Accelerating Exploration of Marine Cloud Brightening Impacts on Tipping Points Using an AI Implementation of Fluctuation-Dissipation Theorem.

Haruki Hirasawa[1], Sookyung Kim[2], Peetak Mitra[2,3], Subhashis Hazarika[2], Salva Ruhling-Cachay[2,4], Dipti Hingmire[1], Kalai Ramea[2], Hansi Singh[1], Philip J. Rasch[5]

[1] University of Victoria, Victoria, BC, Canada
[2] Palo Alto Research Center, Palo Alto, CA, US
[3] Excarta, Palo Alto, CA, US
[4] University of California San Diego, San Diego, CA, US
[5] University of Washington, Seattle, WA, US
Corresponding authors: hhirasawa@uvic.ca, sookim@parc.com

Abstract

Marine cloud brightening (MCB) is a proposed climate intervention technology to partially offset greenhouse gas warming and possibly avoid crossing climate tipping points. The climate impacts of MCB are typically estimated using computationally expensive Earth System Model (ESM) simulations, preventing a thorough assessment of the large possibility space of MCB intervention patterns. Here, we describe an AI model, named AiBEDO, that can be used to rapidly project climate responses via a novel application of the Fluctuation-Dissipation Theorem (FDT). AiBEDO is a Multilayer Perceptron (MLP) model that maps from monthly-mean radiation anomalies to surface climate anomalies at a range of time lags. By leveraging a internal climate noise from a large existing dataset of ESM simulations, we use AiBEDO to construct an FDT operator that successfully projects the pattern of MCB climate responses when evaluated against ESM simulations. We propose that AiBEDO could be used to identify MCB forcing patterns to that reduce tipping point risks while minimizing negative side effects in other parts of the climate.

Introduction

Marine Cloud Brightening

Tipping points in the climate system are critical components of the climate response to anthropogenic warming, as they have the potential to undergo rapid, self-perpetuating, and possibly irreversible changes (McKay et al. 2022). Should warming approach or cross a threshold that activates such a tipping point, a climate intervention might be considered to prevent it. One class of interventions are solar radiation modification (SRM) methods which scatter away a portion of incoming sunlight (A.K.A., solar radiation) to counter some of the effects of greenhouse warming.

Here, we consider one such SRM technique, Marine Cloud Brightening (MCB), in which sea salt aerosols would be injected into marine boundary layer clouds to increase their albedo (Latham et al. 2012). If MCB is to be deployed with the aim of limiting tipping point risk, it is crucial that we carefully determine if MCB actually reduces these risks and rule out MCB scenarios that might cause negative unintended climate changes or exacerbate tipping points (Diamond et al. 2022). Due to the short atmospheric lifetime of tropospheric aerosol particles, MCB interventions would be highly localized. This presents both a substantial challenge and a potential opportunity, as the possibility space of MCB interventions is vast, both in strength and spatial pattern. Thus, a thorough assessment of the feasibility of MCB interventions must consider a wide range of potential scenarios. On the other hand, it may be possible to find specific patterns of MCB intervention that achieve desirable climate effects while minimizing negative side effects.

Typically, the effect of MCB is evaluated using simulations in Earth System Models (ESMs), which are comprehensive, dynamic models of the coupled atmosphere-ocean-land-ice system (Rasch, Latham, and Chen 2009; Jones, Haywood, and Boucher 2009; Stjern et al. 2018). However, ESM simulations are computationally expensive, requiring tens of thousands of core-hours to obtain sufficient sample sizes to assess the impact of a given intervention scenario. Thus, they are impractical as tools to explore a wide range of possible MCB intervention patterns. To accelerate this exploration, we have developed AiBEDO, an AI model that emulates the relationship between atmospheric radiative flux anomalies and resulting surface climate changes. By using AiBEDO to project the climate impact of cloud radiative flux anomalies, we can rapidly evaluate the impact of MCB-like perturbations on the climate.

Fluctuation-Dissipation Theorem

As there are few MCB forcing simulations that have been conducted in the most recent generation of Coupled Model Intercomparison phase 6 (CMIP6) ESMs, we cannot train on an existing repository of ESM responses to MCB forcing. Thus, the design philosophy of AiBEDO borrows from the Fluctuation-Dissipation Theorem (FDT), which posits that under certain conditions, the response of a dynamical system to a perturbation can be inferred from the time-lagged correlation statistics of natural internal fluctuations in the system (Kubo 1966; Leith 1975). It has been argued that the climate is one such system, as it can be represented in a manner that approximately satisfies the requirements of the

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

FDT. Namely that the climate approximately satisfies the Liouville relation and has a stationary and differentiable probability density function (Leith 1975; Gritsun and Branstator 2007; Majda, Abramov, and Gershgorin 2010).

The FDT has been used to estimate the linear response of the climate to a range of forcings: CO_2 doubling and solar radiation perturbations (Cionni, Visconti, and Sassi 2004), atmospheric heating (Gritsun and Branstator 2007), regional ocean heat convergence anomalies (Liu et al. 2018), etc. If the statistics of the dynamical system are Gaussian, the FDT operator $\mathbf{L}$ can be computed by convolving the covariance matrix between the predictor variables $\vec{x}$ and predictand variables $\vec{y}$, $\mathbf{C}_{\vec{y},\vec{x}}(\tau)$, with the autocovariance matrix of $\vec{x}$, $\mathbf{C}_{\vec{x},\vec{x}}(0)$, over time lags τ. The climatological mean response $\langle \delta\vec{y} \rangle = \langle y' \rangle - \langle y \rangle$ (angle brackets indicating the climatological mean) to a constant forcing $\delta\vec{f}$ is then computed as

$$\langle \delta\vec{y} \rangle = \mathbf{L}\delta\vec{f} = \left[\int_0^\infty \mathbf{C}_{\vec{y},\vec{x}}(\tau)\mathbf{C}^{-1}_{\vec{x},\vec{x}}(0)d\tau \right] \delta\vec{f} \quad (1)$$

As FDT is limited to the linear component of the climate response, we seek to use an AI model with the intention of capturing both linear and non-linear components of the response and loosening some of the conditions required by classical FDT (e.g., the need to *a priori* reduce the dimensionality of the system and the assumption of near-Gaussian statistics; see Majda, Abramov, and Gershgorin 2010).

