

SamudrACE: Fast and Accurate Coupled Climate Modeling with 3D Ocean and Atmosphere Emulators

James P. C. Duncan¹, Elynn Wu¹, Surya Dheeshjith², Adam Subel², Troy Arcomano¹, Spencer K. Clark^{1,4}, Brian Henn¹, Anna Kwa¹, Jeremy McGibbon¹, W. Andre Perkins¹, William Gregory³, Carlos Fernandez-Granda^{2,6}, Julius Busecke^{2,5}, Oliver Watt-Meyer¹, William J. Hurlin⁴, Alistair Adcroft³, Laure Zanna^{2,6}, Christopher Bretherton¹

¹Allen Institute for Artificial Intelligence (Ai2), Seattle, USA

²Courant Institute of Mathematical Sciences, New York University, New York, USA

³Princeton University, Princeton, USA

⁴NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, USA

⁵Lamont-Doherty Earth Observatory of Columbia University, New York, USA

⁶Center for Data Science, New York University, New York, USA

Key Points:

- SamudrACE is an AI emulator of the GFDL CM4 coupled global climate model (GCM), trained on 150 years of pre-industrial output.
- Like traditional coupled GCMs, SamudrACE's components were developed independently by distinct teams and coupled via fine-tuning.
- Running on a single NVIDIA H100 GPU, SamudrACE can simulate 800 years per day.

Abstract

Traditional numerical global climate models simulate the full Earth system by exchanging boundary conditions between separate simulators of the atmosphere, ocean, sea ice, land surface, and other geophysical processes. This paradigm allows for distributed development of individual components within a common framework, unified by a coupler that handles translation between realms via spatial or temporal alignment and flux exchange. Following a similar approach adapted for machine learning-based emulators, we present SamudrACE: a coupled global climate model emulator which produces centuries-long simulations at 1-degree horizontal, 6-hourly atmospheric, and 5-daily oceanic resolution, with 145 2D fields spanning 8 atmospheric and 19 oceanic vertical levels, plus sea ice, surface, and top-of-atmosphere variables. SamudrACE is highly stable and has low climate biases comparable to those of its components with prescribed boundary forcing, with realistic variability in coupled climate phenomena such as ENSO that is not possible to simulate in uncoupled mode.

Plain Language Summary

Climate scientists use computer models to understand how different parts of the Earth system, like the atmosphere and ocean, work together. Traditionally, these models are built with separate components that exchange information. We applied this same approach to new, much faster models based on artificial intelligence (AI). We connected an AI atmosphere model (called ACE) to an AI ocean model (called Samudra) to create a new, coupled AI model called SamudrACE, capable of simulating the full Earth climate evolution. Our combined model runs stably for centuries, producing accurate, high-quality climate simulations. A key success is that by linking the ocean and atmosphere, SamudrACE can realistically simulate complex, large-scale climate phenomena like the El Niño-Southern Oscillation (ENSO), something the individual AI models cannot do on their own. This work demonstrates a successful strategy for building fast and powerful AI-based tools to study long-term climate evolution more efficiently.

1 Introduction

The advent and success of machine learning (ML)-based weather prediction (Pathak et al., 2022; Bi et al., 2023; Lam et al., 2023) has led to similarly data-driven global atmosphere emulators trained on output of numerical models, such as the atmosphere-only version of the Ai2 Climate Emulator (ACE) (Watt-Meyer et al., 2023). Since then, atmosphere model emulators have continued to mature and support Atmosphere Model Intercomparison Project (AMIP) (Gates, 1992) compatible simulations (Watt-Meyer et al., 2024; Kochkov et al., 2024; Chapman et al., 2025).

This paper will demonstrate early progress toward the natural next step in this progression, a global climate model (GCM) emulator, which consists of modular coupled atmosphere, sea ice, land, and ocean emulators, capable of running the Coupled Model Intercomparison Program (CMIP) DECK simulation suite (Eyring et al., 2016). This could later be extended to incorporate other components of the Earth system (e.g., biogeochemical processes).

Coupled atmosphere and ocean emulation is needed to learn and generate realistic climate trends (e.g., through the time-evolving spatial patterns of ocean heat uptake and sea-surface temperature rise). It is also needed to generate the variability in physical phenomena that emerge through the realistic interaction and coupled evolution of atmospheric surface forcing and upper ocean response, such as El Niño-Southern Oscillation (ENSO) variability (Zebiak & Cane, 1987).

Several recent papers have incorporated simplified forms of ocean coupling into ML atmospheric emulators, e.g. by use of a physically-based slab ocean model (Clark et al., 2024) expressed in PyTorch (Paszke et al., 2019), or by prognostically emulating sea surface temperature (SST) (Cresswell-Clay et al., 2024), or with the addition of near-surface temperature on a limited number of upper-ocean levels (C. Wang et al., 2024). This has enabled accurate seasonal forecasts (C. Wang et al., 2024) and stable simulation of present-day (Cresswell-Clay et al., 2024) and CO₂-modulated (Clark et al., 2024) equilibrium climate. However, these ocean representations are too simplified to support accurate coupled atmosphere-ocean variability such as ENSO.

To go further toward emulation of climate-coupled phenomena requires resolving the full extent of the ocean in order to simulate ocean circulation and response to atmospheric forcing on annual and decadal timescales. Fortunately, three-dimensional ML ocean emulators have been recently developed for data-driven ocean forecasting on timescales up to 1–2 years (Chen et al., 2023; El Aouni et al., 2024; Xiong et al., 2023; X. Wang et al., 2024) and for longer-running simulations forced by specified time-evolving atmospheric conditions (Dheeshjith et al., 2025). Specifically, Samudra (Dheeshjith et al., 2025) stably emulates GFDL’s Ocean Model v4 (OM4) and reproduces ocean dynamics on decadal timescales when forced from above by the net downward heat flux (at the ocean surface, or beneath sea ice where present) and surface wind stresses. Similarly, successful emulators have been developed for sea ice (Durand et al., 2024).

However, these advances in component model emulation do not necessarily enable their coupling. So far, there has not been a data-driven approach capable of successful 3D coupled emulation of the full vertical extents of the atmosphere and ocean. Indeed, full earth system model emulation presents additional challenges beyond those of emulating individual component models. These include:

- A large increase in the number of variables, with distinct vertical coordinates for the atmosphere and ocean;
- The addition of new internal boundary conditions for coupling components in the vertical direction (sea surface) and the horizontal (land-sea-ice interactions);
- The choice of coupling procedure, which in the context of AI emulators may be done in physical state space or in the latent space of component emulator receptive fields.
- A larger range of scales of spatiotemporal variability than encountered in any one component (e.g., deep ocean heat uptake vs. weather variability);
- Emulation of emergent atmosphere-ocean coupled variability such as ENSO.

In this paper, we present SamudrACE, which is constructed by coupling the ACE2 3D atmosphere emulator to the Samudra 3D ocean emulator which has been extended to predict sea-ice concentration and thickness. Both components are emulated at 1° lat/lon horizontal resolution. We train SamudrACE to emulate a fully coupled 200-year simulation¹ by the GFDL CM4 physics-based coupled GCM with constant pre-industrial greenhouse gas and aerosol concentrations and a repeating annual cycle of insolation (Held et al., 2019).

The resulting trained emulator, SamudrACE:

- Generates stable centuries-long simulations of the coupled atmosphere and ocean with low bias;
- Realistically emulates CM4’s ENSO variability, accurately reproducing the response of the vertical structure of equatorial Pacific currents and the spatial pattern of precipitation to El Niño conditions;

¹ For convenience in the remainder of the text and figures, we use “CM4” interchangeably to denote both the coupled GCM and the 200-year simulation output from that model.)

- Accurately emulates the seasonal cycle of monthly mean sea ice extent and thickness in both the northern and southern hemispheres.

2 Materials and Methods

2.1 Component emulators: Samudra and ACE2

Analogous to CM4’s coupling (Balaji, 2011) of the components AM4/LM4 (Zhao et al., 2018, 2018) to OM4/SIS2 (Adcroft et al., 2019), SamudrACE is composed of two independent component emulators and a coupler that handles communication between the two. Specifically, SamudrACE couples the ACE2 (Watt-Meyer et al., 2024) atmosphere and land surface emulator to the Samudra (Dheeshjith et al., 2025) ocean emulator. To facilitate this coupling, both ACE2 and Samudra are pretrained from random weight initializations on the CM4 simulation outputs with prescribed forcings, a key step that we call “uncoupled pretraining” and describe in detail below. Additionally, we modified Samudra to prognose sea ice variables, including its concentration and mass, which plays a crucial role in SamudrACE’s coupler. After pretraining, ACE2 and Samudra are coupled and fine-tuned in two stages: first only updating Samudra’s weights and then jointly updating both components. SamudrACE is the end result of this pretraining, coupling, and fine-tuning pipeline. Tables S1 and S2 provide a complete listing of the SamudrACE component outputs.

2.2 The SamudrACE coupler

Figure 1c provides a schematic of the SamudrACE coupler. The ACE2 model simulates the atmosphere in 6-hour increments over a 5-day period. At the end of each step, core prognostic variables, such as temperature, are saved as an instantaneous snapshot. In contrast, diagnostic boundary fluxes are calculated as the average value over that same 6-hour interval, with the averaging period ending at the snapshot time.

For each forward step, ACE2 is forced by the sea ice fraction and SST from time t . Once ACE2 completes 20 forward steps the coupler aggregates the 6-hour average boundary fluxes into a single 5 day average. The generated ocean surface fluxes of energy, moisture and momentum are then used to force a single step of Samudra, which evolves the ocean state forward in time by 5 days to catch up to the atmosphere. The new SST state is then prescribed onto the final generated global surface temperature T_S state in preparation to resume atmosphere forward stepping. This coupling loop is then repeated for the length of the simulation.

