM emulation

We use HadCM3 (Gordon, 2000) to simulate climate response to SAI as it is the first GCM not to require flux adjustments to avoid large scale climate drift. In addition, HadCM3 is still used as a baseline model for IPCC reports (IPCC, 2013).

The radiative forcing of sulfate aerosols is emulated in HadCM3 by adjusting the atmospheric optical depth (AOD) in the lower stratosphere, i.e. a larger AOD corresponds to a larger sulfate aerosol concentration. Predominantly zonal winds in the stratosphere are assumed to keep aerosol optical depth zonally uniform to first order, so it is prescribed for each of the model's 73 latitude bands. We also assume that aerosol concentration completely decays within a year and that aerosol concentration is upper-bounded by coagulation effects (Hansen, 2005) and thus capped at $4\overline{\rho}$ in each latitude band, with $\overline{\rho} = m_o A_S^{-1}$, where A_S is the surface area of the lower stratosphere band at an altitude of 20 km (Smith, 2018).

Despite being up to a factor 10^3 faster than many contemporary GCMs, a single HadCM3 year still corresponds to about 15 hours of computation on a generic single-thread

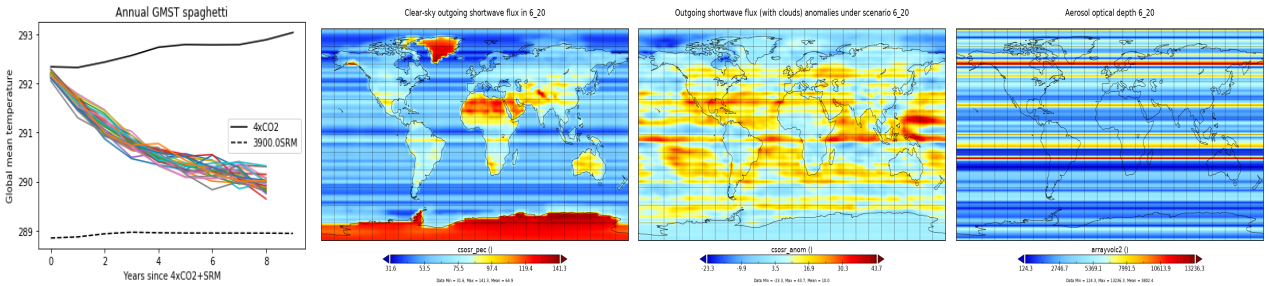

Figure 1. from left: 1. global mean cooling relative to ρ-uniform and zero AOD baselines (full HadCM3 dataset) 2. clear-sky upwards short-wave radiative flux 3. outgoing short-wave radiative flux (with clouds) 4. AOD distribution (2., 3. and 4. at same time / same run)

CPU. In order to employ deep reinforcement learning, we therefore require a fast emulator that can predict next states in a matter of milliseconds.

We approximate the full HadCM3 state s_t at time t by the scalar surface fields *sea ice fraction* $S_t(x, y)$, *surface temperature* $T_t(x, y)$, *depth layer-weighted ocean heat content* $H_t(x, y)$ and *stratospheric aerosol optical depth* $\tau_t(x, y)$. From these quantities, the emulator needs to predict S_{t+1}, H_{t+1} and T_{t+1}, as well as other quantities of interest to the policy optimisation objective, such as local precipitation rates $P_{t+1}(x, y)$. All these quantities are returned from HadCM3 simulations as scalar grids of dimension 73×96.

To emulate HadCM3, we use an encoder-decoder network similar to UNet (Ronneberger, 2015) given HadCM3 output is largely deterministic, which can be augmented with an prediction uncertainty channel. We pre-train the encoder on ImageNet (Mirowski, 2016) and fine-tune the output layers on 2000 output samples of HadCM3 rollouts based on aerosol density distributions drawn randomly from a 73-dimensional Dirichlet distribution with shape parameters $\alpha_k = 1.5$ (to discourage extremes) and output scaling factor $\overline{\rho}$. We reject samples violating the $4\overline{\rho}$ coagulation cap.

