E3SM Webinar

Advancing Earth System Modeling using AI/ML

Dan Lu

Senior Computational Earth Scientist
Oak Ridge National Laboratory

A joint effort with my research group and external partners.

October 23, 2025

Our AI efforts for E3SM

❖ **Emulators**:
- Build an accurate emulator of ELM with limited ensembles
- Improve comp. efficiency of ensemble generation and UQ

❖ **Hybrid ML-process modeling**:
- Use explainable AI to guide ELM development
- Build a hybrid ML-process river model

❖ **Downscaling**:
- Refine coarse-res ensemble outputs into high-res predictions
- Improve flood simulations
- Inform hydropower operations

❖ **Model calibration & UQ**:
- Use generative AI to automate parameter tuning and UQ
- Combine emulator and Bayesian inference for model calibration

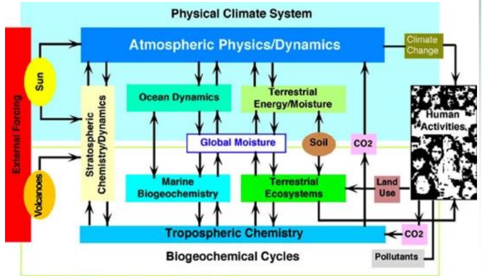
Five paradigms of Earth system modeling

1st Paradigm: Empirical Model

Experiments to explain or empirically describe natural phenomena

2nd Paradigm: Theoretical Model

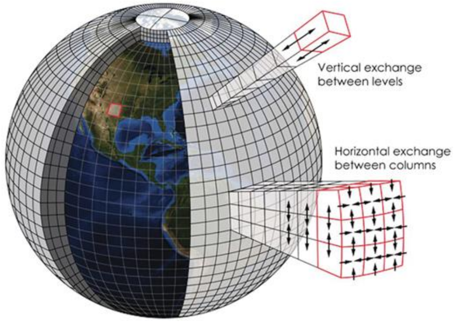
Develop physical laws, theoretical models



(from Earth System Science: An Overview, NASA, 1988)

3rd Paradigm: Computational Model

Computational models, simulating complex, coupled Earth system



4th Paradigm: Data-driven ML Model

A deluge of Earth system data have become available; Derive ML models from observation and simulation data.

5th Paradigm: AI Foundation Model

1600

1950

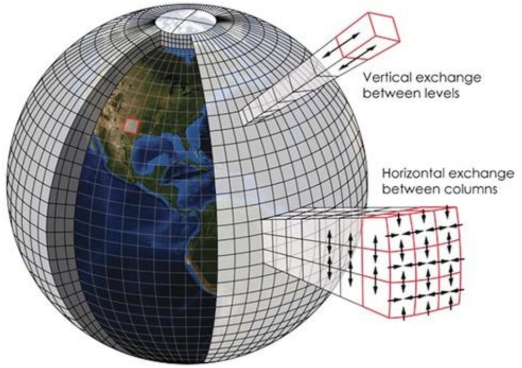
2000

2020

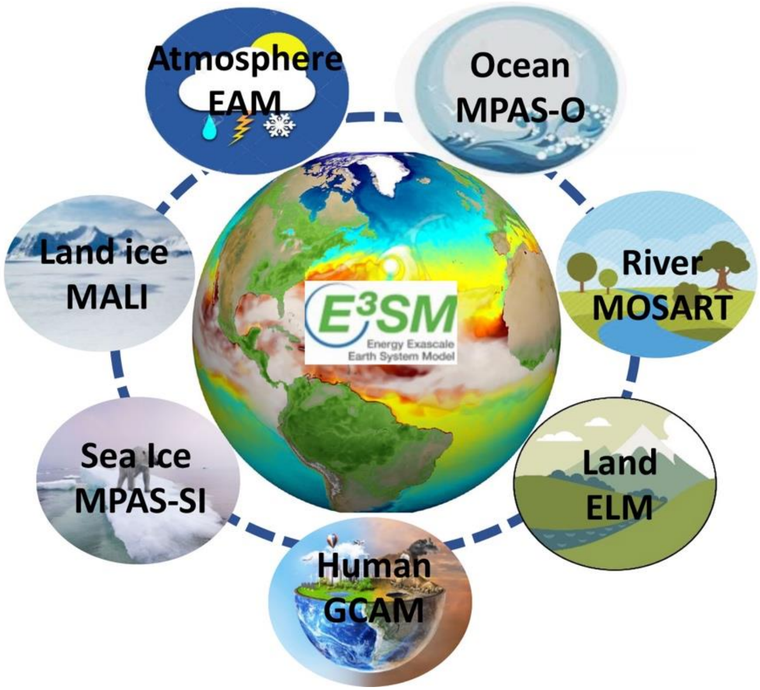
Physics-based Earth system modeling is costly and uncertain