The 6-hour and 5-day timesteps chosen for the atmosphere and ocean emulators reflect their respective timescales of change on the O(100 km) grid scale used in this study. Using a 20-fold longer ocean time step than atmosphere time step introduces complications in training of the coupled emulator, but also enables simulations that have both realistic variability in both components. The SamudrACE coupler is heavily influenced by GFDL’s Flexible Modeling System (FMS) coupler (Balaji, 2011) used in CM4 simulations. However, unlike in typical physics-based coupled models, ACE2 internally predicts the surface flux components, rather than these being computed by the ocean or land surface models.

Our choice of coupling procedure is an important physics-informed design element of SamudrACE. The flow of information at component interfaces is column-local and constrained so that the atmosphere forces the ocean with surface fluxes and the ocean forces the atmosphere with SST and sea ice. Enforcing this exchange of information between components in physical state space via a coupling procedure avoids the exchange of nonphysical latent representations between the two models. ACE2 and Samudra were each designed from the start with coupling in mind by learning the diagnostic variables (surface turbulent and radiative fluxes, plus surface precipitation) needed for a physically justifiable coupling procedure and temperature and salinity budget closure.

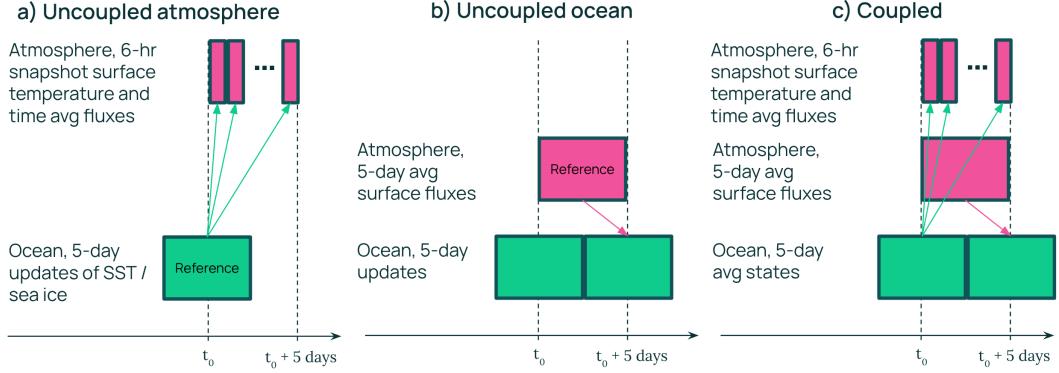


Figure 1. A single 5-day forward step in uncoupled (a, b) and coupled (c) modes. In uncoupled mode, ACE and Samudra are forced by the appropriate CM4 reference fields. Uncoupled ACE is forced by the reference 5-day average SST and sea ice concentration, stepping forward 6 hours at a time until reaching 5 days, at which point the next 5-day average forcing is prescribed. For uncoupled Samudra, we use the 5-day average of the CM4 reference wind stress, precipitation, and surface fluxes as prescribed forcing inputs. In coupled mode, aggregation of the diagnostic 6-hour average surface boundary conditions to 5-day averages is done online as ACE completes 20 forward steps. The generated 5-day average is then passed as input to Samudra, which generates a single 5-day forward step. Samudra’s generated SST and sea ice states will then be used to force ACE in the next iteration of the coupler loop.

2.3 Pretraining ACE2

We follow ACE2’s training protocol (Watt-Meyer et al., 2024) with two additional diagnostic variables – surface zonal and meridional wind stress. Training data come from the atmosphere component output of CM4 at 6-hourly temporal resolution, with the exception of surface temperature over ocean, sea ice fraction f_i , and ocean fraction f_o (derived from f_i). For each 5-day window aligned with the ocean data’s timesteps, the ocean-covered grid cells of 6-hourly T_S are held fixed at their values at the beginning of the 5-day window, while allowing grid cells with sea ice or land to evolve with the usual 6-hour timestep. Similarly, all 20 snapshots of f_i and f_o will be identical to the first snapshot. ACE2 is trained for 50 epochs with a batch size of 16 (707,200 gradient steps) and a learning rate of 10^{-4} . ACE2 pretraining loss and inference RMSE is given in Figure S1b.

2.4 Pretraining SamudraI

Our implementation of Samudra (Dheeshjith et al., 2025) modifies the original protocol as follows:

- 1. Pretraining dataset:** We pretrain Samudra on 5-day mean OM4 output. For consistency, this is taken from the same segment of our reference CM4 simulation as used for ACE2 pretraining.
- 2. Prognostic Variables:** We expand the set of prognostic variables to include sea ice concentration (SIC) and thickness, coining the new emulator SamudraI, in addition to the original potential temperature, salinity, sea surface height, and ocean velocities.
- 3. Forcing Conditions:** The model is forced using surface conditions from ACE above the sea ice. This includes wind stress and a full suite of heat and water fluxes: upward and downward shortwave and longwave radiation, latent and sensible heat fluxes, and precipitation. The original Samudra was forced only by the top-of-ocean heat flux

and surface stress, calculated below any sea ice. All forcings are time-averaged over 5-day blocks.

4. **Time Stepping:** We simplify the model to use a single prior time step as input to predict a single future time step ($P = 1$), reduced from the original two ($P = 2$). The model remains optimized across four predicted steps ($N = 4$).

To achieve a fully coupled global climate emulator, we must also predict sea ice. Inspired by the FMS coupler, we treat sea ice as the interface between the ocean’s slow time step and the atmosphere’s fast time step. SamudraI updates the sea ice concentration state every 5 days, in turn determining where ACE2 is allowed to prognose surface temperature over sea ice for the subsequent 5-day window. We use sea ice concentration – defined as the sea ice grid cell fraction divided by (1 - land grid cell fraction) – in order to avoid over-representing grid points with minimal ocean cover, and find that doing so reduces biases near coastlines where ocean fraction is small. SamudraI is trained for 150 epochs with a batch size of 16 (106,050 gradient steps) and a learning rate of 10^{-4} . SamudraI pretraining loss and inference RMSE is given in Figure S1a.

2.5 Fine-tuning SamudrACE

Once ACE2 and SamudraI pretraining is complete, we select for coupled fine-tuning the checkpoint with the lowest normalized channel-mean root mean square error (RMSE) over all epochs for each respective model. Together, the coupled emulators have a combined total of nearly 600 million parameters.

Fine-tuning of these checkpoints in coupled mode proceeds in two stages. In the first stage, we fine-tune SamudraI by optimizing the loss of four forward steps (20 days) while coupled to the pretrained ACE2 model. Throughout this first phase of coupled fine-tuning, ACE2’s model weights are held fixed and atmospheric fields do not contribute to the training or validation loss. However, as seen in Figure S1c, atmosphere fields do contribute to the “best” checkpoint selection, which is the SamudraI checkpoint leading to the lowest normalized channel-mean RMSE across all ocean and atmosphere fields.

In the second stage, we continue fine-tuning with the best SamudraI checkpoint found in stage one. As the ACE2 model was not updated during the first phase, its stage two initialization is identical to the best checkpoint selected during pretraining. In contrast to the first stage, during the second stage the model weights for ACE2 and SamudraI are jointly optimized, with four 5-day ocean steps and two 6-hour atmosphere steps entering the optimization (Figure S1d). In addition, we switch from a constant learning rate to a cosine-annealing learning rate decay schedule (Loshchilov & Hutter, 2016) with an initial learning rate of 10^{-5} , ten times smaller than that used during pretraining and the first phase of fine-tuning.

2.6 Datasets

Our reference training and evaluation datasets are from a 200-year preindustrial control simulation from GFDL’s Climate Model v4 (CM4) (Held et al., 2019), rerun to save high frequency data, starting from year 151 of the original CMIP6 simulation. CM4 was run with a \sim 100 km resolution C96 atmosphere with 33 terrain-following vertical levels and a 0.25 degree ocean on a tripolar grid with 75 hybrid pressure/isopycnal vertical coordinate levels. Both datasets were conservatively remapped to a 1 degree Gaussian grid with 8 terrain-following atmospheric layers and 19 ocean layers with constant-depth interfaces for emulator training and testing. We use the first 155 years of output for training, the next 5 years for validation, and the remaining 40 years of data are held out for testing.

The reference AM4 atmospheric fields and SIS2 sea ice concentration were output fully consistently with ACE, with instantaneous snapshots of prognostic variables output every

six hours. All surface and top-of-atmosphere fluxes (including precipitation) were output as six-hour time-averages between these snapshots. This enables the surface fluxes to be accumulated over 20 atmospheric steps into 5-day averages suitable for forcing the ocean emulator.

The reference OM4 ocean fields were all output as 5-day averages, rather than instantaneous snapshots. This includes the sea-surface temperature, used to force the ACE atmosphere model. For the purposes of this paper, these 5-day averages are used as an estimate of the ocean state at the midpoint of the averaging interval.

Achieving consistent land, ice, and open ocean masking between the emulators is a crucial step, with important implications for coupling. Just as a traditional coupled GCM must reconcile differences in horizontal resolution when coupling the ocean to the atmosphere, GCM emulators must do so either during data preparation or via an online procedure. In this work, we emulate CM4 with a common 1 degree horizontal resolution for both the atmosphere and ocean and therefore handle mask alignment during the data preparation step. However, due to differences in the pipelines for 6-hourly atmosphere and 5-daily ocean data processing we pretrained ACE2 and SamudraI with distinct versions of sea ice concentration. This in turn led to larger than expected biases in surface temperature near coastlines in polar regions upon initial coupling of pretrained ACE2 to pretrained SamudraI. These biases were removed with the additional coupled fine-tuning of ACE2 weights, as described above. We have since corrected this error in data processing and are in the process of retraining ACE2 and SamudraI.

