Flawed Emergency Intervention: Slow Ocean Response to Abrupt Stratospheric Aerosol Injection

Daniel Pflüger[1], Claudia Elisabeth Wieners[1], Leo van Kampenhout[1], René Wijngaard[1], and Henk A. Dijkstra[2]

[1]Utrecht University
[2]Institute for Marine and Atmospheric research Utrecht

March 07, 2024

Abstract

Given the possibility of irreversible, anthropogenic changes in the climate system, technologies such as solar radiation management (SRM) are sometimes framed as possible emergency interventions. However, little knowledge exists on the efficacy of such deployments. To fill in this gap, we perform Community Earth System Model 2 (CESM 2) simulations of an intense warming scenario on which we impose gradual early-century SRM or rapid late-century cooling (an emergency intervention), both realised via stratospheric aerosol injection (SAI). While both scenarios cool Earth's surface, ocean responses differ drastically. Rapid cooling fails to release deep ocean heat content or restore an ailing North Atlantic deep convection but partially stabilizes the Atlantic meridional overturning circulation. In contrast, the early intervention effectively mitigates changes in all of these features. Our results suggest that slow ocean timescales impair the efficacy of some SAI emergency interventions.

Flawed Emergency Intervention: Slow Ocean Response to Abrupt Stratospheric Aerosol Injection

Daniel Pflüger [1],Claudia E. Wieners[1], Leo van Kampenhout[1], René R. Wijngaard[1], Henk A. Dijkstra[1]

[1]Institute Marine and Atmospheric Research Utrecht, Princetonplein 5, 3584 CC Utrecht, The Netherlands

Key Points:

- Efficacy of SAI impaired by anthropogenic ocean heating
- Deep ocean heating, weakened AMOC and collapsed North Atlantic deep convection only partially addressed by late SAI
- SAI decouples AMOC and GMST, thereby inducing climate states not seen in purely GHG-forced scenarios

Corresponding author: Daniel Pflüger, `d.pfluger@uu.nl`

Abstract

Given the possibility of irreversible, anthropogenic changes in the climate system, technologies such as solar radiation management (SRM) are sometimes framed as possible emergency interventions. However, little knowledge exists on the efficacy of such deployments. To fill in this gap, we perform Community Earth System Model 2 (CESM 2) simulations of an intense warming scenario on which we impose gradual early-century SRM or rapid late-century cooling (an emergency intervention), both realised via stratospheric aerosol injection (SAI). While both scenarios cool Earth's surface, ocean responses differ drastically. Rapid cooling fails to release deep ocean heat content or restore an ailing North Atlantic deep convection but partially stabilizes the Atlantic meridional overturning circulation. In contrast, the early intervention effectively mitigates changes in all of these features. Our results suggest that slow ocean timescales impair the efficacy of some SAI emergency interventions.

Plain Language Summary

Stratospheric aerosol injection (SAI) is a promising, yet controversial proposal to mask the effects of anthropogenic climate change by releasing sunlight-reflecting particles into the atmosphere. Currently, many studies are focusing on the benefits of near future SAI deployments. We, however, investigate SAI as a late *emergency intervention.* To what extent can SAI still help if we continue to heat and destabilize the climate?

In this study, we simulate the impacts of an abrupt, SAI cooling intervention deployed against the backdrop of a climate much hotter than today's. While SAI readily cools Earth's surface, it is challenged by a slow ocean response. Heat trapped below the ocean surface remains a contributor to sea-level rise and important currents weakened by climate change linger in ailing condition. In contrast, an earlier SAI intervention effectively mitigates changes in these features.

Our findings re-emphasize the urgent need for climate action. If anthropogenic heating continues, even an intervention as powerful as SAI will encounter its limits.

1 Introduction

While global heating puts increasing pressure on societies and ecosystems (IPCC, 2022a), current policies are insufficient to prevent 1.5°C or even 2°C of warming (IPCC, 2022b). To mitigate the associated risks, interventions that cool Earth by reflecting sunlight - *Solar Radiation Management* (SRM) - are being explored as complementary measures to emission reductions (National Academies of Sciences, Engineering, and Medicine, 2021). Among several potential schemes, *Stratospheric Aerosol Injection* (SAI) received considerable attention due to its low perceived technical barriers (Smith, 2020), plausible physical effectiveness(Kleinschmitt et al., 2018; Plazzotta et al., 2018). While model studies demonstrate its benefits (Tilmes et al., 2018, 2020; Visioni et al., 2021), including its ability to control global mean surface temperature (GMST), SAI can not address all consequences of rising greenhouse gas (GHG) concentrations and may induce side-effects of its own (Irvine et al., 2016; Zarnetske et al., 2021). Ethical concerns (Svoboda, 2017; Oomen, 2021) lead some to suggest a ban on its research and deployment (Biermann et al., 2022) whereas others suggest further research (Wieners et al., 2023).

It is not enough to ask *whether* SRM should be deployed. Multiple degrees of freedom in SRM deployments implore us to ask *how and to what end* may be SRM used. Currently popular frameworks include *peak-shaving* (Long & Shepherd, 2014; Reynolds, 2019), in which SRM stabilizes GMST, while other measures gradually tackle GHG concentrations. However, there is no guarantee SRM would be deployed in such a well-optimized and *proactive* fashion. In this study, we examine as SRM as an *emergency* intervention instead, to be deployed only after prolonged heating. This notion, adapted from Lockley et al. (2022), naturally arises when SRM deployments are restricted to particularly extreme situations such as rapid climate tipping. To what extent can later deployments reverse the impacts of heating? How would they compare to earlier, proactive interventions?

In this study, we focus on SRM's physical impact on the ocean. There, long response timescales hamper prospects of reversibility under an emergency intervention. Many ocean features are subject to anthropogenic climate change and have profound impacts on humans and ecosystems which elevates the study of them above a purely academic exercise. We are interested in

- ocean heat content (OHC) change, a major contributor to sea-level rise (Church et al., 2013).
- the Atlantic Meridional Overturning Circulation (AMOC) which may weaken (or even collapse) in the future (Weijer et al., 2020), thereby reducing meridional heat transport and modulating regional sea level rise.
- North Atlantic deep convection which may shut down in the future, leading to abrupt cooling and shifts in the jet-stream (Sgubin et al., 2017; Swingedouw et al., 2021).

We consider only SRM scenarios with extreme levels of GHG and aerosol forcing, including abrupt changes thereof. They should not be seen as desirable or politically realistic futures but rather as physical edge cases that provide valuable intuitions and constraints for more cautious scenarios: if an abrupt cooling struggles to reverse certain ocean changes, a slower deployment would likely do so, too. Furthermore, we restrict ourselves to a single SRM implementation: planetary-scale SAI.

2 Methods

Our scenarios are simulated in CESM2 (Danabasoglu et al., 2020) with atmospheric component CAM6 at $1^\circ \times 1^\circ$ horizontal resolution and ocean model POP2 at similar resolution. Ice sheets are non-evolving, which also prohibits calving, but the land model CLM provides glacial run-off fluxes.

SAI is implemented via prescribed aerosol fields: a compromise between physical realism and computational cost. We favored this approach as it may enable computationally cheap simulations capturing longer ocean timescales in the future. Other groups have used similar scaling-based implementations (Visioni et al., 2021).

Schematically, the protocol works as follows:

- Every year, observe the deviation of GMST from the target
- Based on past GMST deviations, infer the level of SAI - expressed in terms of global mean aerosol optical depth (AOD) - which is necessary to achieve the desired target.
- Use the AOD to scale all SAI-related aerosol fields appropriately.
- Feed the scaled fields into CAM6.

The first two steps are implemented via an established feedforward-feedback control algorithm (Kravitz et al., 2016, 2017). Our specific implementation stabilizes GMST as its sole objective, whereas interactive aerosol simulations(MacMartin et al., 2017; Tilmes et al., 2020) can also control other features such as the inter-hemispheric temperature contrast.

The input aerosol fields derive from an interactive aerosol simulation performed by Tilmes et al. (2020) in CESM2-WACCM, more specifically their Geo SSP5 8.5 1.5 scenario. In contrast to CAM6, the improved, albeit more costly, atmospheric component WACCM allows for detailed chemical aerosol dynamics (Danabasoglu et al., 2020). Our prescribed aerosol fields are averaged versions of the WACCM aerosol fields as described in the supplementary material.

We simulate three scenarios based on SSP5-8.5 background concentrations:

- Control (2015-2100): historical spin-up continued by SSP5-8.5
- SAI 2020 (gradual SAI): branch off Control in 2020; stabilise GMST at 1.5°C above pre-industrial conditions; analogous to Geo SSP5-8.5 1.5 by Tilmes et al. (2020)
- SAI 2080 (emergency intervention): branch off Control in 2080, deploy SAI to restore GMST to 1.5°C.

Note that SAI 2080 involves some adjustments to the control algorithm, described in the supplementary material. Otherwise, the initially high deviation from the target GMST can overcharge the feedback controller and risk excessive cooling.