3rd Paradigm: Computational Model

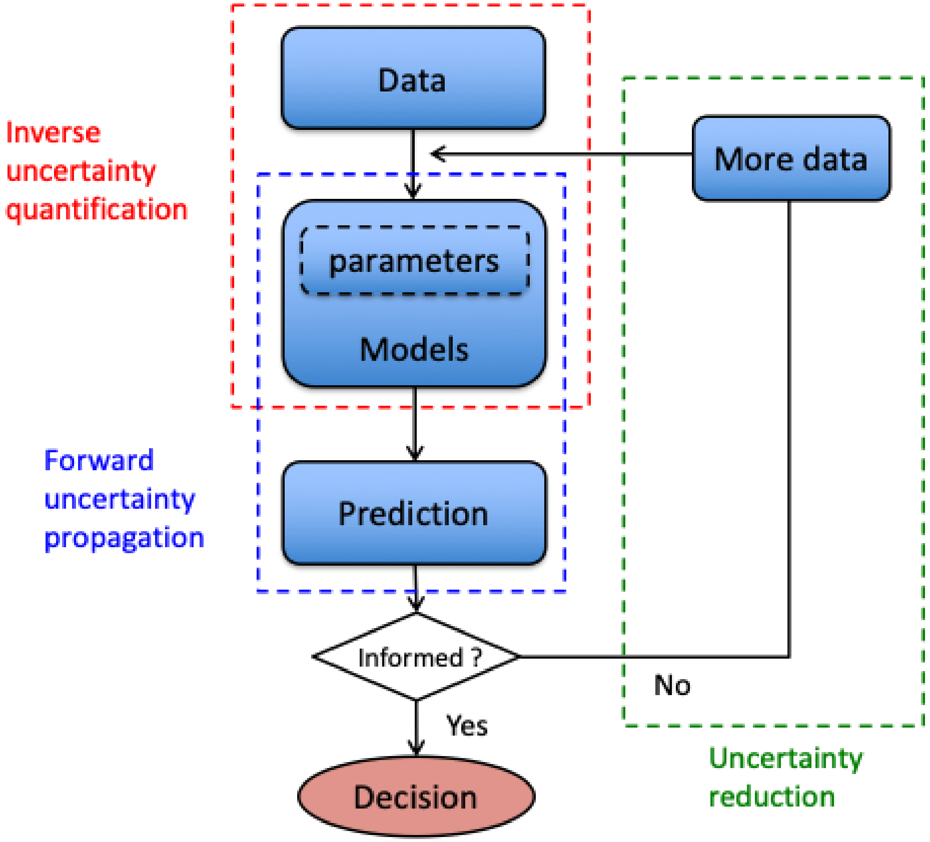
Computational models, simulating complex Earth system



Energy Exascale Earth System Model (E3SM)



Physical-based Earth system modeling

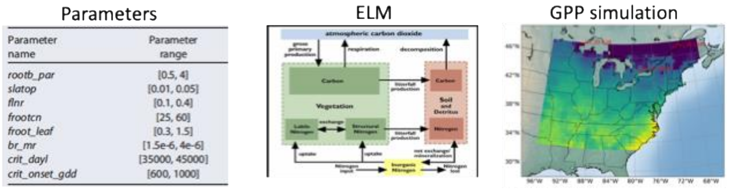


- Uncertainty quantification (UQ) is crucial for physics-based Earth system modeling (ESM) to enhance prediction accuracy and support informed decision-making.

