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 $n_i^f(\mathrm{AOD})$ 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

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

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

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

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

Under SAI 2020, the integrator is simply initialized in $t_{\mathrm{i}} = 2020$. 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.

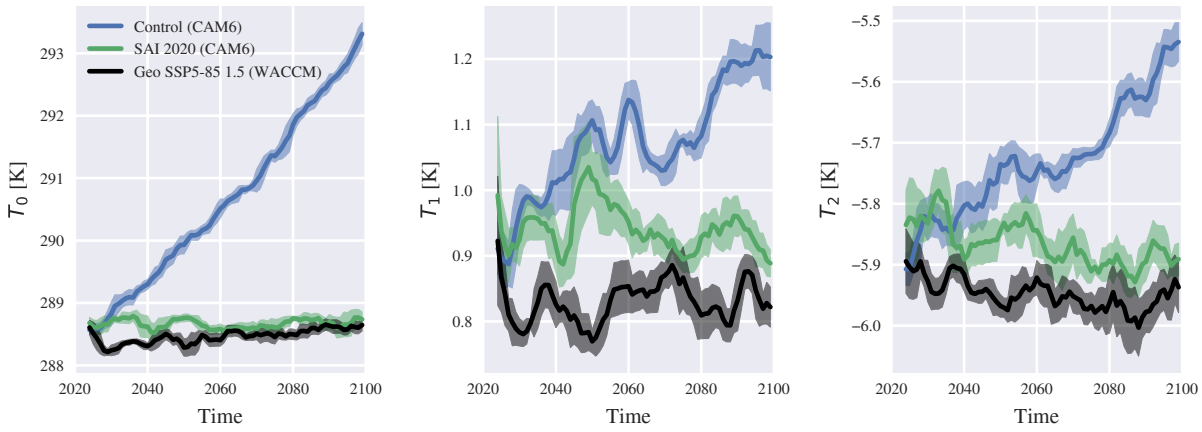

Figure S 1: **A**: Annual mean T_0 (=GMST) for Control, SAI 2020 and Geo SSP5-8.5 1.5 (3rd ensemble member) over time - we applied a five year rolling mean and standard deviation (uncertainty bars). **B-C**: same as **A** but for T_1 and T_2 respectively

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

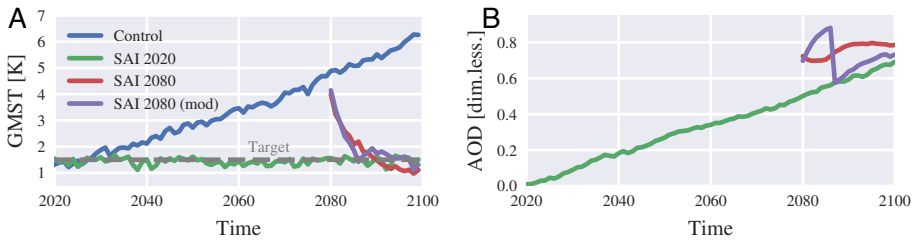

Figure S 2: **A**: Annual mean GMST for all scenarios from main text with addition of modified SAI 2080 scenario **B**: Global and annual mean stratospheric aerosol optical depth in all SAI scenarios (including modified SAI 2080 scenario)

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

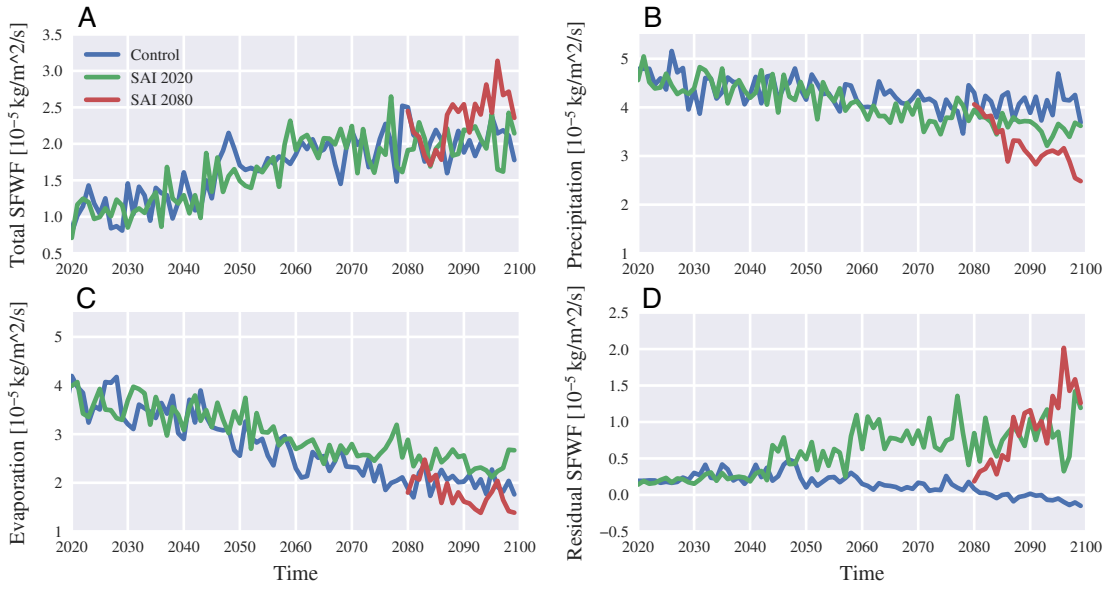

Figure S 3: Mean annual surface freshwater fluxes into total *East* and *West* regions; positive values indicate downward flux except for **C** - **A**: Total flux **B**: Precipitation **C**: Evaporation **D**: Residual = Total flux - (Precipitation - Evaporation); contains sea ice contributions