3 Results

3.1 Temperature Response

In Fig. 1A, we see that the gradual SAI strategy (SAI 2020) indeed stabilises GMST at target level. By contrast, SAI 2080 experiences rapid cooling and even shoots past the target. This undercooling is an artefact of the feedback controller and can be removed by fine-tuning the cooling process (Fig. S2).

Even though GMST is stabilised, total depth OHC accumulates continuously in SAI 2020 (Fig.1B) in agreement with past studies (Fasullo et al., 2018; MacMartin et al., 2022). The warming takes place below the surface and likely stems from deep ocean response timescales (Cheng et al., 2022) combined with the goal of maintaining GMST. As sub-surface layers have not yet adapted to increased surface temperatures, they act as a heat sink for the ocean surface. The induced downward heat flux is then compensated by the feedback controller that allows for a residual top-of-atmosphere radiative imbalance in order to stabilize GMST.

SAI 2080 accumulates more total depth OHC than SAI 2020. The deep tail of OHC in SAI 2080 (Fig.1C) matches that of Control while the near-surface layers are cooled to SAI 2020 levels. Given the short time-frame of SAI 2080, it is not clear whether the vertical OHC distribution has reached equilibrium or deeper layers are simply cooling very slowy.

On the surface, however, both SAI scenarios have comparable OHC anomalies. This suggests that while abrupt SAI readily cools the ocean surface, heat anomalies trapped in deeper layers are more persistent.

Surface temperature responses to SAI are spatially inhomogeneous (Fig. 2). Most strikingly, the subpolar North Atlantic is significantly undercooled in both SAI scenarios. This pattern resembles a *North Atlantic warming hole* known from purely GHG-forced simulations (Drijfhout et al., 2012; Menary & Wood, 2018), which to some extent is also visible in Control. SAI may have merely unmasked this feature rather than induce it. The warming hole is expanded and colder in SAI 2080, while the Southern Hemisphere is warm compared to SAI 2020.

Multi-objective feedback procedures (Kravitz et al., 2017; MacMartin et al., 2017) allow for a more elaborate control of the global temperature pattern including the interhemispheric temperature gradient. Therefore, the asymmetric response of SAI 2080 (Fig. 2E) may be mitigated in a refined control scheme. In our study, however, both SAI

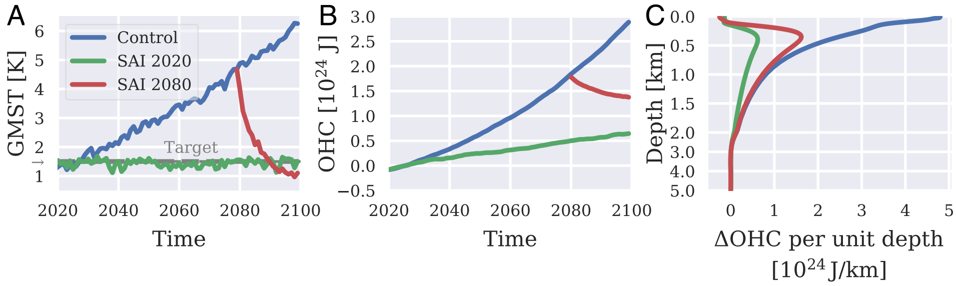

Figure 1. **A**: Annual mean GMST above pre-industrial reference temperature **B**: Change in annual mean total depth OHC relative to 2020-2030 conditions in Control. **C**: Difference in vertical OHC between end-of-simulation (2090-2100) conditions and present-day conditions in Control.

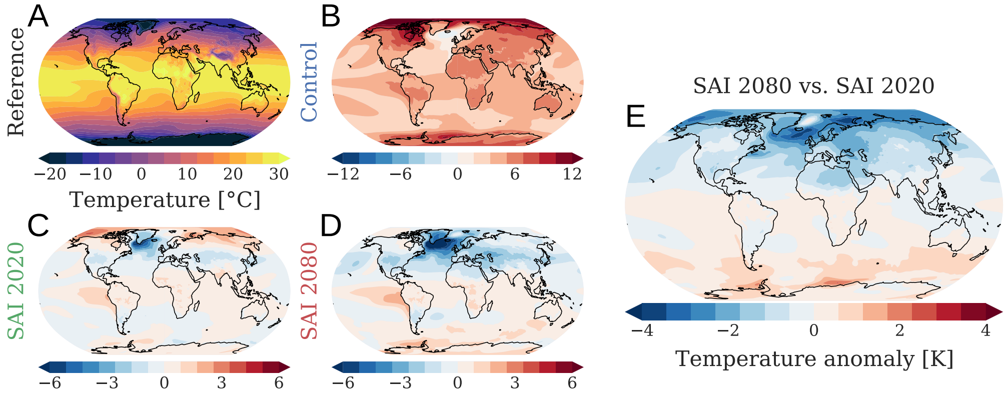

Figure 2. **A**: Reference (2020-2030) annual mean near-surface air temperatures in Control **B-D**: Late-century (2090-2100) temperature changes with respect to the reference for Control, SAI 2020 and SAI 2080 respectively. **E**: Difference between SAI scenarios (**D** minus **C**)

scenarios use spatially identical aerosol patterns which rules out a control of the asymmetry.

3.2 AMOC Response

The AMOC index and meridional heat transport (MHT) roughly halve in Control (Fig. 3A-B). Even the low-emission SSP1-2.6 scenario is projected to lead to similar AMOC index changes. SAI 2020 drastically mitigates but does not halt the AMOC and MHT decline, similar to other studies (Xie et al., 2022; MacMartin et al., 2022). SAI 2080 stabilizes the AMOC index but only has an inconclusive impact on the MHT.

SAI effectively decouples the GMST and the AMOC index (Fig. 3C). This could explain the interhemispheric temperature contrast featured in SAI 2080: a weak AMOC impedes northward heat transport leading to a see-saw temperature pattern (Stocker, 1998; Liu et al., 2020) not masked by heat otherwise present in Control.

To study the spatial pattern of the AMOC, we plot meridional streamfunction changes under all scenarios from 2070-2080 to 2090-2100 (Fig. 4). This choice of time intervals helps to reveal the immediate AMOC response to SAI 2080. Additionally, we subtract

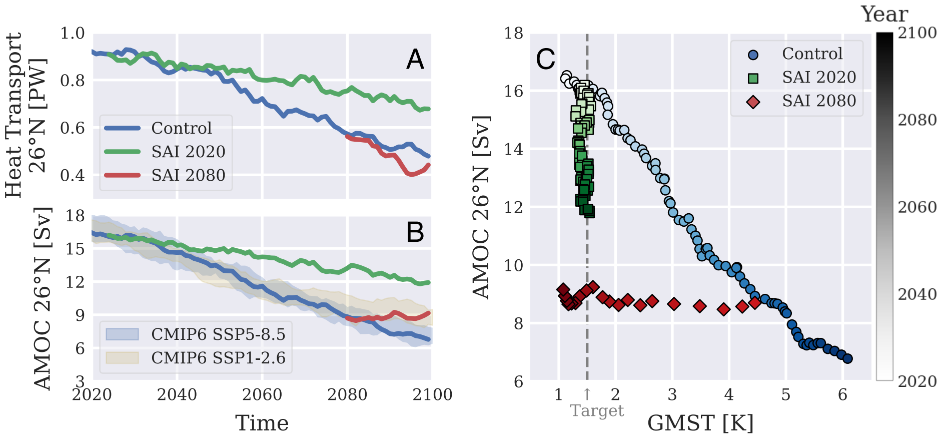

Figure 3. **A**: Annual mean Atlantic northwards heat transport at 26° N where we apply a rolling average over five year periods with backward window **B**: AMOC index defined as the maximum of the annual mean meridional overturning streamfunction at 26° N below 200 m - Partially transparent uncertainty bands depict three CESM2 CMIP6 (Coupled Model Intercomparison Project Phase 6) ensemble members (Danabasoglu, 2019c, 2019d) per GHG concentration pathway. The uncertainty is the ensemble standard deviation. Again, we apply rolling averages over five year periods. **C**: Annual mean GMST vs. AMOC index - The marker saturation denotes the year: light (2020) to dark (2100).

the changes in Control from the ones in the SAI scenarios in an attempt to disentangle GHG from SAI-related impacts.

Fig. 4D reveals a potential feedback in the AMOC stabilization under SAI 2080. Following the deployment, the pattern of relative AMOC strengthening closely mirrors the pre-deployment streamfunction, albeit mostly near the surface and in the northern hemisphere. This suggests that the AMOC response to abrupt SAI is dependent on the AMOC state itself. While a similar observation can be made for SAI 2020 (Fig. 4C), disentangling the forced response from internal feedback is not obvious during the gradual change in aerosol forcing. SAI 2080 gives a much better indication that it is indeed the state of the AMOC which steers its response to SAI.

