Climate Modeling with Spherical Fourier Neural Operators for Stratospheric Aerosol Injections

Nick Masi [*1] **Mason Lee** [*1]

Abstract

We develop a Spherical Fourier Neural Operator (SFNO) emulator for stratospheric aerosol injection (SAI) scenarios, leveraging the ARISE-SAI-1.5 dataset. By incorporating stratospheric aerosol optical depth as a forcing variable, our model captures key climate responses to SAI while maintaining long-term stability. We then show that predictions can be improved by pretraining the SFNO encoder on masked image modeling of climate data as a pretext task. These results indicate that SFNO-based emulators can effectively approximate computationally expensive climate simulations for geoengineering analyses.

1. Introduction

"The global environment crisis is, as we say in Tennessee, real as rain" — Al Gore

Global mean temperatures continue to rise due to anthropogenic greenhouse gas emissions, with projections indicating we may exceed the 1.5°C warming threshold set by the Paris Agreement within decades. Stratospheric Aerosol Injection (SAI) has emerged as a proposed climate intervention strategy that could potentially help limit temperature rise by increasing Earth's albedo. This approach mimics the temporary global cooling effects observed after major volcanic eruptions, which inject sulfate aerosols into the stratosphere. The ARISE-SAI-1.5 project provides a comprehensive modeling framework to study the impacts of SAI implementation using the Community Earth System Model version 2 (CESM2). This dataset simulates the injection of SO_2 at specific latitudes and altitudes, which forms reflective sulfate aerosols in the stratosphere. The simulations show how a feedback control system could theoretically maintain global mean temperature near 1.5°C above pre-industrial levels through strategic aerosol injection (Richter et al., 2022).

However, studying the full range of SAI scenarios and their impacts remains computationally intensive. Climate models like CESM2 require significant computational resources and time to simulate even a single scenario. This limits our ability to explore the parameter space of different injection strategies and analyze the associated uncertainties. Machine learning approaches, particularly those designed for spherical data, offer a promising path forward by enabling rapid emulation of climate model behavior while maintaining physical consistency. In this work, we leverage the Spherical Fourier Neural Operator (SFNO) architecture to develop an emulator of the ARISE-SAI-1.5 simulations. Our emulator can efficiently predict the evolution of key atmospheric variables under different SAI scenarios while capturing both large-scale circulation patterns and local climate impacts. Building on recent advances in ML-based climate modeling (Watt-Meyer et al., 2023), we develop methods to handle SAI-specific variables and ensure long-term stability of predictions. Our code is available at `https://github.com/N-Masi/sfno-sai`.

2. Background

Machine learning approaches for climate modeling have advanced significantly in recent years. Early work focused on direct emulation of physics-based models using convolutional neural networks, but these approaches struggled with long-term stability and physical consistency. More recent architectures like FourCastNet (Pathak et al., 2022) and GraphCast (Lam et al., 2022) have shown promising results for weather prediction up to 2 weeks, but extending ML models to climate timescales remains challenging. Key requirements include conservation of physical quantities, stability over multi-year predictions, and appropriate handling of external forcings.

2.1. ARISE-SAI-1.5 Dataset

The ARISE-SAI-1.5 dataset provides a comprehensive framework for studying Stratospheric Aerosol Injection (SAI) using the CESM2(WACCM6) model (Richter et al.,

[*]Equal contribution [1]Department of Computer Science, Brown University, Providence, RI, USA. Correspondence to: Nick Masi <nmasi@cs.brown.edu>, Mason Lee <mason_lee@brown.edu>.

Submitted to the Brown Self Supervised Learning Workshop (BSSL).

2022). This dataset simulates SAI by injecting SO_2 at four specific latitudes (15°S, 15°N, 30°S, 30°N) at approximately 21.5 km altitude. The simulations span 35 years (2035-2069) and use a controller algorithm to place injections that aim to maintain global mean temperature at 1.5°C above pre-industrial levels.

