SamudrACE: Coupled Climate Simulations with ACE and Samudra

Elynn Wu,* James Duncan, Troy Arcomano, Spencer K. Clark, Brian Henn, Anna Kwa, Jeremy McGibbon, W. Andre Perkins, Oliver Watt-Meyer, Christopher Bretherton
Allen Institute for Artificial Intelligence (Ai2), Seattle, USA

Surya Dheeshjith, Adam Subel, Laure Zanna
New York University, New York, USA

William J. Hurlin
Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, USA

William Gregory, Alistair Adcroft
Princeton University, Princeton, USA

Abstract

We present a coupling of the Ai2 Climate Emulator (ACE) 3D global atmosphere emulator to the Samudra 3D global ocean emulator, both of which are large autoregressive ML models with a combined total of nearly 600 million parameters. A coupled emulator has the potential advantage of accelerating climate change projections under different future scenarios, enabling more iterative and insightful strategies for climate policy, adaptation, and mitigation. The coupled emulator facilitates the exchange of boundary conditions between separate models of the atmosphere and ocean, with prognostic sea ice included among Samudra's outputs. The coupled emulator produces a stable climate with remarkably small climate biases, a good seasonal cycle of sea ice, insignificant temporal climate drift, and realistic ENSO variability. The coupled emulator marks a significant step toward enabling fully coupled climate modeling with emulators.

1 Introduction

The advent and success of machine learning (ML)-based weather prediction has led to similarly data-driven global atmosphere emulators, such as the atmosphere-only version of the Ai2 Climate Model (ACE) [1]. Since then, atmosphere model emulators have continued to mature and support AMIP-style simulations [2, 3]. This paper demonstrates early progress toward the natural next step in this progression – a global climate model emulator which includes modular coupled atmosphere, sea ice, land, and ocean emulators, capable of running the Coupled Model Intercomparison Program (CMIP) DECK simulation suite [4]. 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, including physical phenomena that emerge through the interaction and coupled evolution of atmospheric surface forcing and sea surface response such as El Niño-Southern Oscillation (ENSO) variability. Recently, several papers have incorporated simplified forms of ocean modeling in ML

*elynnw@allenai.org

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

atmospheric emulators and achieved stable and accurate simulation of present-day [5] or present-day and CO_2-enhanced [6] mean climate. However, a better representation of the full extent of the ocean is needed to accurately predict coupled atmosphere-ocean variability. Three-dimensional ML ocean emulators have been recently developed for ocean forecasting on timescales up to 1–2 years [7–10], and for longer-running simulations forced by specified time-evolving atmospheric conditions [11, 12].

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. In this paper, we present an emulator constructed by coupling the ACE version 2 (ACE2) [2] 3D atmosphere emulator (8 layers, 6 hour time step) to the Samudra [12] 3D ocean emulator (19 layers, 5 day time step), extended to predict sea-ice concentration and thickness. Both components are emulated at 1° lat/lon horizontal resolution. We train the coupled emulator with a coupled 200-year simulation by the GFDL CM4 physics-based global climate model, forced with constant pre-industrial greenhouse gas and aerosol concentrations and a repeating annual cycle of insolation. The resulting trained emulator 1) generates stable centuries-long simulations of the coupled atmosphere and ocean with low bias and a 50x speedup over the reference physics-based model (800 SYPD with 1 H100 vs. 16 SYPD with 5535 CPU cores on AMD EPYC 7H12 processors); 2) emulates CM4's ENSO variability, accurately reproducing the spatial pattern of precipitation to El Niño conditions; and 3) accurately emulates the seasonal cycle of sea ice fraction in both the northern and southern hemisphere.

2 Methods

2.1 Dataset

Our reference training and evaluation datasets are from a 200-year pre-industrial control simulation from GFDL's Climate Model v4 (piCM4) [13]. We use the first 155 years of output for training, the next 5 years for validation, and the remaining 40 years of data is held out for testing.

The reference atmosphere fields were output as 6-hour instantaneous snapshots for prognostic variables and 6-hour time-averages for all surface and top-of-atmosphere fluxes (including precipitation). This enables the surface fluxes to be accumulated over 20 atmospheric steps into 5-day averages suitable for forcing the ocean emulator. The reference ocean fields were all output as 5-day averages. This includes the sea-surface temperature and sea ice fraction, which are used to force the atmosphere emulator.