3 Results

SamudrACE can stably emulate CM4 with low bias in the time series of global means and time mean spatial error patterns over a held-out 40-year inference period covering the final part of the simulation. These biases are of similar magnitude to the biases of the uncoupled component emulators. Using a single NVIDIA H100 GPU, a SamudrACE simulation generates around 800 SYPD (3.4 days per second), a $1730 \times$ decrease in energy usage when compared to the CM4 simulation at 16 SYPD (0.068 days per second) using 5535 CPU cores on AMD EPYC 7H12 processors. Beyond the brief summary provided in Sections 3.1 –3.2, we provide a comprehensive accounting of bias and other details in the Supplementary Material, similar to previous analyses of ACE2 and Samudra.

3.1 Climate mean state

Figures 2a-b show that SamudrACE faithfully reproduces CM4’s time-mean precipitation and surface temperature over the 40-year held-out test period. Time mean biases are small for both precipitation and surface temperature around the globe, with absolute maxima of just over 2 mm/day or 3 K, respectively, and small RMSEs. Precipitation biases are concentrated in the tropics while surface temperature biases are pronounced in areas of sea ice and topographic features.

In general, SamudrACE is a highly skilled emulator of climate mean states, not unlike its component ML emulators on their own. SamudrACE’s overall time mean state is accurate in terms of the RMSE of the generated time mean (Figure S2), globally averaged annual mean series (Figures S3 and S4), and near-surface zonal sea water currents (Figures S5 and S6).

The Atlantic Meridional Overturning Circulation (AMOC) is an important measure of time mean state on climate timescales. As a critical pathway for carrying warm water into the North Atlantic, the AMOC plays a vital role in regulating the global climate. Figure S12(a) shows that SamudrACE emulates the vertical structure of CM4’s AMOC with low bias in the 200-year time mean state. Figure S12(b) gives the corresponding

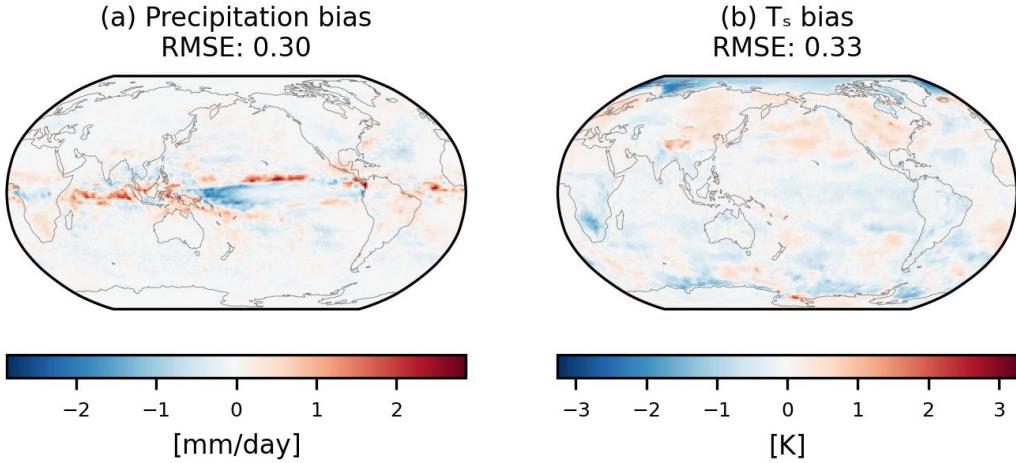


Figure 2. Spatial bias patterns of the generated 40-year time-mean precipitation and surface temperature, computed as the difference between the time mean of the emulated output and the time mean of the reference simulation over the held-out inference period.

time series of AMOC anomalies in uncoupled SamudraI, SamudrACE and CM4. Whereas uncoupled SamudraI lacks sufficient variability in the AMOC anomaly series, SamudrACE shows similar low frequency variability to CM4. As seen in Figure S4, global ocean heat content in SamudrACE is stable over time and does not drift in a way that is inconsistent with CM4.

3.2 Sea ice climatology

Overall, SamudrACE predicts a stable and accurate sea ice climatology with a realistic seasonal cycle. Sea ice extent is determined based on grid points where sea ice fraction exceeds 15%. Sea ice fraction shown here is derived from SamudrACE's native sea ice concentration described in Section 2.4. Figures 3a-b show monthly Northern and Southern hemispheres sea ice extent averaged over the 40-year held-out period. The shading indicates the interannual standard deviation. SamudrACE skillfully simulates the CM4 seasonal cycle with minimal bias, but it has smaller interannual variability compared to the CM4 target in the Southern Hemisphere. We find similarly muted interannual variability in sea ice thickness (Figure S13). Figure 3c-d) shows the time-mean sea ice fraction for the Arctic and Antarctica for the CM4 target, SamudrACE, and the bias. SamudrACE tends to have a larger bias in sea-ice concentration near the edge of where sea ice exists. Notably, the highest positive bias is in the Greenland Sea, corresponding to where SamudrACE also simulates a weaker surface wind stress (not shown).

3.3 ENSO

Figure 4 compares key characteristics of ENSO simulated by CM4 and SamudrACE. SamudrACE shows promisingly similar ENSO variability to CM4, with realistically large amplitude El Niño and La Niña events having a similar pattern of spatial variability and a 3-year peak in the temporal power spectrum.

Panel (a) compares time series of the Niño 3.4 index – SST anomaly with climatology removed, averaged over 5°N - 5°S , 170°W - 120°W (Barnston et al., 1997) – over the full CM4 target (black) with 200-year SamudrACE rollouts, initialized from the CM4 state on January 1 of 5 successive years in the validation period.

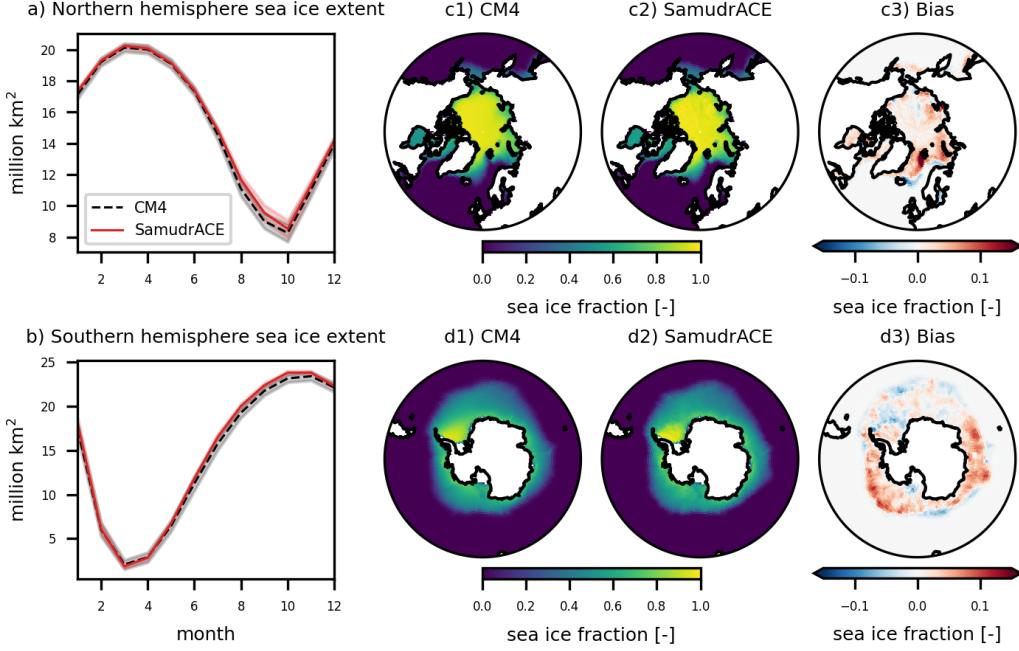


Figure 3. Monthly mean over the 40-year held-out period of (a) Northern and (b) Southern Hemisphere sea ice extent. Shading denotes the interannual standard deviation over 40 years. Panel c-d) shows the time mean sea ice fraction over the same time period for the CM4 target, SamudrACE, and its bias.

The SamudrACE rollouts (colors) have comparable but slightly weaker variability than CM4, with a bias toward sharper warm anomalies (El Niños) and weaker cold anomalies (La Niñas). With an overall minimum of -1.6 K and maximum of 2.1 K, SamudrACE has a somewhat weaker La Niña than seen in CM4 but has comparably large amplitude El Niño events. SamudrACE rollouts also appear more susceptible to sustained decade-long periods of unrealistically weak ENSO activity than CM4, e.g. around rollout years 115 to 135 in the run initialized from 0311-01-01 and centered at year 75 in the 0317-01-01 run. Figure S7 further confirms SamudrACE's relative lack of strong La Niña events.

Figure 4b compares power spectra of the monthly CM4 and SamudrACE Niño 3.4 index. To arrive at the spectra for CM4 (black line), we split the 200-year simulation period into five 40-year segments, computed the spectrum for each segment, and took their average. We similarly computed the spectra of the five 200-year SamudrACE rollouts. With the exception of the rollout initialized from year 0313, the SamudrACE generated power spectrum shows too much power for periods between 2 to 4 years compared to what is observed in the CM4 simulation. This is reflected in Figure S10, where we see that SamudrACE has unrealistically high autocorrelation at 36 months. For all initial conditions, SamudrACE loses low-frequency power relative to CM4 beyond periods of 4 years.