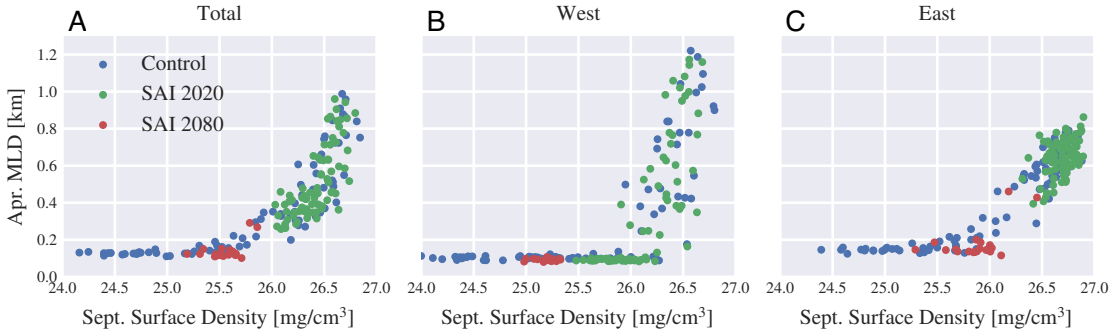

Figure S4: **A-C** April mixed layer depth versus surface density of previous September in respective regions - The density values have an offset of 1000 $\mathrm{mg/cm^3}$.

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

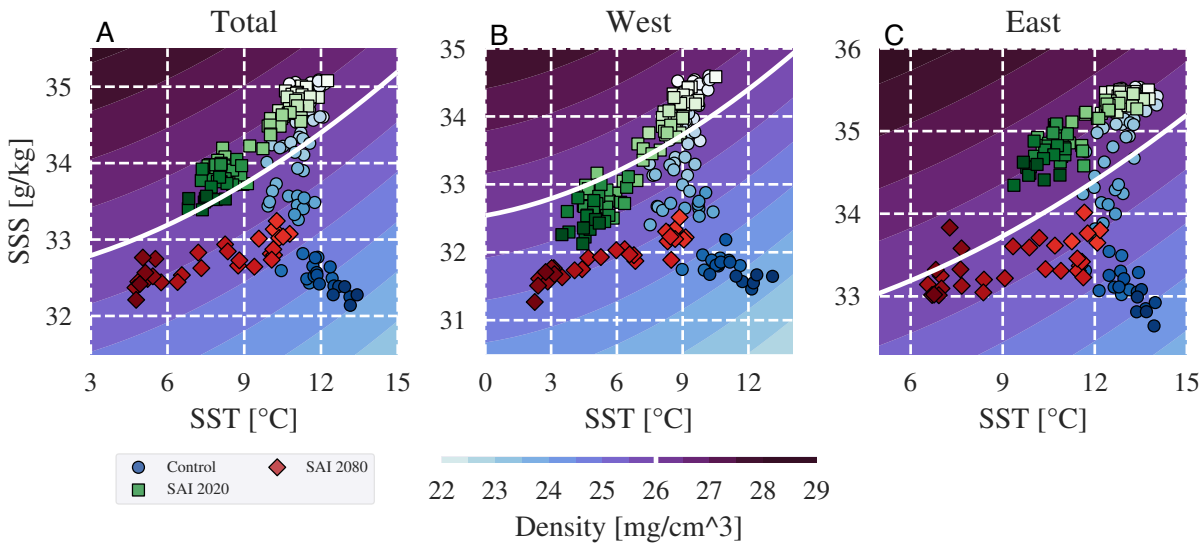

Figure S5: **A-C**: Sea surface salinity and temperature trajectories in all respective regions - Filled contours represent the water density. The singled out white contour is at the critical density of 26 mg/cm^3. Marker saturation represents time and ranges from light (2020) to saturated (2100).

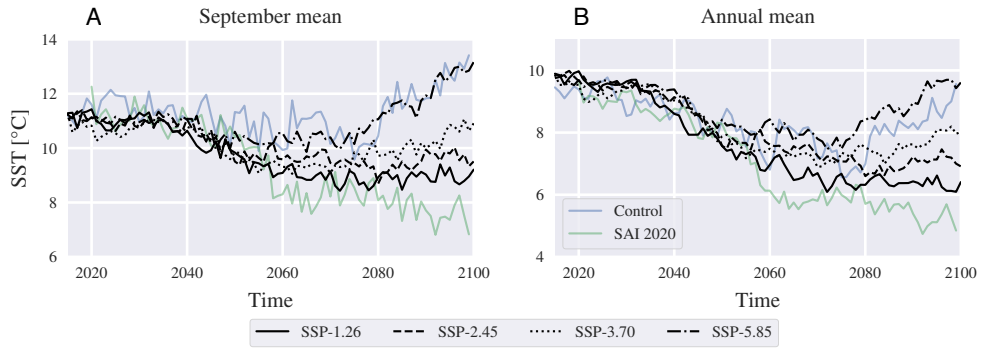

Figure S6: **A**: September mean SST over combined deep convection region - Black lines show CMIP6-CESM2 ensemble means (three members) for different scenarios. Pale colored lines show own simulation results. **B**: Same as A but for annual mean data

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