CAMULATOR VERSION 1: FULLY RESOLVED, FAST EMULATION OF THE COMMUNITY ATMOSPHERE MODEL *

William E. Chapman[‡], John S. Schreck[†], Yingkai Sha[†], David John Gagne II[†], Dhamma Kimpara[†], Laure Zanna[ζ], Judith Berner[‡]

Climate and Global Dynamics (CGD) Laboratory[‡]
Computational and Information Systems (CISL) Laboratory[†]
Courant Institute of Mathematical Sciences, New York University[ζ]

NSF National Center for Atmospheric Research
Boulder, Colorado, USA

{wchapman, schreck, ksha}@ucar.edu

ABSTRACT

We introduce CAMulator, an auto-regressive machine-learned (ML) emulator of the Community Atmosphere Model version 6 (CAM6) that simulates the next atmospheric state given prescribed sea surface temperatures and incoming solar radiation. CAMulator explicitly conserves global dry air mass, moisture, and total atmospheric energy while remaining numerically stable over indefinite climate integrations. It successfully reproduces the annual climatology and key modes of climate variability, including the El Niño–Southern Oscillation, the North Atlantic Oscillation, and the Pacific-North American pattern, though with slightly muted variability. Additionally, CAMulator improves upon some longstanding biases in traditional physics-based models, such as the drizzle problem.

Despite these strengths, when forced with +2K and +4K SST warming experiments, CAMulator exhibits a systematic cold bias in high-latitude regions, particularly in boreal winter, likely due to the absence of interactive land and sea ice. Nonetheless, CAMulator achieves these results with a 500 times speedup over CAM6, making it an efficient alternative for generating large ensembles. CAMulator represents a step toward fast, physically consistent ML-driven climate modeling. Its ability to efficiently emulate CAM6 while maintaining fundamental physical constraints highlights the potential of machine learning in accelerating climate simulations and improving uncertainty quantification.

1 Introduction

Climate models are essential for understanding Earth system dynamics, characterizing extreme events, and projecting future climate scenarios. However, their high computational cost limits their practical applications. One major computational bottleneck of these models is the atmosphere component. Recent advances in machine learning (ML) emulators offer a promising path forward, with the potential to accelerate simulations and, in some cases, improve accuracy. Demonstrating the fidelity of ML-driven emulation is crucial; success in climate model emulation could enhance and enable key modeling activities such as rapid hypothesis testing, large-ensemble generation [1, 2], and comprehensive uncertainty quantification [e.g., 3, 4], ultimately broadening the scope and efficiency of climate research.

Progress in ML-based climate emulation has been rapid, particularly in emulating component models [5, 6, 7] and recently, in coupled systems [8]. The community is beginning to demonstrate that ML-based system models can emulate key features that are required for climate research, such as emergent variability and response to external forcing [7]. However, differences in variable representation, reduced dimensionality, and lower physical resolution—both

* *Citation*: Chapman W. et al. fully resolved, fast emulation of the Community Atmosphere Model

spatially and vertically—pose challenges for integrating these ML-based emulators with other physics-based Earth system components, including ocean, land, cryosphere, high-top atmospheric models, and supermodeling applications [e.g., 9, 10]. These discrepancies hinder seamless coupling with traditional physics-based models, limiting their broader adoption in Earth system simulations.

Here, we introduce CAMulator Version 1, a ML-based emulator that mimics the Community Atmosphere Model version 6 (CAM6) while preserving both vertical and horizontal spatial resolution. CAMulator runs approximately 500 times faster than CAM6 while maintaining key conservation properties, including global dry mass, total water, and energy. In this work, we focus on emulating atmospheric physics during the historical period 1979–2014, a common validation period for CAM6. We conduct this experiment as an Atmospheric Model Intercomparison Project (AMIP) simulation and demonstrate that by initializing a climate state with a weather snapshot, CAMulator can be integrated indefinitely while preserving the statistical properties of the CAM6 climate system. Additionally, we show that CAMulator's response to increases in climate forcing mimics that of CAM6 in the historical record and in some out-of-sample experiments.

For this work, we leverage the NSF National Center for Atmospheric Research (NCAR)-Community Runnable Earth Digital Intelligence Twin (CREDIT) platform [11, 12] for model training and testing. CREDIT is a scientific research platform that provides an efficient framework for the rapid development of auto-regressive models, making it well-suited for Earth system modeling tasks. In our study, CREDIT serves as the backbone for an efficient data pipeline, scalable high-performance computing, model training, and inference. We extend its capabilities from numerical weather prediction emulation to fast climate emulation and demonstrate the necessity of incorporating physical constraints on mass, water, and energy to enhance model fidelity and improve emulation accuracy.

CAMulator successfully reproduces key climate statistics, including low-frequency modes of variability. Although geopotential height is not directly predicted, the model is still able to accurately capture major modes of variability, such as the North Atlantic Oscillation and the Pacific North American pattern, from predicted variables. This suggests that the model preserves physical consistency across key atmospheric fields, even in extended simulations. Additionally, CAMulator mitigates some well-known deficiencies of CAM6, including the persistent drizzle problem [13, 14]. Beyond historical validation, we test CAMulator in out-of-sample scenarios with ocean surface warming of +2K and +4K. The model demonstrates a promising ability to adjust dynamically to these warmer conditions, though its response appears weaker than CAM6 under +4K warming. In this manuscript, we focus on evaluating CAMulator's performance and exploring its potential applications and extensions.

This manuscript is organized as follows: Section 2 describes the data sources used for model training and evaluation. Section 3 outlines the methodology, including model construction and training. Section 4 presents the results, highlighting CAMulator's performance in reproducing key climate statistics and its response to out-of-sample warming scenarios (+2K and +4K ocean surface warming). Finally, we conclude with a discussion of CAMulator's strengths, limitations, and potential extensions for future work.

2 Data

2.1 CAM6 training data

We use the Community Atmosphere Model version 6 (CAM6), the atmospheric component of the Community Earth System Model version 2.1.5 (CESM2), developed by the NSF NCAR [15, 16]. At the training time, CAM6 was the latest supported model release and incorporates advancements in atmospheric physics, cloud microphysics, and boundary layer turbulence while leveraging a finite-volume (FV) dynamical core.

For this study, we run CAM6 in the AMIP mode, where it is forced by observed sea surface temperatures (SSTs) and sea ice concentrations from 1979 to 2014, with these monthly forcing fields linearly interpolated to daily values. The model also accounts for time-evolving aerosol emissions and trace gas concentrations (including CO_2) to ensure consistency with historical atmospheric conditions.

Model simulations use the scientific release resolution of CAM6, with a horizontal grid spacing of 0.9° latitude $\times$ 1.25° longitude and 32 hybrid sigma-pressure levels extending up to 2.26 hPa in the vertical dimension. The archived dataset, described in Table 1, is saved at 6-hourly intervals. Data from 1979 to 2010 is used for training, with 2011 reserved for validation, and 2012–2014 designated for testing.

Flux-form variables (see bold variables in Table 1) and precipitation are treated as accumulated quantities and rescaled so that downward fluxes are positive. Flux-form variables are essential for estimating sources and sinks in the atmospheric moisture and energy budgets. Precipitation (PRECT) represents the total precipitation leaving the column, including both parameterized and large-scale rain and snow. Prognostic variables are stored as 6-hourly averages. The prognostic

Table 1: Description of input and predicted variables for CAMulator. Variables are categorized as prognostic, diagnostic, or dynamic/static forcing variables. *Qtot is a sum of all column moisture both vapor and condensed Qtot = Specific Humidity + Grid Box Snow Amount + Grid Box Rain Amount

Variable	Description	Units	Single Level/Levels	I/O
Prognostic Variables (Input and Output)				
U	Zonal Wind	m/s	32 levels	Input/Output
V	Meridional Wind	m/s	32 levels	Input/Output
T	Temperature	K	32 levels	Input/Output
*Qtot	Specific Total Water	kg/kg	32 levels	Input/Output
Diagnostic Variables (Output Only)				
PRECT	Accumulated Precipitation	m	Single Level	Output
CLDTOT	Total Cloud Cover	fraction	Single Level	Output
CLDHGH	High Cloud Cover	fraction	Single Level	Output
CLDLOW	Low Cloud Cover	fraction	Single Level	Output
CLDMED	Medium Cloud Cover	fraction	Single Level	Output
TAUX	Zonal Wind Stress	N/m²	Single Level	Output
TAUY	Meridional Wind Stress	N/m²	Single Level	Output
U10	10m Wind Speed	m/s	Single Level	Output
QFLX	Surface Moisture **Flux**	m	Single Level	Output
FSNS	Net Solar **Flux** at Surface	J/m²	Single Level	Output
FLNS	Net Longwave **Flux** at Surface	J/m²	Single Level	Output
FSNT	Net Solar **Flux** at TOA	J/m²	Single Level	Output
FLNT	Net Longwave **Flux** at TOA	J/m²	Single Level	Output
SHFLX	Sensible Heat **Flux**	J/m²	Single Level	Output
LHFLX	Latent Heat **Flux**	J/m²	Single Level	Output
Prognostic Surface Variables (Input and Output)				
PS	Surface Pressure	Pa	Single Level	Input/Output
TREFHT	Near-Surface Air Temperature	K	Single Level	Input/Output
Dynamic Forcing Variables (Time-Varying Input Only)				
SOLIN	Incoming Solar Radiation	J/m²	Single Level	Input
SST	Sea Surface Temperature	K	Single Level	Input
Static Forcing Variables (Input Only)				
Surface Geop.	Surface Height	m²/s²	Single Level	Input
Land-Sea Mask	Land Mask	unitless	Single Level	Input

variable Qtot represents the total water content in the model column, including vapor and condensed phases: specific humidity, as well as snow and rain that are present within the column.

Both static and dynamic forcing variables are included as input. Surface geopotential represents topography, while a land-sea mask distinguishes between ocean and land grid points to ensure accurate surface-atmosphere interactions. The model is also provided with dynamically forced (changing in time) variables, including sea surface conditions and incoming solar radiation, updated every 6 hours.

2.2 Data Preprocessing

All variables, except for the Land-Sea Mask, undergo z-score normalization, where the mean and standard deviation are computed based on the training data period (1979–2011). The Land-Sea Mask represents the fraction of land within each grid cell, with values ranging between 0 and 1, and is left unnormalized.

2.3 Model Validation and Observations

The goal of this work is to represent CAM6 with as much fidelity as possible, thus, we primarily analyze the emulation of CAM6 by CAMulator. In some cases, we incorporate reanalysis products to provide qualitative context for the

differences between CAM6, CAMulator, and the assimilated observational datasets. For precipitation, we use NOAA's Global Precipitation Climatology Project (GPCP) product [17], while all other variables are compared against the global ERA5 reanalysis [18]. The specific dataset used for each comparison is indicated in the corresponding figures and text.

3 Methods

3.1 Model Architecture

CAMulator is based on WXFormer, a transformer-based architecture developed at NSF NCAR and described in [12]. Figure 1 shows the CAMulator architecture and workflow. WXFormer utilizes a CrossFormer backbone [19] for multi-scale feature processing and long-range dependency modeling, combined with hierarchical transpose convolutional layers for upsampling in the decoding stages. Standard skip connections, similar to those in a U-Net [20], are incorporated to efficiently transfer feature information from the encoder to the decoder, preserving spatial details.

WXFormer has demonstrated state-of-the-art performance among leading AI-based weather prediction (AIWP) models [12], making it an ideal foundation for climate modeling. To adapt WXFormer for our use case, we introduce several architectural modifications tailored to climate-scale forecasting.

First, we increase the feature embedding size by doubling the dimensionality of the cross embedding layers (CEL) while maintaining computational efficiency (Fig. 1b). The CEL employs a multi-kernel approach, applying four convolutional kernels (4×4, 8×8, 16×16, and 32×32) in parallel, each with a 2×2 stride during the initial processing stage [19]. This multi-scale strategy enhances the model's ability to capture spatial patterns across a range of atmospheric scales.

Additionally, we reduce both the global and local window sizes of the Long Short Distance Attention mechanism (Fig. 1c) to align with the lower-resolution CAM6 use case compared to WXFormer's original $\sim$0.28° ERA5 application (see [12] for further details). This adaptation ensures more effective regional feature extraction while maintaining performance consistency.

With these modifications, CAMulator is comprised of ~751 million trainable parameters, making it a high-capacity climate emulator optimized for long-term prediction. WXFormer takes as input the state of the atmosphere at time step i and predicts the state at $i + 1$ with a 6-hour forecast horizon. The model operates autoregressively, generating multi-step forecasts, allowing it to function as a standalone climate model capable of running indefinitely.

3.2 Conservation Blocks

Following [11] and [7], conservation schemes are applied after the CAMulator output layer to ensure physically consistent model roll-outs (Fig. 1e). The order of application is critical, as each scheme depends on the corrections applied in previous steps. Below, we describe the purpose and implementation of each scheme, with the adjusted variables emphasized. For the direct calculation of these corrections see Appendix section B.

1. **Nonnegative correction**: AI models can produce negative values, which are unphysical for certain variables. For all nonnegative variables (*specific total water, total precipitation, 10-meter windspeed, and [Total, High, Low, Medium] Cloud Cover*), any negative raw outputs are set to zero. While this approach ensures physical consistency, alternative correction methods, such as redistribution, may be beneficial in some cases.
2. **Global dry air mass conservation scheme**: At each forecast step, *surface pressure* is corrected to ensure that global dry air mass remains constant, maintaining consistency with the initial condition.
3. **Global moisture budget conservation scheme**: At each forecast step, *total precipitation* is adjusted to balance the global sum of the total precipitable water tendency (derived from specific total water) and the accumulated net flux of precipitation and evaporation over the previous 6-hour period.
4. **Global total atmospheric energy conservation scheme**: The global atmospheric energy budget is defined as the balance between the tendency of total atmospheric energy and net energy fluxes at the top of the atmosphere and the surface. At each forecast step, *Temperature* is corrected to ensure that the sum of total atmospheric energy tendencies aligns with the net energy sources and sinks over the past 6-hour period.

