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

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

Brookhaven National Laboratory
U.S. DEPARTMENT OF ENERGY

MOTIVATION

Model the Atmosphere, Ocean, and Land as a Coupled Earth System

- **DOE Water-for-Energy Need:** Energy planning, water availability, and drought preparedness require **predictions** of precipitation and hydroclimate variability from **weeks to years**
- **Coupled Predictability:** Long-term forecast skill depends on atmosphere-ocean-land feedbacks
- **Observational Limitation:** The observational record contains too few independent realizations of important climate modes to **learn robust long-term coupled dynamics**
- **Computational Constraint:** Dynamical Earth system models and explicit physical couplers remain the scientific reference, but generating long simulations and large forecast ensembles is **computationally intensive**
- **Disjoint AI Forecasting:** Most current AI weather and climate models are trained for **individual domains**, datasets, resolutions, and timescales, then coupled or **evaluated separately**
- **A unified coupled Earth System Foundation Model does not yet exist**
- **Direct AI Coupling:** A-OASIS tests whether shared representation learning can recover useful coupling structure directly, rather than training separate domain emulators and connecting them afterward

Bounded Scientific Question: How much of the coupling workflow can a unified AI foundation model learn reliably, and where does that strategy fail across variables, resolutions, lead times, and physical constraints?

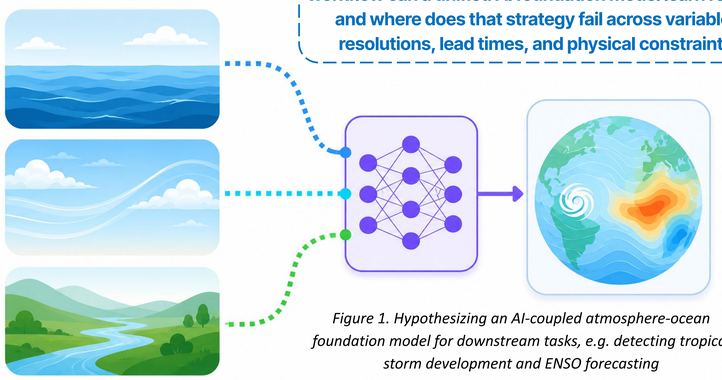


Figure 1. Hypothesizing an AI-coupled atmosphere-ocean foundation model for downstream tasks, e.g. detecting tropical storm development and ENSO forecasting

FOUNDATION MODEL ARCHITECTURE

Adapt Aurora Model [1] to Develop Our Framework

- **Perceiver encoder/decoder:** support **heterogeneous Earth system datasets** with different variables and vertical resolutions (flexible input/output)
  - Standardized representation: climate data are mapped into a common 3D tensor for processing by the shared model backbone
  - Level aggregation: Physical pressure levels are embedded and **compressed into a fixed number of latent vertical levels**
  - Vertical dimensionality reduction: Climate columns with varying numbers of depth levels are compressed into a fixed-size latent representation
- **Latent vectors:** learn **which depth levels contain the most informative features**.
  - Shared latent space: Produce a **compact representation** that is independent of the original dataset resolution, enabling a common backbone across datasets
- **Swin (Shifted Window) Transformer backbone [3]:** Performs self-attention within local windows
  - Shifts the window boundaries between layers to enable **information propagation across neighboring regions**

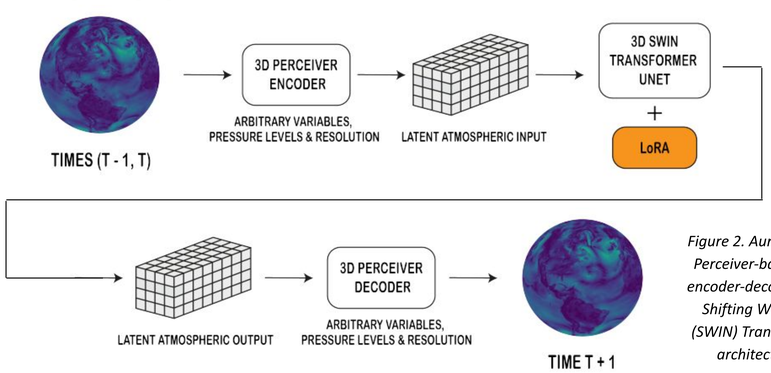


Figure 2. Aurora's [1] Perceiver-based [2] encoder-decoder and Shifting Window (SWIN) Transformer architecture