We define an AiBEDO operator $A_\tau(\vec{x}_i(t))$, which maps the statistical relationship from a given input $\vec{x}_i(t)$ field to an output $\overline{\vec{y}_i(t+\tau)}$ field after some time lag τ,

$$A_\tau(\vec{x}_i(t)) \; : \; \vec{x}_i(t) \; \rightarrow \; \overline{\vec{y}_i(t+\tau)} \, . \quad (2)$$

with i indexing the different initial conditions sampled from internal climate noise. Due to uncertainties in the initial condition $\vec{x}_i$ (from monthly averaging and discrete sampling of the fields) and the chaotic dynamics of the system, there is no unique mapping from a given input $\vec{x}_i(t)$ to a later output $\vec{y}_i(t+\tau)$. Rather, AiBEDO projects the mean of the distribution of possible $\vec{y}_i(t+\tau)$ trajectories after τ months given the initial conditions $\vec{x}_i(t)$. We denote this mean using an overline, and the output of AiBEDO as $\overline{\vec{y}_i(t+\tau)}$. If we consider a case where $\vec{x}_i(t)$ is perturbed by an one-month forcing $\delta\vec{f}'$, the mean evolution becomes $A_\tau(\vec{x}_i + \delta\vec{f}') = \overline{\vec{y}'_i(t+\tau)}$. We can approximate the effect of the forcing at lag τ as

$$\begin{aligned} \overline{\delta\vec{y}_i(t+\tau)} &= \overline{\vec{y}'_i(t+\tau)} - \overline{\vec{y}_i(t+\tau)} \\ &= A_\tau(\vec{x}_i(t) + \delta\vec{f}') - A_\tau(\vec{x}_i(t)) \end{aligned} \quad (3)$$

We assume that the time-mean climate response is equivalent to the mean response across many different initial conditions $\vec{x}_i$ (ergodicity). Thus, we can compute the climate mean lag-τ response to a forcing $\delta\vec{f}(t)$ by averaging over N samples of internal variability $\vec{x}_i$. We then integrate the average lag-τ responses from $\tau = 0$ to some upper limit $\tau = T_{max}$, where the response to a perturbation approximately converges to noise (we choose 60 months), to obtain the climate mean response:

$$\langle \delta\vec{y}(t) \rangle = \sum_{\tau=0}^{T_{max}} \frac{1}{N} \sum_{i=0}^{N} \Big(A_\tau(\vec{x}_i + \delta\vec{f}(t-\tau)) - A_\tau(\vec{x}_i) \Big) \quad (4)$$

This allows us to replace the linear response function of classic FDT with a non-linear AiBEDO response function. Note, we assume there are no non-linearities between the AiBEDO responses at different lags (i.e., that the effect of $\delta\vec{f}(t)$ is not affected by the changes induced by $\delta\vec{f}(t-1)$, $\delta\vec{f}(t-2)$, etc).

In this study, we discuss the model architecture and training data used to construct this novel AI-based approach to FDT, which allows us to leverage existing climate model data to project climate responses to external forcing. We evaluate the performance of AiBEDO when emulating climate noise, and present a comparison of the AiBEDO response to MCB-like perturbations to the responses in the fully-coupled ESM. Finally, we propose strategies for estimating uncertainties in the AiBEDO response and exploring the possibility space of MCB intervention scenarios using AiBEDO. By assessing a wide range of MCB scenarios on a scale not possible with ESM experimentation, we aim to determine optimal scenarios to avoid crossing potential tipping points and rule out scenarios with undesirable impacts on tipping points.

Methods

Model Architecture

Here, we describe the generation of the AiBEDO operator A_τ to map input radiative flux anomalies at time t (input: $\vec{x}_i(t) \in \mathbb{R}^{d \times c_{in}}$) to corresponding output surface climate variable anomalies after a time lag τ (output: $\overline{\vec{y}_i(t+\tau)} \in \mathbb{R}^{d \times c_{out}}$). To tackle this, we formulate the problem as a pixel-wise regression problem, learning a mapping from input fields to output fields, $A_\tau : \mathbb{R}^{d \times c_{in}} \rightarrow \mathbb{R}^{d \times c_{out}}$,where d is the dimension of the data, and c_{in} and c_{out} are input and output channels, respectively, that are comprised of climate variables (listed in Table 1). To train the model A_τ, we minimize L_{mse}, the pixel-wise mean squared loss between the estimated climate response output $\widehat{\vec{y}}$ and the ground-truth climate response $\vec{y}$, averaged over all dimensions of output:

$$\mathcal{L}_{mse} = \frac{1}{c_{out} d} \sum_t \|\widehat{y}_{t+l} - y_{t+l}\|_2^2, \quad (5)$$

Spherical Sampling The ESM data we use here is originally on a regular latitude-longitude grid, which is difficult to utilize for training purposes due to the large differences in grid areas between points near the equator versus those at the poles. Specifically, it is challenging to accurately depict the Earth's rotational symmetry through the use of two-dimensional meshes, leading to inaccurate representations of significant climate patterns in ML models that assume a two-dimensional format of data. For this reason, we utilize a geodesy-aware spherical sampling that converts the 2D latitude longitude grid to a spherical icosahedral mesh. Icosahedral grids are specified at the lowest resolution by defining

twenty equilateral triangles to form a convex polygon, called an icosahedron. The vertices of the icosahedron are equally spaced points on the sphere that circumscribes it. The resolution of the mesh can be increased by dividing each edge of the icosahedron in half and projecting each new point onto the circumscribed sphere. By resampling in this manner, we are able to iteratively increase the resolution on the sphere. Here, we perform bilinear interpolation (non-conservative) from 2-D climate data to a level-5 icosahedral grid whose vertices define a 1-D vector of length d =10242 with a nominal resolution of $\sim$220 km.