The variable names used above match those listed in Table 1. Corrections to surface pressure, total precipitation, and air temperature are applied using multiplicative ratios across all grid cells, computed dynamically at each time step. While this approach enforces conservation, there are no explicit safeguards against overcorrection and no effort to redistribute values in the nonnegative correction, which may warrant further investigation in future work.

It is essential to apply these conservation schemes in the specified order to preserve the theoretical dependencies between mass, moisture, and energy conservation. Further technical details are provided in Appendix B.

3.3 Training

CAMulator was trained in a staged approach to balance stability and conservation constraints. Initially, the model was trained for 113 epochs as a single-step (6-hour) prediction task, minimizing mean squared error (MSE), with conservation block layers entirely omitted to allow for more stable initial learning. After this phase, conservation layers were introduced, and training continued for 76 additional epochs, now as a two-step (12-hour) forecast task. In the two-step training, the loss from each forecasted state was accumulated and used to optimize the model weights. For the single-step pretraining, cosine-annealing schedules were applied with an initial learning rate of $1\mathrm{E}-5$. Both stages used latitude-weighted MSE as a loss function, the AdamW optimizer [21], and batch sizes of 32. The training was conducted on NVIDIA A100 GPUs, each with 40 GB of memory, using Pytorch [22].

To determine the final model, we conducted a validation cycle, which consisted of training on 500 samples followed by a full-year simulation. The resulting climatology was then compared to an archived CAM6 climatology. This process was repeated after each training iteration to assess the model's ability to capture the year-long climatology of precipitation and 2-meter temperature (T2M). The final model was selected based on latitude-weighted mean squared error (RMSE), ensuring the closest match to the CAM6 climatology for these two fields.

3.4 Climate Forecast Inference Options

We explore multiple inference configurations, each defined by the forcing applied to the sea surface temperature (SST) field, while all other dynamic forcing variables remain unchanged. The following four SST scenarios are considered:

1. **Observed SSTs (1979–2013)** – Uses historical SST values from this period as in the CAM6 simulations
2. **Year 2000 Climatological SST** – Applies the **mean SST state** from the year 2000, held constant throughout the simulation.
3. **Year 2000 +2K Climatological SST** – Uses the **year 2000 SST climatology**, with a uniform **+2K temperature increase** applied globally to the SST field.
4. **Year 2000 +4K Climatological SST** – Similar to the previous scenario, but with a **+4K temperature increase** applied globally to the SST field.

In the **1979–2014 observed SST case**, the simulation continuously cycles through these years, allowing for realistic variability. In contrast, **climatological SST scenarios (year 2000, +2K, and +4K)** can be extended indefinitely, making them well-suited for exploring equilibrium climate responses under different baseline ocean conditions. We note that the model was trained exclusively on the 1979–2010 SST state, making the remaining three inference cases (Year 2000, +2K, and +4K climatological SSTs) entirely out-of-sample predictions. To ensure diverse initial conditions, we introduce a stochastic kinetic energy backscatter scheme [23] for the first 15 days to the 1979–2014 SST runs, allowing the model to reach an independent atmospheric state by week 2. This approach mirrors the initialization strategy used in CAM6 simulations, where initial temperature perturbations help generate distinct trajectories [1].

3.5 Computational Speed and Opportunities

CAM6 achieves significant computational throughput on Derecho, a NSF NCAR supercomputer, with 10 CPU nodes delivering 14 simulation years per day at the selected resolution. This performance metric reflects pure compute time, excluding I/O overhead.

In contrast, our ML-based emulator achieves a dramatic 530 times speedup, running at 750 simulation years per day on a single NVIDIA A100 GPU, while including all computational overhead and I/O operations. This acceleration enables high-throughput climate experiments, facilitating long-term scenario projections, ensemble simulations, and uncertainty quantification that would otherwise be computationally prohibitive with traditional numerical models.

3.5.1 Opportunities for Further Optimization

While this speedup is already substantial, further improvements are possible through:

- **Scalability via Ensemble Parallelism**: Expanding ensemble members across multiple GPUs to improve robustness and uncertainty estimation.
- **Memory-Efficient Data Handling**: Leveraging **asynchronous I/O strategies** and distributed storage solutions to further reduce bottlenecks in large-scale climate simulations.

Unlike some deep-learning applications, we avoid mixed precision (FP16/BF16) arithmetic to preserve the conservation properties critical for accurate climate modeling.

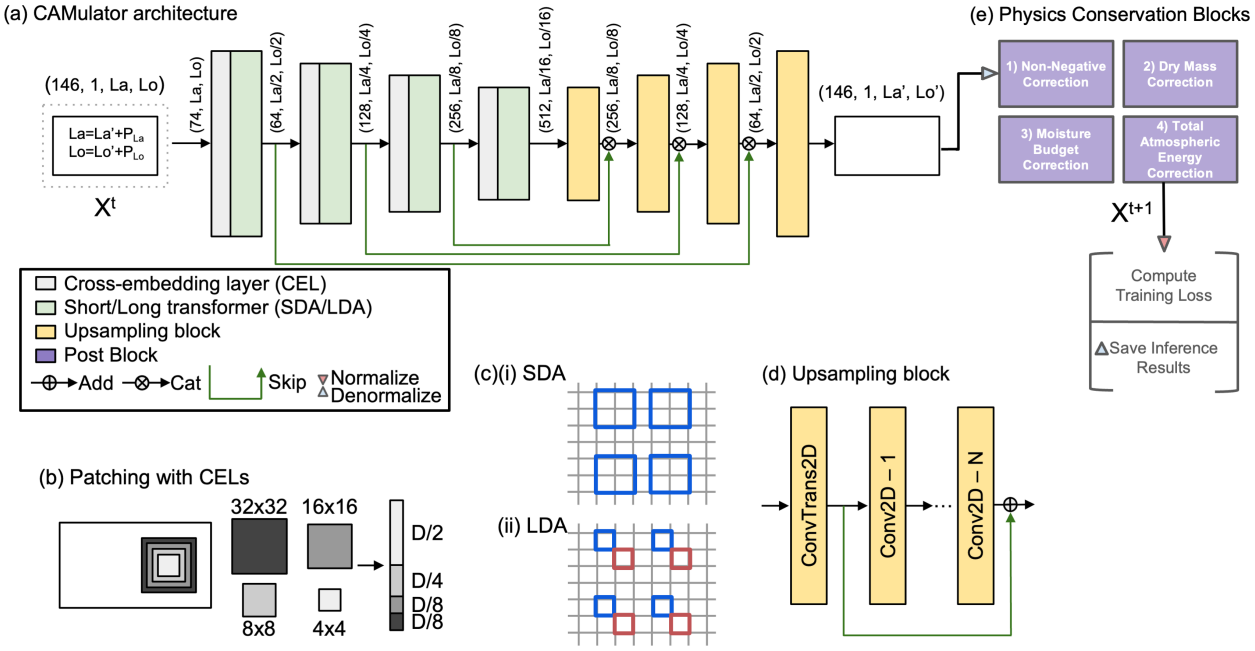

Figure 1: (a) the CAMulator architecture consists of encoding stages using a CrossFormer backbone and decoding stages with hierarchical transpose convolutional layers, with skip connections for improved feature flow. (b) The CEL captures multi-scale features using four convolutional kernels. The LSDA mechanism includes (c)(i) SDA for local interactions and (c)(ii) LDA for global dependencies. (d) The decoder component employs convolutional upsampling blocks with skip connections to progressively increase feature map resolution and maintain spatial information. e) predictions are then de-normalized and past through the four physics conservation blocks prior to loss caclulations

This optimized workflow positions the emulator as a scalable and efficient alternative to traditional GCMs, capable of running thousands of years of simulation within days, opening new avenues for large-ensemble climate forecasting, extreme event attribution, and policy-relevant decision support.

4 Results

4.1 Annual Cycle and Roll-out

Figure 2 illustrates two key aspects of CAMulator's response to SST forcing. The top panel compares a 12-member CAMulator ensemble with the training data from 1979–2013, demonstrating that CAMulator effectively captures the long-term warming trend of total column-integrated heat content. To isolate this trend, the seasonal cycle was removed by regressing out the six leading harmonics, and a 90-day rolling mean was applied to the time series (Fig. 2a). The CAM simulation remains well within the ensemble spread, highlighting the model's ability to reproduce observed interannual variability. The ensemble spread arises from the introduction of stochastic kinetic energy backscatter [SKEBS, e.g., 23], which perturbs initial conditions over the first 15 days before allowing the system to evolve freely. We see a likely underestimation of total heat capacity in the period of 2003 to 2010, though CAM still largely sits within the ensemble spread of CAMulator, this is discussed further in the model deficiencies section.

Figure 2b demonstrates the model's behavior when forced with fixed year-2000 climatological SSTs. In this scenario, CAMulator maintains an indefinite stable rollout with no emergent trend, as the absence of external forcing prevents any sustained warming signal. This suggests that while the model responds effectively to imposed SST trends, it does not generate spurious warming in the absence of a forcing mechanism, and is capable of indefinite fixed climate rollouts. We show an identical figure to Figure 2 in the supplemental material, but for total water path (Fig. S1) and find smilar results.

Notably, across all SST forcing scenarios—including cases where the simulations were forced with SSTs outside of the training distribution—CAMulator has exhibited no signs of numerical instability. To date, no model crashes with our final model configuration have been observed, highlighting its robustness in handling a range of climate conditions.

Figure 2: Column-integrated heat content in CAMulator to different SST forcing scenarios. (Top panel) A 12-member CAMulator ensemble (teal) is compared to the CAM simulation (purple) using observed SSTs from 1979–2013. CAMulator successfully captures the long-term warming trend and interannual variability. To isolate the trend, the seasonal cycle has been removed by regressing out the six leading harmonics, and a 90-day rolling mean has been applied. The ensemble spread arises from stochastic kinetic energy backscatter (SKEBS) perturbations applied during the first 15 days of simulation. (Bottom panel) CAMulator forced with fixed year-2000 climatological SSTs produces a stable long-term simulation with no discernible trend, demonstrating that the model does not introduce artificial warming in the absence of an external forcing mechanism.

Figure 3 shows the conservation properties for global mass, water, and energy of the CAMulator system with the conservative layers activated (Figure 3; CAMulator-phys teal line) and inactive (Figure 3; CAMulator-nophys black line). CAMulator-nophys immediately deviates from the desired conservation properties and quickly settles into it's own errant climatology halfway through the first month of the climate simulation. The residuals are calculated as the previous time-step minus the current $[t_{-1} - t_0]$, meaning and observed positive residual indicates that CAMulator-nophys sheds mass, water, and energy prior to settling into a steady state. We note, that we can still achieve simulation runs with CAMulator-nophys without major instabilities, and indefinite roll-outs are achieved.

Interestingly, enforcing total energy tendency conservation balances the tendency such that it is centered around zero and regularly oscillates with the diurnal cycle (Fig. 3e), whereas sporadic behavior is observed in CAMulator-nophys.

4.2 Annual Mean Biases

We next evaluate the time-averaged annual climatology of the CAMulator simulation over the period 1979–2013. Figure 4a–c shows the zonal mean precipitation, 2m temperature, and zonal wind at the lowest model level for CAM6 and CAMulator, shown in purple and teal, respectively. The zonal mean for two reanalysis products is also shown in dashed gray.

Overall, CAMulator simulates the annual mean state well. Figure 4d–f presents the difference between the annual means (CAMulator - CAM). The largest precipitation errors occur in the tropics, with a wet bias over Central America and a dry bias over the Maritime Continent (Fig. 4d). For T2m, a persistent warm bias is evident over Greenland, particularly in winter, likely due to the lack of ice representation in the model state (Fig. 4e). The largest discrepancies in zonal

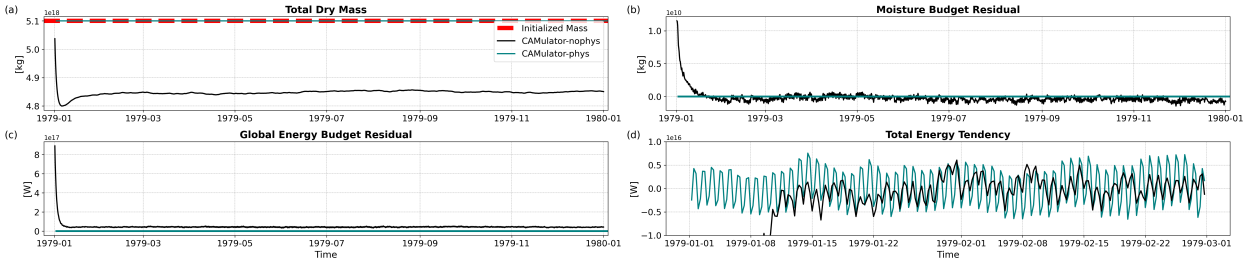

Figure 3: Time series of conservation diagnostics and budget residuals for CAMulator with and without physics conservation blocks. (a) Total dry mass (kg) for CAMulator-phys (teal), CAMulator-nophys (black), with the initialized mass shown as a reference (red dashed line). (b) Moisture budget residual (kg) comparing CAMulator-phys and CAMulator-nophys. (c) Global energy budget residual (W) for CAMulator-phys and CAMulator-nophys. (d) Total energy tendency (W), representing the time derivative of total atmospheric energy, comparing CAMulator-phys and CAMulator-nophys.