The variables we consider from this dataset are described in Table 3. All variables are resolved at 192×288 grid points across the globe (*i.e.*, 192 latitudes & 288 longitudes). For reasons described in subsection G.4, we look at month-scale "snapshots" of these variables (*i.e.*, each step is a month). These can be thought of as 192×288 "images," but this should not be confused to mean that the data is visual data collected from, for example, a satellite. The variables are measurements of physical atmospheric properties; the units of these variables are included in Table 3. To prevent the different scale and magnitude of certain variables dominating the loss function, snapshots are normalized (see Appendix F). Each snapshot has a "channel" for each variable at each vertical level it is resolved at. Some variables have no variance across altitude (*e.g.*, sea surface temperature), but others do (*e.g.*, air temperature). We describe our sampling approach for vertical levels in subsection E.2.

Variables are organized into three key categories: **Prognostic Variables**, which represent state variables such as temperature and winds that evolve according to physical equations and are used as both inputs and outputs; **Forcing Variables**, which include external factors like SO_2 injection rates and solar radiation that drive the evolution of the system; and **Diagnostic Variables**, which are derived quantities, such as precipitation and energy fluxes, computed based on the system's state. For next-timestep prediction, the model inputs are forcing and prognostic variables, and the outputs are prognostic and diagnostic variables.

The dataset consists of a 10-member ensemble, each starting from slightly different initial conditions to sample internal climate variability. The initial conditions of members 006–010 are slight perturbations of those in 001–005, respectively. The Amazon S3 buckets for runs 008 and 009 were unavailable. Because of the dependence between a simulation 00{X} and 00{X+5}, we hold out the two runs without corresponding simulations, 003 and 004, for validation and testing, respectively. Thus, we train on ensemble members 001, 002, 005-007, and 010. Following (Watt-Meyer et al., 2023), we formulate our training objective using a geometric loss function, which approximates the L^2 norm on the sphere:

$$\mathcal{L}(F_\theta[u_n], u_{n+1}) = \sum_{c \in \text{channels}} \left(\frac{\sum_{i \in \text{grid}} |F_\theta[u_n](x_i) - u_{n+1}(x_i)|^2}{\sum_{i \in \text{grid}} |u_{n+1}(x_i)|^2} \right)^{1/2} \quad (1)$$

This loss function computes the relative L^2 error for each channel by normalizing the prediction error with the magnitude of the target. The overall loss is the mean of these relative errors across all channels, promoting balanced performance across different climate variables.

2.2. Spherical Fourier Neural Operators

The Spherical Fourier Neural Operator (SFNO) is a generalization of the Fourier Neural Operator (FNO) architecture. It replaces the Fourier transform used within the model with spherical harmonic transforms (SHT). SHTs are a generalization of the Fourier transform to the spherical domain (Mohlenkamp, 1999; Swarztrauber & Spotz, 2000). SFNOs were first introduced by Bonev et al. (2023), who found it improved climate predictions over FNOs given Earth's spherical geometry.

Watt-Meyer et al. (2023) showed that SFNOs can be used for accurate long-term (at the scale of tens of years) predictions of weather and climate. This work also showed that SFNOs can perform well with all different type of climate variables: prognostic, diagnostic, and forcings. We follow this approach for having forcing and prognostic variables as the model inputs, while diagnostic and prognostic variables are the predicted outputs.

3. Autoregressive Climate Forecasting

We build on the open-source SFNO implementation for climate modeling by Watt-Meyer et al. (2023), known as ACE, retaining their hyperparameters for consistency. Our data preprocessing follows their two-stage normalization strategy (detailed in Appendix F), which first normalizes each field level-wise and then adjusts for temporal residuals. We take ACE and apply it to the novel task of climate forecasting in the presence of SAI forcings. The goal is to have a network that is more computationally efficient and accessible than the supercomputer simulations of ARISE, but that can still perform accurate autoregressive rollout of climate on the scale of decades. To do this, we train the model to make next-timestep predictions of the climate for the next month based on the current monthly climate snapshot. This is a self-supervised task in the sense that the label signal is derived from the timeline of climate data, which is a single ensemble member (*i.e.*, simulation run).