Panel (c) shows a close match between the regressions of the spatial pattern of precipitation on the Niño 3.4 index in CM4 and SamudrACE for the held-out 40-year inference period. This is a key aspect of the spatial structure of ENSO. In the Supplementary Information we further compare the temporal and spatial structure of SST anomalies in SamudrACE and CM4 for the 40-year held-out test period. Despite modestly underestimating La Niña amplitude, SamudrACE generates a fairly similar vertical structure of anomalous warming

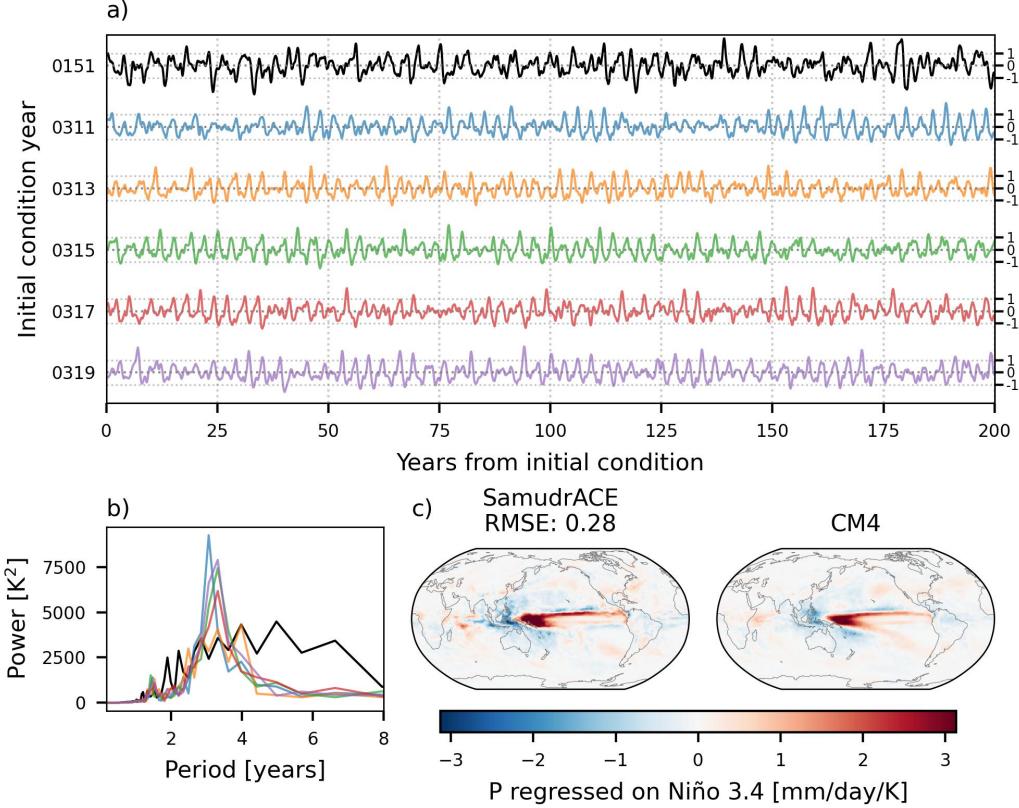


Figure 4. ENSO characteristics in the 200-year CM4 simulation (black) and 5 separate rollouts of SamudrACE starting from different initial conditions (colors): (a) time series of monthly mean Niño 3.4 index; (b) corresponding temporal power spectra, averaged over 5×40 -year segments for each 200-year simulation; (c) regression of the spatial pattern of precipitation on the Niño 3.4 index in CM4 and SamudrACE for the held-out 40-year inference period.

to CM4 in the equatorial Pacific in the two extremes of ENSO conditions (Figures S8 and S9).

3.4 Decadal coupled variability and the Interdecadal Pacific Oscillation

The Interdecadal Pacific Oscillation (IPO) is a pattern of Pacific Ocean variability characterized by multiyear-to-decades-long periods of sustained anomalous SSTs, with important implications for wildlife ecology in the Pacific Northwestern U.S. and Canada (Mantua et al., 1997). Like ENSO, the IPO provides an important feature of emergent coupled variability, but at much lower periodicity. Figure S11 compares annual mean Pacific Ocean SST anomalies by computing the IPO Tripole Index (Henley et al., 2015) for the 200-year CM4 simulation and a 200-year rollout of SamudrACE. SamudrACE has a clear pattern of sustained anomalous SST with transitions between extreme polarities on timescales that are similar to those seen in CM4. However, SamudrACE generally shows less low-frequency IPO variability than CM4.

4 Conclusions

We have presented SamudrACE, a coupled global climate model emulator created by linking the ACE2 3D atmosphere and Samudra 3D ocean emulators. Our work demonstrates a successful strategy for building stable, data-driven Earth system models capable of generating centuries-long fully coupled simulations. SamudrACE maintains low climate biases comparable to its uncoupled component model emulators while running orders of magnitude faster than the traditional coupled numerical model it emulates, GFDL-CM4.

A key achievement of this work is the realistic simulation of emergent climate phenomena that arise directly from atmosphere-ocean interaction. By coupling the two components, SamudrACE can generate realistic ENSO variability, including its characteristic power spectrum and the associated teleconnections to global precipitation patterns, a feat which was not possible for the uncoupled emulators. Similarly, the model produces a stable and accurate seasonal cycle of sea ice extent in both hemispheres. While promising, our analysis also reveals areas for future improvement. In ENSO and other coupled phenomena such as the IPO, the emulator generally underestimates low-frequency variability on time scales longer than 4 years. Future efforts could focus on refining the fine-tuning strategy or model architecture to address these biases.

In our approach which couples the component emulators in physical state space, we found that even very minor differences in the data processing pipelines for atmosphere and ocean fields led to unphysical biases that could only be removed with additional coupled fine-tuning of both components. However, in ongoing work we have already observed that these biases disappear with careful data preparation alone, reducing the cost of coupled fine-tuning. Alternative coupling approaches in component emulator latent space may be able to circumvent these issues, at the cost of lowered physical interpretability. It may also be possible to avoid offline horizontal regridding of ocean and sea-ice outputs by direct emulation of high-resolution fields and online deterministic or learned regridding.

The successful coupled GCM emulation framework of SamudrACE provides a clear pathway toward emulating a complete Earth system by incorporating additional components, such as land and biogeochemical models, opening new avenues for efficient, large-ensemble climate studies. As a natural next step, future work could explore the incorporation a sea-ice emulator able to prognose sea ice concentration at the original 6-hourly temporal resolution used in forcing uncoupled ACE2, and for handling atmosphere-ocean flux and momentum exchange.

Open Research Section

Prior to publication, the source code for uncoupled pretraining, coupled fine-tuning, and evaluation of SamudrACE will be released in a future version of the open source repository <https://github.com/ai2cm/ace>. The SamudrACE model weights will be uploaded as public artifact on Hugging Face and Zenodo. Initial conditions and incoming solar radiation will be included in the Zenodo repository. The full processed 200-year CM4 simulation will be made publicly available.

Acknowledgments

Ai2 is supported by the estate of Paul G. Allen. This research received support through Schmidt Sciences, LLC, under the M²LInES project. We thank all members of the M²LInES team for helpful discussions and their support throughout this project. This research was also supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. We also thank members of the NOAA Geophysical Fluid Dynamics Laboratory and Princeton University community for their support, specifically V.

Ramaswamy for allocating GFDL computing resources for the CM4 reference simulation, Mitch Bushuk, Huan Guo, Pu Lin, and Sergey Malyshev for their help as we discussed which diagnostics to output from the CM4 run, and Linjiong Zhou for constructive comments on an earlier draft of this manuscript. Lastly, we thank Nathaniel Cresswell-Clay and co-authors of Cresswell-Clay et al. (2024) for sharing SST outputs from the DLESyM coupled atmosphere and SST emulator. We acknowledge the use of AI tools in the preparation of this work and manuscript, specifically, for code generation and for stylistic edits of the manuscript draft for improved clarity and succinctness, using the following models: Google Gemini 2.5 Flash and 2.5 Pro, and Anthropic Claude Sonnet 4 and Opus 4.1.

Conflict of Interest

The authors declare there are no conflicts of interest for this manuscript.

References

- Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., ... others (2019). The gfdl global ocean and sea ice model om4. 0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, 11(10), 3167–3211.
- Balaji, V. (2011). The flexible modeling system. In *Earth system modelling-volume 3: Coupling software and strategies* (pp. 33–41). Springer.
- Barnston, A. G., Chelliah, M., & Goldenberg, S. B. (1997). Documentation of a highly enso-related sst region in the equatorial pacific. *Atmosphere-Ocean (Canadian Meteorological & Oceanographic Society)*, 35(3).
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970), 533–538.
- Chapman, W. E., Schreck, J. S., Sha, Y., Gagne II, D. J., Kimpara, D., Zanna, L., ... Berner, J. (2025). Camulator: Fast emulation of the community atmosphere model. *arXiv preprint arXiv:2504.06007*.
- Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj climate and atmospheric science*, 6(1), 190.
- Clark, S. K., Watt-Meyer, O., Kwa, A., McGibbon, J., Henn, B., Perkins, W. A., ... Harris, L. M. (2024). Ace2-som: Coupling to a slab ocean and learning the sensitivity of climate to changes in co _2. *arXiv preprint arXiv:2412.04418*.
- Cresswell-Clay, N., Liu, B., Durran, D., Liu, A., Espinosa, Z. I., Moreno, R., & Karlbauer, M. (2024). A deep learning earth system model for stable and efficient simulation of the current climate. *arXiv preprint arXiv:2409.16247*.
- Dheeshjith, S., Subel, A., Adcroft, A., Busecke, J., Fernandez-Granda, C., Gupta, S., & Zanna, L. (2025). Samudra: An ai global ocean emulator for climate. *Geophysical Research Letters*, 52(10), e2024GL114318.
- Durand, C., Finn, T. S., Farchi, A., Bocquet, M., Boutin, G., & Ólason, E. (2024). Data-driven surrogate modeling of high-resolution sea-ice thickness in the arctic. *The Cryosphere*, 18(4), 1791–1815.
- El Aouni, A., Drillet, Y., Drevillon, M., & van Gennip, S. (2024). Glonet-mercator's end-to-end neural ocean forecasting system. In *Agu fall meeting abstracts* (Vol. 2024, pp. OS34A-02).
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958.
- Gates, W. L. (1992). An ams continuing series: Global change-amip: The atmospheric model intercomparison project. *Bulletin of the American Meteorological Society*, 73(12), 1962–1970.