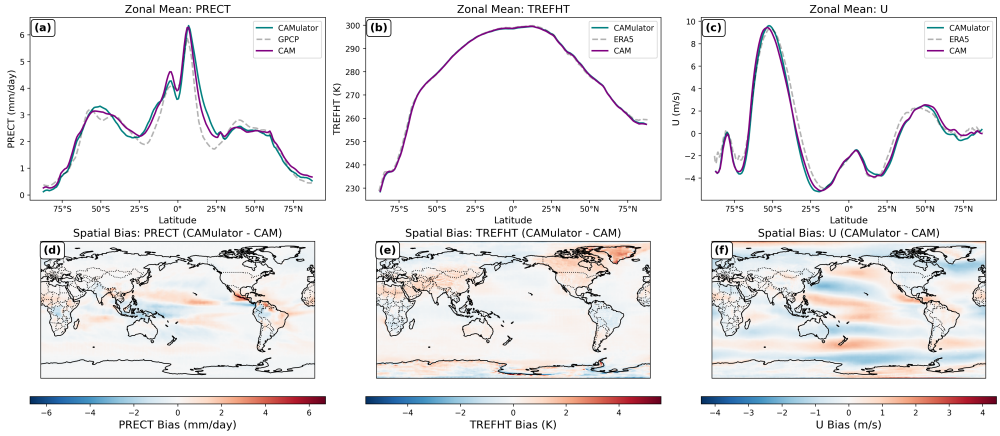

Figure 4: Zonal mean and spatial bias analysis of key climate variables. The top row (panels a–c) displays the zonal mean values for (a) precipitation rate (PRECT, mm/day), (b) near-surface air temperature (TREFHT, K), and (c) zonal wind component (U, m/s), comparing simulations from CAMulator (teal) and CAM (purple). Reanalysis products (GPCP [1981-2010] and ERA5 [1979-2010]) are shown shown in grey dash. The bottom row (panels d–f) presents the corresponding spatial biases (CAMulator - CAM) for (d) PRECT, (e) TREFHT, and (f) U, highlighting regional differences. Biases are computed as the annual mean differences between the two simulations and are visualized using a diverging colormap (red: CAMulator > CAM, blue: CAMulator < CAM)

winds appear in the storm track regions, where slight shifts in the locations of peak wind magnitudes are observed (Fig. 4f). For every field, the spatial annual climatological RMSE is computed and displayed in the supplemental material (Figs. 2S-16S), we additionally compute the annual RMSE values which are is in appendix table A.1.

4.3 Modes of Variability

Evaluating a climate model's performance requires not only assessing its accuracy in representing climatological averages but also its ability to capture lower-frequency climate modes [e.g., 24]. Given the vast number of modes of variability identified in the literature, a comprehensive analysis is impractical. Instead, we focus on three principal and well-documented modes used in major climate model evaluations: the Pacific North American Pattern (PNA), the North Atlantic Oscillation (NAO), and the El Niño–Southern Oscillation (ENSO) precipitation response. As these modes typically peak in boreal winter (defined here as December–February (DJF)), our analysis will center on their wintertime behavior. This section focuses on the representation of the PNA and NAO, while the following section will examine ENSO.

Figure 5 shows the regression of the 500 hPa geopotential height (Z500) anomaly onto the leading principal component for the PNA and NAO regions in DJF, comparing CAMulator and CAM. Panels (a) and (b) illustrate the PNA-associated Z500 anomalies for CAMulator and CAM, respectively, while panels (c) and (d) display the NAO-related anomalies.

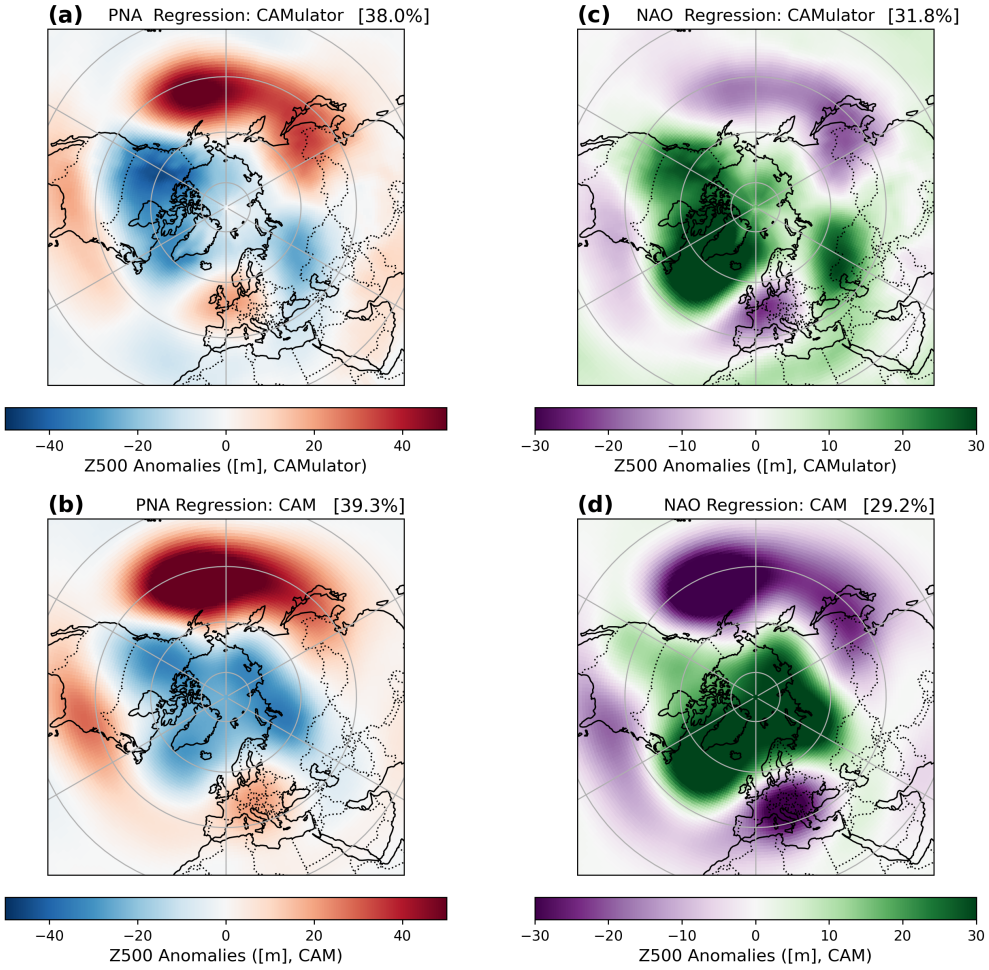

Figure 5: Regression of the anomalous Z500 field on the leading principal during DJF for the PNA region (left column, 20°–85°N, 120°E–120°W) and the NAO (right column, 20°–80°N, 90°W–40°E) over the years 1979–2013. Results are shown for CAMulator (top row) and CAM (bottom row). The explained variance of each mode is shown in the top right of each panel

Both patterns exhibit a close match to the CAM6 variability in terms of explained variance, indicating that these modes are well represented. Although the CAMulator patterns appear slightly muted in amplitude, they qualitatively capture the general structure of the CAM6 patterns and remain within the expected range of variability over a 30-year simulation [25]. Notably, geopotential height is not directly predicted by CAMulator; rather, it is reconstructed by summing thicknesses across half-model levels to account for temperature and moisture variations between grid cells. The fact that CAMulator successfully reproduces the spatial structure of Z500 anomalies suggests that the physical coherence between surface pressure (PS), temperature (T), and total moisture (Qtot) has been well preserved, lending credibility to the model's internal consistency.

4.3.1 PNA & NAO

4.3.2 ENSO Precipitation Response

ENSO is the dominant mode of tropospheric climate variability on interannual timescales, exerting a strong influence on global precipitation patterns. To assess how well CAMulator replicates the ENSO-related precipitation response, we analyze composite precipitation differences (El Niño minus La Niña) for the eight strongest ENSO events, identified based on absolute Niño3.4 SST anomalies over DJF and compare the response to CAM6 (Fig. 6a–b).

Both models capture the characteristic precipitation enhancements in the central and eastern Pacific, particularly near the International Date Line. However, CAMulator exhibits a slightly muted response in this region, a difference that

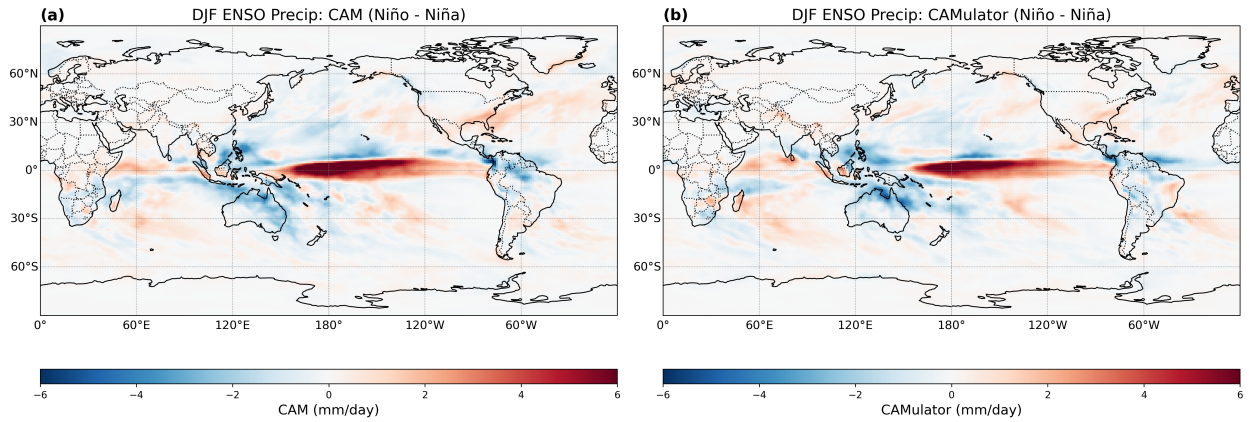

Figure 6: Composite difference in DJF precipitation (mm/day) during ENSO events (El Niño minus La Niña) for the eight strongest ENSO years on record, as identified by the Niño3.4 index. Results are shown for CAM (left) and CAMulator (right), with composites constructed using simple averaging. The Niño3.4-based El Niño years (December) include 1979, 1982, 1986, 1987, 1991, 1994, 1997, and 2002, while La Niña years (December) include 1983, 1984, 1988, 1995, 1998, 1999, 2000, and 2007. The color scale represents precipitation anomalies, where red indicates increased precipitation during El Niño relative to La Niña, and blue indicates reduced precipitation.

may be attributable to internal variability. Notably, this muted response is consistent with the slight underrepresentation of the PNA and NAO modes, suggesting a potential model tendency to underpredict low-frequency variability, which is notable in another ML-based climate model [8]. The precipitation derived from ENSO-driven teleconnections over North America are well captured, with drier conditions over the southern U.S. and wetter conditions in the Pacific Northwest. Similarly, the Maritime Continent response aligns closely between the models, indicating that CAMulator effectively represents the large-scale precipitation shifts associated with ENSO.

Overall, CAMulator successfully reproduces the spatial patterns of the CAM6 ENSO precipitation response, showing strong qualitative agreement despite some regional amplitude differences.

4.4 Extremes and Precipitation

Beyond representing mean climate and large-scale modes of variability, a climate model's ability to simulate extreme events is a crucial measure of its fidelity. Extremes (e.g. heatwaves, heavy precipitation events, or droughts) shape ecosystems, infrastructure resilience, and societal vulnerability. These high-impact phenomena arise from nonlinear interactions between atmospheric and oceanic processes, making their accurate representation a stringent test of a model's physical consistency and predictive skill. Machine learning models, particularly those trained to minimize mean squared error, often struggle to capture extremes, as they tend to favor the mean state over rare, high-magnitude events. However, CAMulator does not exhibit a strong tendency to underpredict extremes. We hypothesize that this may be due to its training methodology, which is limited to only two prediction time steps (out to 12 hours), allowing it to better preserve variability. Supporting this, CAMulator's kinetic energy and potential temperature spectra (see Supplemental Fig. 1S) show minimal smoothing at smaller spatial scales, a common issue in data-driven models [see, 26] that can lead to weakened variability.

Figure 7 presents the annual maximum 6-hourly 2m temperature and precipitation at each grid point for CAM6 and CAMulator, with differences shown in the bottom row. Overall, CAMulator successfully captures the spatial patterns of extreme events, but some notable biases emerge. For 2m temperature, CAMulator overpredicts Arctic extreme temperatures, particularly in regions dominated by sea ice and land ice. This bias is expected, as CAMulator lacks explicit sea-ice and land-ice interactions—a limitation that will be addressed in future versions. For precipitation, CAMulator underpredicts extremes in the deep tropics while overpredicting them in the mid-latitudes. Notably, the region off the coast of Japan exhibits a strong positive bias in extreme precipitation, suggesting that CAMulator may be overestimating the intensity of tropical cyclones in this region.