Preliminary simulation results suggest that emulator training would likely benefit from auxiliary tasks (Liebel, 2018) related to cloud cover prediction (see Figure 1).

4. Reinforcement learning setting

GCM emulator states s_t and sequential aerosol injections conditioned thereon together form a Markov Decision Process (Bellman, 1957). At each time step, which amounts to a year of simulated climate, the agent decides how much aerosol to inject into each of 73 evenly spaced latitude bands, overall selecting an action $\mathbf{u}_t \in \mathbb{R}_+^{73}$. The environment then returns a scalar reward r_t as feedback to the agent. Optimal injection policies $\pi(\mathbf{u}_t|\mathbf{s}_t)$ are then learnt by maximizing the expected future-discounted cumulative reward $R_t = \sum_{t=0}^{T} \gamma^t r_t$, where $\gamma \in [0, 1]$ is a discount factor and $T = 10$ corresponds to an episode length of 10 years.

For reasons of simplicity and robustness, we employ an off-policy deep Q-learning (Mnih, 2015) approach and discretise the action space using $n_u = 10$ bins of equal size for each latitude band. As the resulting joint action space is large (10^{73}), we factorise the joint state-action value function. This could either be achieved using low-dimensional embeddings (Dulac-Arnold, 2015) or, more robustly, by factorising the joint value function using techniques originally developed for cooperative multi-agent settings (Sunehag, 2018; Rashid, 2018). Value function network architecture is based on a convolutional encoder similar to the one used by the GCM emulator.

A simple choice for a bounded team reward function r_t that discourages extreme changes in regional climate is

$$- \max_{x,y \in A} \left[\alpha_P |\Delta_P^t(x, y)| + \alpha_T |\Delta_T^t(x, y)|\right]$$

where Δ_P^t is the difference between the regional precipitation rate and its pre-industrial average (similarly Δ_T^t for surface temperature), A is the earth's surface grid and $\alpha_P, \alpha_T > 0$ are scalar hyperparameters. More advanced reward functions might be weighted by additional local factors, including population density, local resilience factors and other aspect of eco-socio-economic interest and skewed according to measures such as climate adaptability and regional climate specifics.

To ensure physical consistency and robustness, SAI control policies learnt within the emulator are subsequently cross-verified in HadCM3 and/or other GCE models.

5. Conclusion and Outlook

We propose the study of optimal SAI control as a high-dimensional control problem using a fast GCM emulator and deep reinforcement learning.

We believe that DRL may become an important tool in the study of SAI and other geoengineering approaches, such as marine cloud brightening, over the next decade.

References

Ban-Weiss, G. A. and Caldeira, K. Geoengineering as an optimization problem. *Environmental Research Letters*, 5(3):034009, 2010.

Bellman, R. Dynamic Programming, 1957.

Brenowitz, N. D. e. a. Prognostic Validation of a Neural Network Unified Physics Parameterization. *Geophysical Research Letters*, 45(12):6289–6298, 2018. ISSN 1944-8007. doi: 10.1029/2018GL078510.

Castruccio, S. e. a. Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs. *Journal of Climate*, 27(5):1829–1844, October 2013. ISSN 0894-8755. doi: 10.1175/JCLI-D-13-00099.1.

Crutzen, P. J. Albedo enhancement by stratospheric sulfur injections: A contribution to resolve a policy dilemma? *Climatic Change*, 77(3):211, Jul 2006. ISSN 1573-1480. doi: 10.1007/s10584-006-9101-y.

Düben, P. D. e. a. Challenges and design choices for global weather and climate models based on machine learning. 2018.