We train and validate a number of ablations of ACE by changing hyperparameters and the set of variables used. We only test (*i.e.*, find the loss from autoregressive rollout on the held out test ensemble member) the model variants without SST included as a forcing. For one, our validation showed that regardless of model variant, performance decreased when it was included as input (see Figure 2). But fundamentally, a forcing is something that is known so that it does not

need to be modeled. SAI is a forcing because it is controlled and produced by humans. Future SST states are not strictly known when forecasting decades ahead. The results from testing the different model variants are shown in Table 1. The standard ACE model is found to perform best.

Model variant	Autoregressive Rollout Performance
base ACE	**0.4089**
no AOD	0.5629
dropout 0.1	0.5008
dropout 0.2	0.5156
$sf = 16$	1.08256

Table 1. Loss at the final step of the autoregressive rollout after 419 months. Bold indicates the best performing model.

The different ablations we performed to get the different model variants are described in subsection 3.1. The decrease in performance associated with the introduction of stochasticity into ACE is consistent with the findings of Cachay et al. (2024). The loss of best model's autoregressively-predicted month snapshot roughly stabilizes about 7 years into the rollout, hovering around the ≈ 0.4 loss reported. For the first year it has loss of ≈ 0.25, and then the loss steadily increases until the relative stabilization. The full progression of loss can be seen in Appendix H as Figure 3.

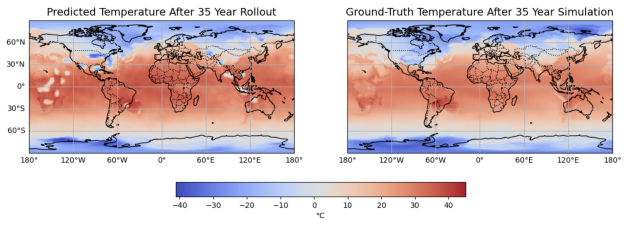

Figure 1. Comparison of surface temperature (TS) across the globe after 35 years of simulation, starting with the initial conditions in ensemble member 004. Left is after 35 years of autoregressive, monthly rollout from the best forecasting model (ACE with AOD without SST). Right is the gold-standard from the ARISE-SAI-1.5 supercomputer simulation.

3.1. Ablation Experiments

To systematically isolate the impacts of our forcing variables and hyperparameter settings, we conducted a set of focused ablation experiments. All ablations used the same monthly next-timestep prediction task.

No AOD. We omitted stratospheric aerosol optical depth (AOD) to determine whether the network could infer aerosol forcing purely from prognostic variables like temperature, winds, and humidity. AOD ordinarily provides a direct measure of aerosol-induced radiative perturbations; its removal forces the model to learn any aerosol signal implicitly, revealing how crucial explicit AOD forcing is for modeling SAI impacts.

Dropout. Dropout is applied to the SFNO's (i) encoder and (ii) MLPs (which are within the Fourier layers). This is the same way that Cachay et al. (Cachay et al., 2024) introduce stochasticity into ACE. They term this model variant ACE-STO.

Scale Factor ($sf = 16$). We modified the SFNO's `scale_factor` to 16, reducing the number of spherical harmonic modes retained. This ablation tests the sensitivity of climate emulation to frequency cutoff constraints, as a higher `scale_factor` places greater emphasis on low-frequency modes and potentially diminishes the network's ability to resolve fine-scale aerosol patterns. Each variant was trained for the same number of epochs, using the same loss function (Equation 1) and optimizer settings, to isolate differences arising solely from the ablation. The training script and hyperparameter configuration are provided in the Appendix.