AI/ML techniques for fast physics-based ESM and UQ

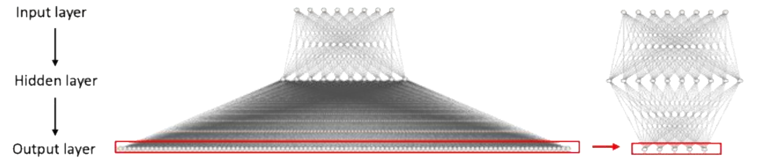
Surrogate Modeling

Build a fast surrogate of expensive model using ensemble simulations



Parameter name	Parameter range
rootb_par	[0.5, 4]
slatop	[0.01, 0.05]
flnr	[0.1, 0.4]
frootcn	[25, 60]
froot_leaf	[0.3, 1.5]
br_mr	[1.5e-6, 4e-6]
crit_dayl	[35000, 45000]
crit_onset_gdd	[600, 1000]

Build a surrogate model of ELM to simulate GPP in US



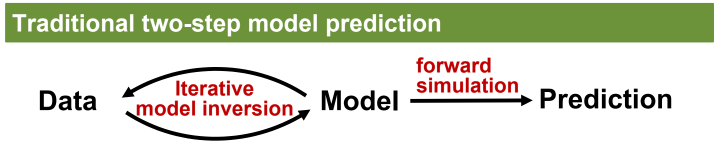
Neural network + dimension reduction build accurate surrogate with limited ELM ensembles

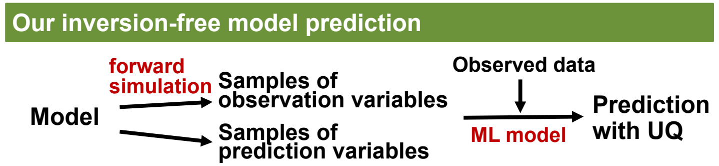
- Reduces single model run time
- Evaluating the surrogate in UQ reduces the total time
- Use NNs to build a surrogate

Lu, et al., *JAMES* 2018
Lu, et al., *GMD* 2019
Fan, Lu, et al., *J. of G. Gas Con.* 2024

Inversion-Free Prediction

Directly learn obs-pred relationship and generate predictions from observed data


Traditional two-step model prediction

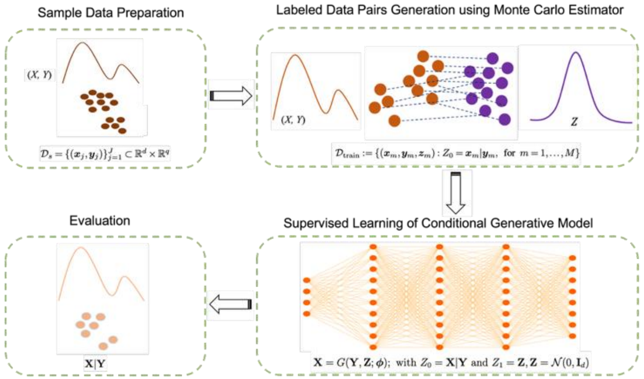

Our inversion-free model prediction

- Avoids expensive, iterative inverse modeling
- Combines offline simulation with online updating
- Enables fast data assimilation

Lu, et al., *ICDM* 2020
Lu, et al., *NeurIPS* 2021
Lu, et al., *Frontier in Energy Res.* 2022

Generative AI

Diffusion model generates samples for both forward and inverse UQ by evaluating NN



- Uses NN to estimate a generator of the target distribution
- Evaluates the NN to generate samples for UQ
- Comp. and storage efficient