3.3 North Atlantic Deep Convection

We now focus on deep convection processes in the North Atlantic. Using mixed layer depth as a proxy for deep convection, we identify two regions, *East* and *West*, where the mixed layer depth in April (the month with the deepest mixed layer) exceeds 550 m (Fig. 5A). This threshold depth was chosen as it is sufficiently large to distinguish deep convection from regular mixed-layer conditions and small enough to provide a good signal-to-noise ratio. The regions are separated longitudinally by the southern tip of Greenland.

In Control, deep convection in *West* ceases around 2050, followed by a shutdown in *East* around 2060. SAI 2020 prevents the shutdown in *East*, but only postpones the shutdown in *West* by about a decade. The *West* shutdown is not as definite as in the case of Control and isolated years with deep convection still occur. For SAI 2080, deep convection remains absent in both regions with the exception of a single outlier year for *East*.

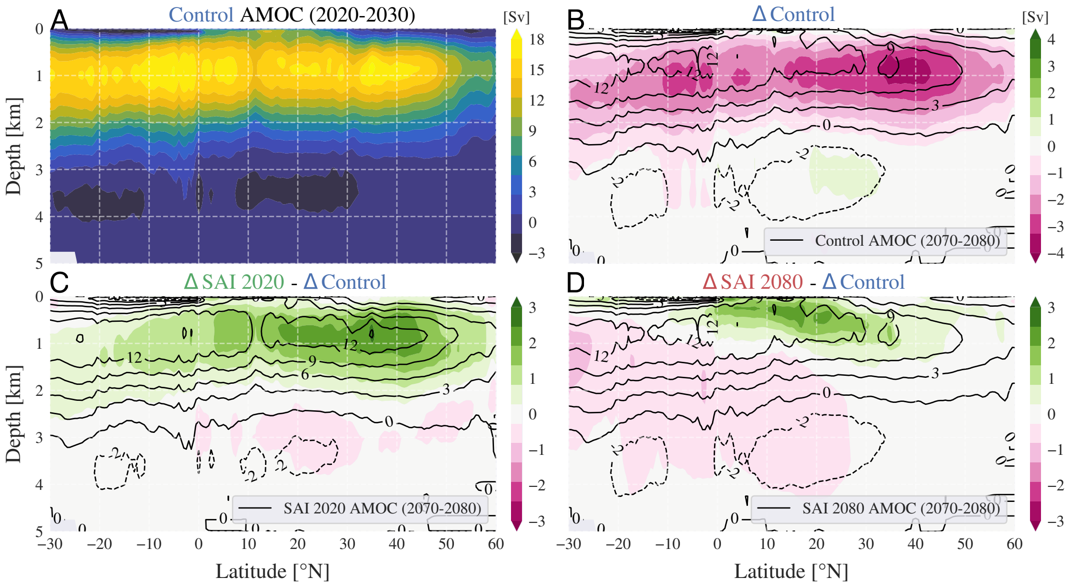

Figure 4. **A**: AMOC streamfunction in Control averaged over 2020-2030. In **B-D**, for any simulation X, Δ X is the mean over 2090-2100 minus the mean over 2070-2080. **B**: Change in AMOC streamfunction under Control - Black contour lines show the mean streamfunction over 2070-2080 for Control while the shading indicates Δ Control. **C**: Change in AMOC streamfunction in SAI 2020 relative to Control - Black contour lines show the mean streamfunction over 2070-2080 for SAI 2020 while the shading indicates Δ SAI 2020 - Δ Control. **D**: Analogous to **C** but for SAI 2080.

Why does cooling in SAI 2080 not revive deep convection before 2100? We address this question by studying the ocean stratification over both deep convection regions. Deep convection in April is inhibited if the surface density in the previous September has been too low, i.e. the water column was too stratified (Fig. S4). Thus, surface density serves as a proxy for favorable convection conditions.

The sea surface density is determined by temperature and salinity, seen in Fig. 5D-F. In all scenarios, final salinities are well below reference conditions. SAI 2020 roughly halves the decline with respect to Control. This difference becomes very noticeable mid-century simultaneously with the *East* and *West* shutdown in Control. SAI 2080 does not fundamentally alter the trajectory of Control apart from a transient increase in salinity that correlates with an isolated year of deep convection. Therefore, freshening contributes to density loss in all scenarios.

In the case of Control, temperature trends are non-monotonous (Fig. 5D) and do not lead to a denser surface. What appears to be a weak cooling trend is mostly masked by inter-annual variability and eventually superseded by intense heating. Typically, deep convection shutdown induces a rapid cooling (Sgubin et al., 2017; Swingedouw et al., 2021) which is not obvious from Fig. 5D . It can, however, be detected by using a CESM2 SSP5-8.5 ensemble and switching to an annual-mean rather than a single-month perspective (Fig. S6).

SAI 2020 shows an overall cooling trend dominated by a quick decline at time of *West* shutdown. Former observation could indicate AMOC weakening whereas latter phenomenon is again consistent with abrupt cooling during deep convection collapse (Sgubin et al., 2017; Swingedouw et al., 2021). In SAI 2080, the cooling is more drastic (Fig. 5D), perhaps a result of full deep convection shutdown and a weakened AMOC. These temperature drops have a positive effect on density and thereby convection, albeit not sufficient to bring SAI 2080 densities to SAI 2020 levels (Fig. 5F). Therefore, the salinity deficit of SAI 2080 with respect to SAI 2020 (Fig. 5E) presents a clear obstacle to restarting deep convection.

Recognizing the importance of salinity changes, we sketch a possible mechanism behind SAI 2080's failure to spur convection. Firstly, all scenarios see an increase in surface freshwater forcing (Fig. S3) which contributes to a gradual salinity loss. This weakens convection and consequently the AMOC. Subsequently, weak AMOC and convection reduce salt transport into the subpolar gyre reinforcing the salinity decline (Kuhlbrodt et al., 2007). While SAI 2020 mitigates these feedbacks early on, SAI 2080 arrives only after substantial freshening. Closing the density gap via cooling then runs into 'diminishing returns': density gains are less than proportional to temperature decreases (Fig. S5).

Another potentially important factor not included in this analysis is Greenland runoff. It likely contributes to fresher subpolar gyre conditions in SAI 2080. Additionally, Arctic outflows also supply freshwater to the deep convection regions and could vary depending on the scenario (Li et al., 2021).

4 Discussion

In our simulations, the quick drop in GMST due to abrupt SAI is contrasted by a slow ocean response. Gradual early-century SAI, on the other hand, retains an ocean state much closer to the present-day reference. Elevated OHC, weak AMOC and absent deep convection coupled with a lower GMST presents a (transient) climate state unknown from purely GHG-forced scenarios.

Note that our scenarios are extreme cases with a high signal-to-noise ratio, rather than desirable or plausible futures. More cautious protocols typically deploy SAI in tandem with emission mitigation to limit a temporary temperature overshoot (National Academies

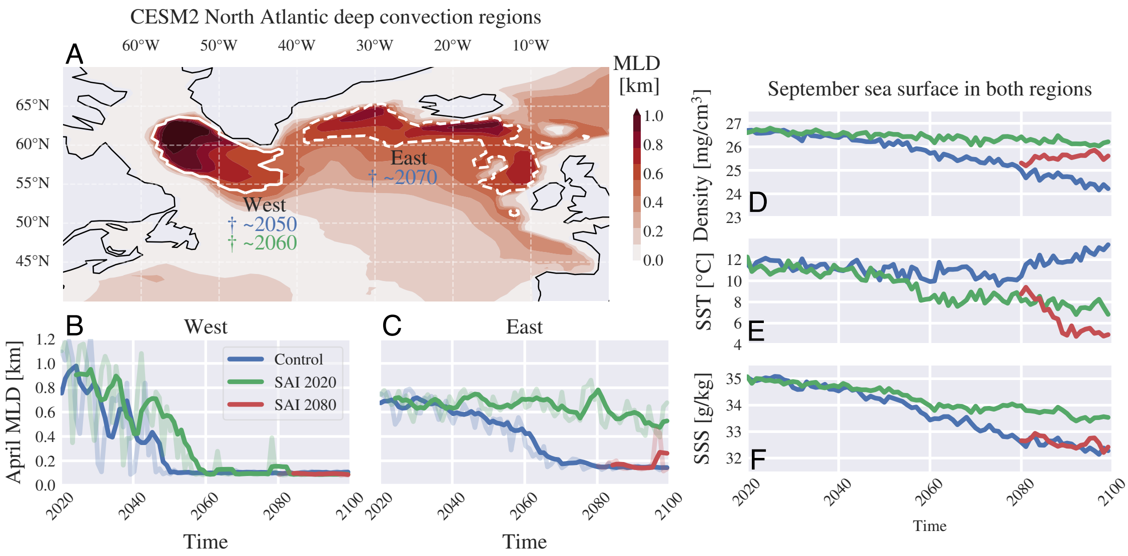

Figure 5. **A**: North Atlantic April mixed layer depths in CESM2 (2020-2030) - *East* and *West* are enclosed by solid and dashed lines respectively. Shutdown dates are denoted with a cross and colored according to scenario (blue: Control, green: SAI 2020). **B-C**: April mixed layer depths in *West* and *East* respectively - Solid lines are five year rolling means (with backward window) applied to the data shown by transparent lines. **D-F**: September mean sea surface density, temperature and salinity over the total *East* and *West* region

of Sciences, Engineering, and Medicine, 2021). If a cooling scenario were actually considered, a ramp-up of SAI would be more sensible than the sudden deployment in SAI 2080. Such a gradual approach would enable a fine-tuning of the injection scheme based on observations.