4. Masked Image Modeling

To improve the next-timestep prediction in section 3, we implement masked image modeling (MIM) as a self-supervised pre-training strategy for climate representation learning. To develop our approach we consulted Hondru et al. (2024). In MIM, our model uses forcings and prognostic variables as input, similar to the experiments in section 3, but also outputs these same variables. That is, forcings are included as an output and diagnostics are not. This is because the MIM task fundamentally aims to reconstruct the input as the output. This allows the model to develop meaningful representations of climate snapshots in the latent space. Forcing variables cannot be excluded because then the encoder would have an incorrect dimensionality when transferred to the ACE model.

For all MIM tasks we use the ACE model with AOD and without SST as forcings, as this performed best on the validation set (see Figure 2). It also happened to be the best performing model on the training set as reported section 3.

4.1. Climate Data Reconstruction

We adopt the SimMIM framework's masking strategy (Xie et al., 2022): randomly masking non-overlapping patches such that a certain ratio of the full image is masked. To perform the visible signal reconstruction task, loss is calculated as the ℓ_1 loss of only the masked patches. Following Pathak et al. (2016), we implement masking by zeroing normalized

variable values rather than using tokenizing the image—an approach better suited for continuous climate fields.

Our implementation explores two critical masking variants: **Consistent masking** which applies identical mask patterns across all variable channels, and **Channel-wise masking** which randomly masks different patches across each channel while maintaining consistent masking ratios. This dual approach allows us to evaluate whether maintaining spatial consistency across variables improves the learned representations for climate modeling tasks.

We furthermore consider two more methods for ablation: the patch size (16 or 32) and masking ratio (0.3 or 0.5). Because this is a pretraining task, a comparison of performance on this task alone is not germane. The validation loss of all 8 approaches is reported in Appendix I, and all 8 pretrained encoders are assessed when transferred back to the autoregressive task in subsection 4.2.

4.2. Downstream Impact

We take the encoder from the SFNO of all 8 ablations discussed in subsection 4.1 and transfer them (unfrozen) to a new ACE model that gets finetuned on the next-timestep prediction task. The results are detailed in Table 2. They indicate that pretraining on the MIM task can improve SFNO performance on simulating climate in the presence of SAI forcings. However, not every masking strategy will confer benefits. While the MIM work we drew upon found optimal performance with patch sizes of 32×32 and a masking ratio of 0.5, climate modeling with SAI present got worse under that masking strategy. However, climate modeling was improved with smaller patches and less dropout. In general, we did not observe a significant trend in the impact of consistent *vs.* channel-wise masking.

	Masking	Patch Size	Dropout Rate	Rollout Loss
Masking Strategy	Consistent	16×16	0.3	**0.406073**
			0.5	0.424582
		32×32	0.3	0.426842
			0.5	0.414779
	Channel-wise	16×16	0.3	**0.408463**
			0.5	0.423820
		32×32	0.3	0.426626
			0.5	0.415463

Table 2. Loss at the final step of the autoregressive rollout after 419 months. The best performing hyperparameters for ACE found in section 3 were used. Encoders trained for image reconstruction under the 8 different masking strategies in this table were transferred to ACE to produce these results. Bold indicates better-than-baseline (*i.e.*, ACE) performance.

5. Future Work

Oommen et al. (2024) introduce a method to better attend towards the high frequency modes in FNOs. Because in section 3 we found that performance decreases in the SAI case when filtering out higher frequencies, we suspect that SAI has a high frequency effect due to its subtle nature. We think the techniques in (Oommen et al., 2024) could thus improve our results if integrated with ACE. Spherical DYffusion (Cachay et al., 2024) is a novel framework to improve the long term autoregressive capabilities of SFNO models by introducing a second SFNO which acts as an "interpolator." This is another method which could improve climate modeling in the SAI case. It was shown to help

To further ground and justify the autoregressive rollout capabilities of the model in general, our approach should be assessed in the unperturbed cases, *i.e.*, with historical data.

To prove that the benefits of pretraining on MIM aren't simply from more training in general, regardless of task, future work could perform a baseline of further training the encoder on the same next-timestep prediction task. In general, future work should include more training steps as our approach had far fewer than the original ACE paper (Watt-Meyer et al., 2023). Constraints of limited available compute prevented us from being able to do this in this work.