Lu, et al., *JGR: ML and Comp.* 2024
Fan, Lu, et al., *JH* 2024
Fan, Lu, et al., *SoftwareX* 2025

Generative AI method (DBUQ) for efficient parameter calibration

- Objective: draws samples to approximate posterior distribution of parameter X given observed y,

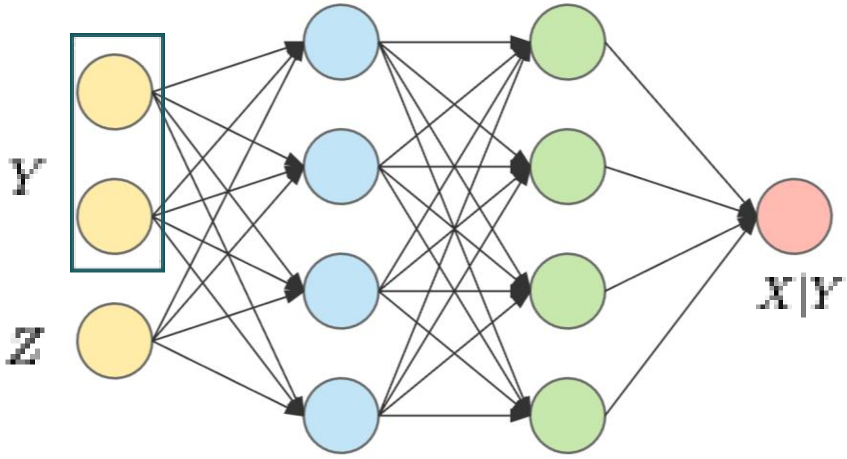
$$p(X|Y=y) \propto p(Y=y|X)p(X)$$

- Our diffusion-based UQ (DBUQ) formulates a generative model *F* to draw the target samples ,

$$X|Y \approx F(Y,Z;\theta)$$

- A neural network (NN) is trained to estimate *F*;
- After training, the NN evaluate Z to quickly generate desired parameter posterior samples

$$X|Y \text{ at } Y=y$$



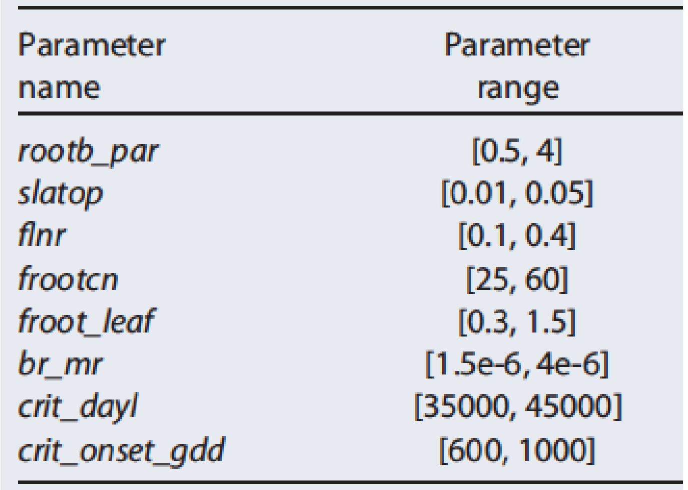
❖ Use a NN to learn the relationship between [Y, Z] and X|Y;
  - X|Y is the parameter of interest;
  - Y is the observation variable;
  - Z is the standard Gaussian variable.

❖ The generation of target samples of X|Y is both computationally and memory efficient;

❖ For any given observational data, the NN generates corresponding parameter posterior samples for UQ without re-running optimization, thus achieving amortized Bayesian inference.