Machine Learning Method In this work, we utilize a Multi-Layer Perceptron (MLP) model. MLP models have proven to be effective for spatio-temporal modeling of ESM data (Park, Yoo, and Nadiga 2019; Wang et al. 2014). MLP is a representative structure of Deep Neural Networks (DNNs) in which an input and an output layer are inter-connected with multiple hidden layers. Each node in a given layer is fully connected with all nodes in the previous layer. The connection between any two nodes represents a weighted value that passes through the connection signal between them. A non-linear activation function is used in each node to represent non-linear correlation in the connection between nodes. The operation between consecutive layers is defined as multiplication between nodes in previous layer and corresponding weight parameters, and applying activation function. Here, we use MLP with 4-hidden layers and 1024 nodes in each layer with layer normalization (Ba, Kiros, and Hinton 2016). We use Gaussian error linear units (Gelu) activation in each layer (Hendrycks and Gimpel 2016). We combine MLP with the spherical sampling approach to create an S-MLP architecture to generate A_τ. A schematic of our S-MLP model architecture is shown in Figure 1.

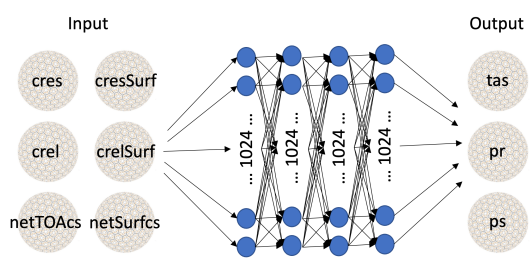

Figure 1: Schematic view of the Spherical Multi-Layer Perceptron (S-MLP) model used in this study.

Training Data

Because the signal-to-noise ratio in short-term climate fluctuations is small, FDT requires a large amount of training data. We use a subset of the Community Earth System Model 2 Large Ensemble (CESM2-LE) as a source of internal climate variations (Rodgers et al. 2021) (Table 2), specifically the 50 ensemble members in which historical simulations are forced with smoothed biomass burning emissions between 1997 and 2014. Each of these 50 ensemble members is forced identically, but is initialized with different initial conditions, meaning that individual members differ only in the chaotic fluctuations internal to the climate system. As such, the CESM2-LE is one of the largest data sets of single-ESM CMIP6-generation simulations for training and testing our model, as it provides a total of nearly 100,000 months of data.

We use a set of six input variables and three output variables. These variables are listed in Table 1. The data are preprocessed by subtracting the ensemble mean of the LE at each grid point, month, and year of the historical time series. This removes both the seasonal cycle and long term secular trends in the data, leaving only monthly fluctuations internal to the system. We then bilinearly remap the data from the original 2D latitude-longitude ESM grid to the spherical icosahedral grid for use by the AI model using Climate Data Operators (cdo, Schulzweida 2022).

Validation Dataset

To validate AiBEDO's ability to plausibly model the climate response to MCB-like perturbations, we compare the AiBEDO response to responses from a novel set of fully dynamic, coupled CESM2 simulations (Hirasawa et al. Submitted). These simulations are summarized in Table 2. MCB forcing is imposed by increasing in-cloud liquid cloud droplet number concentrations to 600cm^{-3} within three selected regions in the northeast Pacific, southeast Pacific, and southeast Atlantic, together and separately in SSP2-4.5 simulations (Shared Socioeconomic Pathway 2, 4.5Wm^{-2} forcing). The effect of MCB is then calculated by taking the difference between the perturbed simulations and the baseline SSP2-4.5 simulations. In addition to the coupled CESM2 simulations, we have conducted "fixed-sea surface temperature" (fixed SST) simulations, wherein the MCB-like forcing is imposed in the model with SSTs held to climatological values. These are used to calculate the effective radiative forcing (ERF) due to the MCB forcing (Forster et al. 2016). AiBEDO is perturbed ($\delta\vec{f}$) with the annual mean `cres`, `crel`, `cresSurf`, `crelSurf`, `netTOAcs`, and `netSurfcs` anomaly fields from the year-2000 MCB Perturbed minus year-2000 Control fixed-SST simulations. Thus, we can compare AiBEDO and CESM2 responses to the same MCB ERF. Note that it is crucial that $\delta\vec{f}$ is computed using fixed SST simulations, as radiation anomalies computed this way do not include radiative feedbacks, which are considered to be part of the response rather than the forcing. In principle, the effects of these feedbacks are encoded in the mappings AiBEDO has learned. Thus, using the ERF to perturb AiBEDO avoids "double counting" the effect of the radiative feedbacks.

In order to calculate the response to the radiative perturbations, we first run AiBEDO on 480 randomly sampled months of preprocessed CESM2 internal variability radiation anomalies to obtain a control ensemble of AiBEDO outputs. Then, we run AiBEDO on the same 480-month sample, but with the MCB radiation perturbations added to the input fields, to obtain a perturbed ensemble of AiBEDO

Variable	Description	Role in AiBEDO
cres	Net TOA shortwave cloud radiative effect	input
crel	Net TOA longwave cloud radiative effect	input
cresSurf	Net Surface shortwave cloud radiative effect	input
crelSurf	Net Surface longwave cloud radiative effect	input
netTOAcs	Net TOA clear-sky radiative flux	input
netSurfcs	Net surface clear-sky radiative flux plus all-sky surface heat flux	input
lsMask	Land fraction	input
ps	Surface pressure	output
tas	Surface air temperature	output
pr	Precipitation	output

Table 1: Name, description, and use by AiBEDO of variables derived from CESM2 LE historical smoothed biomass burning monthly mean data. Thus, $c_{in} = 7$ channels and $c_{out} = 3$ channels. All radiative and heat fluxes at the surface and top of atmosphere (TOA) are positive down.