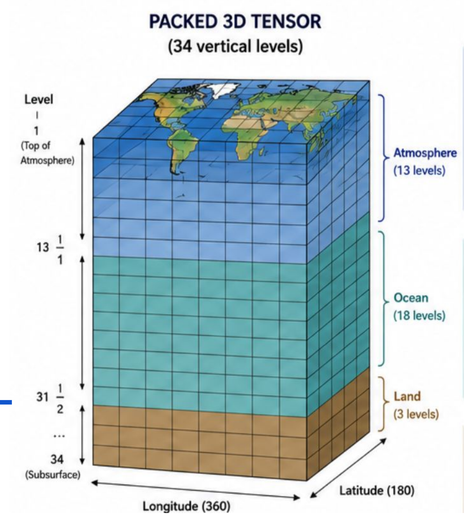
METHODOLOGY

CMIP6 HISTORICAL & SIMULATED DATA 1850 → 2014
↓
RAW DATA INGESTION, HARMONIZATION, & STORAGE
↓
PREPROCESSING & PACKED 34-LEVEL REPRESENTATION
↓
WINDOWING & SAMPLE CONSTRUCTION
↓
MASKING, NORMALIZATION
↓
TRAINING TIME BATCHING
↓
A-OASIS MODEL PRETRAINING
↓
VALIDATE, CHECKPOINT, OUTPUT, & POST PROCESSING

Data Lifecycle & Preparation

- Harmonized CESM2 [4] ocean, land, and atmosphere datasets (a total of 15 variables in 2D or 3D shapes) onto a common spatial grid
- **CESM2**: 1980 months | **CTRL**: 14,400 months of historical/simulated data (ctrl forcing fixed at pre-industrial 1850) | 180° latitude x 360° longitude
- Converted NetCDF archives to the **Zarr** format for scalable, parallel I/O
- Generated **domain-specific masks** to identify valid ocean, land, and atmospheric grid points
- Computed dataset-wide **summary statistics** for variable normalization (std. deviation, average)

A unified 3D tensor on a common 1° global grid



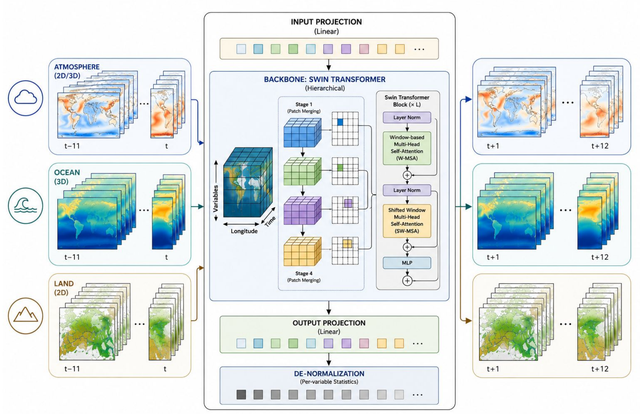
Unified Earth System Representation

- Extended the Aurora-architecture based model to a coupled ocean-land-atmosphere model
- Introduced a **packed vertical representation** with 34 shared vertical levels:
  - ***Ocean: 20 depth levels***
  - ***Land: 5 soil levels***
  - ***Atmosphere: 9 pressure levels***
- Mapped each variable into its corresponding vertical region

A-OASIS MODEL

A-OASIS Architecture

- Modified encoder to **jointly process all domains** within a shared latent representation
- Updated decoder to **reconstruct predictions for all three domains** from the packed representation
- Expanded training targets from ocean-only variables to coupled ocean-land-atmosphere prediction



Training Pipeline

- Implemented a repository-local OASIS runtime supporting packed metadata, masking, normalization, and distributed training
- Increased temporal context from 2 to 12 historical months, **enabling 12-month input → 12-month prediction forecasting**