Apply DBUQ to improve ELM parameter calibration

- Problem: Use DBUQ to estimate 8 ELM parameters;
- Observation: Annual averaged latent heat flux (LH) for 5 years at the Missouri Ozark AmeriFlux site in 2006-2010;
- Prior sample: 1000 samples from ELM simulation $\mathcal{D}_{\text{prior}} = \{(x_j, y_j)\}_{j=1}^{J}$
- Two case studies:
  - Synthetic case for method verification
  - Real observations application
- Compare DBUQ with MCMC for performance evaluation



Parameter name	Parameter range
rootb_par	[0.5, 4]
slatop	[0.01, 0.05]
flnr	[0.1, 0.4]
frootcn	[25, 60]
froot_leaf	[0.3, 1.5]
br_mr	[1.5e-6, 4e-6]
crit_dayl	[35000, 45000]
crit_onset_gdd	[600, 1000]

DBUQ

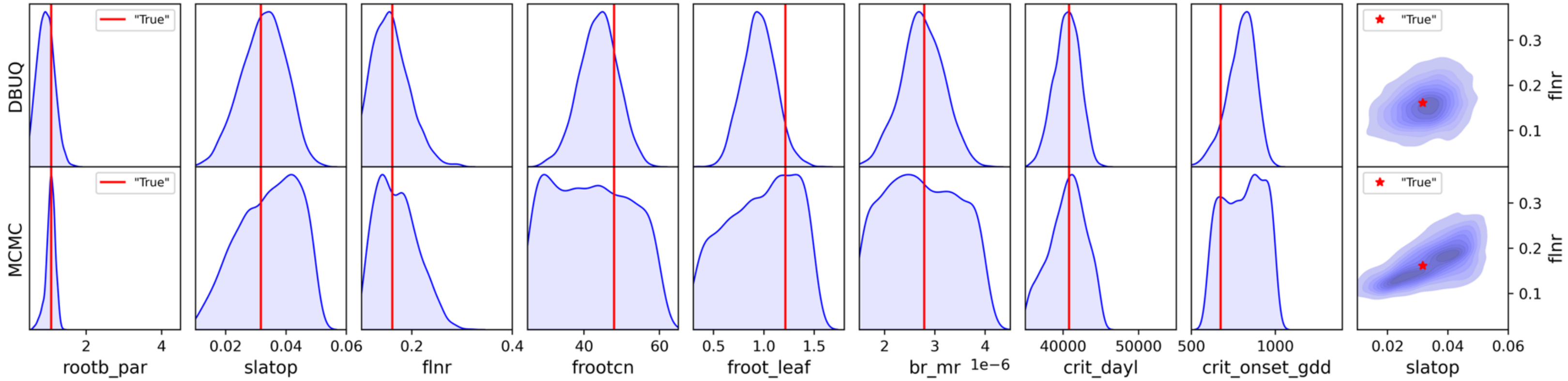
- Input: 1000 ELM samples $\mathcal{D}_{\text{prior}} = \{(x_j, y_j)\}_{j=1}^{J}$
- Output: **a trained generator** that quickly produces target samples for any observation;
- Computing time: < 10 min for solving both cases
- **Particularly suitable for site-specific ELM calibration at a global scale** due to its computational efficiency and amortized inference.

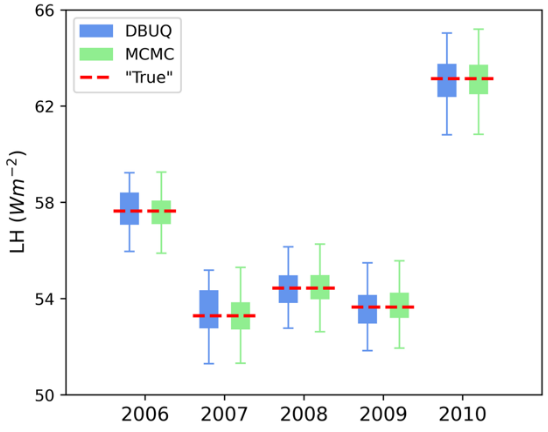
Surrogate + MCMC

- Input: 1000 ELM samples $\mathcal{D}_{\text{prior}} = \{(x_j, y_j)\}_{j=1}^{J}$
- Procedure: build an emulator using these samples, and then perform MCMC on the emulator;
- Output: **a set of posterior samples;** For a different observation, we need to re-run MCMC;
- Computing time: ~ 5 hours for one case to generate the same number of posterior samples as DBUQ.