Besides the high forcings, our scenarios also involve a limited SAI scheme. As our implementation relies on a single degree of freedom, we can only meet a GMST target but not control other aspects of the temperature pattern. More control parameters, on the other hand, may be beneficial to prevent a interhemispheric temperature asymmetry which risks a displacement of the ITCZ (Broccoli et al., 2006; Bischoff & Schneider, 2016). Still, restoring the meridional temperature pattern in SAI 2080 would come with problems of its own: less cooling over the North Atlantic further endangers deep convection.

As for our results, a mitigating effect of SAI on AMOC decline was already known in multiple models and scenarios (Tilmes et al., 2018, 2020; Xie et al., 2022; MacMartin et al., 2022) but not in the case of late-century abrupt deployment. To our knowledge, no studies have been performed on the effect of SAI on deep convection shutdown either. Model dependencies are certain as deep convection shutdown is not a universal phenomenon in CMIP6 (Swingedouw et al., 2021). In fact, the absence of a warming hole in another SAI study using the UKESM1 model (Henry et al., 2023) could indicate a deep convection more stable than that of CESM2.

It is worth pointing out similarities between our abrupt SAI case and rapid negative emission scenarios (Schwinger et al., 2022). Removal of GHG after prolonged heating can lead to an interhemispheric temperature asymmetry if the timescale of extraction is shorter than that of the AMOC recovery. Therefore, the possibility of SAI to manage the interhemispheric temperature gradient is an advantage compared to GHG removal.

A major questions remains open: do the climates of both SAI scenarios eventually converge? This question cannot be answered without extending the simulations, which is outside the scope of this study. When extrapolating our results, the OHC difference is expected to lessen due to residual ocean warming in SAI 2020. Whether the gap fully closes may also depend on the AMOC and deep convection because of their impact on ocean heat uptake (Marshall & Zanna, 2014). As for deep convection, the aforementioned salinity deficit in SAI 2080 inhibits convergence of the SAI scenarios. Nevertheless, should some years of deep convection arise in SAI 2080 (e.g. as a result of natural variability), salt import would be strengthened, thereby improving long-term prospects of a revival.

5 Summary

In this study, we presented model results of a late-century SAI emergency intervention that aims to restore surface temperatures under simultaneous GHG forcing. By comparing our findings with a gradual early-century SAI scenario, we show that abrupt late-century SAI is less effective at mitigating changes in OHC, the AMOC and North Atlantic deep convection.

Firstly, abrupt SAI failed to release heat trapped in deeper ocean layers. Even an early onset of SAI only mitigates but does not halt OHC accumulation. Both results are linked to slow ocean equilibration times and the target of GMST stabilization.

Secondly, abrupt SAI partially stabilized a weakened AMOC, albeit not halting the decline of northward heat transport. Under earlier SAI, the AMOC decline is mitigated in both, volume and heat transport. As a result, the scenarios reach drastically different AMOC states despite comparable GMST. A weaker AMOC may contribute to the observed undercooling of the northern hemisphere in the emergency intervention scenario. This, in turn, may be relevant for the choice of injection pattern.

Thirdly, a shutdown of North Atlantic deep convection could not be reversed with rapid, SAI-induced cooling. We suspect that a weakened AMOC, absence of convective feedback, fresher surface conditions and a sub-proportional density response of water to cooling pose an obstacle for restarting deep convection. An early intervention, on the other hand, retains more salt in the North Atlantic, hence the partial stabilization of deep convection.

Our findings reveal limitations of an SAI emergency deployment: reversing ocean changes after they occur is less feasible than preventing them in the first place. In this context, proactive SAI deployments may be beneficial. Delaying climate action - this includes emission mitigation - in the hope of a later rescue through SAI will come at a price.

6 Open Research

Our CESM2-CAM6 SAI implementation (Pflüger, 2023b), including the input aerosol fields we used, analysis tools (Pflüger, 2023a) and the notebooks used to generate figures (Pflüger, 2023c) can be found on public GitHub repositories. The simulation output required to create all figures is stored in a Zenodo repository (Pflüger et al., 2024). More simulation data can be made available upon reasonable request. The CMIP6 data used for comparison in Fig. 3 is publicly available (Danabasoglu, 2019c, 2019d).

Acknowledgments

We thank Daniele Visioni, Doug MacMartin and Ben Kravitz for sharing their feedback controller code, our colleague Michael Kliphuis for providing his AMOC streamfunction tools and Simone Tilmes for providing CESM2-WACCM data. Furthermore, we thank Jasper de Jong and Michiel Baatsen for fruitful discussions on the feedback control of SAI 2080.

References

Astrom, K. J., & Rundqwist, L. (1989). Integrator windup and how to avoid it. In *1989 American Control Conference.*

Biermann, F., Oomen, J., Gupta, A., Ali, S. H., Conca, K., Hajer, M. A., ... VanDeveer, S. D. (2022). Solar geoengineering: The case for an international non-use agreement. *WIREs Climate Change*. doi: 10.1002/wcc.754

Bischoff, T., & Schneider, T. (2016). The equatorial energy balance, ITCZ position, and double-ITCZ bifurcations. *Journal of Climate*. doi: 10.1175/JCLI-D-15-0328.1

Broccoli, A. J., Dahl, K. A., & Stouffer, R. J. (2006). Response of the ITCZ to northern hemisphere cooling. *Geophysical Research Letters*. doi: 10.1029/2005GL024546

Cheng, L., Von Schuckmann, K., Abraham, J. P., Trenberth, K. E., Mann, M. E., Zanna, L., ... Lin, X. (2022). Past and future ocean warming. *Nature Reviews Earth & Environment*. doi: 10.1038/s43017-022-00345-1

Church, J. A., White, N. J., Domingues, C. M., Monselesan, D. P., & Miles, E. R. (2013). Sea-level and ocean heat-content change. In *International geophysics* (Vol. 103, pp. 697–725). Elsevier. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/B9780123918512000271` doi: 10.1016/B978-0-12-391851-2.00027-1

Danabasoglu, G. (2019a). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp126* [Dataset]. `http://dx.doi.org/10.22033/ESGF/CMIP6.7746`. Earth System Grid Federation.

Danabasoglu, G. (2019b). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp245* [Dataset]. `http://dx.doi.org/10.22033/ESGF/CMIP6.7748`. Earth System Grid Federation.

Danabasoglu, G. (2019c). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp370* [Dataset]. Earth System Grid Federation. doi: 10.22033/ESGF/CMIP6.7753

Danabasoglu, G. (2019d). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp585* [Dataset]. Earth System Grid Federation. doi: 10.22033/ESGF/CMIP6.7768

Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020). The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*. doi: 10.1029/2019MS001916

Drijfhout, S., van Oldenborgh, G. J., & Cimatoribus, A. (2012). Is a decline of AMOC causing the warming hole above the north atlantic in observed and modeled warming patterns? *Journal of Climate*. doi: 10.1175/JCLI-D-12-00490.1

Fasullo, J. T., Tilmes, S., Richter, J. H., Kravitz, B., MacMartin, D. G., Mills, M. J., & Simpson, I. R. (2018). Persistent polar ocean warming in a strategically geoengineered climate. *Nature Geoscience*. doi: 10.1038/s41561-018-0249-7

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G., ... others (2005). Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*.

Henry, M., Haywood, J., Jones, A., Dalvi, M., Wells, A., Visioni, D., ... Tye, M. R.

(2023). Comparison of UKESM1 and CESM2 simulations using the same multi-target stratospheric aerosol injection strategy. *Atmospheric Chemistry and Physics*, *23*(20), 13369–13385. doi: 10.5194/acp-23-13369-2023

IPCC. (2022a). *Climate change 2022: Impacts, adaptation and vulnerability.* Cambridge University Press. doi: 10.1017/9781009325844

IPCC. (2022b). *Climate change 2022: Mitigation of climate change.* Cambridge University Press.

Irvine, P. J., Kravitz, B., Lawrence, M. G., & Muri, H. (2016). An overview of the earth system science of solar geoengineering. *WIREs Climate Change*, *7*(6), 815–833. doi: 10.1002/wcc.423

Kleinschmitt, C., Boucher, O., & Platt, U. (2018). Sensitivity of the radiative forcing by stratospheric sulfur geoengineering to the amount and strategy of the SO_2 injection studied with the LMDZ-S3A model. *Atmospheric Chemistry and Physics*. doi: 10.5194/acp-18-2769-2018

Kravitz, B., MacMartin, D. G., Mills, M. J., Richter, J. H., Tilmes, S., Lamarque, J., ... Vitt, F. (2017). First simulations of designing stratospheric sulfate aerosol geoengineering to meet multiple simultaneous climate objectives. *Journal of Geophysical Research: Atmospheres*. doi: 10.1002/2017JD026874

Kravitz, B., MacMartin, D. G., Wang, H., & Rasch, P. J. (2016). Geoengineering as a design problem. *Earth System Dynamics*. doi: 10.5194/esd-7-469-2016

Kuhlbrodt, T., Griesel, A., Montoya, M., Levermann, A., Hofmann, M., & Rahmstorf, S. (2007). On the driving processes of the atlantic meridional overturning circulation. *Reviews of Geophysics*.