2.2 Coupler

Figure 1 provides a schematic of the coupler. Briefly, ACE2 steps the atmosphere forward 6 hours at a time for 5 days with diagnostic boundary fluxes represented as 6-hour time averages. For each forward step, ACE2 is forced by sea ice fraction and sea surface temperature. 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 boundary condition is then used to force a single step of Samudra, which evolves the ocean state forward in time by 5 days. The new sea surface temperature and sea ice fraction are then used to force the atmospheric emulator. This coupling loop is repeated for the length of the simulation.

2.3 Training

Pretraining checkpoint selection and fine-tuning strategy have non-negligible impacts on the characteristics of the coupled emulator.

Pretraining ACE2 We follow ACE2's training protocol [2] 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, and ocean fraction, which are held fixed at the beginning of the 5-day window. ACE2-CM4 is trained for 50 epochs with a batch size of 16 (707,200 gradient steps) at a learning rate of 10^{-4}.

Pretraining Samudra We follow Samudra's training protocol [12] except 1) we force the model with all surface heat and water fluxes that are predicted by ACE2: upward and downward shortwave

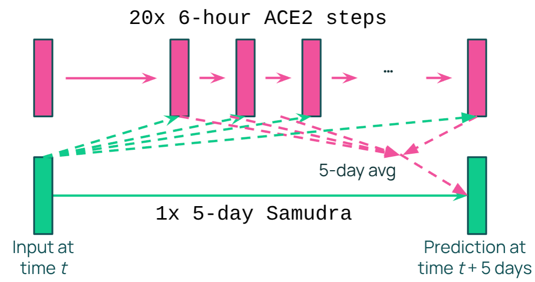

Figure 1: A single forward step of the coupled emulator. Forced by 5-day average SST and sea ice concentration, ACE2 steps forward 6 hours at a time for 5 days. Wind stress, precipitation, and surface fluxes are averaged and input as forcings to Samudra, which takes a single 5-day step forward.

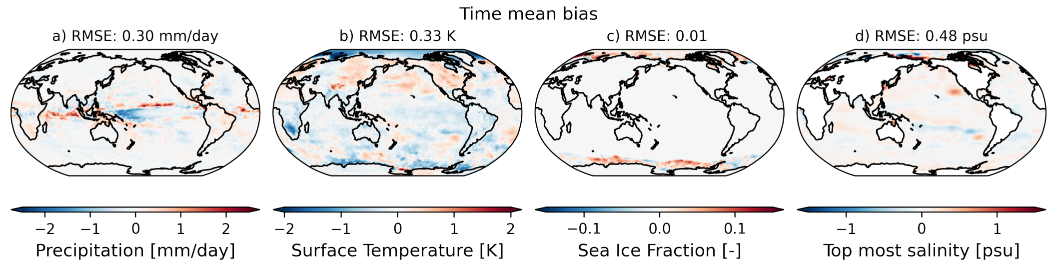

Figure 2: Time-mean bias of precipitation, surface temperature, sea ice fraction, and sea-surface salinity over the 40-year held-out inference period between the coupled emulator and the piCM4 reference model.

and longwave radiative fluxes, latent and sensible heat fluxes, precipitation, and wind stress forcing from above the sea ice; 2) we use a single input/output state; and 3) we add sea ice concentration and thickness as prognostic variables. Samudra is trained for 150 epochs with a batch size of 16 (106,000 gradient steps) at a learning rate of 10^{-4}.

Coupled fine-tuning Once ACE2 and Samudra pretraining is complete, we select the checkpoints with the lowest normalized channel-mean RMSE over autoregressive rollouts for each model for coupled fine-tuning. Together, the coupled emulators have a combined total of nearly 600 million parameters. Lastly, we fine-tune the ocean (FTO) in coupled mode. In this step only the MSE of the ocean fields contribute to the training loss. ACE2's weights are held fixed while Samudra adapts to its outputs by optimizing on one coupled forward steps (20 days total). Coupled fine-tuning is trained for 20 epochs with a batch size of 16 (14,140 steps gradient steps) at a learning rate of 10^{-4}.