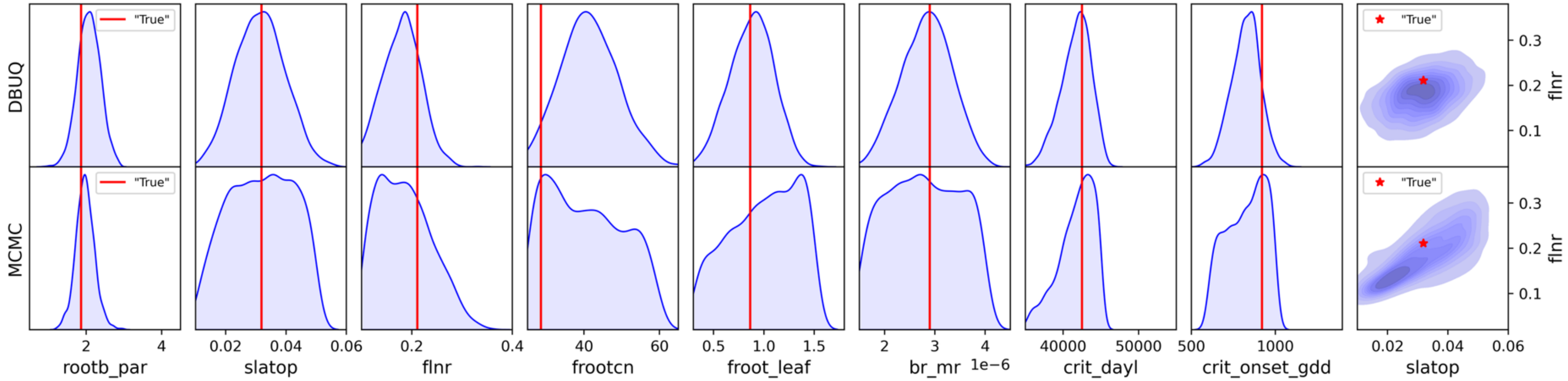
DBUQ accurately and efficiently estimated ELM parameters