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., Vertenstein, M., et al. (2018, February). *Technical description of version 5.0 of the community land model (clm)* (Technical Report). National Center for Atmospheric Research. `https://www2.cesm.ucar.edu/models/cesm2/land/CLM50_Tech_Note.pdf`.

Li, F., Lozier, M. S., Holliday, N. P., Johns, W. E., Le Bras, I. A., Moat, B. I., ... De Jong, M. F. (2021). Observation-based estimates of heat and freshwater exchanges from the subtropical north atlantic to the arctic. *Progress in Oceanography*, *197*, 102640. doi: 10.1016/j.pocean.2021.102640

Liu, W., Fedorov, A. V., Xie, S.-P., & Hu, S. (2020). Climate impacts of a weakened atlantic meridional overturning circulation in a warming climate. *Science Advances*. doi: 10.1126/sciadv.aaz4876

Lockley, A., Xu, Y., Tilmes, S., Sugiyama, M., Rothman, D., & Hindes, A. (2022). 18 politically relevant solar geoengineering scenarios. *Socio-Environmental Systems Modelling*. doi: 10.18174/sesmo.18127

Long, J. C. S., & Shepherd, J. G. (2014). The strategic value of geoengineering research. In B. Freedman (Ed.), *Global environmental change.* Springer Netherlands. doi: 10.1007/978-94-007-5784-4_24

MacMartin, D. G., Kravitz, B., Tilmes, S., Richter, J. H., Mills, M. J., Lamarque, J., ... Vitt, F. (2017). The climate response to stratospheric aerosol geoengineering can be tailored using multiple injection locations. *Journal of Geophysical Research: Atmospheres*. doi: 10.1002/2017JD026868

MacMartin, D. G., Visioni, D., Kravitz, B., Richter, J., Felgenhauer, T., Lee, W. R., ... Sugiyama, M. (2022). Scenarios for modeling solar radiation modification. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.2202230119

Marshall, D. P., & Zanna, L. (2014). A conceptual model of ocean heat uptake under climate change. *Journal of Climate*. doi: 10.1175/JCLI-D-13-00344.1

Menary, M. B., & Wood, R. A. (2018). An anatomy of the projected north atlantic warming hole in CMIP5 models. *Climate Dynamics*. doi: 10.1007/s00382-017-3793-8

National Academies of Sciences, Engineering, and Medicine. (2021). *Reflecting sunlight: Recommendations for solar geoengineering research and research* *governance*. National Academies Press. doi: 10.17226/25762

Oomen, J. (2021). *Imagining climate engineering: dreaming of the designer climate*. Routledge, Taylor & Francis Group.

Pflüger, D. (2023a). *cesm_tools* [Software]. Retrieved 2023-12-11, from https://github.com/daniel-pflueger/cesm_tools

Pflüger, D. (2023b). *imau_cesm2_prescribed_aerosols* [Software]. Retrieved 2023-12-11, from https://github.com/daniel-pflueger/imau_cesm2_prescribed_aerosols

Pflüger, D. (2023c). *sai_ocean_figures_2023* [Notebook]. Retrieved 2023-12-11, from https://github.com/daniel-pflueger/sai_ocean_figures_2023

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 (Dataset)* [Dataset]. doi: 10.5281/zenodo.10606102

Plazzotta, M., Séférian, R., Douville, H., Kravitz, B., & Tjiputra, J. (2018). Land surface cooling induced by sulfate geoengineering constrained by major volcanic eruptions. *Geophysical Research Letters*. doi: 10.1029/2018GL077583

Reynolds, J. L. (2019). Solar geoengineering to reduce climate change: a review of governance proposals. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*(2229). doi: 10.1098/rspa.2019.0255

Schwinger, J., Asaadi, A., Goris, N., & Lee, H. (2022). Possibility for strong northern hemisphere high-latitude cooling under negative emissions. *Nature Communications*. doi: 10.1038/s41467-022-28573-5

Sgubin, G., Swingedouw, D., Drijfhout, S., Mary, Y., & Bennabi, A. (2017). Abrupt cooling over the north atlantic in modern climate models. *Nature Communications*. doi: 10.1038/ncomms14375

Smith, W. (2020). The cost of stratospheric aerosol injection through 2100. *Environmental Research Letters*. doi: 10.1088/1748-9326/aba7e7

Stocker, T. F. (1998). The seesaw effect. *Science*. doi: 10.1126/science.282.5386.61

Svoboda, T. (2017). *The ethics of climate engineering: solar radiation management and non-ideal justice*. Routledge, Taylor & Francis Group.

Swingedouw, D., Bily, A., Esquerdo, C., Borchert, L. F., Sgubin, G., Mignot, J., & Menary, M. (2021). On the risk of abrupt changes in the north atlantic subpolar gyre in CMIP6 models. *Annals of the New York Academy of Sciences*. doi: 10.1111/nyas.14659

Tilmes, S., MacMartin, D. E., Lenaerts, J. T. M., Van Kampenhout, L., Muntjewerf, L., Xia, L., … Robock, A. (2020). Reaching 1.5 °c and 2.0 °c global surface temperature targets using stratospheric aerosol geoengineering. *Earth System Dynamics*. doi: 10.5194/esd-11-579-2020

Tilmes, S., Richter, J. H., Kravitz, B., MacMartin, D. G., Mills, M. J., Simpson, I. R., … Ghosh, S. (2018). CESM1(WACCM) stratospheric aerosol geoengineering large ensemble project. *Bulletin of the American Meteorological Society*. doi: 10.1175/BAMS-D-17-0267.1

Visioni, D., MacMartin, D. G., Kravitz, B., Boucher, O., Jones, A., Lurton, T., … Tilmes, S. (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*. doi: 10.5194/acp-21-10039-2021

Visioni, D., MacMartin, D. G., Kravitz, B., Lee, W., Simpson, I. R., & Richter, J. H. (2020). Reduced poleward transport due to stratospheric heating under stratospheric aerosols geoengineering. *Geophysical Research Letters*. doi: 10.1029/2020GL089470

Weijer, W., Cheng, W., Garuba, O. A., Hu, A., & Nadiga, B. T. (2020). CMIP6 models predict significant 21st century decline of the atlantic meridional overturning circulation. *Geophysical Research Letters*. doi: 10.1029/

2019GL086075

Wieners, C. E., Hofbauer, B. P., De Vries, I. E., Honegger, M., Visioni, D., Russchenberg, H. W. J., & Felgenhauer, T. (2023). Solar radiation modification is risky, but so is rejecting it: a call for balanced research. *Oxford Open Climate Change*. doi: 10.1093/oxfclm/kgad002

Xie, M., Moore, J. C., Zhao, L., Wolovick, M., & Muri, H. (2022). Impacts of three types of solar geoengineering on the atlantic meridional overturning circulation. *Atmospheric Chemistry and Physics*. doi: 10.5194/acp-22-4581-2022

Zarnetske, P. L., Gurevitch, J., Franklin, J., Groffman, P. M., Harrison, C. S., Hellmann, J. J., ... Yang, C.-E. (2021). Potential ecological impacts of climate intervention by reflecting sunlight to cool earth. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1921854118

Supplementary Material:
Flawed Emergency Intervention: Slow Ocean Response to Abrupt Stratospheric Aerosol Injection

Daniel Pflüger[1], Claudia E. Wieners[1], Leo van Kampenhout[1], René R. Wijngaard[1], and Henk A. Dijkstra[1]

[1]Institute Marine and Atmospheric Research Utrecht, Princetonplein 5, 3584 CC Utrecht, The Netherlands

February 7, 2024

Variable name	Description	Normalization
AODVISstdn	Aerosol optical depth	Global mean
SAD_AERO	Surface aerosol density	Total aerosol surface area
so4mass_a1	Aerosol mass concentration of aerosol mode one	Total mass
so4mass_a2	Aerosol mass concentration of aerosol mode two	Total mass
so4mass_a3	Aerosol mass concentration of aerosol mode three	Total mass
diamwet_a1	Aerosol wet diameter of aerosol mode one	Root mean square
diamwet_a2	Aerosol wet diameter of aerosol mode second	Root mean square
diamwet_a3	Aerosol wet diameter of aerosol mode three	Root mean square

Table 1: Prescribed and scaled aerosol fields in CESM2-CAM6 with description of respective normalizing approach

1 Methods

1.1 Prescribed aerosols

Our SAI implementation is based on prescribed aerosol fields. This means that the variables representing stratospheric aerosols are predetermined, non-interactive and serve as boundary conditions for CAM6. We obtain these fields by processing output from CESM2-WACCM simulations (Tilmes et al., 2020) of a SAI 2020-like scenario. The processed output is scaled according to the desired level of cooling, and ultimately fed into CAM6.