Pathak et al. (2016) show that the loss function we borrowed (subsection 4.1) for image reconstruction has a tendency to produce "blurry" or smoothed-out reconstructions rather than high resolution images. To address this they jointly train a discriminative model to add an adversarial term to the training loss. It is left to future work to see if the same resolution problem exists in masked climate data reconstruction, and whether it would benefit from such a technique.

6. Conclusion

In this paper we showed that Spherical Fourier Neural Operators (SFNOs) can be used for stable, decadal-scale climate modeling in stratospheric aerosol injection (SAI) contexts. We first show this by training an SFNO on next-timestep prediction of monthly climate state. This task is self-supervised, drawing from ARISE-SAI-1.5 climate simulations. We then show that this performance can be improved by pretraining on image reconstruction as another self-supervised pretext task. We explore a number of ablations in model choice and masking strategy that impact the performance of autoregressive rollout.

References

Bonev, B., Kurth, T., Hundt, C., Pathak, J., Baust, M., Kashinath, K., and Anandkumar, A. Spherical fourier neural operators: Learning stable dynamics on the sphere, 2023. URL https://arxiv.org/abs/2306.03838.

Cachay, S. R., Henn, B., Watt-Meyer, O., Bretherton, C. S., and Yu, R. Probabilistic emulation of a global climate model with spherical dyffusion, 2024. URL https://arxiv.org/abs/2406.14798.

Hondru, V., Croitoru, F. A., Minaee, S., Ionescu, R. T., and Sebe, N. Masked image modeling: A survey, 2024. URL https://arxiv.org/abs/2408.06687.

Lam, R. et al. Graphcast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*, 2022.

Mohlenkamp, M. J. A fast transform for spherical harmonics. *Journal of Fourier Analysis and Applications*, 5:159–184, 1999. URL https://doi.org/10.1007/BF01261607.

Oommen, V., Bora, A., Zhang, Z., and Karniadakis, G. E. Integrating neural operators with diffusion models improves spectral representation in turbulence modeling, 2024.

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., and Efros, A. A. Context encoders: Feature learning by inpainting. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544, 2016. doi: 10.1109/CVPR.2016.278. URL https://arxiv.org/abs/1604.07379.

Pathak, J. et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Richter, J. H., Visioni, D., MacMartin, D. G., Bailey, D. A., Rosenbloom, N., Dobbins, B., Lee, W. R., Tye, M., and Lamarque, J.-F. Assessing Responses and Impacts of Solar climate intervention on the Earth system with stratospheric aerosol injection (ARISE-SAI): protocol and initial results from the first simulations. *Geoscientific Model Development*, 15(22): 8221–8243, November 2022. ISSN 1991-959X. doi: 10.5194/gmd-15-8221-2022. URL https://gmd.copernicus.org/articles/15/8221/2022/. Publisher: Copernicus GmbH.

Swarztrauber, P. N. and Spotz, W. F. Generalized discrete spherical harmonic transforms. *J. Comput. Phys.*, 159 (2):213–230, April 2000. ISSN 0021-9991. doi: 10.1006/jcph.2000.6431. URL https://doi.org/10.1006/jcph.2000.6431.

Watt-Meyer, O., Dresdner, G., McGibbon, J., Clark, S. K., Henn, B., Duncan, J., Brenowitz, N. D., Kashinath, K., Pritchard, M. S., Bonev, B., Peters, M. E., and Bretherton, C. S. Ace: A fast, skillful learned global atmospheric model for climate prediction, 2023. URL https://arxiv.org/abs/2310.02074.

Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., and Hu, H. SimMIM: A simple framework for masked image modeling. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. URL https://arxiv.org/abs/2111.09886.

A. Author Contributions

Masi had the idea for MIM, wrote the code for autoregressive and MIM training & testing, gathered all experimental data, created figures, and helped serialize the ARISE data.

