Stochastic ensemble coupled modeling of pre-industrial control and 1% increasing CO_2 scenarios with physically conservative machine learning emulators

James P. C. Duncan[1], Elynn Wu[1], Surya Dheeshjith[2], Adam Subel[2], Spencer K. Clark[1,3], Laure Zanna[2,4], Alistair Adcroft[5], Christopher Bretherton[1], Oliver Watt-Meyer[1]

[1] The Allen Institute for Artificial Intelligence
[2] Courant Institute of Mathematical Sciences, New York University
[3] NOAA/Geophysical Fluid Dynamics Laboratory
[4] Center for Data Science, New York University
[5] Princeton University

Abstract

Machine learning (ML) global climate model (GCM) emulators provide faithful compressed representations of traditional numerical GCMs, enabling new climate science frontiers via high-quality and stable centuries-long simulations of order 1000x less energy intensive than their non-ML counterparts. Far from replacing existing products, pure ML emulators inhabit a complementary niche that benefits from a diverse range of high-quality datasets. Through their flexible training and deployment, ML emulators promise novel Earth system representations which borrow strength from physically consistent numerical models, historically-accurate reanalysis products, and operational weather forecasts.

We present SamudrACE, a first-of-its-kind full-depth coupled GCM ML emulator which couples an 8-level atmosphere emulator (ACE) to a 19-level ocean/ice emulator (Samudra). On a single H100 GPU, SamudrACE can simulate 1500 years per day of 6-hourly atmosphere and 5-daily ocean outputs at 1° horizontal resolution. Every 5 days, SamudrACE couples its components in physical space by passing the 5-day average of ACE's 6-hour surface energy, momentum, and moisture flux outputs to Samudra, which in turn passes its 5-day sea surface temperature and sea ice fraction step back to ACE. Through stochastic noise conditioning, SamudrACE can provide an ensemble of outputs spanning the full coupled earth system. We describe physical consistency constraints on atmosphere-ocean energy exchanges, showing their importance for generalization beyond the training data to unseen forcing scenarios in a reference simulation from GFDL's CM4.0 coupled climate model. Training on novel numerical model simulations with randomly varying CO_2 concentrations further improves emulator physical robustness to idealized forcing scenarios in the CMIP DECK.