# A-OASIS: Building an AI-Coupled Atmosphere-Ocean Foundation Model

Nia Maheshwari[1], Guang Zhao[2], Shuyan Ding[3], Zhengyu Liu[3], Wei Xu[2]
[1]University of Tennessee, Knoxville; [2]Brookhaven National Laboratory; [3]The Ohio State University

***Abstract*—U.S. Department of Energy (DOE) water-for-energy planning requires predictions of precipitation and hydroclimate variability from weeks to years, yet these horizons depend on coupled atmosphere-ocean-land feedback that are difficult to learn from observations. Established dynamical Earth system models and explicit couplers remain the scientific reference, but producing long simulations and forecast ensembles is computationally intensive. This work in progress asks a bounded question: after careful, domain-guided data harmonization, how much of the coupling workflow can a unified AI foundation model learn directly, and where does that strategy cease to be reliable? We present the Atmosphere-Ocean Artificial System for Integrated Simulation (A-OASIS), an Aurora-derived model that jointly encodes atmosphere, ocean, and land states in one latent space rather than training and coupling separate emulators. A packed, masked vertical representation reconciles heterogeneous coordinates while retaining native metadata. Initial experiments demonstrate stable full resolution distributed training and decreasing validation loss with increased model capacity. These results establish implementation feasibility, driving the next step of validating forecast capability.**

***Keywords*—*Earth system foundation models, direct AI coupling, data harmonization, hydroclimate prediction, distributed training***

## I. Background and Introduction

The DOE Predicting U.S. Water for Energy challenge emphasizes weeks-to-years information for energy system planning, water availability, and drought preparedness. A motivating downstream target is Southwestern United States monthly precipitation at 12-24 month lead times, whose skill depends on El Niño Southern Oscillation variability and remote atmosphere-ocean teleconnections. Current machine learning systems are strongest at daily-to-weekly weather prediction, while longer horizons require persistent ocean memory, cross-domain feedbacks, and substantially more training data than observations alone provide.

A-OASIS explores a complementary alternative to the established *train-or-simulate-first, couple-afterward* workflow. We do not propose that a learned coupler should replace validated physical modeling across all regimes. Instead, simultaneous training on harmonized multi-domain states is treated as a controlled scientific experiment: can shared representation learning recover useful coupling structure, and under which variables, resolutions, lead times, and physical constraints does it fail?

## II. Data Harmonization And Packed Representation

The current prototype uses monthly Community Earth System Model 2 (CESM2) fields with 11 historical simulation models each having 1,980 months spanning 1850-2014 and one preindustrial control model for 14,400 months. Except one historical model data is reserved for testing, the remained are used to training and validation. Ocean, land, and atmosphere variables are remapped to a common 180×360 latitude-longitude grid, converted from NetCDF archives to chunked Zarr stores, and normalized using training set statistics. Domain specific masks identify ocean cells, valid soil layers, and atmospheric pressure levels above the local surface; masked values remain excluded from the loss.

Harmonization is part of the coupling hypothesis, not merely data engineering. A shared monthly cadence suppresses unresolved weather noise and reduces temporal scale mismatch, while common spatial indexing enables collocated cross-domain learning. To preserve heterogeneous vertical coordinates, A-OASIS packs 20 ocean depths, 5 soil levels, and 9 atmospheric pressure levels into a 34-slot axis. The packed axis is an indexing device rather than a physical coordinate; each variable occupies only its assigned slots, and metadata retains native level identity. This design avoids separate domain encoders that could weaken semantically important surface and teleconnection pathways.

## III. Model And Training

A Perceiver style encoder compresses surface and packed 3D fields into a fixed latent representation. A hierarchical shifted-window Swin Transformer evolves that state across spatial scales, and decoder heads reconstruct all domains. Distributed pretraining runs on two nodes with four A100 GPUs each using DeepSpeed. Each sample uses 12 consecutive monthly states to predict the next 12 months under a variable-balanced masked MSE objective, preventing fields with larger valid domains from dominating optimization.

We established a pretraining pipeline and demonstrated stable full-resolution training on harmonized CESM2 data. The result shows that the unified model can be optimized end to end over jointly represented states.

## IV. Results

We completed a full epoch pretraining run for the Medium (M) configuration (420M parameters). The run achieved a validation/test loss of 0.0911/0.1082. To quantify model capacity scaling, we configured four smaller model sizes within the compute budget. Under a common experimental setting and training data amount, each pretraining ran 2,750 steps with a decaying learning rate scheduler, the validation MSE loss decreased monotonically from 0.1151 to 0.1107, 0.1060, and 0.1026. Relative to XXXS, the S configuration contains 14.3 times more trainable parameters and reduces loss by ~10.9%. Improvement is observed at every scale transition, establishing a clear positive relationship between model capacity and validation performance across the tested range. The test data performance is similar.

## V. Future Work

The next phase will expand from the prototype toward full-size model pretraining (1.5B), forecast rollout, and downstream prediction of US sea surface related phenomena.