1.1.1 Processing of WACCM output

Let $F^{\text{in}}(t, d, x)$ be an CESM2-WACCM stratospheric aerosol field at year t, day of the year d and position x (e.g. longitude, latitude, altitude or a combination thereof). We process this field in three steps: normalization, averaging and fitting.

Normalization

Firstly, we normalize the field. The choice of normalization depends on the type of field and can either be physically motivated or mathematically abstract, see also Table 1. For example, a reasonable way to normalize a mass concentration field is via its spatial integral, the total mass. What matters, is that the norm - or amplitude - behaves monotonically in the overall SAI intensity. That way, amplitudes of different fields can later be mapped onto each other in the fitting step.

In any case, we obtain an amplitude $n(t)$ of $F^{\text{in}}(t, d, x)$ for every year t. This also gives a normalized field $\hat{F}^{\text{in}}(t, d, x) = F^{\text{in}}(t, d, x)/n(t)$

The normalized field carries information about the spatial and seasonal distribution of the aerosol field while ignoring its amplitude.

Averaging

Secondly, we average the normalized field over multiple years. In our case, we decided to use the years 2070-2100 in which the CESM2-WACCM simulation has accumulated a large aerosol burden, providing an accurate starting point for our SAI 2080 scenario. The averaged field is defined by $\bar{F}(d, x) = \frac{1}{t_f - t_i} \sum_{t=t_i}^{t_f} \hat{F}^{\text{in}}(t, d, x)$.

It is crucial to perform the normalization step before averaging. Otherwise, the terms of the sum may be different in magnitude. When computing an average of a mass concentration field, for instance, later years with higher SAI intensity would dominate the sum.

Fitting

After performing the prior steps for all fields, now indexed by i, we obtain the averaged fields $\bar{F}_i(d, x)$ and amplitudes $n_i(t)$. These amplitudes are all related to each other. It makes physical sense, for example, that a higher global mean AOD comes along with higher total aerosol mass. This means that can we can designate a reference amplitude - here: global mean AOD denoted as AOD(t) - from which other amplitudes are derived.

We establish these relationships by simple power-law fits of the form $y(x) = ax^b + c$ where a, b, c are fit parameters, x is AOD and y a target amplitude (evaluated in the same year t). Finally, we obtain fitted amplitudes $n_i^f(\text{AOD})$ for all fields i expressed solely in terms of AOD.

1.1.2 Scaling of aerosol fields

The process above gives averaged fields $\bar{F}_i(d, x)$ which can be scaled in amplitude depending on the desired AOD. This gives the scaled fields

$$F_i(t, d, x) = n_i^f(\text{AOD}(t))\bar{F}_i(d, x) \tag{1}$$

AOD(t) itself is determined in the context of the simulation scenario. Below, we describe a control algorithm that chooses AOD(t) such that a GMST stabilization can be achieved.

1.2 Feedback-Feedforward Algorithm

We control the GMST by adjusting the aerosol shading, parameterised by the AOD n. For that purpose, we use a feedback-feedforward algorithm that has become common in SAI modelling.

The algorithm starts from an informed guess of the expected AOD necessary for a specific level of cooling. This so-called feedforward could for example come from estimates of aerosol sensitivity of radiative forcing (Hansen et al., 2005). In our case, we use tweaked estimates from aforementioned CESM2-WACCM runs.

On top of the feedforward, proportional-integral feedback adds a correction based on the deviation $\Delta T(t)$ of the GMST from the target. As their names suggest, the proportional and integral components of the feedback introduce corrections directly proportional to $\Delta T(t)$ as well as proportional to the discrete sum $\sum_{t'=t_i}^{t} \Delta T(t')$.

In total, we have

$$\text{AOD}(t) = \underbrace{k_{\text{ff}}(t - t_{\text{ff}})}_{\text{feedforward}} + \underbrace{k_{\text{p}}\Delta T(t)}_{\text{proportional}} + \underbrace{k_{\text{i}} \sum_{t'=t_i}^{t} \Delta T(t')}_{\text{integral}} \tag{2}$$

where $k_{\text{ff}}, k_{\text{p}}, k_{\text{i}}$ and t_{ff} are constants.

Under SAI 2020, the integrator is simply initialized in $t_{\text{i}} = 2020$. To avoid a large integral term - an 'integrator windup' (Astrom & Rundqwist, 1989) - during cooling in SAI 2080, we have considered multiple options described in the next subsection.

Note that the feedforward was adjusted when going from SAI 2020 to SAI 2080, see also Table 2. The updated parameters were obtained by using the output of the trained SAI 2020 controller.

1.3 Validation and limitations

As demonstrated by the main article, our implementation can successfully be run in CESM2-CAM6. We can evaluate the level of physical realism by comparison with CESM2-WACCM. This is possible because our SAI 2020 scenario mirrors the Geo SSP5-8.5 1.5 case implemented in CESM2-WACCM (Tilmes et al., 2020).

Scenario	k_{ff}	t_{ff}	k_{p}	k_{i}
SAI 2020	0.0103	2020	0.028	0.028
SAI 2080	0.0096	2028	0.028	0.028
SAI 2080 (mod)	0.0096	2028	0.028	0.028

Table 2: Feedforward-feedback parameters for all scenarios assuming that time is given in units of years and temperature in units of Kelvin.

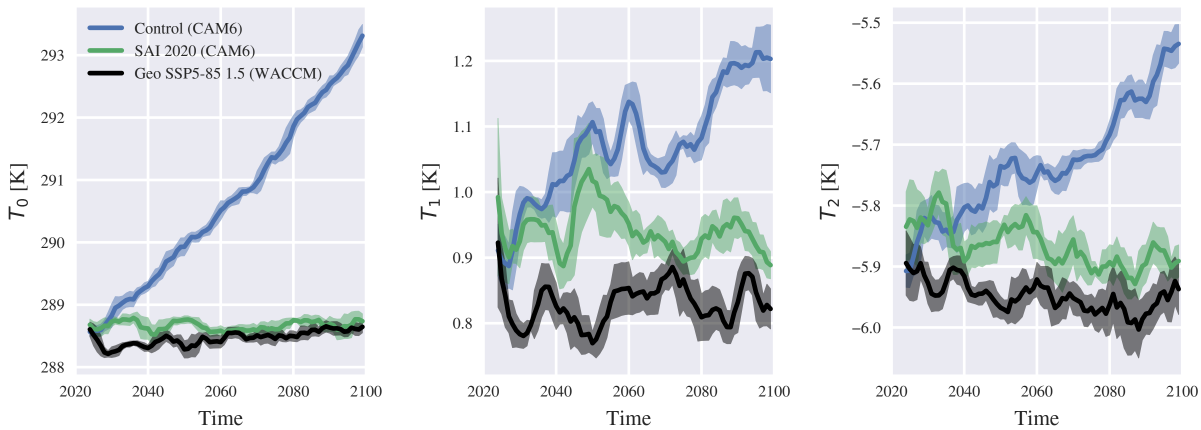

Figure S 1: **A**: Annual mean T_0 (=GMST) for Control, SAI 2020 and Geo SSP5-8.5 1.5 (3rd ensemble member) over time - we applied a five year rolling mean and standard deviation (uncertainty bars). **B-C**: same as **A** but for T_1 and T_2 respectively

Similar to CESM2-WACCM, we see that SAI 2020 mitigates AMOC decline. The CESM-CAM6 AMOC index decrease of roughly 30% in the period 2020-2100 matches that of CESM2-WACCM in the same period (Tilmes et al., 2020). Note that we evaluate AMOC at 26°N rather than at the maximum (around 35°N) as done by Tilmes et al., which poses no problem due to the similar scaling.

As already demonstrated in the main text, our approach can indeed stabilize GMST. Additionally, our prescribed aerosol fields lead to a similar cooling pattern as in the original CESM2-WACCM runs. The zonal-mean surface temperature can be expressed by(Kravitz et al., 2016)

$$T_0 = \frac{1}{A}\int_{-\pi/2}^{\pi/2} \mathrm{dlat} \int_0^{2\pi} \mathrm{dlon}\, T(\mathrm{lon},\mathrm{lat}) \tag{3}$$

$$T_1 = \frac{1}{A}\int_{-\pi/2}^{\pi/2} \mathrm{dlat} \int_0^{2\pi} \mathrm{dlon}\, T(\mathrm{lon},\mathrm{lat}) \sin(\mathrm{lat}) \tag{4}$$

$$T_2 = \frac{1}{A}\int_{-\pi/2}^{\pi/2} \mathrm{dlat} \int_0^{2\pi} \mathrm{dlon}\, T(\mathrm{lon},\mathrm{lat}) \frac{1}{2}(3\sin^2(\mathrm{lat}) - 1) \tag{5}$$

$$\tag{6}$$

where $T(\mathrm{lon},\mathrm{lat})$ is the (near-)surface air temperature depending on longitude and latitude and A is Earth's surface area. T_0, T_1, T_2 can be intuitively understood as GMST, inter-hemispheric (positive values: NH is warmer than SH) and equator-pole temperature difference (positive values: poles hotter than equator). Fig. S1 shows that T_1 and T_2 trends are successfully mitigated in SAI 2020.