3 Results

Climatology The coupled emulator is able to stably emulate piCM4 with minimal bias. Figure 2 shows maps of the time-mean biases of precipitation, surface temperature, sea ice fraction, and sea-surface salinity over the 40 year held-out test period. Precipitation biases are concentrated in the tropics while surface temperature biases are pronounced in areas of sea ice and topographic features. The emulator simulates the piCM4 seasonal cycle of sea ice fraction well, but slightly underestimates interannual variability in the Southern Hemisphere (Figure S1). These biases have similar magnitudes to those of the uncoupled component emulators.

ENSO The coupled emulator shows promising ENSO variability, a feature not possible to capture by single component emulator alone. Following Section 2.2.2 in [2], we compute the response to ENSO by regressing the surface precipitation map onto the Niño3.4 index for the 40-year test period (Figure 3); the match to piCM4 is excellent. When initialized from the coupled state at the start of piCM4 and freely run for 200 years (this seed corresponds to Figure 4 FTO RS1), the emulator

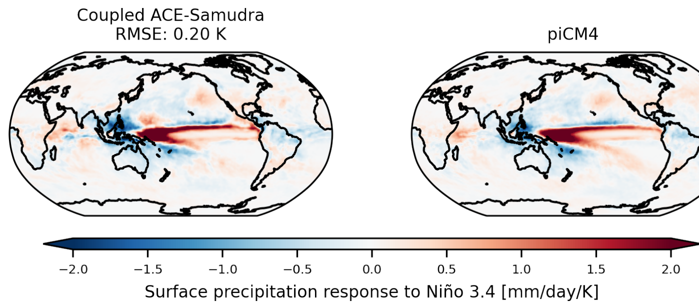

Figure 3: Maps of regression coefficients of the coupled emulator and reference piCM4 surface precipitation against the Niño3.4 index over the 40-year held-out test period.

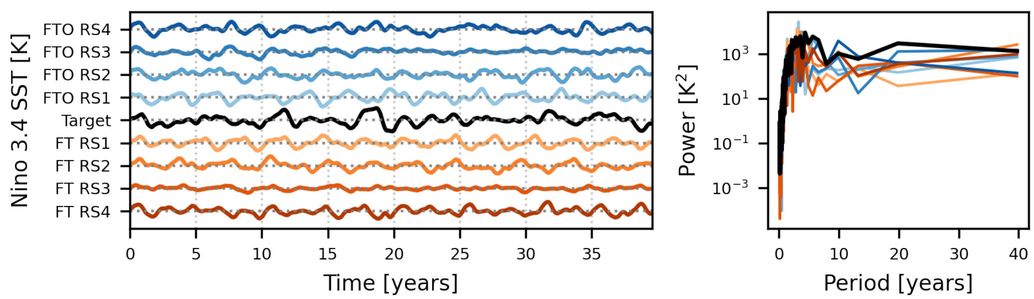

Figure 4: Time series of monthly Niño3.4 index with coupled fine-tuning where only ocean MSE contribute to training losses (FTO) or both ocean and atmosphere MSE are considered (FT) for 4 different random seed each and their corresponding temporal power spectra.

has comparable but slightly weaker variability than piCM4 (Figure S2), with slightly sharper cold anomalies (La Niñas) and weaker warm anomalies (El Niños).

Impact of coupled fine-tuning on ENSO For comparison, Figure 4 also shows time series of Niño3.4 index and temporal power spectra using a different coupled fine-tuning strategy, FT. In FT, weights are optimized in both the atmosphere and ocean emulators. FT and FTO share the same starting point for the individual atmosphere and ocean pretraining checkpoint. For one random seed (RS3) FT appears to have almost no low frequency ENSO variability, in contrast to the corresponding FTO run.

4 Conclusions and future work

Our work demonstrates a successful strategy for building stable, data-driven Earth system models capable of generating centuries-long fully coupled simulations. The coupled emulator maintains low climate biases comparable to its uncoupled component models while running orders of magnitude faster than the reference numerical model. A key new achievement of this work is the realistic simulation of emergent climate phenomena that arise directly from atmosphere-ocean interaction. By coupling the two components, the emulator can generate realistic ENSO variability including the associated teleconnections to global precipitation patterns – a feat which is not possible for the uncoupled emulators. Similarly, the model produces a stable and accurate seasonal cycle of sea ice fraction in both hemispheres.