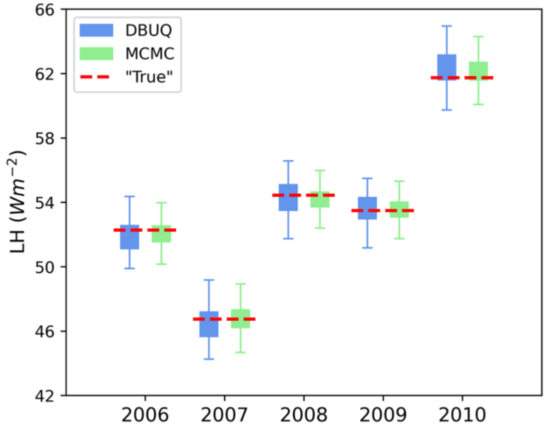
Synthetic case I





Synthetic case II

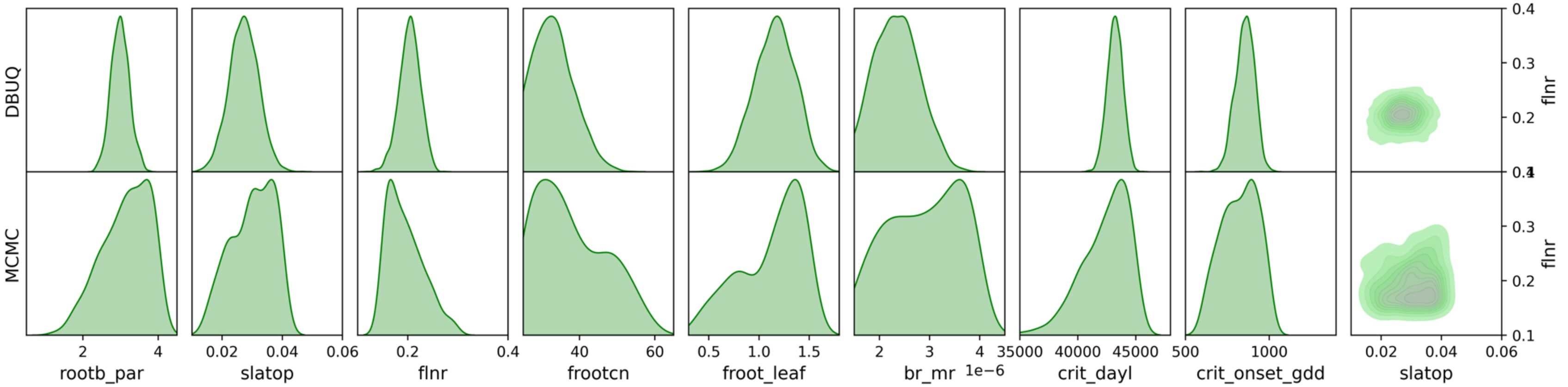


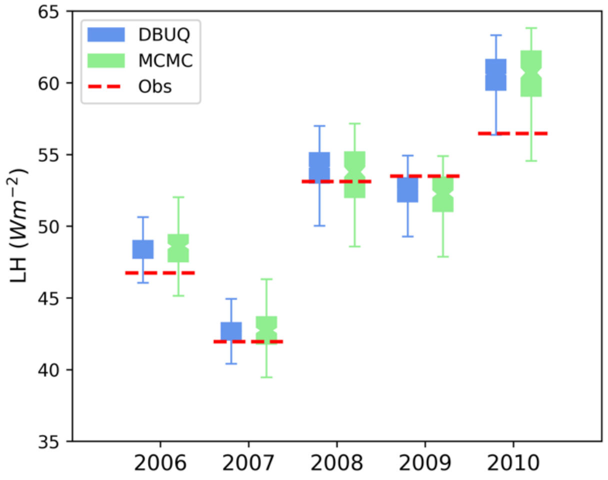


- ❖ DBUQ shows high accuracy in approximating the parameter posterior distributions.
- ❖ DBUQ demonstrates an accurate model calibration, as the prediction samples simulated from the parameter posterior samples are closely around the “true” observation.

DBUQ accurately and efficiently calibrated ELM

Real observation case





- ❖ DBUQ again shows high accuracy in approximating parameter posterior distributions.
- ❖ It showed accurate calibration, with prediction samples tightly enclosing the observations.
- ❖ DBUQ achieves comparable accuracy with MCMC with significantly less computational time.
  - DBUQ: 10 mins for all the three case studies;
  - MCMC: 5 hours for one case study;

https://github.com/patrickfan/GenAI4UQ

- Lu et. al, *JGR--Machine Learning and Computation*, 2024.
- Fan, M., Lu, D., et al., *SoftwareX* 2025

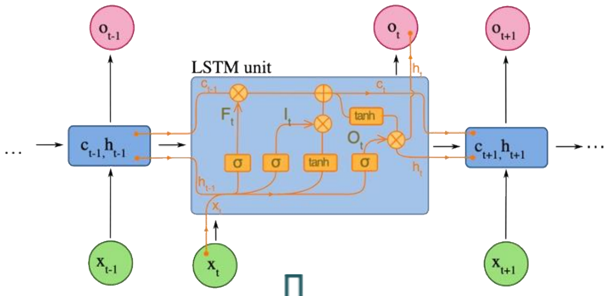
Data-driven Earth system modeling lacks trustworthiness

4th Paradigm:
Data-driven ML model

A deluge of Earth system data have become available; Derive ML models from observation and simulation data.