A clear limitation of our approach is the inability to follow multiple climate objectives.

Since only a single parameter, AOD, can be altered, it is not possible to directly adjust T_1 and T_2. This becomes very obvious in SAI 2080 where a strong inter-hemispheric temperature contrast exists. The cooling pattern which stabilized T_1 under strong AMOC conditions in 2020 is no longer appropriate in SAI 2080. If our method were used to efficiently produce ensembles of SAI 20XX-like scenarios, a single CESM2-WACCM run with multiple objectives would suffice to generate the necessary aerosol patterns. That way, our implementation can still save computation time.

More subtle, the assumption that the aerosol fields do not change their spatial (and intra-annual temporal) pattern depending on the level of SAI is not generally valid. As increasing aerosol burdens heat the stratosphere, they alter the circulation and hence the aerosol distribution (Visioni et al., 2020). This detail is captured by our chosen averaging interval of 2070-2100. We implicitly use aerosol fields consistent with higher aerosol concentrations and reduced polewards transport. As it is unclear how the stratosphere responds to even higher aerosol burdens, our approach should be constricted to GHG concentrations not higher than SSP5-8.5 in 2100.

A manuscript performing a deeper evaluation - also including atmospheric responses - is currently in preparation in collaboration with our colleagues Jasper de Jong and Michiel Baatsen.

1.4 Design choices in SAI 2080

In its original form, the control algorithm described above will lead to a drastic undercooling when used in a scenario like SAI 2080. This is because the GMST error at deployment time is 'remembered' by the integrator and therefore adds to the prescribed AOD until all traces of the initial perturbations are removed. That, in turn, can only happen if GMST drops below the target such that negative contributions can enter the integrator. Eventually, this process removes the initial undercooling. When exactly that is, is not obvious. Given the short simulated timeframe of SAI 2080, it makes sense to think of alternatives. We considered three different approaches:

- **Slow equilibration without integrator** (not successfully implemented): The integrator is turned off during the initial cooling phase which means that the feedforward dictates the AOD (the proportional feedback is very small in our case). As a result, the cooling process is slower. There is also no guarantee that the GMST target can be reached because the feedforward may be inaccurate. Since technical limitations originally prevented us from simulating beyond 2100, we ruled out this approach. If implemented successfully, however, this approach requires little ad-hoc assumptions and is therefore a candidate for a generalized protocol.

- **Conditional integrator (SAI 2080)** : The integrator is turned off initially but gets activated after GMST is within 0.5K of the target or, as a fail-safe measure, six years have passed. Latter condition helps to speed up the cooling process if it is too sluggish but risks an integrator windup. In fact, this occurred during SAI 2080 and lead to the slight undercooling.

- **Integrator reset (SAI 2080 (mod))** : The integrator is always on but the summed error term is reset once the target GMST is reached. That way, the integrator still speeds up cooling but cannot induce an undercooling. A downside is the intermittently higher AOD which may produce transient atmospheric effects such as a stronger precipitation decrease. We can not rule out that this has some impact on the ocean (e.g. by altering surface salinities) but the importance of the

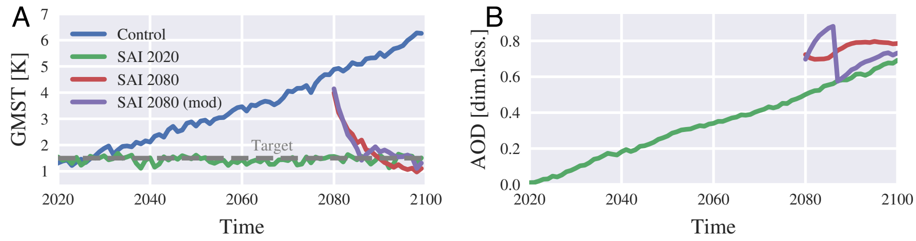

Figure S 2: **A**: Annual mean GMST for all scenarios from main text with addition of modified SAI 2080 scenario **B**: Global and annual mean stratospheric aerosol optical depth in all SAI scenarios (including modified SAI 2080 scenario)

transient phase should wane over a longer duration. This is why SAI 2080 (mod) is our candidate for future extension studies.

Fig. S2 shows how the integrator reset in SAI 2080 (mod) resolves the issue of undercooling while at the same time introducing an AOD discontinuity a few years after deployment.

2 Surface Freshwater Fluxes in Deep Convection Regions

Multiple drivers are responsible for fresher conditions in deep convection regions. While we have not disentangled all possible contributions, we can rule out surface freshwater flux (SFWF) being the distinguishing feature between scenarios. SFWF consists of precipitation, evaporation, sea ice melt/growth and runoff terms. Fig. S3 shows that SFWF increases in all scenarios. While Control and SAI 2020 have similar values throughout the simulation, SAI 2080 induces slightly fresher conditions.

The remarkably similar SFWF are unexpected because SAI has a distinct impact on the hydrological cycle (Fig. S3B-C). The decline in atmospheric freshwater flux turns out to be compensated by increased sea ice melting (Fig. S3D). Apparently, the cool SAI conditions allow for sea ice import and subsequent melting in the deep convection regions. The negative residual fluxes at the end of Control are an artefact of the implementation of ice runoff fluxes in the land model (Lawrence et al., 2018, Ch. 13.5, p. 145).

3 Stratification and Mixed Layer

The deep convection season in the North Atlantic depends on a pre-conditioning, i.e. a weak stratification after summer. Fig. S4 makes it clear that high sea surface densities (here used as a proxy for stratification) in September enable deep mixed layers in the following April. More specifically, deep convection is enabled for sea surface densities beyond a critical value of around 26 mg/cm^3. Beyond that point, there is a large variability in mixed layer depths.

Fig. S5 explains the sea surface density dynamics in terms of temperature and salinity. We see that salinities in *West* fall enough to place both, SAI 2020 and SAI 2080, well below the critical density. In *East*, SAI 2020 manages to stay above the critical threshold as cooling balances the effects of freshening. Branching off from Control, SAI 2080 experiences a cooling shock that brings densities very close to the line of critical density.

Note that the lines of equal density in Fig. S5 are convex which is a consequence of the nonlinear equation of state for sea water. The thermal expansivity of water decreases with

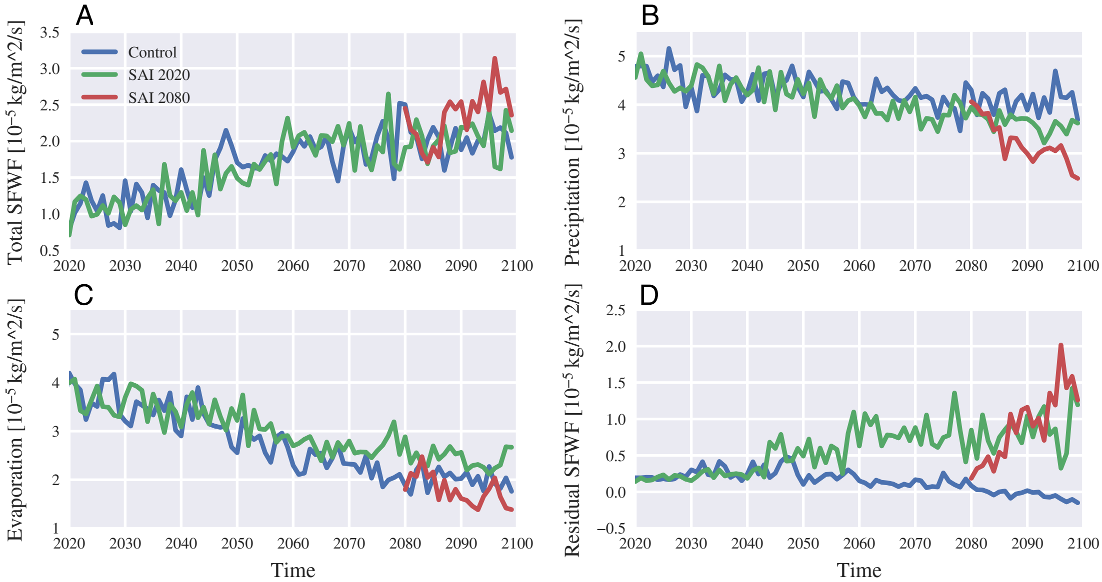

Figure S 3: Mean annual surface freshwater fluxes into total *East* and *West* regions; positive values indicate downward flux except for **C** - **A**: Total flux **B**: Precipitation **C**: Evaporation **D**: Residual = Total flux - (Precipitation - Evaporation); contains sea ice contributions