Experiment	Role	Forcing	Time span	N
Historical LE	training, testing, validation	historical	1850 - 2015	50
Y2000 Control	perturbation	Year 2000 Fixed SST	1 - 20	N/A
Y2000 MCB Perturbed	perturbation	Year 2000 Fixed SST + MCB in NEP, SEP, and SEA	1 - 10	N/A
SSP2-4.5 LE	response validation	SSP2-4.5	2015 - 2100	17
SSP2-4.5 + ALL MCB	response validation	SSP2-4.5 + MCB in NEP, SEP, and SEA	2015 - 2065	3
SSP2-4.5 + NEP	response validation	SSP2-4.5 + MCB in NEP	2015 - 2065	3
SSP2-4.5 + SEP	response validation	SSP2-4.5 + MCB in SEP	2015 - 2065	3
SSP2-4.5 + SEA	response validation	SSP2-4.5 + MCB in SEA	2015 - 2065	3

Table 2: CESM2 simulations used to train and verify AiBEDO. NEP, SEP, SEA denote regions where 600cm^{-3} CDNC MCB forcing is imposed, where NEP - Northeast Pacific (0 to 30N; 150W to 110W), SEP - Southeast Pacific (30S to 0; 110W to 70W), SEA - Southeast Atlantic (30S to 0; 25W to 15E). Note the fixed SST simulations use constant climatological conditions, so we do not note specific years for these simulations.

outputs. The impact of the MCB perturbations is estimated as the mean difference between the control and perturbed AiBEDO outputs. This is repeated for the different time lags. This methodology ensures that the input anomaly fields in the simulations are not too different from the model training data set. Running AiBEDO with just the radiation perturbations results in artifacts, as the near-zero anomalies outside the perturbation regions are mean the fields are unlike any that the model was trained on.

Model Training and Inference

The decoupled weight decay regularization optimization method, AdamW (Loshchilov and Hutter 2017) was utilized to train our model in an iterative manner. The learning rate was initially set to 2×10^{-4} and exponentially decayed at a rate of 1×10^{-6} per epoch. The model is trained on five CESM2 LE simulations at a time until the mean squared error does not reduce for five successive training epochs (batch size 10), at which point a new set of five simulations is used. This process is repeated until all 50 ensemble members have been used to train the model. We find that the mode can be trained for more epochs for short lags relative to longer lags due to the greater signal-to-noise at shorter lags. Our S-MLP models have $\sim$ 108M trainable parameters, and it takes around 1 minute per single epoch for training. The model inference takes an average of 0.5 seconds per data point to generate a prediction.

Results

Emulation of Climate Noise

We validate the baseline performance of AiBEDO for emulating the connection between input radiative fluxes to the output surface climate variables (i.e. equation 2) for a sample of preprocessed CESM2 data in Fig. 2 a-f. The reference data is from CESM2's CMIP6 contribution, and thus was not included in the training dataset but uses the same ESM and very similar forcing boundary conditions. We evaluate the emulation by computing the time root mean squared error (RMSE) of the AiBEDO output time series with the corresponding lagged CESM2 output time series at each grid point. We find that the RMSE is generally highest in regions where internal variability is also high, such as high tas (Fig. 2a) and ps (Fig. 2e) RMSE at high latitudes and high pr (Fig. 2c) RMSE in the tropics. Thus, we compute the ratio of the RMSE and the CESM2 standard deviation in time: smaller values identify regions where AiBEDO performs best relative to the internal climate noise. This ra-