- Held, I., Guo, H., Adcroft, A., Dunne, J., Horowitz, L., Krasting, J., ... others (2019). Structure and performance of gfdl's cm4. 0 climate model. *Journal of Advances in Modeling Earth Systems*, 11(11), 3691–3727.
- Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K. (2015). A tripole index for the interdecadal pacific oscillation. *Climate dynamics*, 45(11), 3077–3090.
- Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., ... others (2024). Neural general circulation models for weather and climate. *Nature*, 632(8027), 1060–1066.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... others (2023). Learning skillful medium-range global weather forecasting. *Science*, 382(6677), 1416–1421.
- Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the american Meteorological Society*, 78(6), 1069–1080.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... others (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., ... others (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- Wang, C., Pritchard, M. S., Brenowitz, N., Cohen, Y., Bonev, B., Kurth, T., ... Pathak, J. (2024). Coupled ocean-atmosphere dynamics in a machine learning earth system model. *arXiv preprint arXiv:2406.08632*.
- Wang, X., Wang, R., Hu, N., Wang, P., Huo, P., Wang, G., ... others (2024). Xihe: A data-driven model for global ocean eddy-resolving forecasting. *arXiv preprint arXiv:2402.02995*.
- Watt-Meyer, O., Dresdner, G., McGibbon, J., Clark, S. K., Henn, B., Duncan, J., ... Bretherton, C. S. (2023). ACE: A fast, skillful learned global atmospheric model for climate prediction.
doi: 10.48550/arXiv.2310.02074
- Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., ... Bretherton, C. S. (2024). Ace2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *arXiv preprint arXiv:2411.11268*.
- Xiong, W., Xiang, Y., Wu, H., Zhou, S., Sun, Y., Ma, M., & Huang, X. (2023). Ai-goms: Large ai-driven global ocean modeling system. *arXiv preprint arXiv:2308.03152*.
- Zebiak, S. E., & Cane, M. A. (1987). A model el niño–southern oscillation. *Monthly Weather Review*, 115(10), 2262–2278.
- Zhao, M., Golaz, J.-C., Held, I., Guo, H., Balaaji, V., Benson, R., ... others (2018). The gfdl global atmosphere and land model am4. 0/lm4. 0: 1. simulation characteristics with prescribed ssts. *Journal of Advances in Modeling Earth Systems*, 10(3), 691–734.

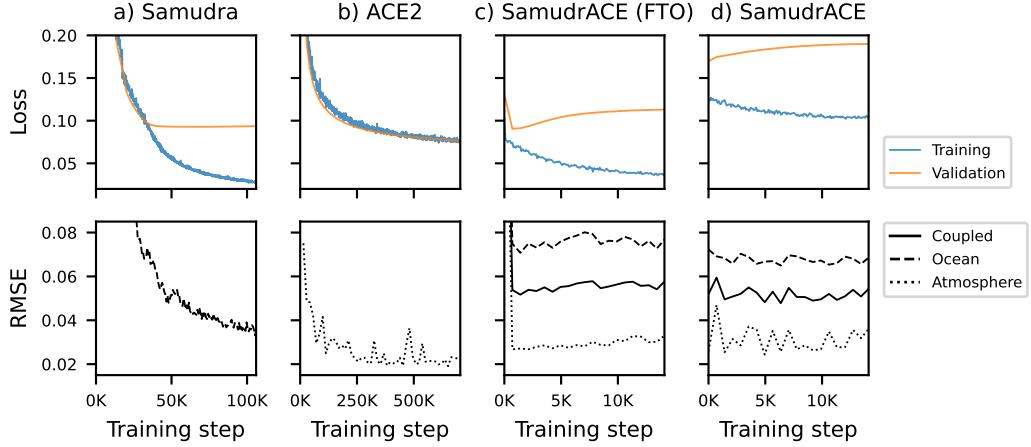


Figure S1. Training and validation loss and channel-mean RMSE. Samudra and ACE2 are pretrained in uncoupled mode for 150 epochs (106,050 steps) and 50 epochs (707,200 steps), respectively, and both achieve minimum channel-mean RMSE and validation loss late in training. After pretraining, ocean-only coupled fine-tuning (“SamudrACE (FTO)”) was carried out for 20 epochs (14,140 steps), during which the Samudra checkpoint with the lowest RMSE was coupled to the pretrained and fixed ACE2 checkpoint with lowest RMSE. At initialization the coupled pretrained emulators result in large validation loss and coupled RMSE (0.13 and 0.23, respectively, at training step 0). After 2 completed epochs the ocean-only coupled fine-tuning achieved its lowest coupled RMSE of 0.052, averaged over all ocean and atmosphere channels. This checkpoint was then further fine-tuned for an additional 20 epochs, updating both Samudra and ACE2 (“SamudrACE”), and reached the lowest coupled RMSE of 0.048 after 9 completed epochs. RMSEs are averaged across all channels, where the “Coupled” RMSE (solid black line) is the weighted average of the “Ocean” (dashed black line) and “Atmosphere” (dotted black line) channel-mean RMSEs in coupled fine-tuning runs.

Table S1. Description of SamudrACE sea ice and ocean variables shown in the left column of Figure S2. All variables are prognostic (input and output) and 5-day time averaged. Three dimensional variables have 18 levels, denoted by the subscript $k \in \{0 \dots, 18\}$. Additional variables beyond what was included in Dheeshjith et al. (2025) have descriptions highlighted in bold.

Symbol	Description	Units
<i>SIC</i>	Sea ice fraction of the ocean surface	[0, 1]
<i>HI</i>	Sea ice thickness	m
<i>SST</i>	Sea surface temperature	K
<i>ZOS</i>	Sea surface height above the geoid	m
$\theta_{o,k}$	Sea water velocity in the northward direction	°C
$s_{o,k}$	Sea water salinity	PSU
$u_{o,k}$	Sea water velocity in the eastward direction	m/s
$v_{o,k}$	Sea water velocity in the northward direction	m/s

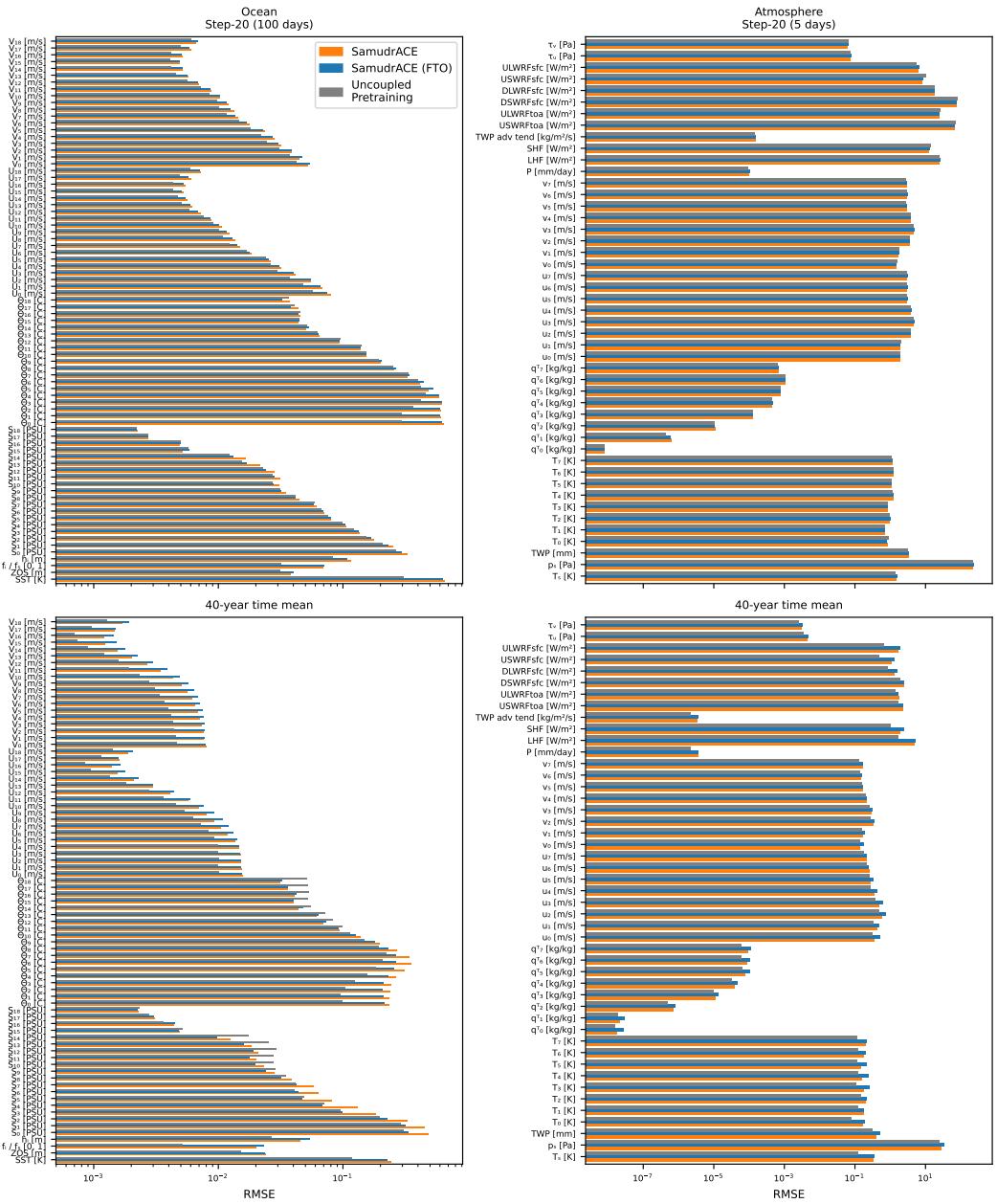


Figure S2. Step-20 and 40-year RMSEs for the time-mean spatial structure of all oceanic and atmospheric output fields.