4.5 Rain Amount Distribution

In this section we look at the similarities of the modeled rain amount and rain frequency distributions [e.g., 14], using logarithmically distributed rain-rate bins following [27]. See the appendix section A.3 for the exact form of

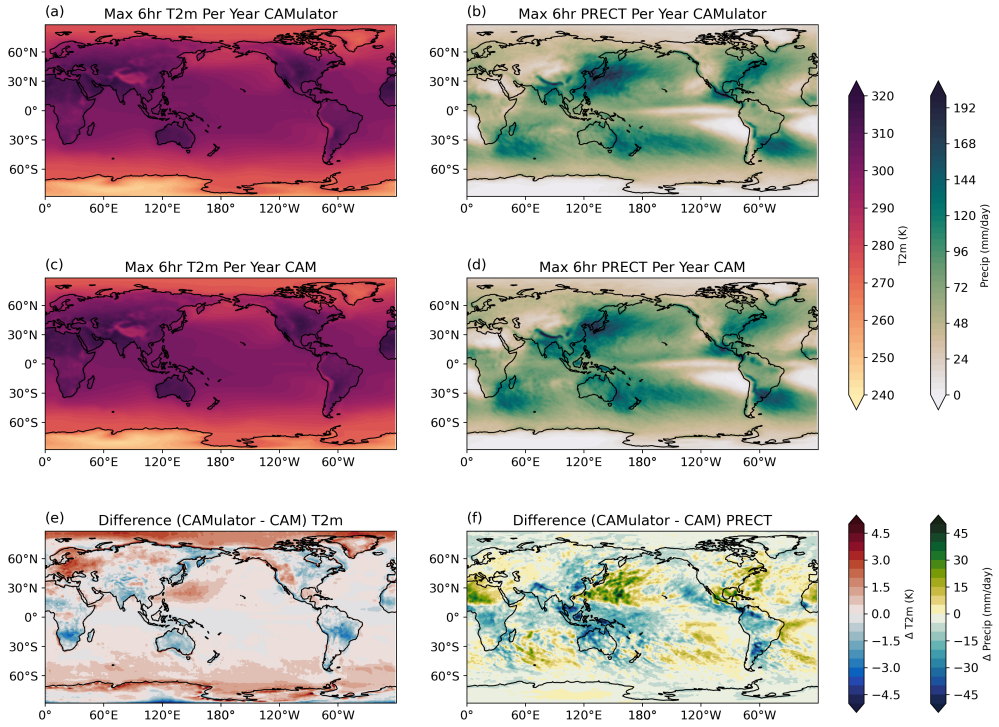

Figure 7: Climatology of the annual maximum 6-hourly average two-meter temperature (T2m, left column) and precipitation rate (PRECT, right column) over the period 1979–2014. The top row shows results from CAMulator, while the middle row presents results from CAM. The bottom row displays the differences (CAMulator - CAM), highlighting areas where CAMulator simulates warmer or colder extreme temperatures (red and blue shading, respectively) and higher or lower extreme precipitation rates* (green and blue shading, respectively). Color bars indicate absolute values for T2m (K) and precipitation (mm/day) in the top and middle rows, while the bottom row represents their respective differences.

the calculation. The global distributions of CAM and CAMulator, along with the distribution observed in the daily GPCP dataset spanning years 1997-2012 are shown in Figure 8. The rain amounts in CAM and CAMulator are nearly identical, each peaking at around 10 mm day^{-1} and are in agreement for all rain rates (Fig 8a). However, the biases in CAM6 when compared to the GPCP product (i.e. narrower distribution focused on a higher rain rate) persist in CAMulator. This is expected as CAMulator was trained as a CAM6 emulator.

At larger rain rates (> 10 mm day^{-1}) the rain frequency distributions align fairly well (Fig 8b). However, at lighter rain rates, between 0.1 and 3 mm day^{-1}, there is a noted reduction in CAMulator rain frequency, bringing the distribution closer the the GPCP. Additionally, the dry-day frequency is nearly double that of CAM, and much more similar to GPCP. These are known biases in CAM6 that have persisted through most of the generations of the CAM models. As discussed in [11], we attribute this improvement to the drizzle as a property of the conservation schemes, which can tackle this problem because they are operated based on the global sum rather than gridpoint-wise quantities. Said another way, in the moisture budget conservation scheme, the global sum of total precipitation must match with other conservation properties. If the AIWP model overestimates drizzle and underestimates dry areas, its global sum of total precipitation will violate the conservation laws, and corrections will be applied–this correction is assigned to close the conservation budget only, and it almost surely will increase the mean squared error training loss–so the AIWP model will be penalized by its drizzle heavily.

4.6 +2K and +4K runs

To evaluate the CAMulator's response to SST perturbations beyond its training distribution, we impose climatological SSTs for the year 2000 as a baseline forcing (+0K). Figure 9 presents the global mean temperature at the lowest model level (Fig. 9a), along with the vertically integrated heat content for the lower (850–1000 hPa; Fig. 9b) and upper troposphere (200–850 hPa; Fig. 9c). To assess the model's extrapolation capabilities under out-of-distribution forcing, we introduce uniform SST anomalies of +2K and +4K across all ocean grid cells and compare the CAMulator's

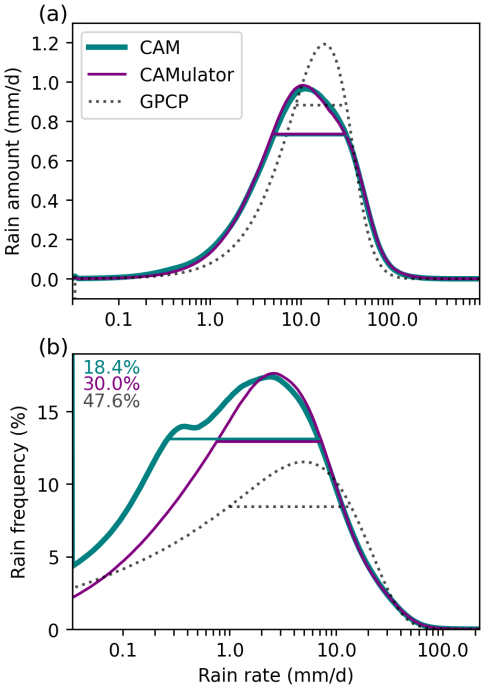

Figure 8: Global climatological distributions of daily rainfall from CAM (teal), CAMulator (purple), and GPCP observations (dotted black). (Top) Rain amount (mm day^{-1}) as a function of rain rate. (Bottom) Rain frequency distribution (%) as a function of rain rate. The dry-day frequency is indicated in the top left of the bottom panel, with values for CAM (teal), CAMulator (purple), and GPCP (black). GPCP serves as an observational reference dataset and is coarsened prior to computing distributions.

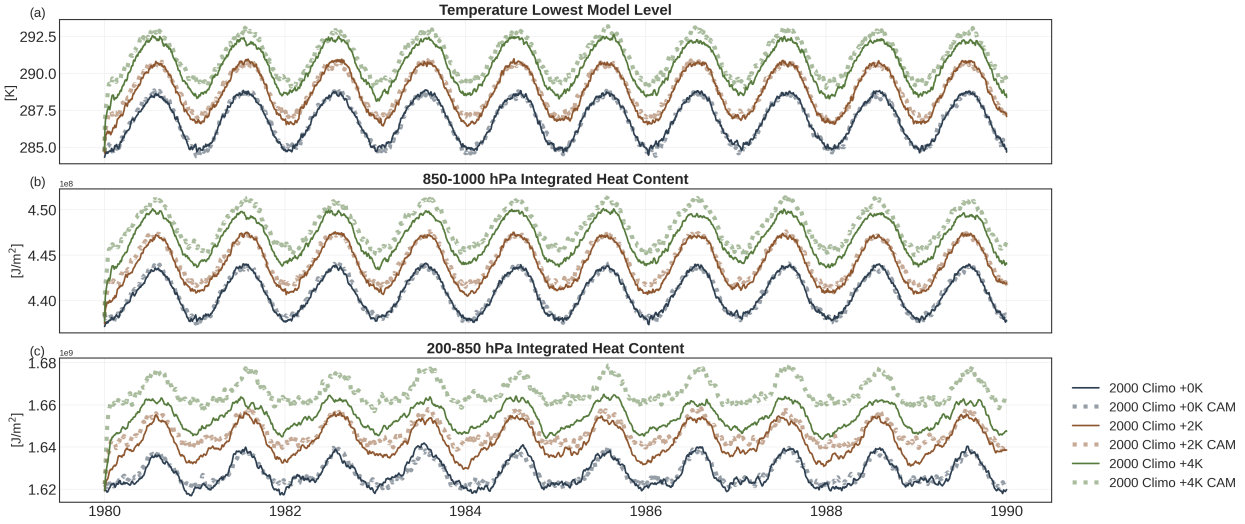

Figure 9: Time series of Global average Surface Temperature (a) and Column-integrated heat content (b,c) for simulations driven by different sea surface temperature (SST) climatologies: 2000 SST climatology (dark blue), 2000+2K (brown), and 2000+4K (green). The results from CAM (dashed lines) and CAMulator (solid lines) are shown. Integrated heat content is shown in two layers: (b) 850–1000 hPa and (c) 200–850 hPa.

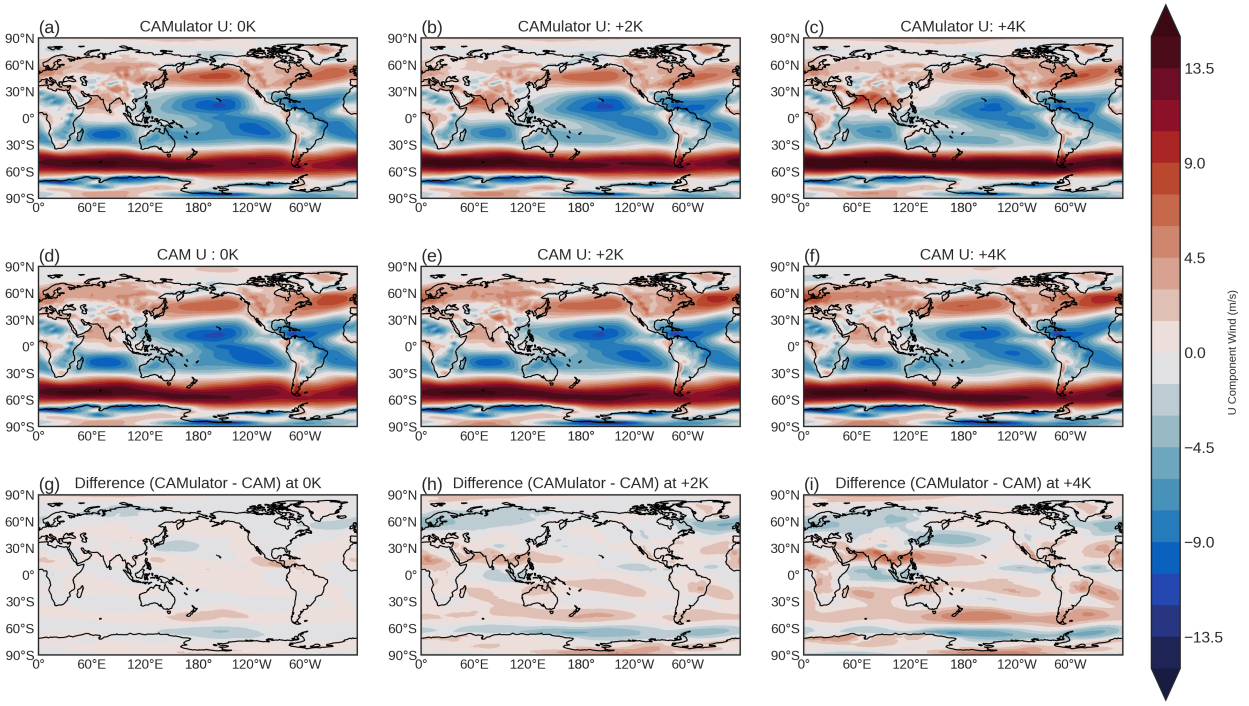

Figure 10: Annual mean zonal wind at the second model level in CAM (middle row) and CAMulator (top row) for three sea surface temperature (SST) climatologies: 2000 (left column), 2000+2K (middle column), and 2000+4K (right column). The bottom row shows the difference (CAMulator - CAM), highlighting deviations between the two models. The color scale represents zonal wind speed (m/s), with positive/negative values (red/blue) indicating stronger/weaker values in CAMulator.

response to that of CAM6. Notably, these warming states lie well beyond the training data, testing the model's ability to generalize under extreme forcing scenarios.

The surface temperature response in the lowest model level exhibits the strongest agreement with CAM6, particularly in the +0K and +2K cases, where the seasonal cycles are well-aligned. However, discrepancies emerge under stronger warming, with a systematic cold bias in the CAMulator during boreal winter in both the +2K and +4K simulations. This bias suggests that CAMulator fails to fully capture key feedback mechanisms governing seasonal temperature variations in high-latitude regions.

As shown in Supplemental Figure 17S, the largest errors in 2-meter temperature occur in polar and sea-ice-dominated regions, indicating that the CAMulator lacks a robust representation of ice-atmosphere interactions. Specifically, the model has no explicit knowledge of sea-ice melt dynamics, surface albedo feedbacks, or key phase transition processes that modulate energy exchange in these environments. In contrast, CAM6 dynamically represents the retreat and expansion of ice cover, which strongly regulates the surface energy budget in these regions. The absence of such processes in the CAMulator likely leads to an unrealistic seasonal persistence of cold anomalies during boreal winter.

In the lower troposphere (Fig. 9b), the integrated heat content exhibits a seasonal amplitude shift relative to CAM6, which may reflect errors in boundary layer processes or the representation of heat fluxes from the surface. Similarly, upper-tropospheric heat content (Fig. 9c) shows systematic biases that may be linked to discrepancies in heating distribution via convective processes or the large-scale advection of heat anomalies.