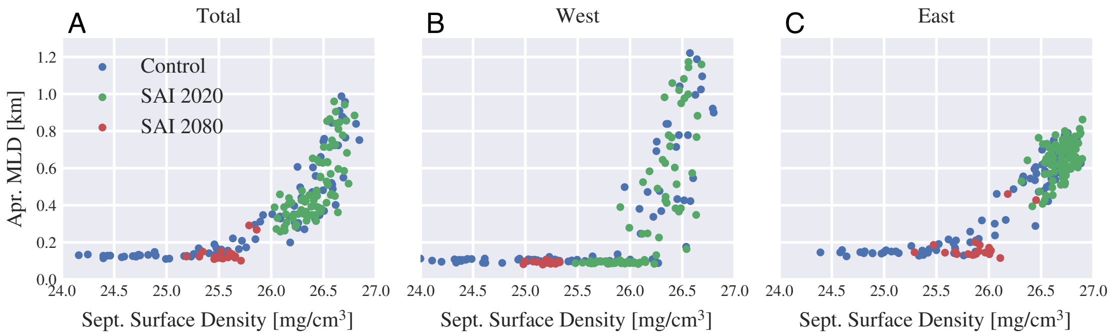

Figure S4: **A-C** April mixed layer depth versus surface density of previous September in respective regions - The density values have an offset of $1000\,\mathrm{mg/cm^3}$.

lower temperatures: the cooler the initial temperature, the weaker the density gain for any given temperature drop. If the equation of state were linear, (i.e. density depending linearly on temperature and salinity) abrupt cooling could have restarted deep convection in *East*.

4 CMIP6-CESM2 North Atlantic SST

Abrupt sea surface cooling is a key feature in deep convection shutdown and has been observed in CESM2 (and CESM1) before (Sgubin et al., 2017; Swingedouw et al., 2021). As shown by Fig. S6, this phenomenon is best seen in annual mean SSTs. Our main text, on the other hand, only shows September SST from which a rapid cooling in Control is not obvious. Note that inter-annual variability makes it hard to discern temperature trends, in particular in the case of Control. For that reason, we add a three member CESM2 ensemble from CMIP6 (Danabasoglu, 2019a, 2019b, 2019c, 2019d) to help filter

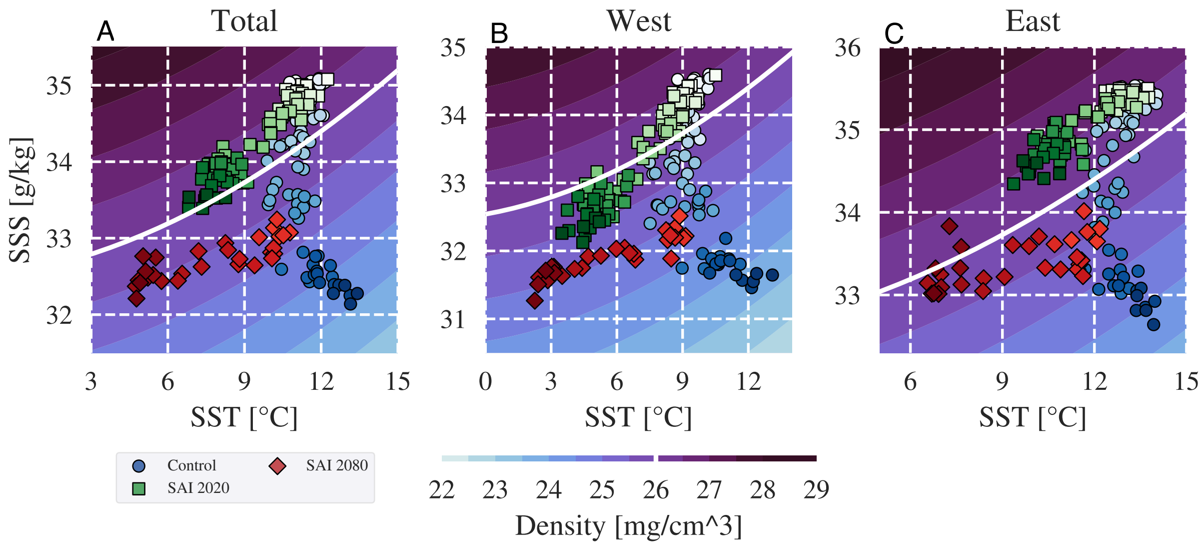

Figure S5: **A-C**: Sea surface salinity and temperature trajectories in all respective regions - Filled contours represent the water density. The singled out white contour is at the critical density of 26 mg/cm^3. Marker saturation represents time and ranges from light (2020) to saturated (2100).

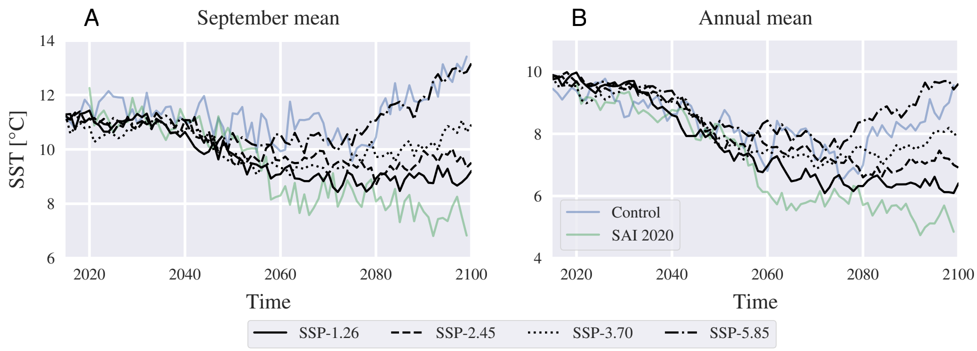

Figure S6: **A**: September mean SST over combined deep convection region - Black lines show CMIP6-CESM2 ensemble means (three members) for different scenarios. Pale colored lines show own simulation results. **B**: Same as A but for annual mean data

out this variability

Note that abrupt cooling takes place in a wide range of SSP scenarios (Fig. S6), indicating that a CESM2 deep convection shutdown may already be locked in. In fact, the SST evolution of the different SSP scenarios only diverges around mid-century. The fact that SSTs decline in SAI 2020 (Fig. S6) despite the partial stabilization of deep convection could imply that a shutdown of *West* alone is sufficient for the cooling to occur.

In total, an abrupt cooling in the North Atlantic in CESM2 is not preventable by either strong emission mitigation (SSP1-2.6) or SAI 2020. While it is plausible that a proactive SAI intervention with a lower GMST target could stabilize *West*, this would also cool the North Atlantic below present-day conditions. How that SAI-induced cooling would compare to the convection-loss-induced cooling is not obvious.

References

Astrom, K. J., & Rundqwist, L. (1989). Integrator windup and how to avoid it. In *1989 American Control Conference.*

Danabasoglu, G. (2019a). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp126* [Dataset]. http://dx.doi.org/10.22033/ESGF/CMIP6.7746. Earth System Grid Federation.

Danabasoglu, G. (2019b). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp245* [Dataset]. http://dx.doi.org/10.22033/ESGF/CMIP6.7748. Earth System Grid Federation.

Danabasoglu, G. (2019c). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp370* [Dataset]. http://dx.doi.org/10.22033/ESGF/CMIP6.7753. Earth System Grid Federation.

Danabasoglu, G. (2019d). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp585* [Dataset]. http://dx.doi.org/10.22033/ESGF/CMIP6.7768. Earth System Grid Federation.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G., ... others (2005). Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*.

Kravitz, B., MacMartin, D. G., Wang, H., & Rasch, P. J. (2016). Geoengineering as a design problem. *Earth System Dynamics*. doi: 10.5194/esd-7-469-2016

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., Vertenstein, M., et al. (2018, February). *Technical description of version 5.0 of the community land model (clm)* (Technical Report). National Center for Atmospheric Research. https://www2.cesm.ucar.edu/models/cesm2/land/CLM50_Tech_Note.pdf.

Sgubin, G., Swingedouw, D., Drijfhout, S., Mary, Y., & Bennabi, A. (2017). Abrupt cooling over the north atlantic in modern climate models. *Nature Communications*. doi: 10.1038/ncomms14375

Swingedouw, D., Bily, A., Esquerdo, C., Borchert, L. F., Sgubin, G., Mignot, J., & Menary, M. (2021). On the risk of abrupt changes in the north atlantic subpolar gyre in CMIP6 models. *Annals of the New York Academy of Sciences*. doi: 10.1111/nyas.14659

Tilmes, S., MacMartin, D. E., Lenaerts, J. T. M., Van Kampenhout, L., Muntjewerf, L., Xia, L., ... Robock, A. (2020). Reaching 1.5°c and 2.0°c global surface temperature targets using stratospheric aerosol geoengineering. *Earth System Dynamics*. doi: 10.5194/esd-11-579-2020

Visioni, D., MacMartin, D. G., Kravitz, B., Lee, W., Simpson, I. R., & Richter, J. H. (2020). Reduced poleward transport due to stratospheric heating under stratospheric aerosols geoengineering. *Geophysical Research Letters*. doi: 10.1029/2020GL089470