Ocean Foundation Model

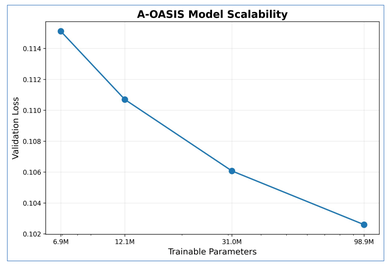
MODEL DESIGN CHALLENGES

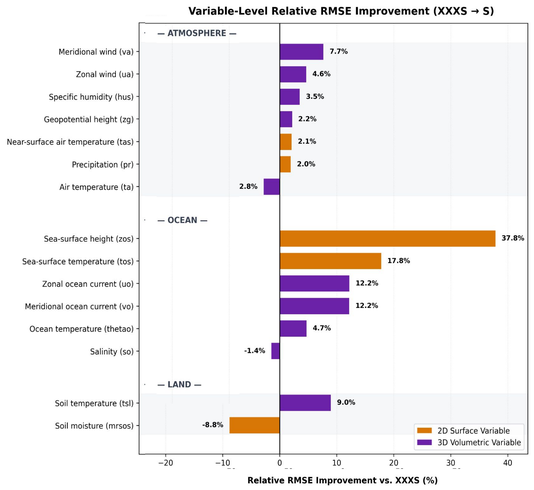
- Original encoder **assumed 1 vertical axis** for all variable entries
- Strict architectural assumption in encoder/decoder: variables must be jointly stackable with a **single shared vertical level axis** (failed with current dataset)
- Multi-system 3D variables with **different channel counts** in the same atmospheric branch violated key assumption
- Multiple encoders (one encoder per group of variables) may not preserve vital semantic relationships
- **Risk missing out on critical ocean-atmosphere interactions**

RESULTS

Scalability Study: Is Bigger, Better?

- View impact of model size on validation loss.
- Conducted foundation model pre-training across five model scales
- Scale metric is maximum trainable parameters
  - **XXXS, XXS, XS, S, M**
- Visualizing change in **relative RMSE** by variable





- A full epoch pretraining run of an **M-sized** model achieved a test loss of 0.0911 (420 M parameters)
- Pretraining configuration
  - **History:** 12 months
  - **Prediction Horizon:** 12 months
  - **Optimizer Steps:** 2750
  - **Learning Rate:** 5×10−5
  - Trainable parameters are the model weights and biases that are optimized during training

Figure 3. Relative changes in RMSE are shown for all predicted variables, grouped by Earth system domain. Highlighting how individual variables respond to increasing model capacity beyond the aggregate validation-loss trend.

DISCUSSION

Future Work

- Results demonstrate the benefit of scalability for Earth System modeling
- Introduce more accurate comparisons to baseline works and current state-of-the-art
- Evaluate downstream predictive tasks, e.g. precipitation projection + ENSO sea surface temp.
- Study forecast capability boundary to determine maximum temporal prediction capacity to verify our hypothesis, perform autoregressive evaluations

REFERENCES

1. Bodnar, C. *et al.*, "A Foundation Model for the Earth System," *arXiv:2405.13063* [physics.ao-ph], 2024.
2. Jaegle, A. *et al.*, "Perceiver IO: A General Architecture for Structured Inputs & Outputs," *arXiv:2107.14795* [cs.LG], 2022.
3. Liu, Z. *et al.*, "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," *arXiv:2103.14030* [cs.CV], 2021.
4. Petrie, R. *et al.*, "Coordinating an Operational Data Distribution Network for CMIP6 Data," *Geosci. Model Dev.*, vol. 14, pp. 629–644, 2021. doi:10.5194/gmd-14-629-2021.

ACKNOWLEDGEMENTS

This work is supported in part by the US Department of Energy SciDAC ImPACTS project and SciDAC RAPIDS project under Contract No. DE-SC0012704 FWP No. CC127 and No. CC122 respectively, as well as Brookhaven National Laboratory's SULI program.