While promising, our analysis also reveals areas for future improvement. The emulator exhibits somewhat weaker El Niño events than the reference model and underestimates the interannual variability of Southern Hemisphere sea ice. Future efforts could focus on refining the fine-tuning strategy or model architecture to address these biases. The successful framework of our coupled emulator could enable its use for generating large ensembles of coupled climate simulations, and provides a template for emulating additional earth system components, such as land and biogeochemistry, opening new avenues for efficient climate studies.

References

[1] Oliver Watt-Meyer, Gideon Dresdner, Jeremy McGibbon, Spencer K. Clark, Brian Henn, James Duncan, Noah D. Brenowitz, Karthik Kashinath, Michael S. Pritchard, Boris Bonev, Matthew E. Peters, and Christopher S. Bretherton. ACE: A fast, skillful learned global atmospheric model for climate prediction. 2023. doi: 10.48550/arXiv.2310.02074.

[2] 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. *npj Climate and Atmospheric Science*, 2025. doi: 10.1038/s41612-025-01090-0.

[3] William E Chapman, John S Schreck, Yingkai Sha, David John Gagne II, Dhamma Kimpara, Laure Zanna, Kirsten J Mayer, and Judith Berner. Camulator: Fast emulation of the community atmosphere model. *arXiv preprint arXiv:2504.06007*, 2025.

[4] Veronika Eyring, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, Ronald J. Stouffer, and Karl E. Taylor. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 2016. doi: 10.5194/gmd-9-1937-2016.

[5] 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.

[6] Spencer K Clark, Oliver Watt-Meyer, Anna Kwa, Jeremy McGibbon, Brian Henn, W Andre Perkins, Elynn Wu, Christopher S Bretherton, and Lucas M Harris. Ace2-som: Coupling to a slab ocean and learning the sensitivity of climate to changes in co2. *arXiv preprint arXiv:2412.04418*, 2024.

[7] Lei Chen, Xiaohui Zhong, Feng Zhang, Yuan Cheng, Yinghui Xu, Yuan Qi, and Hao Li. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, 2023. doi: 10.1038/s41612-023-00512-1.

[8] Anass El Aouni, Quentin Gaudel, Charly Regnier, Simon Van Gennip, Olivier Le Galloudec, Marie Drevillon, Yann Drillet, and Jean-Michel Lellouche. GLONET: Mercator's end-to-end neural Global Ocean forecasting system. arXiv, 2025. doi: 10.48550/arXiv.2412.05454.

[9] Wei Xiong, Yanfei Xiang, Hao Wu, Shuyi Zhou, Yuze Sun, Muyuan Ma, and Xiaomeng Huang. Ai-goms: Large ai-driven global ocean modeling system. *arXiv preprint arXiv:2308.03152*, 2023.

[10] Xiang Wang, Renzhi Wang, Ningzi Hu, Pinqiang Wang, Peng Huo, Guihua Wang, Huizan Wang, Senzhang Wang, Junxing Zhu, Jianbo Xu, et al. Xihe: A data-driven model for global ocean eddy-resolving forecasting. *arXiv preprint arXiv:2402.02995*, 2024.

[11] Zijie Guo, Pumeng Lyu, Fenghua Ling, Jing-Jia Luo, Niklas Boers, Wanli Ouyang, and Lei Bai. ORCA: A Global Ocean Emulator for Multi-year to Decadal Predictions. 2024. doi: 10.48550/arXiv.2405.15412.

[12] Surya Dheeshjith, Adam Subel, Alistair Adcroft, Julius Busecke, Carlos Fernandez-Granda, Shubham Gupta, and Laure Zanna. Samudra: An AI Global Ocean Emulator for Climate. *Geophysical Research Letters*, 2025. doi: 10.1029/2024GL114318.