Lee had the original idea to apply SFNOs to the task of modeling SAI climate data, added data normalization, helped procure the ARISE data, and provided compute.

B. Acknowledgments

We would like to acknowledge all of the insightful guidance we received from Professor Randall Balestriero throughout the duration of this project. This, in combination with his instruction throughout the semester, was critical to our success. We would additionally like to thank Professor Daniele Visioni for his tremendous guidance and the generosity to lend his domain expertise. Lastly, thank you to all of our classmates in CSCI 2952X.

C. Generative AI Usage

ChatGPT was used to assist in writing code for data normalization as well as formatting some figures. It was also used to help debug at various points. All code was verified by the authors, and no generative AI was used in the process of writing this paper.

D. Appendix Overview

The rest of this appendix provides supplementary materials, methodological details, and additional analyses that support the main findings presented in our paper. The appendix is organized into several sections covering different aspects of our research methodology and results.

E. Variables and Vertical Levels

This section provides detailed information about the variables and vertical levels used in our study.

E.1. Variable Definitions

Table 3 presents a comprehensive list of variables used in this study. With 12 vertical levels, this results in a set of $3 + (4 \cdot 12) + 2 = 53$ input variables (52 when SST is not used) and $3 + (4 \cdot 12) + 3 = 53$ predicted output variables from the model. ARISE symbols are the same as those in WACCM6. Variables with a k subscript indicate they vary over altitude, and thus take different values at the 12 different vertical levels we sample.

E.2. Vertical Level Structure

We adopt and extend the 8 vertical levels used in the original Spherical DYffusion model to better resolve the stratosphere where SAI occurs. Table 4 details our 12-level vertical structure, which includes additional levels in the lower stratosphere. The levels are strategically chosen to cover the altitude range from the surface (1000 hPa) to 0.6 hPa (approximately 80 km), capturing key atmospheric layers relevant to SAI.

F. Data Normalization

Our normalization framework, extending (Watt-Meyer et al., 2023), addresses the multi-scale nature of climate data through a two-stage process optimized for both spatial and temporal coherence.

F.1. Mathematical Framework

Consider a climate field $X_{t,l,y,x}$ indexed by time t, vertical level l, and spatial coordinates (y, x). Our procedure proceeds in two stages:

F.1.1. Stage 1: Full-Field Normalization

Level-wise normalization is given by:

$$X^{ff}_{t,l,y,x} = \frac{X_{t,l,y,x} - \mathbb{E}_{t,y,x}[X_{t,l,y,x}]}{\sqrt{\mathbb{E}_{t,y,x}[(X_{t,l,y,x} - \mu_l)^2]}}. \tag{2}$$

F.1.2. Stage 2: Temporal Scaling

Next, we compute a residual scaling based on temporal differences:

$$\sigma^{ff}_{\Delta,l} = \sqrt{\mathbb{E}_{t,y,x}\Big[\big(X^{ff}_{t+1,l,y,x} - X^{ff}_{t,l,y,x}\big)^2\Big]}. \tag{3}$$

A global reference scale σ_{ref} unifies scaling across variables:

$$\sigma_{ref} = \exp\Big(\frac{1}{N_l}\sum_l \log\big(\sigma^{ff}_{\Delta,l}\big)\Big). \tag{4}$$

The final normalized field is then:

$$X^{norm}_{t,l,y,x} = \frac{X^{ff}_{t,l,y,x}}{\sigma^{ff}_{\Delta,l}/\sigma_{ref}}. \tag{5}$$

F.2. Assumptions

For data normalization, we make the simplifying assumption that variables will have the same distribution in all ensemble members. However, to completely faithfully reproduce the results of (Watt-Meyer et al., 2023), we should normalize variables based on the statistics across all ensemble members rather than normalizing the data of each run based on