Overall, these results underscore the limitations of the CAMulator in extrapolating beyond its training data, particularly in cryospheric regions where ice-related feedbacks play a critical role. Future work should assess whether incorporating explicit representations of land-ice/sea-ice state changes, or at least indirect constraints on polar energy fluxes, could mitigate these systematic biases and improve the model's ability to handle out-of-distribution SST forcings.

CAMulator's ability to replicate large-scale low-level wind patterns under different SST forcing scenarios is evaluated in Figure 10. The top two rows show the zonal wind component (U) at 850 hPa for CAMulator (Figs. 10a–c) and CAM6 (Figs. 10d–f) across the +0K, +2K, and +4K SST warming states, while the bottom row (Figs. 10g–i) presents the difference between CAMulator and CAM6.

Overall, CAMulator demonstrates notable skill in capturing the large-scale structure of the low-level winds, particularly in the Southern Ocean. The core features of the midlatitude westerlies, including their equatorward contraction under warming, are well preserved. The close alignment with CAM6 suggests that CAMulator effectively learns the statistical relationship between SST forcing and the strength and position of the Southern Hemisphere storm tracks. Given that Southern Ocean winds are strongly tied to meridional temperature gradients and baroclinic wave activity and that this relationship is well represented in the training dataset, it is unsurprising that CAMulator generalizes well in this region.

However, the North Atlantic sector presents a key divergence from CAM6. Despite increasing SSTs, there is little evidence that the North Atlantic westerlies strengthen in response to warming in CAMulator, whereas CAM6 shows a more pronounced intensification. This discrepancy may reflect limitations in how CAMulator extrapolates the response of dynamically driven modes such as the North Atlantic Oscillation (NAO). Since the NAO is strongly influenced by internal atmospheric variability, rather than a direct SST-driven forcing, its response may be more difficult for a data-driven model to capture—especially in out-of-distribution scenarios. Additionally, CAMulator lacks an explicit representation of transient eddies and their feedbacks on mean flow, which could contribute to weaker NAO-like variability under warming.

The difference maps (Figs. 10g–i) further highlight systematic deviations, particularly in the North Atlantic and over regions of strong land-sea contrast, such as the western boundary currents. While these biases remain relatively small in magnitude compared to temperature biases, they suggest that CAMulator does not fully capture the dynamical adjustments that drive localized changes in low-level winds. Given that the model is trained purely on CAM6 output, this implies that some aspects of the wind field may be inherently more difficult for a ML-based emulator to reproduce. This could be due to a lack of direct SST-to-wind causality in the training dataset.

Overall, these results indicate that CAMulator successfully replicates large-scale low-level wind features, particularly in regions where wind anomalies are tightly coupled to SST changes. However, for dynamically driven patterns such as the NAO, its response diverges from CAM6, potentially due to challenges in learning internally generated variability from training data alone. Future work should assess whether incorporating additional predictors—such as large-scale pressure anomalies or eddy kinetic energy—could improve CAMulator's ability to mimic dynamically driven circulation responses to warming.

5 Model Deficiencies, Future Work, and Conclusions

In this work, we introduced CAMulator, an auto-regressive, machine-learned (ML) emulator of the Community Atmosphere Model version 6 (CAM6). CAMulator is forced by sea surface temperatures (SSTs) and incoming solar radiation and is explicitly constrained to conserve global mass, water, and energy. It exhibits numerical stability over indefinite roll-outs and accurately reproduces the surface and integrated atmospheric heating response to SST variations. The annual climatology is well captured, and dominant modes of variability, such as ENSO, the NAO, and the PNA, emerge naturally, though with slightly muted amplitudes in some cases. Additionally, CAMulator's physical constraints alleviate some of the drizzle problem commonly found in traditional climate models (Fig. 8). Beyond these physical attributes, CAMulator is computationally efficient and differentiable, making it a promising tool for a range of scientific applications.

5.1 Model Deficiencies and Future Improvements

Despite these strengths, several key deficiencies remain, highlighting areas for future development.

- **High-latitude biases due to missing cryospheric processes:** The absence of interactive sea and land ice, and it's limited representation in the training data of current climate leads to a persistent cold bias, particularly in boreal winter. This bias is most pronounced in later periods and becomes apparent in seasonal temperature cycles (Supplemental Fig. 18S). The lack of explicit ice-feedbacks and phase transition processes likely contributes to this discrepancy, especially in polar regions.
- **Muted ENSO-related variability:** While CAMulator successfully captures the broad-scale response to ENSO events, it underestimates their magnitude. This is particularly evident in the integrated atmospheric heat content during strong ENSO events, such as the 1997/98 El Niño (Fig. 2). Improved representation of coupled ocean-atmosphere interactions and internal variability could enhance this response.
- **Challenges in extrapolating to extreme SST perturbations:** As discussed in the +2K and +4K warming experiments, CAMulator's response to out-of-distribution SST forcing diverges from CAM6. This suggests that, while the model generalizes well within its training range, it struggles with conditions that require dynamically consistent extrapolation. Future improvements to training strategies, including exposure to a broader range of climate states, may mitigate this issue.

To address these limitations, several promising research directions should be explored:

- **Coupling CAMulator with an interactive ocean, sea ice, and land model:** Incorporating dynamic surface processes would improve the representation of feedback mechanisms, particularly in polar regions, and help resolve biases associated with missing cryospheric states.
- **Enhancing variability through stochastic parameterizations:** Learning with? a Stochastic Kinetic Energy Backscatter Scheme (SKEBS) or a similar approach could improve CAMulator's ability to represent subgrid-scale variability, potentially addressing the muted ENSO and NAO responses.
- **Exploring supermodeling approaches:** One promising avenue is to integrate CAMulator with multiple ML-based emulators coupled to a single dynamical model, allowing for dynamic state corrections and improved climate change projections. Such a framework could enhance the representation of future warming scenarios while maintaining computational efficiency.

5.2 Broader Implications and Future Outlook

CAMulator's ability to rapidly emulate atmospheric states while conserving fundamental physical properties makes it an attractive tool for accelerating climate simulations, uncertainty quantification, and data assimilation. Furthermore, its differentiability opens new opportunities for inverse modeling and sensitivity analyses, potentially improving parameter estimation in Earth system models.

As ML-driven climate modeling continues to evolve, hybrid approaches that combine data-driven methods with traditional physics-based models will likely become increasingly valuable. CAMulator provides a foundation for such efforts, demonstrating that machine-learned emulators can maintain physical fidelity while offering substantial computational advantages. By incorporating additional coupling mechanisms and improving variability representation, future iterations of CAMulator could play a critical role in next-generation climate modeling, bridging the gap between computational efficiency and physical realism.

Acknowledgments

This project is supported by Schmidt Sciences, LLC. We thank all the scientists and administrators who contributed to the development of CESM2. Additionally, this work was supported by the U.S. National Science Foundation National Center for Atmospheric Research (NSF NCAR), which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation.

Open Science

Place Github Here

Appendix Material

A Metric Calculations

A.1 Climatological RMSE

The weighted annual climatological root mean squared error is calculated as:

$$\mathrm{RMSE}(t_i, t_l) = \sqrt{\frac{1}{N_\phi N_\lambda} \sum_{i=1}^{N_\phi} \sum_{j=1}^{N_\lambda} \left\{ w(i) \left[\overline{CAM6(i,j)} - \overline{CAMulator(i,j)} \right]^2 \right\}} \tag{1}$$

Where N_ϕ and N_λ are the number of latitude (ϕ) and longitude (λ) grid cells, respectively. $w\,(i) = \cos(\phi_i)$ is the latitude weighting coefficient, which is normalized such that $\sum_i w(i) = 1$, the over bar indicates that each field was first averaged in time over the 30 year simulation run. Table A.1 shows the RMSE calculated for each single level variable.

Short Name	Long Name	Units	Globally Averaged RMSE
PRECT	Precipitation Rate	mm/day	0.5766
CLDTOT	Total Cloud Cover	fraction	0.0402
CLDHGH	High Cloud Cover	fraction	0.0391
CLDLOW	Low Cloud Cover	fraction	0.0340
CLDMED	Medium Cloud Cover	fraction	0.0296
TAUX	Zonal Wind Stress	N/m²	0.0110
TAUY	Meridional Wind Stress	N/m²	0.0055
U10	10m Wind Speed	m/s	0.3062
QFLX	Surface Moisture Flux	kg/m²/s	0.0474
FSNS	Net Solar Flux at Surface	W/m²	6.4269
FLNS	Net Longwave Flux at Surface	W/m²	2.6681
FSNT	Net Solar Flux at TOA	W/m²	5.8499
FLNT	Net Longwave Flux at TOA	W/m²	3.9247
SHFLX	Sensible Heat Flux	W/m²	2.1538
LHFLX	Latent Heat Flux	W/m²	5.4987

Table 2: Spatial Root Mean Square Errors (RMSE) for climate variables, including their names, descriptions, and measurement units. The RMSE values represent the deviation of CAMulator's 30-year climatological mean from the climatology of the base CAM simulation.

A.2 Atmospheric Heat Capacity

The atmospheric heat capacity is calculated by integrated temperature in the vertical and taking a cosine-latitude weighted sum.

$$\frac{Cp}{g}\int_{P_1}^{P_0} T\,dP \tag{2}$$

Where Cp is the dry air specific heat capacity and is set to 1004 $[Jkg^{-1}K]^{-1}$

A.3 Rain amount and Rain Frequency

The calculation of rain amount and rain frequency follows the method outlined in [14], but we summarize the approach here for completeness. The rain-rate distribution is constructed using logarithmically spaced bins, where each bin is 7% wider than the previous one, with its center shifted accordingly. Only rain rates exceeding 0.03 mm day^{-1} are included, while dry days are defined by a precipitation threshold of 0.0321 mm day^{-1}.

At each grid point, we first compute a histogram of rain rates and normalize it by the total number of days to obtain the rain frequency distribution. The total precipitation amount within each bin is then summed to construct the rain amount distribution. Finally, these distributions are averaged globally using an area-weighted approach to produce the global-mean distributions.

To formalize this, we define the probability distribution of rain amount p and rain frequency f for each dataset, using daily rain accumulation r from model output or gridded observations. The bin edges are denoted as T_i^l (left) and T_i^r (right), with the bin centers defined as $T_i^c = (T_i^l + T_i^r)/2$. The transformation of the distribution is expressed mathematically as follows:

$$p_i(T_i^c) = \frac{1}{\Delta \ln R}\int_{\ln T_i^l}^{\ln T_i^r} p(\ln r)d\ln r = \frac{1}{\Delta \ln T}\sum_{\text{gridpts}} r(T_i^l < r < T_i^r)\frac{A_{\text{gridpt}}}{A_{\text{total}}}, \tag{3}$$

$$f_i(T_i^c) = \frac{1}{\Delta \ln T}\int_{\ln T_i^l}^{\ln T_i^r} f(\ln r)d\ln r = \frac{1}{\Delta \ln T}\sum_{\text{gridpts}} \frac{N_d(T_i^l < r < T_i^r)}{\sum N_d}\frac{A_{\text{gridpt}}}{A_{\text{total}}}, \tag{4}$$

$$F_d = \frac{1}{\sum N_d}\sum_{\text{gridpts}} N_d(r=0)\frac{A_{\text{gridpt}}}{A_{\text{total}}}, \tag{5}$$

where A is the grid cell area, and N_d represents the number of days in the dataset. The bin width is set as

$$\Delta \ln T = \frac{(T_{i+1} - T_i)}{T_i} = 7.67\%, \tag{6}$$

ensuring sufficient resolution to capture the distribution of rain rates.

B Model Grid and Conservation Schemes

CAMulator leverages a hybrid sigma-pressure coordinate system, where vertical integrals are computed by first determining the local pressure value. The pressure at each model level k is given by:

$$P(i, j, k) = A_k P_0 + B_k PS(i, j) \tag{7}$$

where $P(i, j, k)$ represents the pressure at level k for a given latitude-longitude point (i, j). The terms A_k (hPa) and B_k (dimensionless fraction) are reference coefficients defining the hybrid coordinate system, while P_0 (1000.0 hPa) is the reference pressure. The surface pressure, $PS(i, j)$, varies across grid points and is used to determine the pressure levels dynamically.

For a quantity $S(z)$ that varies with height z, its mass-weighted vertical integral can be converted to a pressure level integral using the hydrostatic equation:

$$\int_0^\infty \rho S dz = \frac{1}{g} \int_{p_s}^0 S dp \approx \frac{1}{g} \int_{p_s}^{p_0} S dp \tag{8}$$

where S represents the variable of interest at the midpoint of the (i, j, k)-the grid box, and g is the acceleration due to gravity. We conduct similar corrections to [11], except applied on sigma-hybrid pressure levels, we repeat the derivation of those calculations below for completeness.

B.1 Global dry air mass conservation

The evolution of dry air mass within a given atmospheric column is determined by the divergence of the vertically integrated dry air mass flux:

$$\frac{1}{g} \frac{\partial}{\partial t} \int_{p_1}^{p_0} (1 - q) dp = -\nabla \cdot \frac{1}{g} \int_{p_1}^{p_0} [(1 - q)\, \mathbf{v}] dp \tag{9}$$

where $\mathbf{v}$ represents velocity, and q corresponds to total atmospheric moisture, approximated by specific total water (see Table 1).