[13] I. M. Held, H. Guo, A. Adcroft, J. P. Dunne, L. W. Horowitz, J. Krasting, E. Shevliakova, M. Winton, M. Zhao, M. Bushuk, A. T. Wittenberg, B. Wyman, B. Xiang, R. Zhang, W. Anderson, V. Balaji, L. Donner, K. Dunne, J. Durachta, P. P. G. Gauthier, P. Ginoux, J.-C. Golaz, S. M. Griffies, R. Hallberg, L. Harris, M. Harrison, W. Hurlin, J. John, P. Lin, S.-J. Lin, S. Malyshev, R. Menzel, P. C. D. Milly, Y. Ming, V. Naik, D. Paynter, F. Paulot, V. Ramaswamy, B. Reichl,

T. Robinson, A. Rosati, C. Seman, L. G. Silvers, S. Underwood, and N. Zadeh. Structure and Performance of GFDL's CM4.0 Climate Model. *Journal of Advances in Modeling Earth Systems*, 2019. doi: 10.1029/2019MS001829.

Appendix

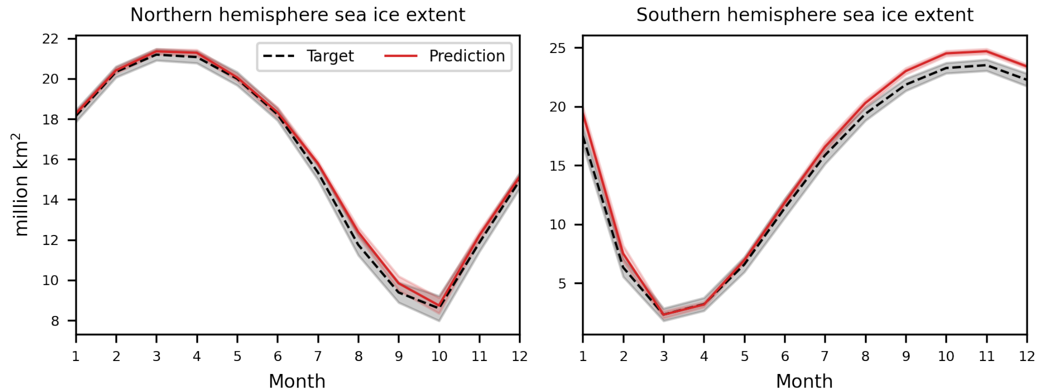

Figure S1: Monthly mean over the 40 year held-out period of Northern and Southern Hemisphere sea ice extent. Shading denotes the interannual standard deviation over 40 years.

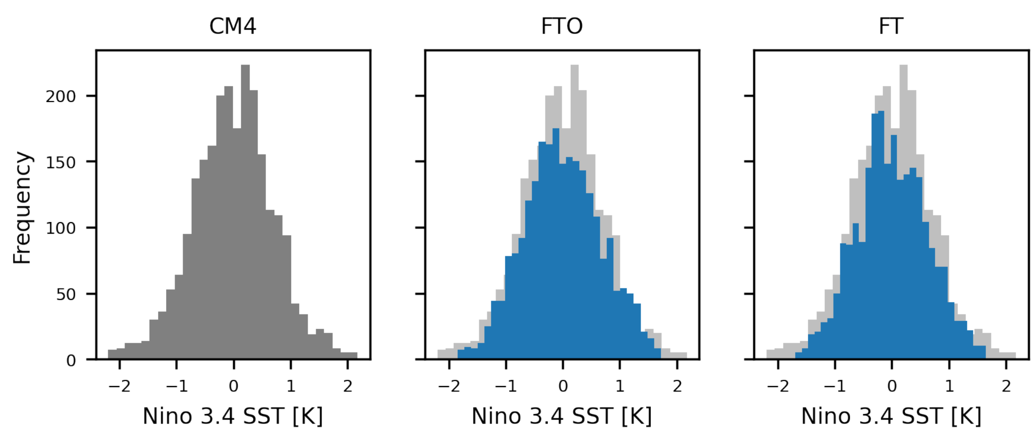

Figure S2: Histogram of Niño3.4 index for CM4 and two different coupled fine-tuning strategy: FTO and FT. In FTO, only ocean fields contribute to training losses and only ocean weights are optimized. In FT, both ocean and atmosphere are optimized.