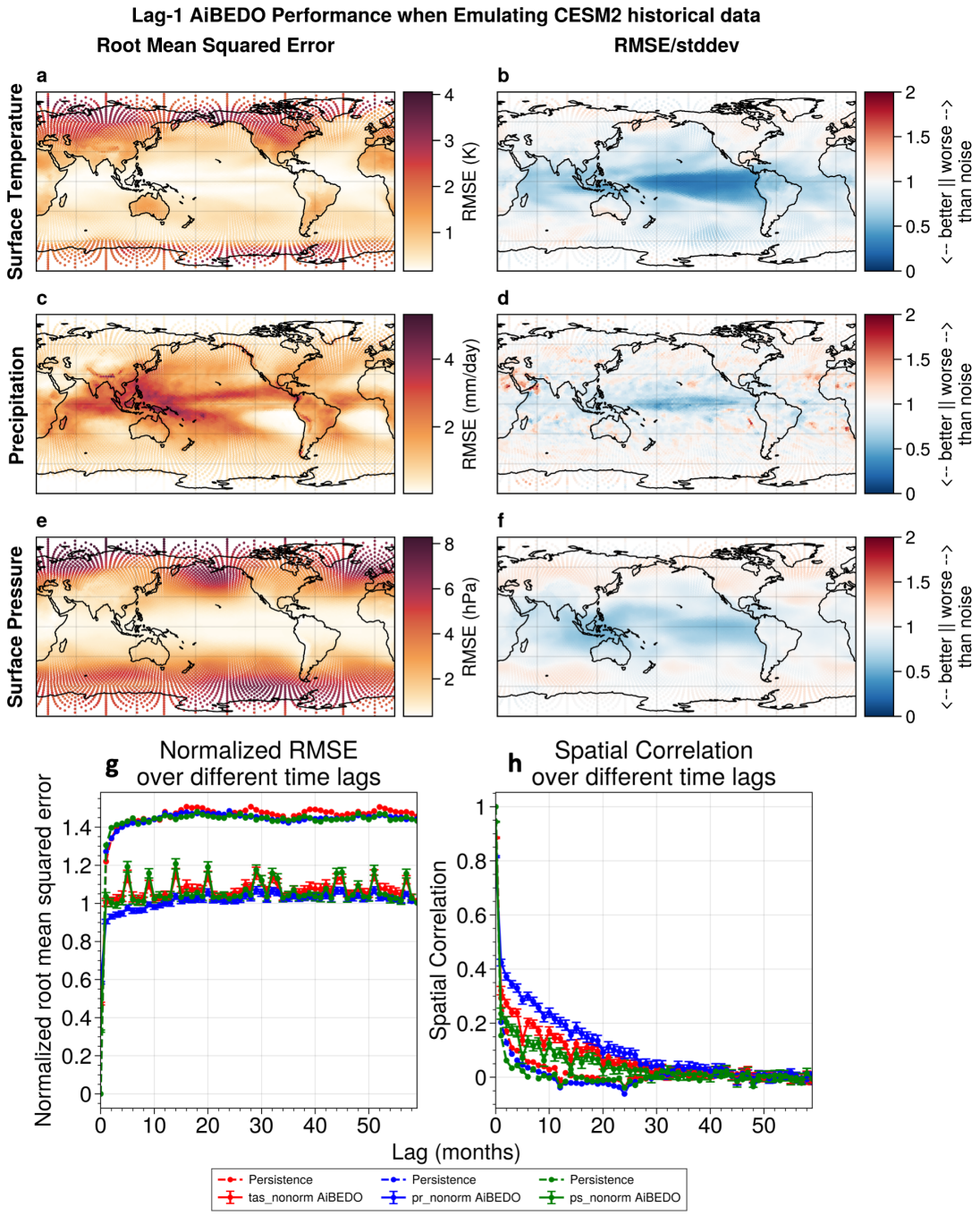

Figure 2: One-month lag AiBEDO compared to CESM2 LE data (a-f). Panels a,c,e show the root mean squared error (RMSE) computed in the time dimension at each icosahedral spherical grid point calculate across 480 months. Panels b,d,f show the ratio of RMSE to the standard deviation of the preprocessed data. Panels a,b show RMSE and RMSE over standard deviation for surface temperature (`tas`), c,d show precipitation (`pr`), and e,f show surface pressure (`ps`). Panels g,h show the normalized RMSE (g) and correlations (h) computed along the spatial dimension in solid lines for the three output variables from different AiBEDO lag models. Dashed lines show the normalized RMSE and correlation computed assuming that the anomaly at month 0 remains the same over time (i.e. the persistence null hypothesis). RMSE here is normalized by the climatological spatial standard deviation of the output variable anomalies.

tio indicates that for all three output variables (Fig. 2b,d,f), AiBEDO performs substantially better in the tropics and subtropics and over oceans, with the tropical Pacific in particular being well represented (this may be a result of the high variance explained by the El Nino-Southern Oscillation). The ratio is near or greater than 1 for much of the mid and high latitudes and over land, especially for `pr`, which may be a consequence of the removal of the seasonal cycle and less direct radiation-surface climate connections in these regions, as surface climate is strongly controlled by synoptic variability in these regions. There are a few regions where AiBEDO has a greater than 1 ratio, such as for `pr` in the parts of the subtropics (Fig. 2d) and for `ps` for the midlatitudes (Fig. 2f).

Fig. 2 g,h shows the spatial RMSE (normalized by the standard deviation) and spatial correlation scores for different versions of AiBEDO trained at different lags respectively. As expected, the predictive skill of the model decreases as lag increases. Notably, we find that the model outperforms persistence consistently across time lags, indicating AiBEDO has learned information beyond the simple memory of 0-month temperature anomalies. We see that AiBEDO performs better than background climate noise even at relatively long time 36-month time lags with best performance for precipitation, followed by temperature and surface pressure. Because the normalized RMSE becomes approximately one after 24 months and the correlation drops to zero at 48-months for all three variables, the 60 lag months upper limit likely captures the response timescales required for the time lag integration.

Response to MCB Perturbations

To validate that AiBEDO can plausibly project climate responses to MCB-like perturbations, we compare the CESM2 coupled model responses to those from the lag-integrated AiBEDO responses (i.e. equation 4) for radiative flux anomalies computed from fixed-SST MCB simulations). To compute the lag integral we use Simpson's rule integration. Fig. 3 shows the CESM2 and AiBEDO responses for the three output variables. We find that AiBEDO is able to reproduce the pattern of climate response to MCB, with correlation scores of 0.78 for `tas`, 0.60 for `pr`, and 0.7 for `ps`. However, there are substantial discrepancies in the magnitude of the responses, with AiBEDO generally projecting larger anomalies than CESM2. This is reflected in the relatively high RMSE when comparing the fields. This magnitude discrepancy may be a result of the missing seasonal information in the training data, which has been noted as a potential source of error in past applications of FDT (Majda and Wang 2010; Gritsun 2010).

Nevertheless, AiBEDO successfully identifies key remote teleconnected responses to the MCB forcing, specifically the La Niña-like `tas` signal in the Pacific, with strong cooling in the tropical Pacific and warming in the midlatitudes east of Asia and Australia, as well as cooling over low-latitude land regions. AiBEDO also reproduces key `pr` changes: it projects drying in northeast Brazil, central Africa, and southern North America and Europe and wetting in west Africa, south and southeast Asia, Australia, and central America. Using these responses, we can estimate the tendency of MCB impacts to affect key regional tipping points. For example, Amazon and West African `pr` changes indicate increased risk of Amazon dieback and West African greening, respectively (Zemp et al. 2017; McKay et al. 2022). The general cooling of the tropical ocean suggests a reduced risk of coral dieoff tipping points. Cooling at high northern latitudes indicates reduced risks to winter sea ice and North American and Eurasian permafrost tipping points.

We also assess the temperature impact of MCB forcing in the individual NEP, SEP, and SEA regions compared to CESM2 simulations with equivalent regional forcing (Fig. 4). We find that AiBEDO performance is weaker when considering these regional perturbations than when all three regions are perturbed together. In particular, AiBEDO's performance declines considerably when projecting the SEA forcing temperatsure response, with a global spatial correlation score of 0.43. AiBEDO correlation scores are better for NEP at 0.63 and best for SEP at 0.73. The weak NEP correlation is due to AiBEDO's too-strong La Niña-like response in the Pacific, possibly indicating that the model over-learns from the El Niño-Southern Oscillation at the expense of other modes of variability. Nevertheless, AiBEDO correctly attributes climate responses to the different forcing regions in several key regions. For example, it correctly identifies that SEP forcing causes La Niña-like cooling and increases in South Asian, West Africa, and Australian rainfall. It also correctly finds that SEA forcing causes tropical Pacific warming and Amazon drying (not shown).