Table 3. List of Variables Used in the Study

Variable Symbol	ARISE Field Name	Description	Units	Category
AOD_{550} or AOD	$AODVISstdn$	Signal of the aerosol injections by controller	-	Forcing
T_{SST}	SST	Sea surface temperature	K	Forcing
$DSWRFtoa$	$SOLIN$	Downward shortwave radiative flux at TOA	W/m^2	Forcing
T_k	T	Air temperature at level k	K	Prognostic
q_k	Q	Specific humidity at level k	kg/kg	Prognostic
u_k	U	Zonal wind at level k	m/s	Prognostic
v_k	V	Meridonal wind at level k	m/s	Prognostic
p_s	PS	Surface pressure	hPa	Prognostic
T_s	TS	Surface temperature of land or sea-ice (radiative)	hPa	Prognostic
SHF	$SHFLX$	Surface sensible heat flux	W/m^2	Diagnostic
LHF	$LHFLX$	Surface latent heat flux	W/m^2	Diagnostic
-	$PRECT$	Total (convective and large-scale) precipitation rate	m/s	Diagnostic

Table 4. Vertical Levels Used in the Study

Level k	p_k [hPa]	Approx. Altitude [km]	Description
0	0.6	∼80	Mesosphere
1	10	∼30	Upper stratosphere
2	20	∼26	Mid-stratosphere
3	30	∼24	Lower stratosphere (SAI region)
4	50	∼21	Lower stratosphere (SAI region)
5	70	∼18	Tropopause
6	120	∼15	Upper troposphere
7	200	∼12	Mid-troposphere
8	400	∼7	Lower troposphere
9	600	∼4	Near surface
10	800	∼2	Near surface
11	1000	0	Surface

the statistics of that run alone. However, the amount of memory that would be required to load in all of this data at once to compute these complete statistics was not available.

G. Further Discussion on SAI Variables and Formulation

In this section, we provide additional considerations regarding the choice of SAI variables used in our model and potential pitfalls in capturing the SAI signal accurately.

G.1. Variables Capturing the SAI Signal

A key challenge lies in identifying which atmospheric or surface fields most directly capture the forcing from injected aerosols. While sulfur dioxide (SO_2) concentrations can appear as an attractive candidate, they do not represent the ultimate cause of the climate forcing. Sulfate aerosols formed from SO_2 oxidation—and specifically their column loading or optical depth—are the actual drivers of the radiative forcing. Hence, for many climate-emulator tasks, **stratospheric aerosol optical depth** (AOD) at a specific wavelength (*e.g.*, AODVISstdn at 550 nm) directly encodes the SAI effect on incoming shortwave radiation. Therefore, including AOD as a forcing input variable may better capture the resulting climate perturbation compared to using raw SO_2 concentrations alone.

G.2. Temporal and Vertical Scale Considerations

Because the process of SO_2 converting to sulfate aerosols occurs over timescales on the order of days to weeks, forcing signals stored in daily SO_2 fields may not translate directly into a clear climate forcing if used at monthly or longer prediction horizons. By contrast, AOD (or integrated sulfate burden) better encapsulates the time-averaged effect of SAI aerosols for monthly predictions.

G.3. Potential Confounders

Even when including an AOD variable to capture the SAI forcing, some prognostic fields (such as sea surface temperature, SST) can themselves encode partial information about the radiative forcing history. Models trained to predict next-month climate from a current-month state may discover that certain large-scale circulation or temperature patterns already account for the presence of aerosols, leading the network to learn redundant signals if AOD is also provided. This can occasionally yield worse validation metrics if the model overfits or conflates the roles of forcing and state variables.

Additionally, distinguishing tropospheric aerosols from stratospheric aerosols in the dataset is essential when the focus is explicitly on stratospheric injections. Burden variables that integrate aerosol mass vertically can dilute or mask the specific stratospheric signal unless the region of interest (the lower stratosphere) is carefully isolated.

G.4. Formulation Caveats

Daily vs. Monthly Timescale. Our modeling choices revolve around monthly timesteps, but SAI forcings can fluctuate on shorter timescales (especially if injections are adapted via feedback controllers). A model trained to emulate a month-ahead state may not need daily aerosol data, as short-term weather variability dominates at the daily scale, whereas aerosol-induced radiative forcing emerges more clearly on monthly or seasonal timescales.