Dulac-Arnold, G. e. a. Deep Reinforcement Learning in Large Discrete Action Spaces. *arXiv:1512.07679 [cs, stat]*, December 2015. arXiv: 1512.07679.

et al, L. S. J. Assessing the controllability of arctic sea ice extent by sulfate aerosol geoengineering. *Geophysical Research Letters*, 42(4):1223–1231, 2015a. doi: 10.1002/2014GL062240.

et al, X. Y. Impacts, effectiveness and regional inequalities of the geomip g1 to g4 solar radiation management scenarios. *Global and Planetary Change*, 129:10 – 22, 2015b. ISSN 0921-8181. doi: 10.1016/j.gloplacha.2015.02.010.

Gabriel, C. J. e. a. Stratospheric geoengineering impacts on el nino/southern oscillation. *Atmospheric Chemistry and Physics*, 15(20):11949–11966, 2015. doi: 10.5194/acp-15-11949-2015. URL https://www.atmos-chem-phys.net/15/11949/2015/.

Gordon, C. e. a. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16(2):147–168, February 2000. ISSN 1432-0894. doi: 10.1007/s003820050010.

Hansen, J. e. a. Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, 110(D18), September 2005. ISSN 0148-0227. doi: 10.1029/2005JD005776.

IPCC. Fifth Assessment Report IPCC. Technical report, 2013.

Jones, A. e. a. The impact of abrupt suspension of solar radiation management (termination effect) in experiment g2 of the geoengineering model intercomparison project (geomip). *Journal of Geophysical Research: Atmospheres*, 118(17):9743–9752, 2013. doi: 10.1002/jgrd.50762.

Keith, D. W. Photophoretic levitation of engineered aerosols for geoengineering. *Proceedings of the National Academy of Sciences*, 107(38):16428–16431, September 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1009519107.

Liebel, L. e. a. Auxiliary Tasks in Multi-task Learning. *arXiv:1805.06334 [cs]*, May 2018. arXiv: 1805.06334.

MacMartin, D. e. a. Managing trade-offs in geoengineering through optimal choice of non-uniform radiative forcing. *Nature Climate Change*, 3:365–368, 04 2013.

MacMartin, D. G. e. a. Geoengineering: The world's largest control problem. In *2014 American Control Conference*, pp. 2401–2406, June 2014. doi: 10.1109/ACC.2014.6858658.

Mirowski, P. e. a. Learning to Navigate in Complex Environments. *arXiv:1611.03673 [cs]*, November 2016. arXiv: 1611.03673.

Mnih, V. e. a. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, February 2015. ISSN 1476-4687. doi: 10.1038/nature14236.

O'Gorman, P. A. e. a. Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events. *Journal of Advances in Modeling Earth Systems*, 10(10):2548–2563, 2018. ISSN 1942-2466. doi: 10.1029/2018MS001351.

Pitari, G. e. a. Stratospheric ozone response to sulfate geoengineering: Results from the Geoengineering Model Intercomparison Project (GeoMIP). *Journal of Geophysical Research: Atmospheres*, 119(5):2629–2653, 2014. ISSN 2169-8996. doi: 10.1002/2013JD020566.

Rashid, T. e. a. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. *arXiv:1803.11485 [cs, stat]*, March 2018. arXiv: 1803.11485.

Rasp, S. e. a. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, September 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1810286115.

Ricke, K. L. e. a. Regional climate response to solar-radiation management. *Nature Geoscience*, 3(8):537–541, August 2010. ISSN 1752-0908. doi: 10.1038/ngeo915.

Ronneberger, O. e. a. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*, May 2015. arXiv: 1505.04597.

Smith, W. e. a. Stratospheric aerosol injection tactics and costs in the first 15 years of deployment. *Environmental Research Letters*, 13(12):124001, November 2018. ISSN 1748-9326. doi: 10.1088/1748-9326/aae98d.

Sunehag, P. e. a. Value-Decomposition Networks For Cooperative Multi-Agent Learning. *arXiv:1706.05296 [cs]*, June 2018. arXiv: 1706.05296.