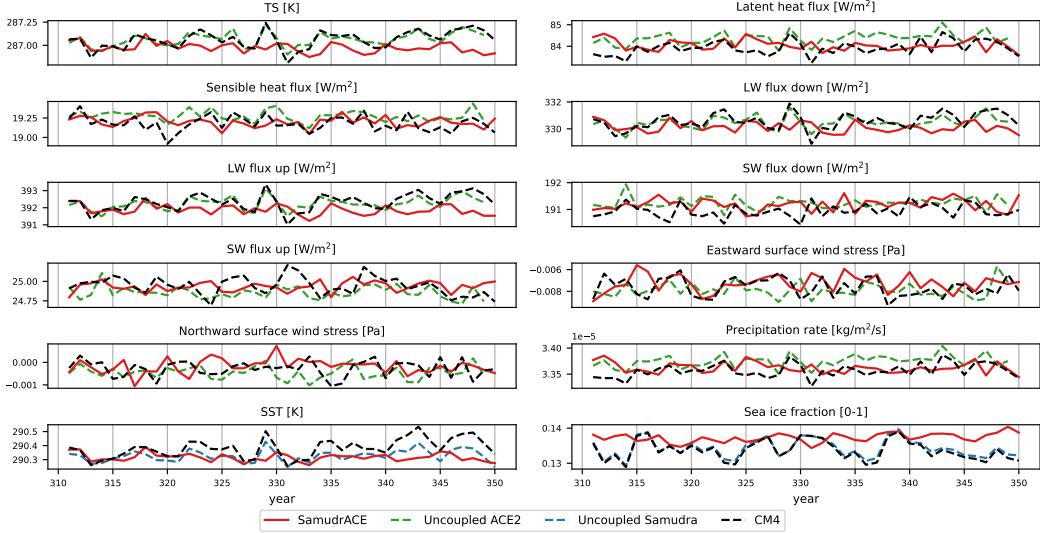


Figure S3. Annual and global mean time series over the 40-year test period for key surface variables involved in the exchange of information between the atmosphere and ocean components of the coupled model. In uncoupled mode, the strong effect of prescribed boundary conditions is apparent for several variables, particularly T_s , longwave fluxes, SST , and sea ice fraction. In coupled mode, the generated boundary conditions in SamudrACE quickly decorrelate from CM4 as the simulation progresses, allowing for increased internal variability and the emergence of coupled physical phenomena.

Table S2. Description of SamudrACE atmosphere output variables shown in the right column of Figure S2. Three dimensional variables have 8 levels, denoted by the subscript $k \in \{0 \dots, 7\}$.

Symbol	Description	Units	Time	Prognostic?
T_k	Air temperature	K	6-hour snapshot	Yes
T_s	Skin temperature of land or sea-ice	K	Snapshot	Yes
q_k^T	Specific total water (vapor + condensates)	g/kg	Snapshot	Yes
U_k	Windspeed in eastward direction	m/s	Snapshot	Yes
V_k	Windspeed in northward direction	m/s	Snapshot	Yes
p_s	Atmospheric pressure at surface	hPa	Snapshot	Yes
RSW	Upward shortwave radiative flux at TOA	W/m^2	Mean	No
OLR	Upward longwave radiative flux at TOA	W/m^2	Mean	No
USW_{sfc}	Upward shortwave radiative flux at surface	W/m^2	Mean	No
ULW_{sfc}	Upward longwave radiative flux at surface	W/m^2	Mean	No
DSW_{sfc}	Downward shortwave radiative flux at surface	W/m^2	Mean	No
DLW_{sfc}	Downward longwave radiative flux at surface	W/m^2	Mean	No
LHF	Surface latent heat flux	W/m^2	Mean	No
SHF	Surface sensible heat flux	W/m^2	Mean	No
P	Surface precipitation rate (all phases)	mm/day	Mean	No
$\frac{\partial TWP}{\partial t} _{adv}$	Tendency of total water path from advection	mm/day	Mean	No
τ_u	Eastward surface wind stress	Pa	Mean	No
τ_v	Northward surface wind stress	Pa	Mean	No

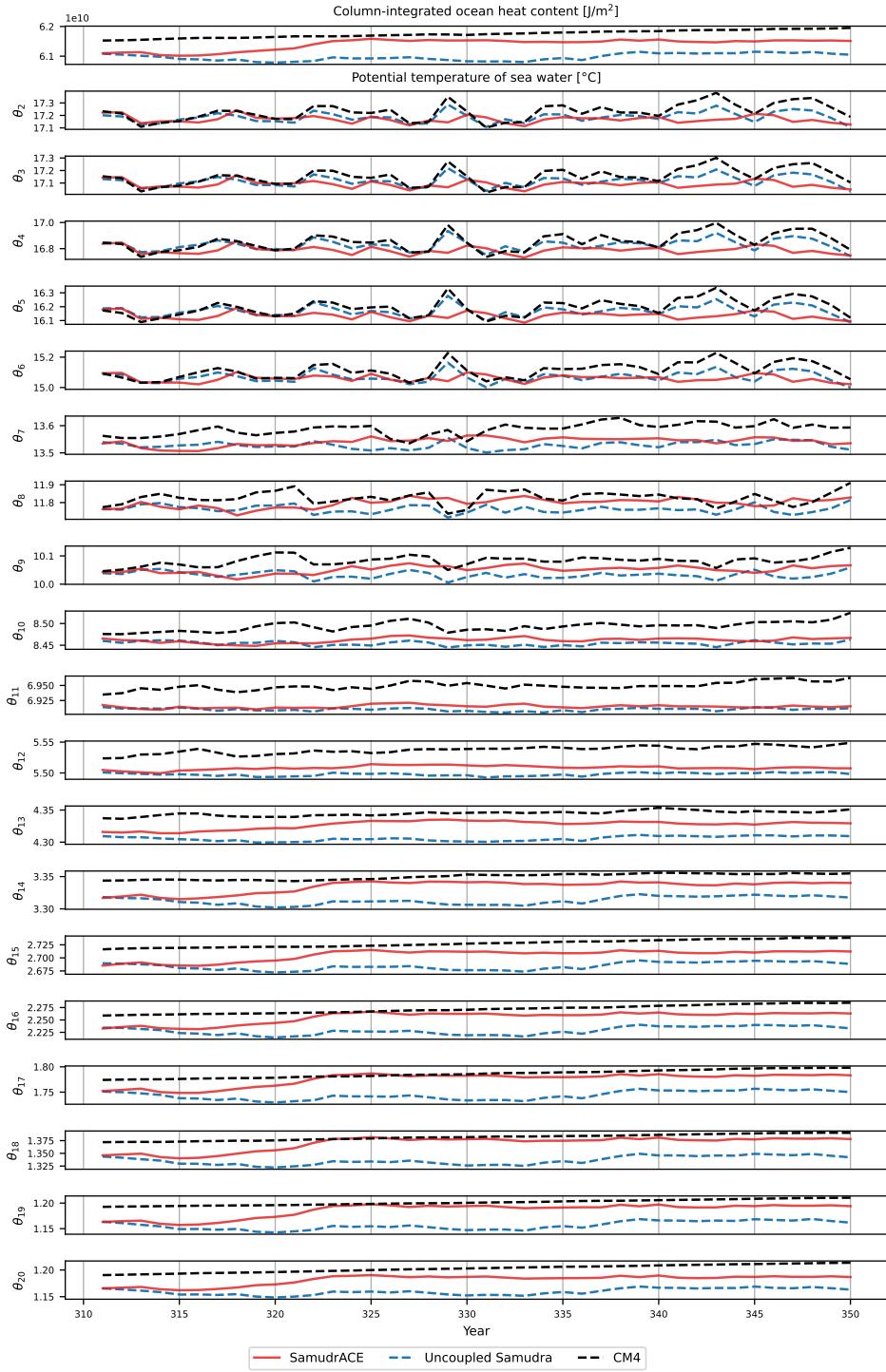


Figure S4. Annual mean time series of column-integrated ocean heat content (OHC) and potential temperature at all 19 vertical ocean levels. SamudrACE substantially reduces the time-mean bias of OHC compared to uncoupled SamudraI. This improvement is primarily concentrated in the deeper ocean layers below a depth of 650 meters ($\theta_{o,10}$).