In all four cases, AiBEDO performs better in the tropics relative to higher latitudes and better over oceans (Fig. 4b) than over land (Fig. 4c). Furthermore, we find significant negative correlations, such as at high southern latitudes over Antarctica where the MCB effect is weak and internal variability is strong, which may indicate the model is overfitting to noise in these regions. Thus, the MCB projection performance is highest in the regions where AiBEDO emulation skill is the highest (Fig. 2b), indicating that the ability of the model to correctly project climate responses to MCB forcing is related to its ability to emulate internal variability.

Discussion

In this study, we present a novel framework for rapidly projecting climate responses to forcing by using a non-linear AI model in place of the classical linear response function used in FDT, which we name AiBEDO. AiBEDO is a MLP model with spherical sampling that maps the relationship between monthly-mean radiative flux anomalies and surface climate variable anomalies. The model successfully emulates the connection between variations in radiative fluxes and surface climate variables out to lags of several months. We verify AiBEDO's projections for the case of MCB by comparison to fully coupled CESM2 MCB responses and find that our model is able to skillfully project the pattern of surface temperature, precipitation, and surface pressure response to MCB. We argue the model has sufficient skill to to be useful in estimating the effects of MCB interventions on key regional climate indices that can be used to estimate the impact of interventions on tipping points. For

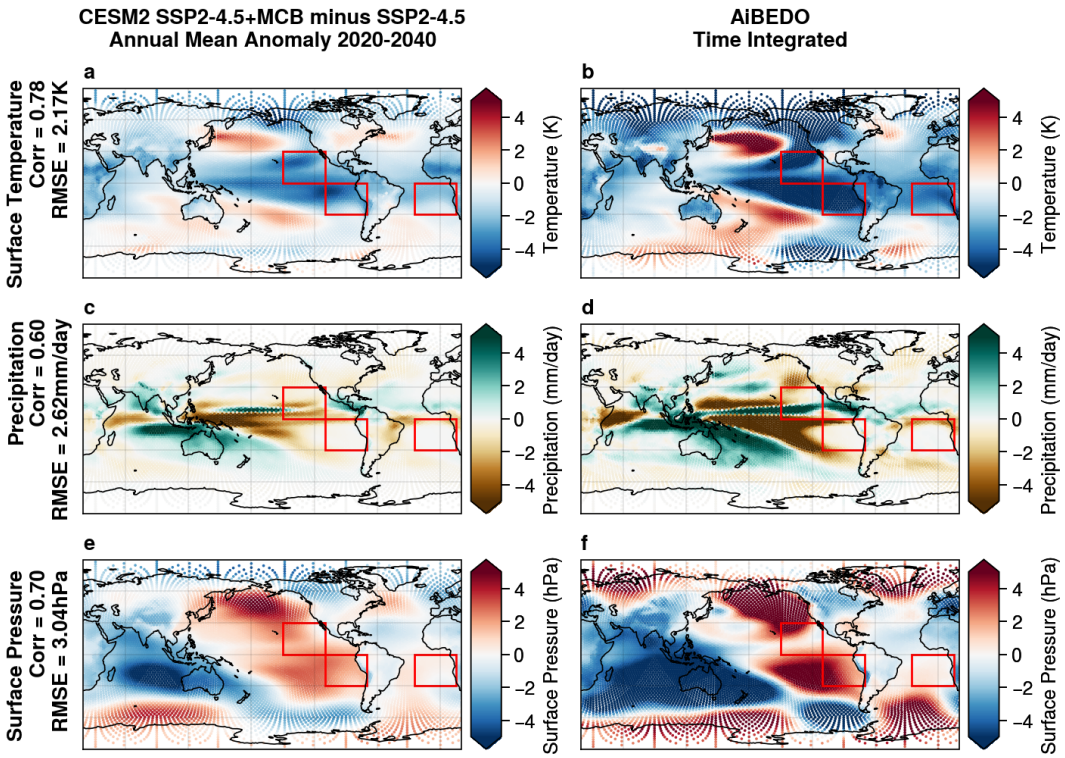

Figure 3: Annual mean temperature (top row - a,b), precipitation (middle row - c,d), and surface pressure (bottom row - e,f) anomalies due to a constant MCB-like forcing for CESM2 (left column) and AiBEDO (right column). Spatial correlation scores and spatial RMSE between the CESM2 and AiBEDO anomalies are displayed in the figure labels on the left side.

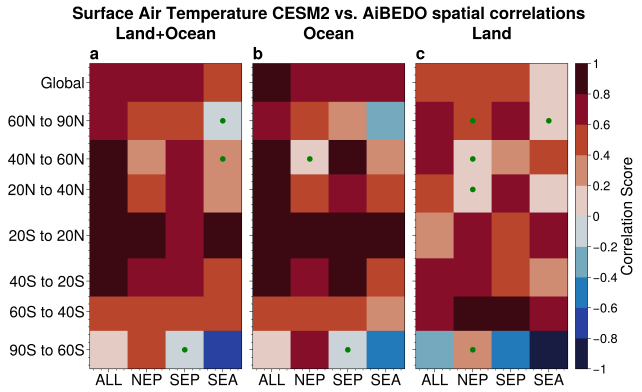

Figure 4: Correlation scores between CESM2 and AiBEDO `tas` responses to MCB forcing for both land and ocean (a), just ocean (b), and just land (c) in different latitude bands for MCB forcing in all regions (ALL), in the northeast Pacific (NEP), in the southeast Pacific (SEP), and in the southeast Atlantic (SEA). Non-significant correlation scores (5 to 95 percent bootstrap resampling range includes 0) are indicated with green dots.

example, AiBEDO projections reproduce rainfall decreases in the southeast Amazon and increases in the Sahel found in reference CESM2 MCB simulations, indicative of increased risks to tipping points associated with Amazon dieback and Sahel greening.