For a global sum, if the atmosphere is considered incompressible, the divergence term in equation (9) becomes zero. Consequently, the total global dry air mass ($\langle M_d \rangle$) remains conserved over time (henceforth, $\langle\rangle$ denotes a global sum):

$$\frac{\partial}{\partial t} \langle M_d \rangle = \frac{\partial}{\partial t} \left\langle \frac{1}{g} \int_{p_1}^{p_0} (1 - q) dp \right\rangle = \epsilon_d \tag{10}$$

Where ϵ_d is the residual term that violates the global dry air mass conservation.

For two forecast time steps separated by $\Delta t = t_1 - t_0$, where t_0 corresponds to the initial analyzed state and t_1 denotes a subsequent validation time, equation (12) can be reformulated as:

$$\frac{\partial}{\partial t} \langle M_d \rangle = \langle M_d\,(t_0) \rangle - \langle M_d\,(t_1) \rangle = \epsilon_d \tag{11}$$

During the correction stage, PS can be adjusted to ensure $\epsilon_d = 0$ using a multiplicative ratio. For this correction, the contribution of global dry air mass from coefficients A and B are estimated as follows:

$$\begin{aligned}\langle M_A\rangle &= \text{SUM}\left[\frac{1}{g}\sum_{i_l=0}^{N_l-1}\Delta A_{i_l}\left(1-q\right)_{i_l}\right], \quad \Delta A_{i_l} = A_{i_l} - A_{i_l-1} \\ \langle M_B\rangle &= \text{SUM}\left[\frac{p_s}{g}\sum_{i_l=0}^{N_l-1}\Delta B_{i_l}\left(1-q\right)_{i_l}\right], \quad \Delta Bi_l = B_{i_l} - B_{i_l-1}\end{aligned} \tag{12}$$

Where $\langle M_A\rangle$ and $\langle M_B\rangle$ are global dry air mass components spread to A and B, respectively. When computed on t_1, they are denoted as $\langle M_A\left(t_1\right)\rangle$ and $\langle M_B\left(t_1\right)\rangle$.

The correction of p_s is defined as follows:

$$p_s^*\left(t_1\right) = p_s\left(t_0\right)\frac{\langle M_d\left(t_1\right)\rangle - \langle M_A\left(t_1\right)\rangle}{\langle M_B\left(t_1\right)\rangle} \tag{13}$$

Where $\langle M_d\left(t_0\right)\rangle$ is the total amount of global dry air mass calculated from the initial condition. $p_s^*\left(t_1\right)$ is the corrected p_s on t_1. The same multiplicative correction is applied to p_s on all grid cells.

B.2 Global moisture budget conservation scheme

For an air column, the time tendency of total precipitable water (M_v) is governed by the balance between the vertically integrated moisture flux divergence, evaporation, and total precipitation:

$$\frac{\partial}{\partial t}M_v = \frac{1}{g}\frac{\partial}{\partial t}\int_{p_1}^{p_0} q dp = -\nabla\cdot\frac{1}{g}\int_{p_1}^{p_0}(\mathbf{v}q)dp - E - P \tag{14}$$

Here, q represents the specific total water content, while E and P correspond to evaporation and total precipitation, respectively, with units of $\mathrm{kg}\cdot\mathrm{m}^{-2}\cdot\mathrm{s}^{-1}$.

On a global scale, the divergence term in equation (14) vanishes, implying that the global sum of M_v is primarily modulated by the spatially averaged evaporation ($\langle E\rangle$) and precipitation ($\langle P\rangle$), with an additional residual term ϵ_m indicating conservation errors:

$$-\left\langle\frac{\partial M_v}{\partial t}\right\rangle - \langle E\rangle - \langle P\rangle = \epsilon_m \tag{15}$$

By convention, downward precipitation results in a positive $\langle P\rangle$, whereas $\langle E\rangle$ is typically negative due to upward evaporation.

To enforce moisture budget closure, precipitation is adjusted during the correction step via a multiplicative factor:

$$P^*\left(t_1\right) = P\left(t_1\right)\frac{\langle P^*\left(t_1\right)\rangle}{\langle P\left(t_1\right)\rangle}, \quad \langle P^*\left(t_1\right)\rangle = -\left\langle\frac{M_v\left(t_1\right) - M_v\left(t_0\right)}{\Delta t}\right\rangle - \langle E\left(t_1\right)\rangle \tag{16}$$

where $\langle P^*\left(t_1\right)\rangle$ denotes the globally adjusted precipitation necessary to achieve moisture conservation. This correction is uniformly applied across all grid points.

B.3 Global total atmospheric energy conservation scheme

For a given air column, its vertically integrated total atmospheric energy (A) is defined as follows:

$$A = \frac{1}{g}\int_{p_1}^{p_0}(C_pT + L_vq + \Phi_s + k)dp \tag{17}$$

The terms on the right side of the equation (17) are thermal energy, latent heat energy, potential energy, and kinetic energy, respectively. L_v is the latent heat of vaporization, and Φ_s is the geopotential at the surface. Kinetic energy (k) is defined

as $k = 0.5\,(\mathbf{v} \cdot \mathbf{v})$. The specific heat capacity of air at constant pressure (C_p) is defined as $C_p = C_{pd}(1-q) + C_{pv}q$. The formulation of equation (17) has some limitations, which will be addressed in a separated section below.

The tendency of A is determined by the divergence of vertically integrated moist static energy ($h = C_pT + L_vq + \Phi$), kinetic energy, and other energy sources and sinks:

$$\frac{\partial}{\partial t}A = -\nabla \cdot \frac{1}{g}\int_{p_1}^{p_0} \mathbf{v}\,(h+k)dp = R_T - F_S \tag{18}$$

Where R_T and F_S are net radiation and energy fluxes on the top of the atmosphere and the surface.

$$\begin{aligned} R_T &= \mathrm{TOA}_{\mathrm{net}} + \mathrm{OLR} \\ F_S &= R_{\mathrm{short}} + R_{\mathrm{long}} + H_s + H_l \end{aligned} \tag{19}$$

Where $\mathrm{TOA}_{\mathrm{net}}$ is the top-of-atmosphere net solar radiation, OLR is outgoing longwave radiation. R_{short}, R_{long}, H_s, and H_l are the surface net solar radiation, surface net longwave radiation, surface net sensible heat flux, and surface net latent heat flux, respectively. Frictional heating is ignored in F_S.

For global sum, the divergence term in equation (18) is zero, and the global sum of the tendency of A is balanced by its energy sources and sinks, subject to a residual term:

$$\langle R_T \rangle - \langle F_S \rangle - \left\langle \frac{\partial A}{\partial t} \right\rangle = \epsilon_A \tag{20}$$

Here, the net energy flux is computed as $\langle R_T \rangle - \langle F_S \rangle$ because both terms have downward as positive. The downward on the top of the atmosphere means the energy goes "into" the atmosphere, but the downward on the surface mean the energy "leaves" the atmosphere. This is different from equation 15 where both sources and sinks are at the surface.

The air temperature (T) can be corrected to ensure thermal energy (C_pT) closes the energy budget, forcing $\epsilon_A = 0$:

$$\begin{aligned} \langle A^*(t_1) \rangle &= \langle A(t_0) \rangle + \Delta t \langle R_T \rangle - \langle F_S \rangle, \quad \gamma = \frac{\langle A^*(t_1) \rangle}{\langle A(t_1) \rangle} \\ T^*(t_1) &= \gamma T(t_1) + \frac{\gamma - 1}{C_p}\left[L_v q(t_1) + \Phi_s + k(t_1)\right] \end{aligned} \tag{21}$$

Where $\langle A^*(t_1) \rangle$ is the corrected global sum of total atmospheric energy, γ is the multiplicative correction ratio. The same γ is applied to T at all grid cells and pressure levels.

Supplemental Material

B.4 Global Spectra

The verification of the global energy spectrum is computed using spherical harmonic transforms. For a given forecasted or analyzed field $F(\phi, \lambda)$, it can be represented using spherical harmonic functions $Y(\phi, \lambda)$ as orthonormal basis and spherical harmonic coefficients (a):

$$F(\phi, \lambda) = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} a_l^m Y_l^m(\phi, \lambda) \tag{22}$$

Where degree l represents the total angular frequency of Y. m is the zonal wave number. The energy spectrum of F at a given m is the sum of magnitudes of a in all degrees with $l \geq m$:

$$P(m) = \sum_{l \geq m} \|a_l^m\|^2 \tag{23}$$

The kinetic energy ($\mathrm{m}^2 \cdot \mathrm{s}^{-2}$) and potential temperature energy (K^2) spectrum on 500 hPa pressure level were computed and as functions of m, t_i, and t_l. The result is averaged on t_i. Comparing $P(m)$ on forecasts and the ERA5 target,

the ability of weather prediction models to represent the energy transfer across scales can be verified. In addition, the energy spectrum provides a measure of the effective resolution of AI NWP models, which helps identify the smoothing effect of neural-network-based computations and model training.

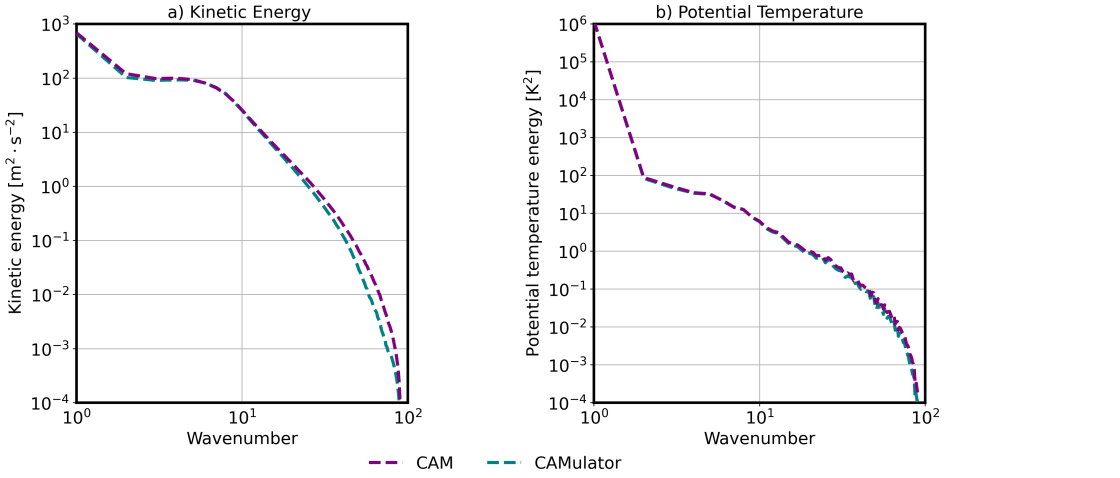

Figure 1S: Global mean kinetic energy (a) and potential temperature (b) energy spectra for CAMulator (teal), and CAM (Purple) at 500 hPa calculated for years 1979-2010.

B.5 Annual Climatolgical RMSE

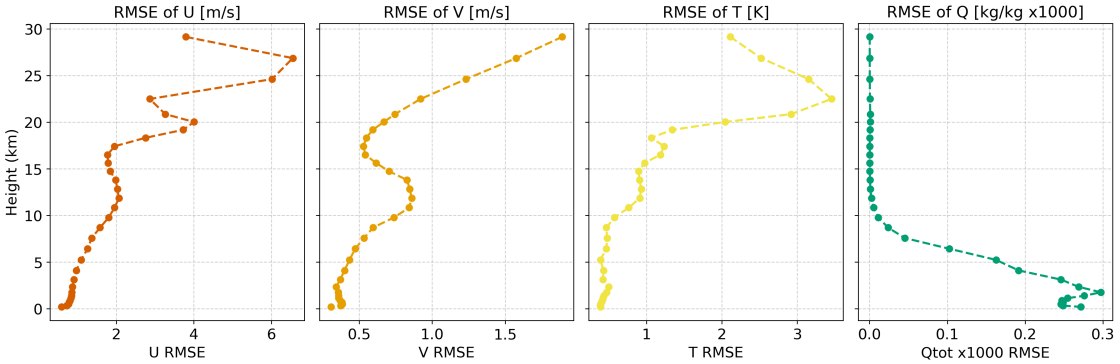

Figure 2S: Annual climatological RMSE by level for U V T and Q for CAMulator simulation run from 1979-2010 computed against the CAM climatology

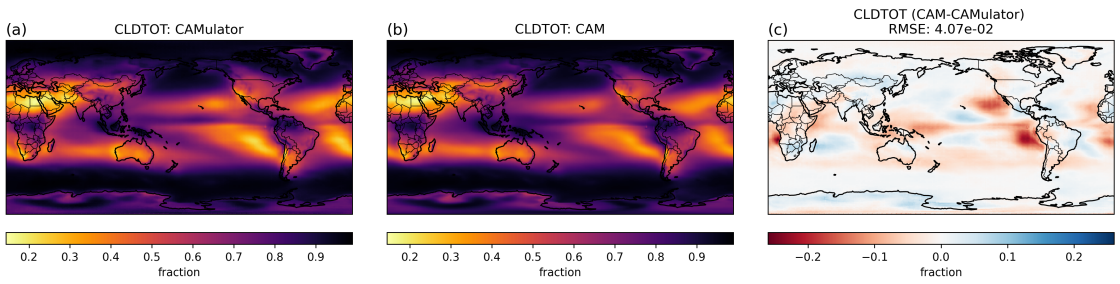

Figure 3S: Annual climatology of the total cloud cover fraction in CAMulator (a), CAM (b) and the difference (c)