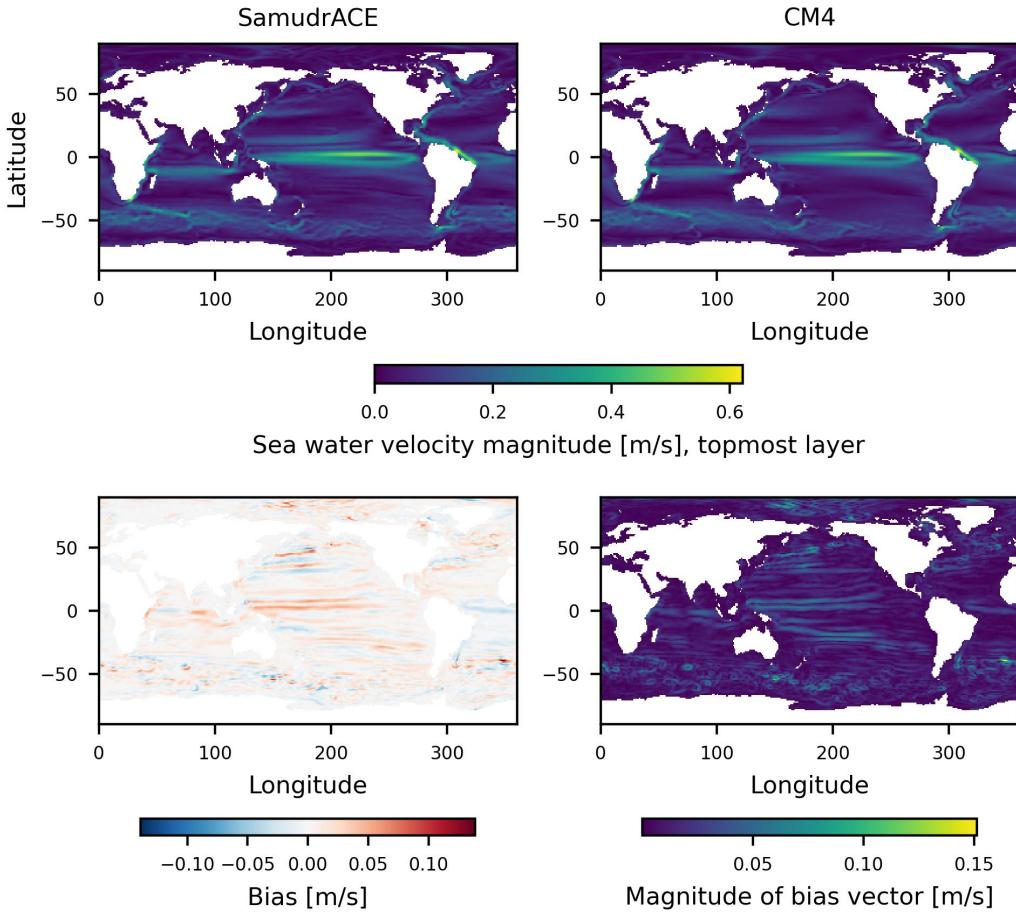


Figure S5. 40-year time-mean generated and target sea water velocity magnitude for the coupled emulator and CM4 target (first row), together with the magnitude bias map (generated - target), and map of bias vector magnitude (second row). Eddy-like bias patterns can be observed in the Southern Ocean, Northern Pacific, and North Atlantic.

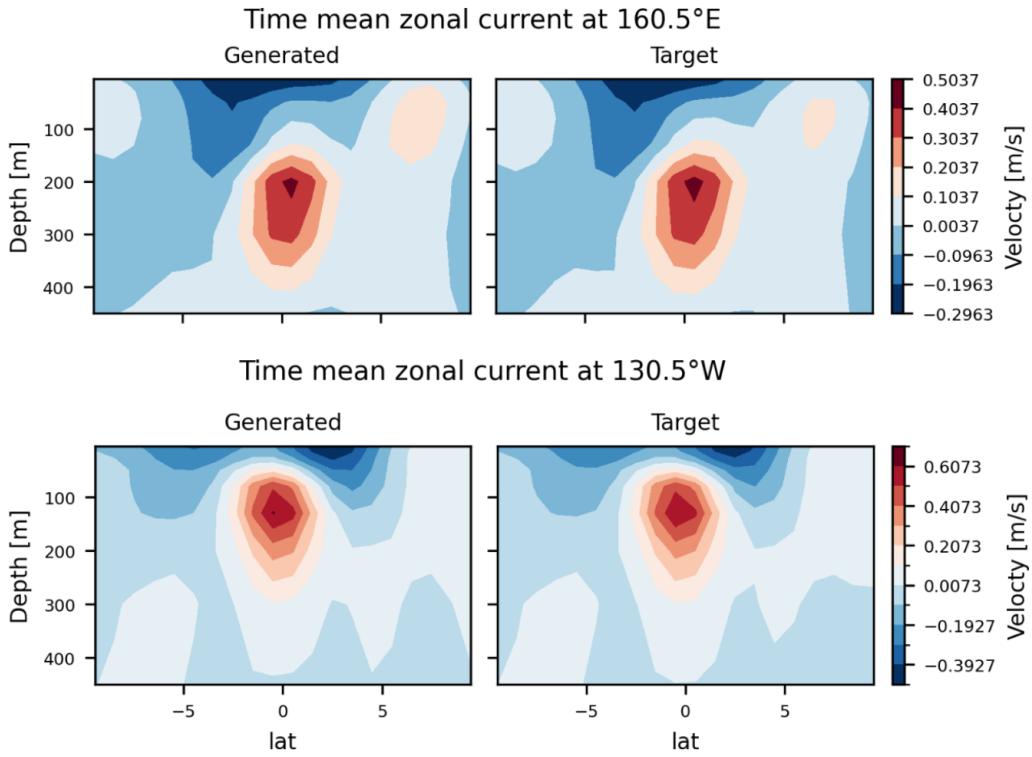


Figure S6. Vertical and meridional structure of tropical zonal currents in the Pacific Ocean. Time-mean zonal sea water velocity from the surface to 400 meters depth for two vertical slices centered at the equator in the middle western Pacific at 160.5°E and 130.5°W . SamudrACE closely emulates the time average near-surface current in the equatorial Pacific.

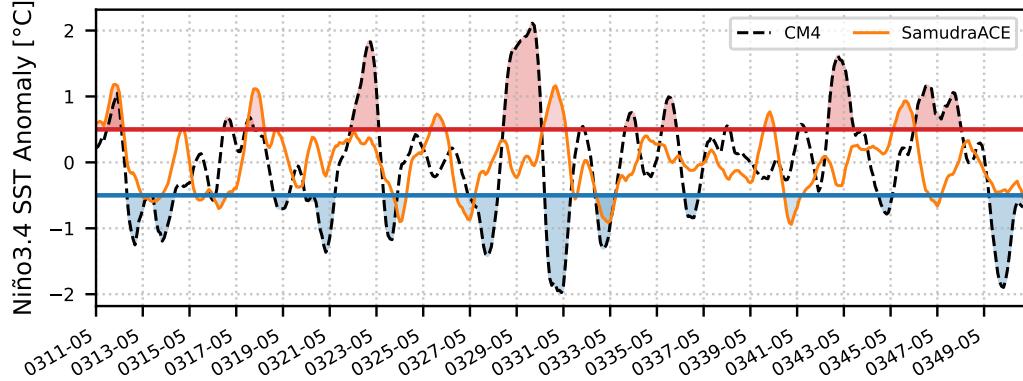


Figure S7. Niño 3.4 SST anomalies over the 40-year held-out inference period. The black dashed line is CM4 and the orange line is SamudrACE. Events with anomalies of ± 0.5 lasting longer than 5 months are highlighted with shading. There were 6 El Niño and 7 La Niña events in the generated data and 9 El Niño and 11 La Niña events in the target data.

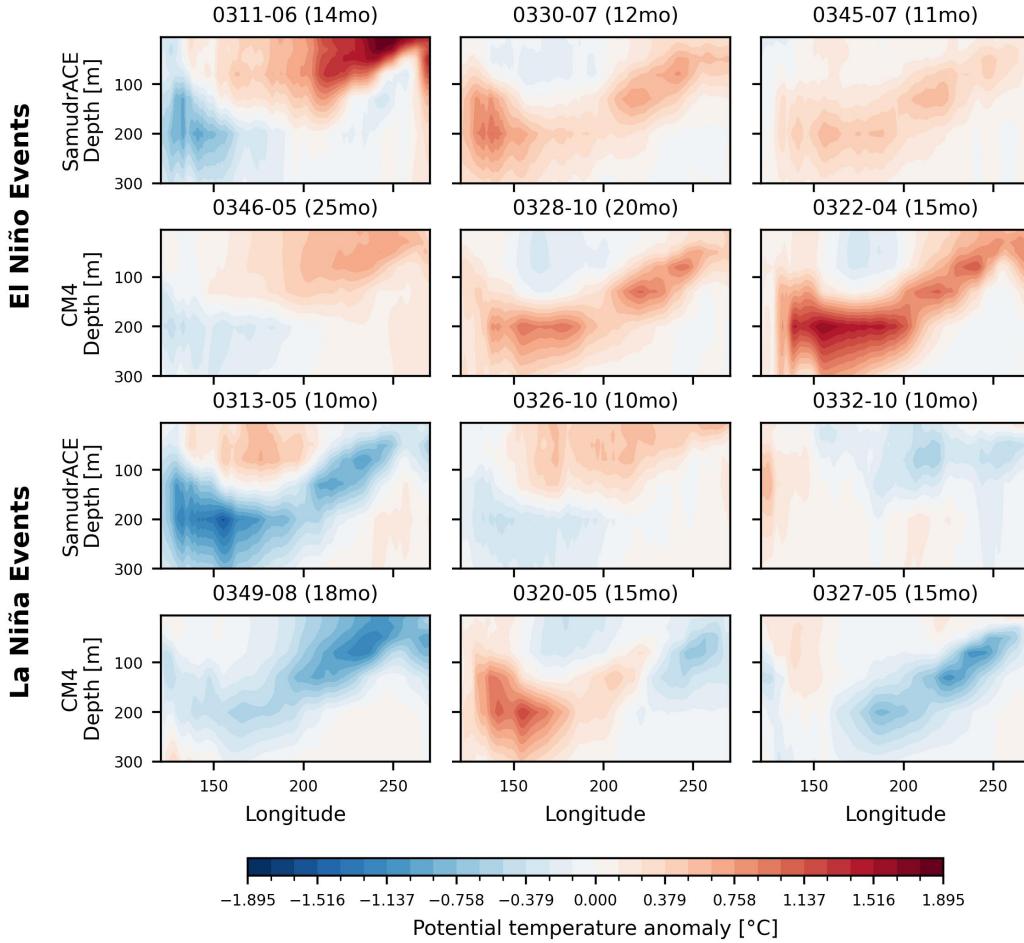


Figure S8. Time-mean potential temperature of sea water θ_o anomaly profiles in the equatorial Pacific. For each model (SamudrACE and CM4) and ENSO condition (El Niño and La Niña) we select the top three events of Figure S8 in terms of duration, compute the θ_o anomaly for each month with respect to the 40-year time-mean, and visualize the time average taken over the duration of each event. We give the event initialization and duration in months above each panel.