This is a novel application of Fluctuation-Dissipation theory to climate data using AI methods. FDT provides a physical-theory informed framework with which we can compute AI-driven climate projections for cases where no existing forcing-response data exists. We use a generalization of linear FDT with which we can use a non-linear model to generate mean climate responses to radiative flux anomalies. Notably we use a large single-ESM ensemble of climate model data, which is crucial for AiBEDO to successfully learn the mapping between climate variables, particularly as the time lag increases. Thus, large ensembles like the CESM2-LE are vital resources for training models like AiBEDO. This produces a novel AI model that can plausibly project the impact of MCB on climate, opening the possibility of exploring forcing scenarios on a vastly larger scale than is possible with ESMs.

We note that while we have selected radiative flux variables as inputs and surface climate variables as outputs here, in principle AI-FDT can be applied to any set of inputs and outputs for which there is sufficient signal-to-noise for a model to learn. An AiBEDO projection can be generated in $\mathcal{O}(10^2)$ processor-seconds while just one of the

coupled CESM2 MCB simulations we performed here required $\mathcal{O}(10^9)$ processor-seconds. Thus, AI models of this kind have the potential to serve as tools with which existing ESM datasets can be leveraged to generate first look estimates of climate responses to forcing, prior to undertaking computationally expensive new ESM simulations.

Future Work

To provide practical information about climate responses to forcing, it is crucial that we are able to estimate the uncertainty in our projections. Here we only consider the uncertainty due to internal variability in the input data when running AiBEDO, but we must also consider uncertainty due to the underlying training dataset. In particular, because ESMs are only an approximation of the real world, the internal variability and forced response differ between ESMs. In climate modeling, this uncertainty is quantified by considering an ensemble of different ESMs, which is made possible by the Coupled Model Intercomparison Project (CMIP). Thus, we plan to develop an analogous ensemble of AiBEDO models trained on internal fluctuations from different ESMs. Because of the large data requirements of training AiBEDO, we must use single-model initial condition Large Ensembles, of which there exist several from CMIP5 and CMIP6 ESMs (Deser et al. 2020), such as the MPI-ESM1.1 Grand Ensemble (Maher et al. 2019) and the CanESM2 Large Ensemble (Kushner et al. 2018).

Furthermore, though we have verified AiBEDO performance in the response to MCB here (which is largely a shortwave cloud perturbation), AiBEDO includes longwave and clearsky input variables. Thus, AiBEDO may be able to project responses to greenhouse gas and anthropogenic sulphate forcings (both tropospheric pollution and stratospheric injections). We therefore plan to apply AiBEDO to these forcings as well by perturbing the model with ERFs computed from fixed SST simulations with these emissions (Forster et al. 2016).

Using the rapid generation of projections enabled by AiBEDO, we aim to develop a method for optimizing MCB forcing patterns to achieve regional climate targets, drawing from the robust existing body of AI-based optimization methods. This will allows us to explore an array of possible MCB scenarios to find which ones may produce desirable regional outcomes. For example, which MCB forcing pattern might achieve the greatest global mean cooling while minimizing drying in the Amazon? Or which patterns minimize polar amplification? This exploration will accelerate the generation of policy-relevant MCB forcing scenarios and allow estimates of the scenario uncertainty in MCB intervention impacts, which is arguably the largest uncertainty in SRM generally (MacMartin et al. 2022).

Acknowledgements

We thank Linda Hedges of Silver Lining and Brian Dobbins of the National Center for Atmospheric Research for their assistance with AWS and CESM2 computing. The development of AIBEDO is funded under the DARPA AI-assisted Climate Tipping-point Modeling (ACTM) program under award DARPA-PA-21-04-02. AiBEDO training and CESM2 simulations were performed using Amazon Web Services (AWS) computing resources thanks to a generous computing grant from Amazon. We thank the CESM2 Large Ensemble Community Project and supercomputing resources provided by the IBS Center for Climate Physics in South Korea. Documentation of this project can be found at https://aibedo.readthedocs.io/en/latest/index.html.

References

Ba, J. L.; Kiros, J. R.; and Hinton, G. E. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.

Cionni, I.; Visconti, G.; and Sassi, F. 2004. Fluctuation dissipation theorem in a general circulation model: FDT IN A GENERAL CIRCULATION MODEL. *Geophysical Research Letters*, 31(9): n/a–n/a.

Deser, C.; Lehner, F.; Rodgers, K. B.; Ault, T.; Delworth, T. L.; DiNezio, P. N.; Fiore, A.; Frankignoul, C.; Fyfe, J. C.; Horton, D. E.; Kay, J. E.; Knutti, R.; Lovenduski, N. S.; Marotzke, J.; McKinnon, K. A.; Minobe, S.; Randerson, J.; Screen, J. A.; Simpson, I. R.; and Ting, M. 2020. Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, 10(4): 277–286.

Diamond, M. S.; Gettelman, A.; Lebsock, M. D.; McComiskey, A.; Russell, L. M.; Wood, R.; and Feingold, G. 2022. To assess marine cloud brightening's technical feasibility, we need to know what to study—and when to stop. *Proceedings of the National Academy of Sciences*, 119(4): e2118379119.

Forster, P. M.; Richardson, T.; Maycock, A. C.; Smith, C. J.; Samset, B. H.; Myhre, G.; Andrews, T.; Pincus, R.; and Schulz, M. 2016. Recommendations for diagnosing effective radiative forcing from climate models for CMIP6: RECOMMENDED EFFECTIVE RADIATIVE FORCING. *Journal of Geophysical Research: Atmospheres*, 121(20): 12,460–12,475.

Gritsun, A.; and Branstator, G. 2007. Climate Response Using a Three-Dimensional Operator Based on the Fluctuation–Dissipation Theorem. *Journal of the Atmospheric Sciences*, 64(7): 2558–2575.

Gritsun, A. S. 2010. Construction of response operators to small external forcings for atmospheric general circulation models with time periodic right-hand sides. *Izvestiya, Atmospheric and Oceanic Physics*, 46(6): 748–756.