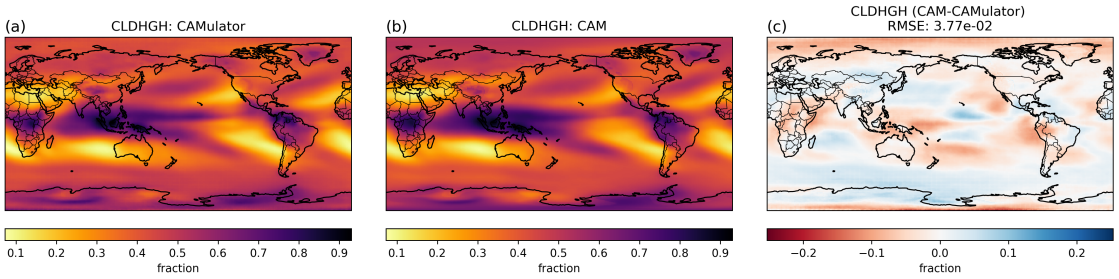

Figure 4S: As in 3S but for high cloud fraction

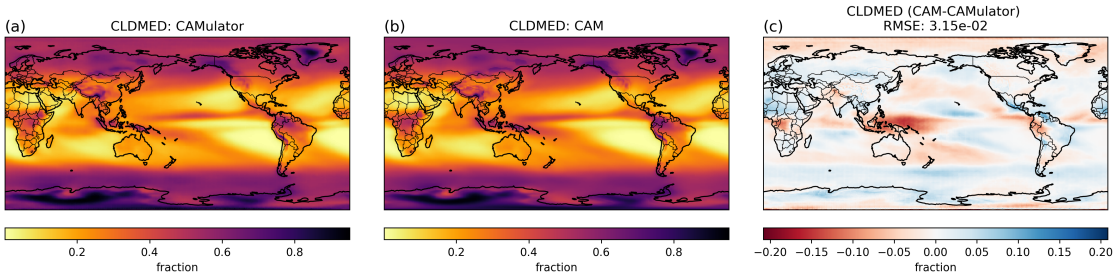

Figure 5S: As in 3S but for medium cloud fraction

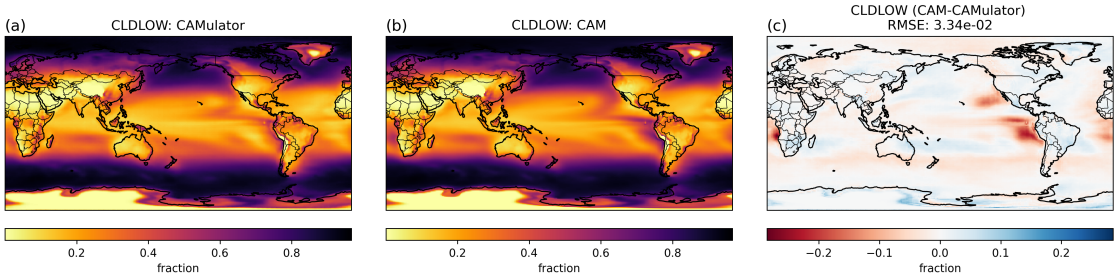

Figure 6S: As in 3S but for low cloud fraction

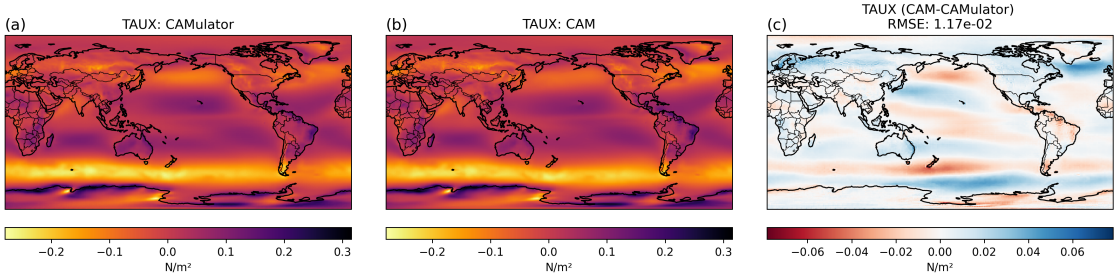

Figure 7S: As in 3S but for zonal surface wind stress

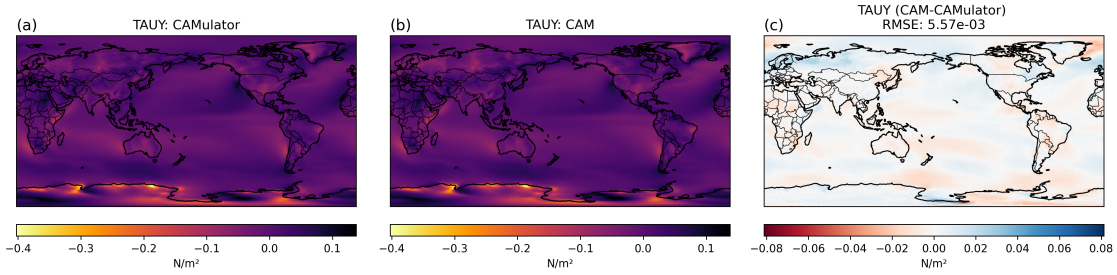

Figure 8S: As in 3S but for meridional surface wind stress

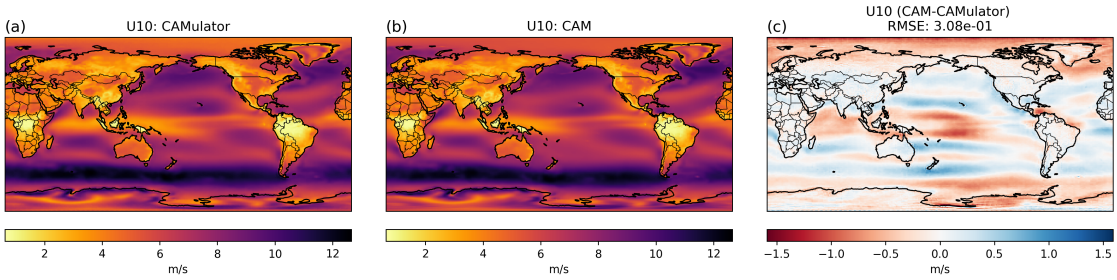

Figure 9S: As in 3S but for 10-meter wind magnitude

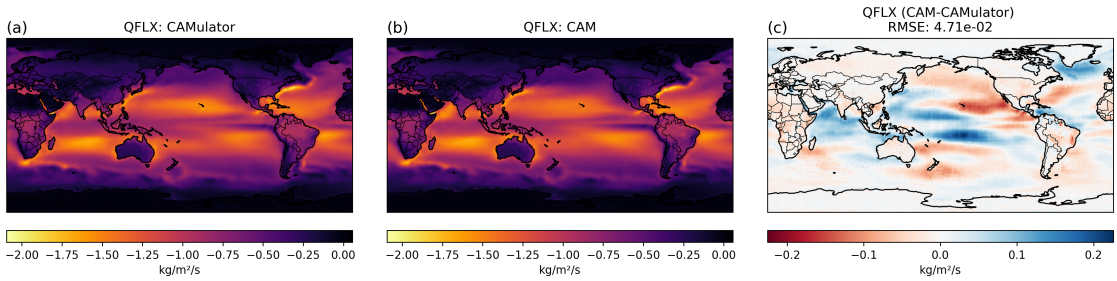

Figure 10S: As in 3S but for surface evaporation

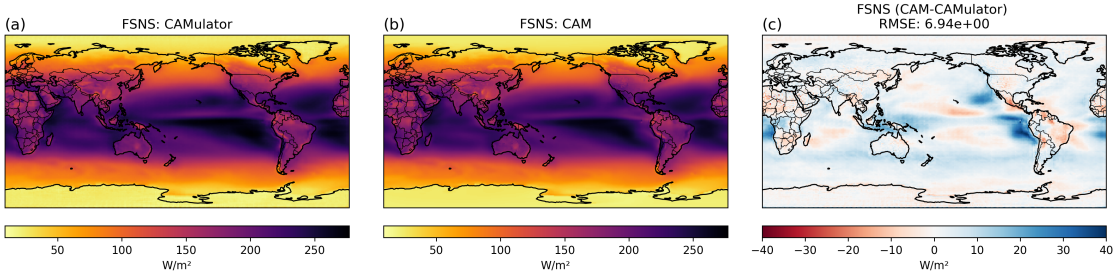

Figure 11S: As in 3S but for near-surface shortwave radiation

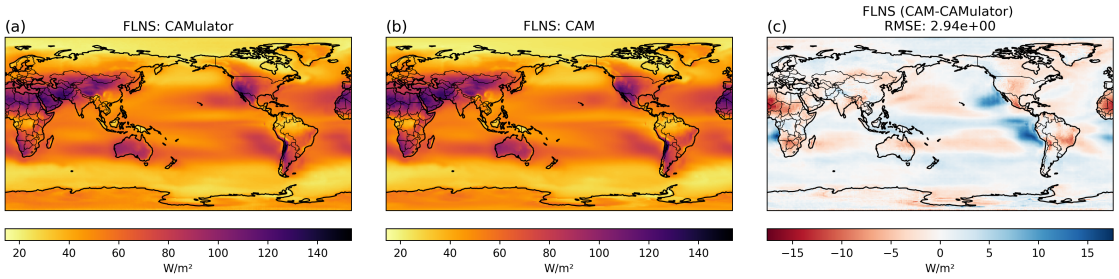

Figure 12S: As in 3S but for near-surface longwave radiation

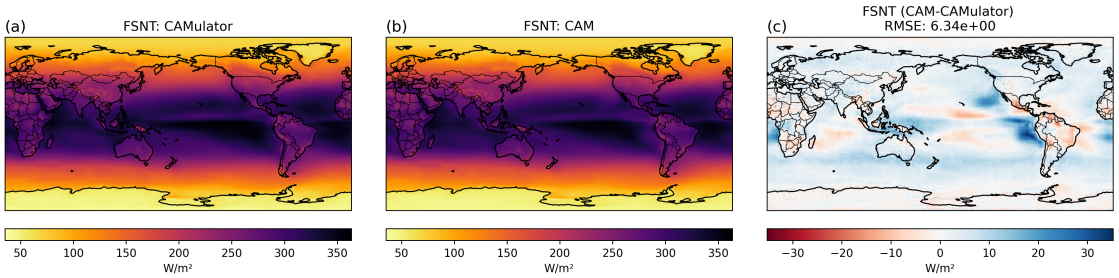

Figure 13S: As in 3S but for model top shortwave radiation

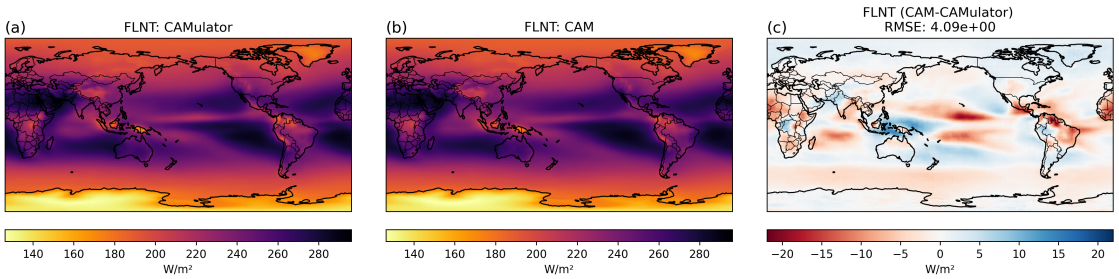

Figure 14S: As in 3S but for model top longwave radiation

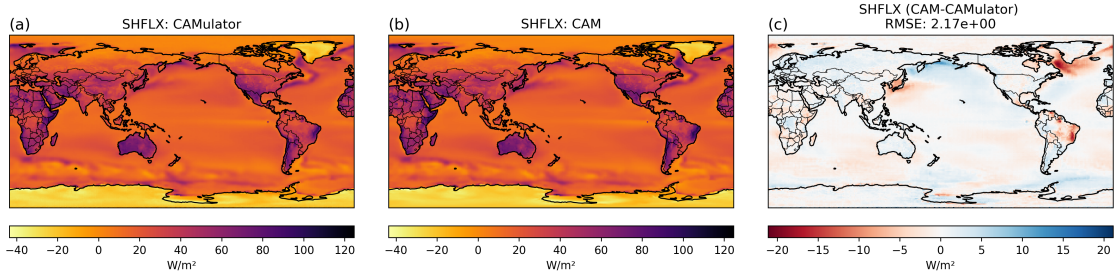

Figure 15S: As in 3S but for surface heat flux

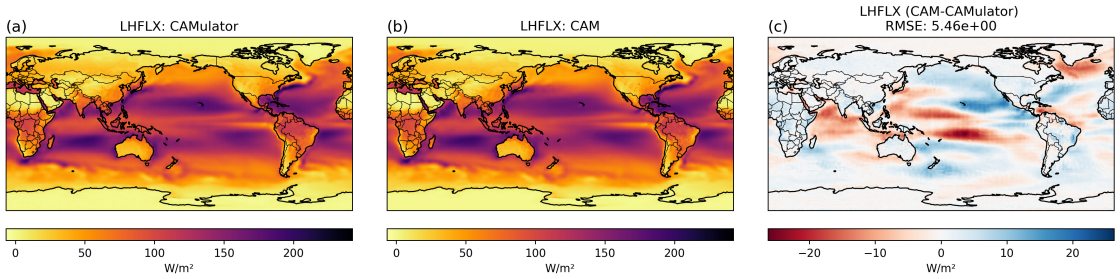

Figure 16S: As in 3S but for latent heat flux

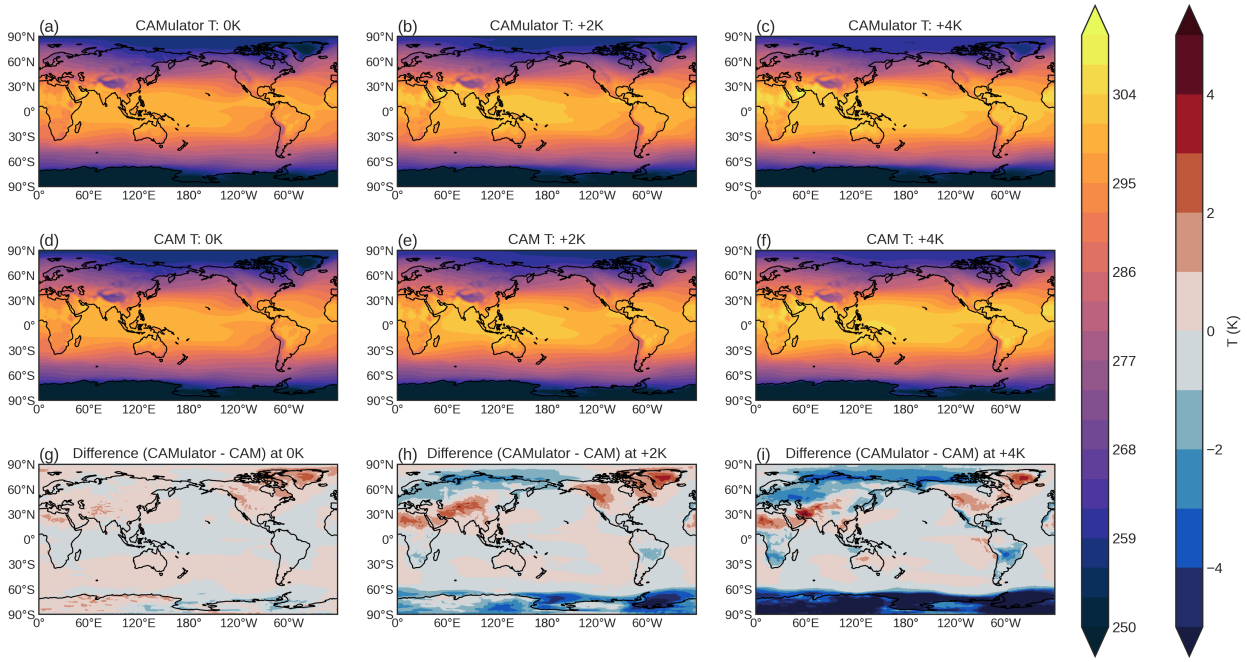

Figure 17S: Annual mean temperature at the bottom model level in CAM (middle row) and CAMulator (top row) for three sea surface temperature (SST) climatologies: 2000 (left column), 2000+2K (middle column), and 2000+4K (right column). The bottom row shows the difference (CAMulator - CAM), highlighting deviations between the two models. The color scale represents temperature (K), with positive/negative values (red/blue) indicating stronger/weaker values in CAMulator.

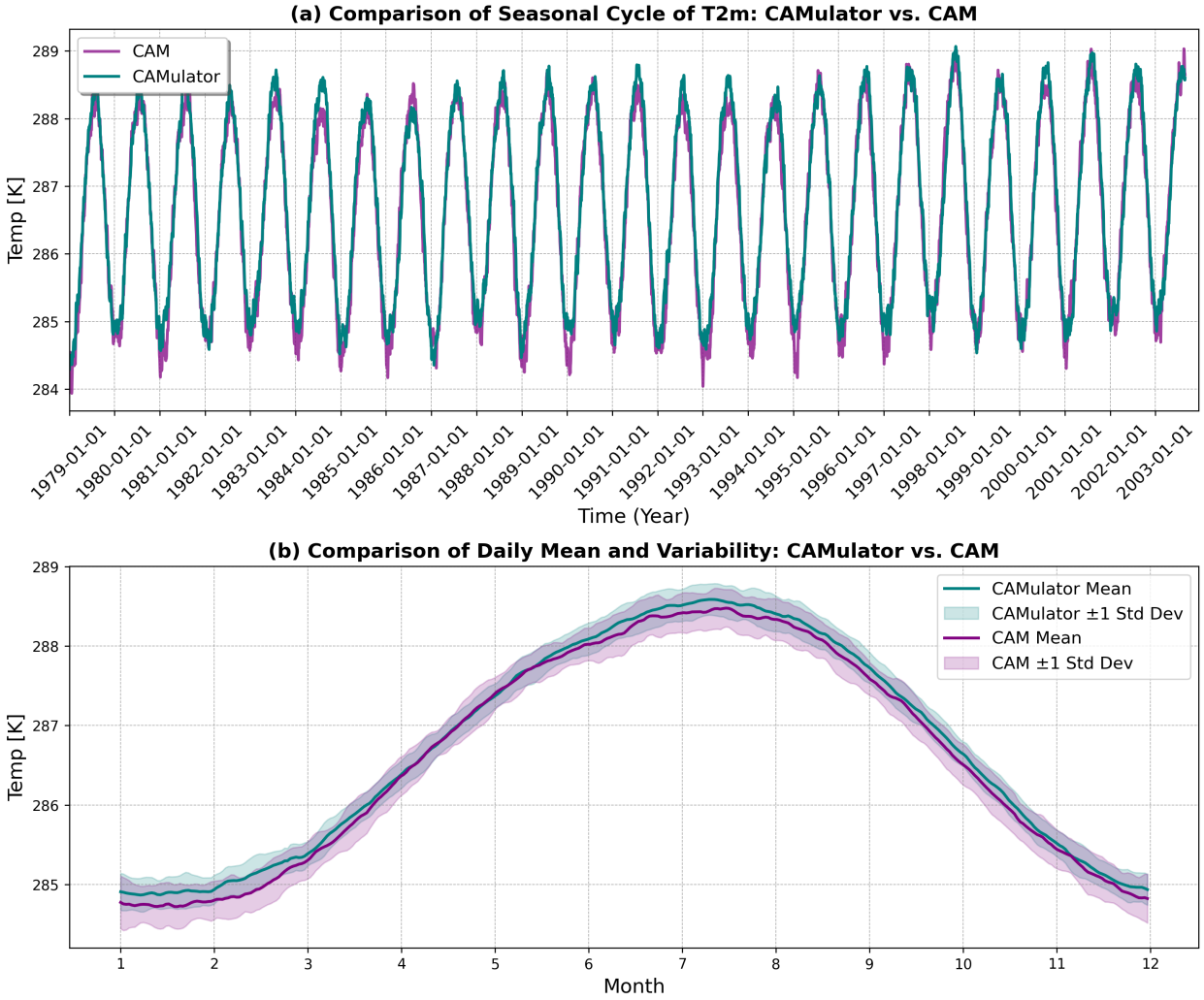

Figure 18S: Time series of the latitude-weighted global average two-meter temperature (T2m) from CAM (purple) and CAMulator (teal). (Top) Seasonal cycle of T2m from 1979 to 2003, showing interannual variations. (Bottom) Climatological annual cycle of T2m, computed as the multi-year mean, with ±1 standard deviation shading representing interannual variability in CAM (purple) and CAMulator (teal). The latitude-weighting accounts for the cosine of latitude to ensure an accurate global mean representation.

References

[1] Jennifer E Kay, Clara Deser, A Phillips, A Mai, Cecile Hannay, Gary Strand, Julie Michelle Arblaster, SC Bates, Gokhan Danabasoglu, James Edwards, et al. The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349, 2015.

[2] Nicola Maher, Sebastian Milinski, and Ralf Ludwig. Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics*, 12(2):401–418, 2021.

[3] Clara Deser, Isla R Simpson, Karen A McKinnon, and Adam S Phillips. The northern hemisphere extratropical atmospheric circulation response to enso: How well do we know it and how do we evaluate models accordingly? *Journal of Climate*, 30(13):5059–5082, 2017.

[4] Danielle Touma, Samantha Stevenson, Daniel L Swain, Deepti Singh, Dmitri A Kalashnikov, and Xingying Huang. Climate change increases risk of extreme rainfall following wildfire in the western united states. *Science advances*, 8(13):eabm0320, 2022.

[5] Surya Dheeshjith, Adam Subel, Alistair Adcroft, Julius Busecke, Carlos Fernandez-Granda, Shubham Gupta, and Laure Zanna. Samudra: An ai global ocean emulator for climate. *arXiv preprint arXiv:2412.03795*, 2024.

[6] Oliver Watt-Meyer, Gideon Dresdner, Jeremy McGibbon, Spencer K Clark, Brian Henn, James Duncan, Noah D Brenowitz, Karthik Kashinath, Michael S Pritchard, Boris Bonev, et al. Ace: A fast, skillful learned global atmospheric model for climate prediction. *arXiv preprint arXiv:2310.02074*, 2023.

[7] Oliver Watt-Meyer, Brian Henn, Jeremy McGibbon, Spencer K Clark, Anna Kwa, W Andre Perkins, Elynn Wu, Lucas Harris, and Christopher S Bretherton. Ace2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *arXiv preprint arXiv:2411.11268*, 2024.