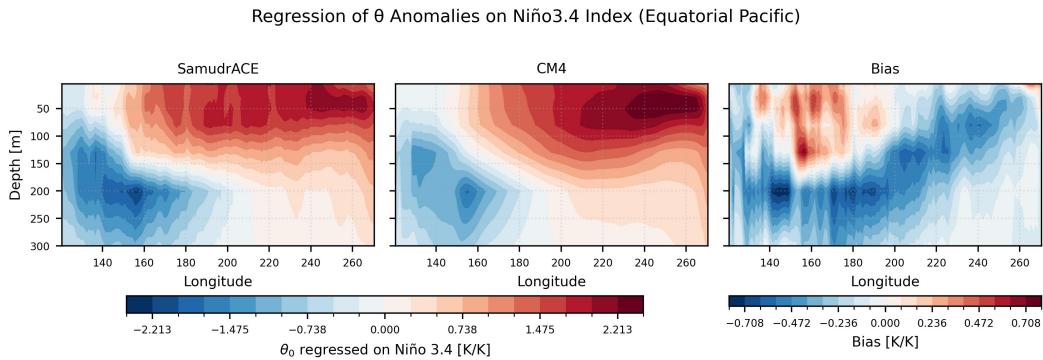


Figure S9. Regression of monthly mean θ_o anomalies on the Niño 3.4 index, respectively for generated and target outputs, over the 40-year held-out test dataset. Biases in the response have a similar spatial structure to CM4's La Niña.

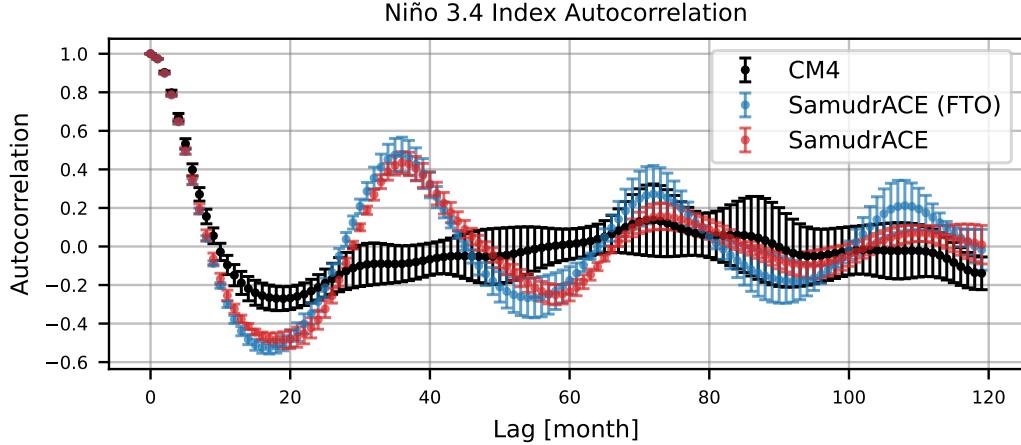


Figure S10. Autocorrelation (lagged correlation) of the Niño 3.4 index for CM4 (black), SamudrACE (red), and FTO (blue). Uncertainty is expressed as the height of the bars above and below each dot, where the uncertainty is the standard deviation over the 5 ensemble members of SamudrACE and FTO. For CM4, we use a boot-strapping method where we randomly sample 25 different 500 month long segments from the single CM4 simulation.

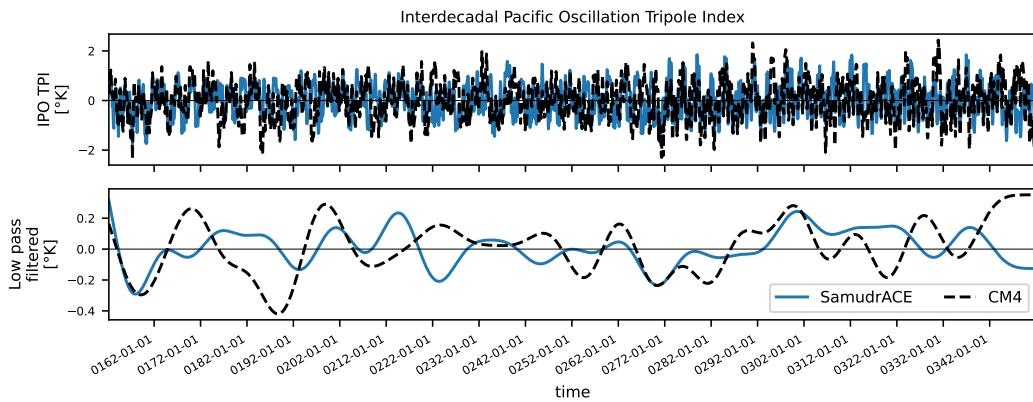


Figure S11. Interdecadal Pacific Oscillation Tripolar Index (Henley et al., 2015). The upper panel shows the unfiltered IPO TPI. Following Henley et al. (2015) we apply a low-pass filter with cutoff of 13 years to extract decadal variability, shown in the lower panel. The filter is Chebyshev Type I of order 5 and passband ripple of 0.5 dB. We apply the filter in both the forward and reverse directions, resulting in a filtering with zero phase response and squared amplitude response.

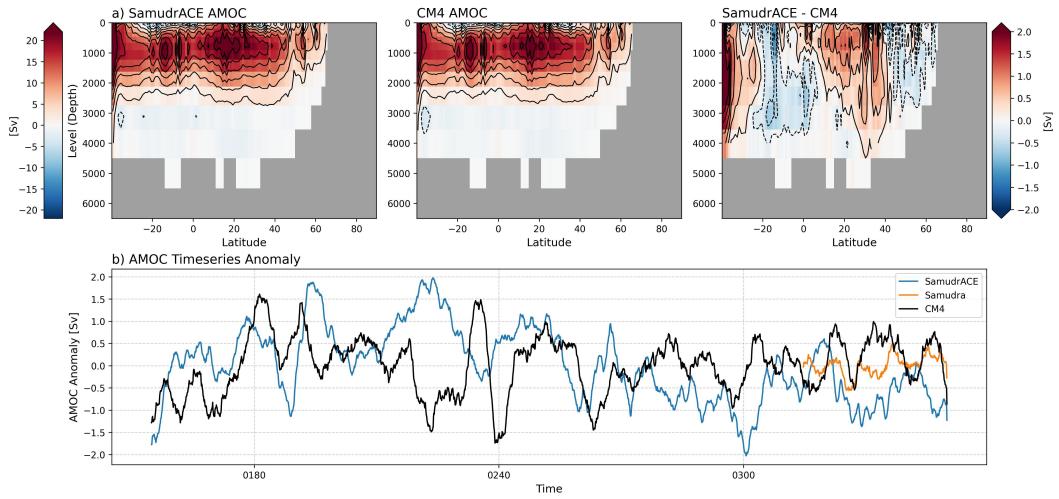


Figure S12. Atlantic Meridional Overturning Circulation from a 200-year rollout in SamudrACE and CM4. Panel a shows the time mean of the AMOC generated with SamudrACE (left), the reference CM4 (middle), and the bias between the two (right). Panel b shows the time series of the AMOC strength anomaly, computed as the deviation from seasonal climatology of the maximum of the AMOC streamfunction between $28 - 32^\circ N$. Before computing the anomaly, we first apply a 5-year rolling mean to filter out high frequencies from the timeseries. For this panel we show the timeseries from CM4, SamudrACE, and for uncoupled Samudra prior to coupled finetuning and using CM4 boundary conditions (final 40 years). When computing the AMOC, we take the depth integral from the bottom of the ocean up to the top and include the contribution from dynamic sea level at the upper ocean cell.

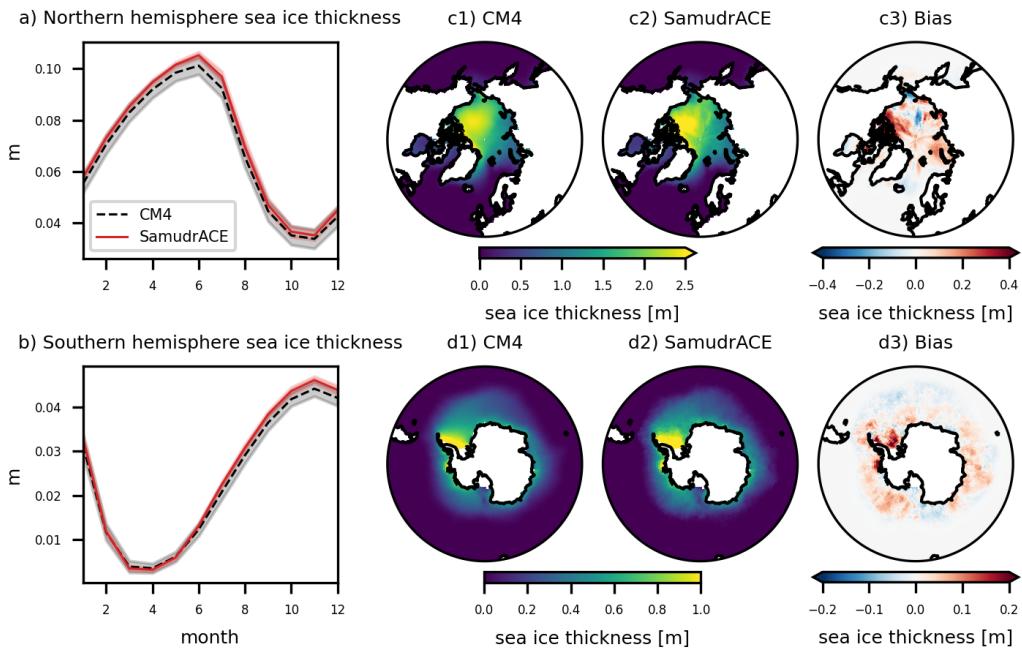


Figure S13. Monthly mean over the 40-year held-out period of (a) Northern and (b) Southern Hemisphere sea ice thickness. Shading denotes the interannual standard deviation over 40 years. Panel c-d) shows the time mean sea ice thickness over the same time period for the CM4 target, SamudrACE, and its bias.