Hendrycks, D.; and Gimpel, K. 2016. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

Hirasawa, H.; Hingmire, D.; Singh, H.; Rasch, P. J.; Kim, S.; Hazarika, S.; Mitra, P.; and Ramea, K. Submitted. Impact of Regional Marine Cloud Brightening Interventions on Climate Tipping Points. *Authorea Preprints for Geophysical Research Letters*.

Jones, A.; Haywood, J.; and Boucher, O. 2009. Climate impacts of geoengineering marine stratocumulus clouds. *Journal of Geophysical Research: Atmospheres*, 114(D10): 2008JD011450.

Kubo, R. 1966. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1): 255.

Kushner, P. J.; Mudryk, L. R.; Merryfield, W.; Ambadan, J. T.; Berg, A.; Bichet, A.; Brown, R.; Derksen, C.; Déry, S. J.; Dirkson, A.; Flato, G.; Fletcher, C. G.; Fyfe, J. C.; Gillett, N.; Haas, C.; Howell, S.; Laliberté, F.; McCusker, K.; Sigmond, M.; Sospedra-Alfonso, R.; Tandon, N. F.; Thackeray, C.; Tremblay, B.; and Zwiers, F. W. 2018. Canadian snow and sea ice: assessment of snow, sea ice, and related climate processes in Canada's Earth system model and climate-prediction system. *The Cryosphere*, 12(4): 1137–1156.

Latham, J.; Bower, K.; Choularton, T.; Coe, H.; Connolly, P.; Cooper, G.; Craft, T.; Foster, J.; Gadian, A.; Galbraith, L.; Iacovides, H.; Johnston, D.; Launder, B.; Leslie, B.; Meyer, J.; Neukermans, A.; Ormond, B.; Parkes, B.; Rasch, P.; Rush, J.; Salter, S.; Stevenson, T.; Wang, H.; Wang, Q.; and Wood, R. 2012. Marine cloud brightening. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1974): 4217–4262.

Leith, C. 1975. Climate Response and Fluctuation Dissipation. *Journal of the Atmospheric Sciences*, 32: 2022–2026.

Liu, F.; Lu, J.; Garuba, O.; Leung, L. R.; Luo, Y.; and Wan, X. 2018. Sensitivity of Surface Temperature to Oceanic Forcing via q-Flux Green's Function Experiments. Part I: Linear Response Function. *Journal of Climate*, 31(9): 3625–3641.

Loshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

MacMartin, D. G.; Visioni, D.; Kravitz, B.; Richter, J.; Felgenhauer, T.; Lee, W. R.; Morrow, D. R.; Parson, E. A.; and Sugiyama, M. 2022. Scenarios for modeling solar radiation modification. *Proceedings of the National Academy of Sciences*, 119(33): e2202230119.

Maher, N.; Milinski, S.; Suarez-Gutierrez, L.; Botzet, M.; Dobrynin, M.; Kornblueh, L.; Kröger, J.; Takano, Y.; Ghosh, R.; Hedemann, C.; Li, C.; Li, H.; Manzini, E.; Notz, D.; Putrasahan, D.; Boysen, L.; Claussen, M.; Ilyina, T.; Olonscheck, D.; Raddatz, T.; Stevens, B.; and Marotzke, J. 2019. The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. *Journal of Advances in Modeling Earth Systems*, 11(7): 2050–2069.

Majda, A.; and Wang, X. 2010. Linear response theory for statistical ensembles in complex systems with time-periodic forcing. *Communications in Mathematical Sciences*, 8(1): 145–172.

Majda, A. J.; Abramov, R.; and Gershgorin, B. 2010. High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability. *Proceedings of the National Academy of Sciences*, 107(2): 581–586.

McKay, D. I. A.; Staal, A.; Abrams, J. F.; Winkelmann, R.; Sakschewski, B.; Loriani, S.; Fetzer, I.; Cornell, S. E.; Rockström, J.; and Lenton, T. M. 2022. Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science*, 377: 6611.

Park, J. H.; Yoo, S.; and Nadiga, B. 2019. Machine learning climate variability. In *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada*, 8–14.

Rasch, P. J.; Latham, J.; and Chen, C.-C. J. 2009. Geoengineering by cloud seeding: influence on sea ice and climate system. *Environmental Research Letters*, 4(4): 045112.

Rodgers, K. B.; Lee, S.-S.; Rosenbloom, N.; Timmermann, A.; Danabasoglu, G.; Deser, C.; Edwards, J.; Kim, J.-E.; Simpson, I. R.; Stein, K.; Stuecker, M. F.; Yamaguchi, R.; Bódai, T.; Chung, E.-S.; Huang, L.; Kim, W. M.; Lamarque, J.-F.; Lombardozzi, D. L.; Wieder, W. R.; and Yeager, S. G. 2021. Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, 12(4): 1393–1411.

Schulzweida, U. 2022. CDO User Guide (2.1.0).

Stjern, C. W.; Muri, H.; Ahlm, L.; Boucher, O.; Cole, J. N. S.; Ji, D.; Jones, A.; Haywood, J.; Kravitz, B.; Lenton, A.; Moore, J. C.; Niemeier, U.; Phipps, S. J.; Schmidt, H.; Watanabe, S.; and Kristjánsson, J. E. 2018. Response to marine cloud brightening in a multi-model ensemble. *Atmospheric Chemistry and Physics*, 18(2): 621–634.

Wang, J.; Zhang, W.; Wang, J.; Han, T.; and Kong, L. 2014. A novel hybrid approach for wind speed prediction. *Information Sciences*, 273: 304–318.

Zemp, D. C.; Schleussner, C.-F.; Barbosa, H. M. J.; Hirota, M.; Montade, V.; Sampaio, G.; Staal, A.; Wang-Erlandsson, L.; and Rammig, A. 2017. Self-amplified Amazon forest loss due to vegetation-atmosphere feedbacks. *Nature Communications*, 8(1): 14681.