[8] Nathaniel Cresswell-Clay, Bowen Liu, Dale Durran, Andy Liu, Zachary I Espinosa, Raul Moreno, and Matthias Karlbauer. A deep learning earth system model for stable and efficient simulation of the current climate. *arXiv preprint arXiv:2409.16247*, 2024.

[9] Francine Schevenhoven, Noel Keenlyside, François Counillon, Alberto Carrassi, William E Chapman, Marion Devilliers, Alok Gupta, Shunya Koseki, Frank Selten, Mao-Lin Shen, et al. Supermodeling: improving predictions with an ensemble of interacting models. *Bulletin of the American Meteorological Society*, 104(9):E1670–E1686, 2023.

[10] William Eric Chapman, Francine Schevenhoven, Judith Berner, Noel Keenlyside, Ingo Bethke, Ping-Gin Chiu, Alok Gupta, and Jesse Nusbaumer. Implementation and validation of a supermodelling framework into cesm version 2.1. 5. *EGUsphere*, 2024:1–21, 2024.

[11] Yingkai Sha, John S Schreck, William Chapman, and David John Gagne II. Improving ai weather prediction models using global mass and energy conservation schemes. *arXiv preprint arXiv:2501.05648*, 2025.

[12] John Schreck, Yingkai Sha, William Chapman, Dhamma Kimpara, Judith Berner, Seth McGinnis, Arnold Kazadi, Negin Sobhani, Ben Kirk, and David John Gagne II. Community research earth digital intelligence twin (credit). *arXiv preprint arXiv:2411.07814*, 2024.

[13] Graeme L Stephens, Tristan L'Ecuyer, Richard Forbes, Andrew Gettelmen, Jean-Christophe Golaz, Alejandro Bodas-Salcedo, Kentaroh Suzuki, Philip Gabriel, and John Haynes. Dreary state of precipitation in global models. *Journal of Geophysical Research: Atmospheres*, 115(D24), 2010.

[14] Angeline G Pendergrass and Dennis L Hartmann. Two modes of change of the distribution of rain. *Journal of Climate*, 27(22):8357–8371, 2014.

[15] Peter A Bogenschutz, Andrew Gettelman, Cecile Hannay, Vincent E Larson, Richard B Neale, Cheryl Craig, and Chih-Chieh Chen. The path to cam6: Coupled simulations with cam5. 4 and cam5. 5. *Geoscientific Model Development*, 11(1):235–255, 2018.

[16] Andrew Gettelman, David N Bresch, Chihchieh C Chen, John E Truesdale, and Julio T Bacmeister. Projections of future tropical cyclone damage with a high-resolution global climate model. *Climatic Change*, 146:575–585, 2018.

[17] Robert F Adler, George J Huffman, Alfred Chang, Ralph Ferraro, Ping-Ping Xie, John Janowiak, Bruno Rudolf, Udo Schneider, Scott Curtis, David Bolvin, et al. The version-2 global precipitation climatology project (gpcp) monthly precipitation analysis (1979–present). *Journal of hydrometeorology*, 4:1147–1167, 2003.

[18] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

[19] Wenxiao Wang, Wei Chen, Qibo Qiu, Long Chen, Boxi Wu, Binbin Lin, Xiaofei He, and Wei Liu. Crossformer++: A versatile vision transformer hinging on cross-scale attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[21] I Loshchilov. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[22] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[23] Judith Berner, GJ Shutts, M Leutbecher, and TN Palmer. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ecmwf ensemble prediction system. *Journal of the Atmospheric Sciences*, 66(3):603–626, 2009.

[24] Adam S Phillips, Clara Deser, and John Fasullo. Evaluating modes of variability in climate models. *Eos, Transactions American Geophysical Union*, 95:453–455, 2014.

[25] Isla R Simpson, Julio Bacmeister, Richard B Neale, Cecile Hannay, Andrew Gettelman, Rolando R Garcia, Peter H Lauritzen, Daniel R Marsh, Michael J Mills, Brian Medeiros, et al. An evaluation of the large-scale atmospheric circulation and its variability in cesm2 and other cmip models. *Journal of Geophysical Research: Atmospheres*, 125(13):e2020JD032835, 2020.

[26] S Rasp, S Hoyer, A Merose, I Langmore, P Battaglia, T Russel, A Sanchez-Gonzalez, V Yang, R Carver, S Agrawal, et al. Weatherbench 2: A benchmark for the next generation of data-driven global weather models, arxiv. *arXiv preprint arXiv:2308.15560*, 2023.

[27] IG Watterson and MR Dix. Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution. *Journal of Geophysical Research: Atmospheres*, 108(D13), 2003.