Predicting Far-Future States. For longer horizons (multiple years to decades), certain forcing variables (notably SST) are not strictly "known" inputs, as future ocean states would themselves be partly determined by the evolving climate. Treating SST as an external forcing can artificially constrain the system unless the objective is only to emulate a known simulation (rather than produce truly out-of-sample forecasts). If the intent is general predictive modeling, one may need additional coupling mechanisms or different ways to handle unknown future SST fields.

Resolution of Aerosol Processes. Models that wish to emulate the *full* chain from SO_2 injection to aerosol formation must capture the chemistry and microphysics involved. Directly using AOD bypasses those intermediate processes, potentially simplifying the emulator's scope but losing interpretability regarding aerosol composition and distribution. Conversely, omitting AOD in favor of raw SO_2 could fail to capture the correct stratospheric forcing. While direct injection metrics (like SO_2 fluxes) can be helpful diagnostic variables, the emergent SAI forcing is typically more accurately captured by stratospheric aerosol optical depth. Ensuring consistency between these forcing variables and the slower-evolving prognostic fields (e.g. temperature, humidity) remains a key design challenge.

H. Autoregressive Rollout Supplemental Material

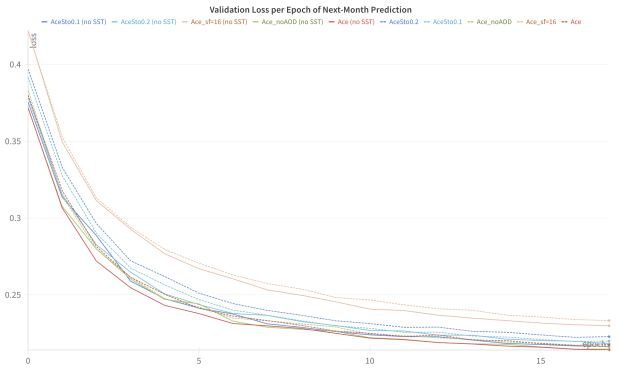

Figure 2. Validation loss of the five ACE variations we consider when they are also trained with SST as a forcing variable. For all variations, SST decreases performance.

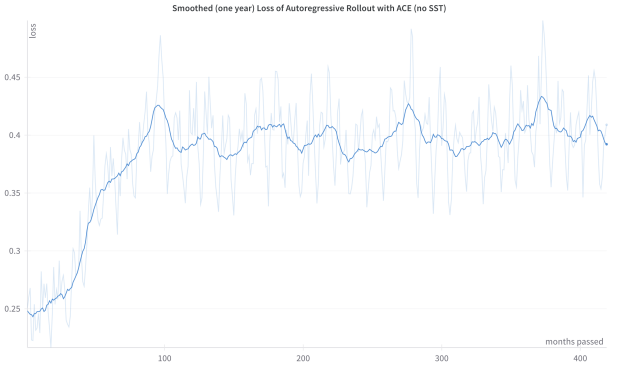

Figure 3. Loss over time as autoregressive rollout progresses using the best model found through validation: ACE without SST as a forcing variable. The light blue line is the actual observed loss, the dark blue line is the running average of the loss over 12 months. Predictions extend 419 months, covering the same 35 year span as the ARISE simulations.

I. Masked Image Modeling Supplemental Material

	Masking	Patch Size	Dropout Rate	Validation Loss on Reconstruction
Masking Strategy	Consistent	16×16	0.3	0.032281
			0.5	0.058225
		32×32	0.3	0.039728
			0.5	0.068489
	Channel-wise	16×16	0.3	0.032267
			0.5	0.058197
		32×32	0.3	0.039725
			0.5	0.068495

Table 5. Validation loss of the ACE model (without SST) on the image reconstruction task (*i.e.*, predicting the masked portions of the 52 input channels) after completing all training epochs. The encoder from these were transferred to ACE on the rollout task.