Long Short-Term Memory (LSTM)

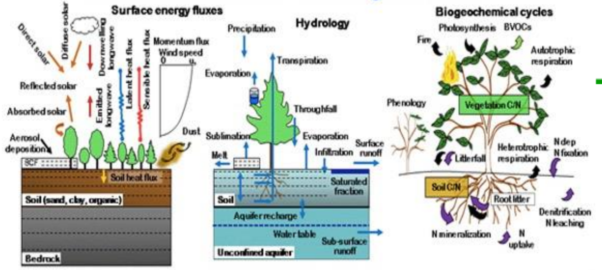
LSTM network learns system dynamics from observations of environmental drivers and carbon/water fluxes to predict future carbon/water fluxes



LSTM simulates a mapping for the inputs over time to an output to consider the memory effect of drivers.

Input: Observation of environmental drivers

Terrestrial Ecosystem



Output: Observation of carbon/water flux

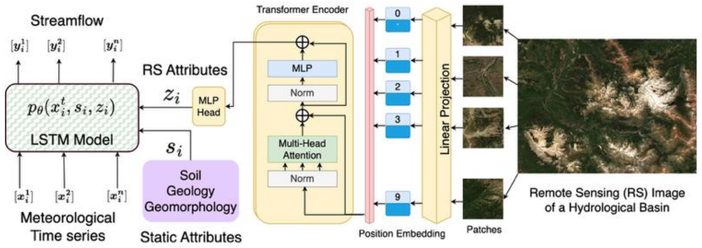
❖ ML models have shown success in ESM, but they have challenges for trustworthy prediction:

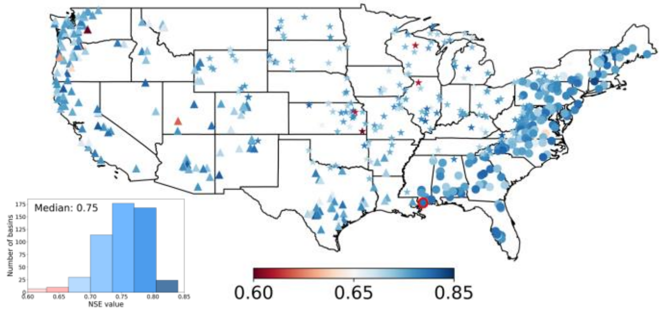
- How can we ensure that ML solutions generalize across space and time?
- How do we verify that models are making accurate predictions for the right reasons?
- How can we guarantee the reliability of predictions under changing environmental conditions?

Multimodal, explainable, reliable AI for data-driven ESM

Multimodal AI

- Vision transformer for satellite image
- LSTM for meteorological time series
- Integrate with static attributes to improve prediction





NSE value: = 1.0 means perfect fit; > 0.65 suggests good prediction

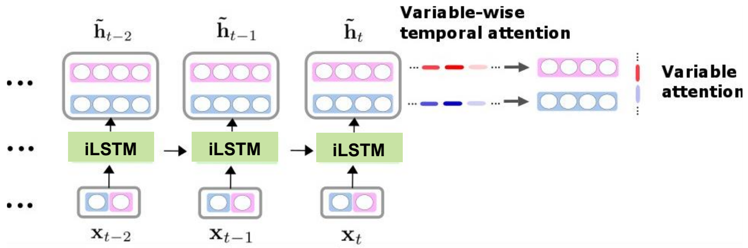
❖ Use advanced AI models to integrate diverse data to improve accuracy and generalizability.

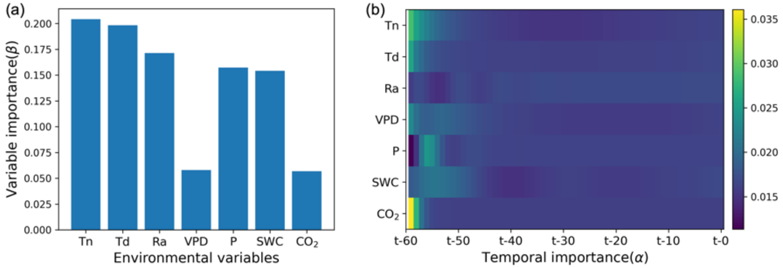
Tayal, Lu, et al., *ERL* 2024; *ICLR* 2024

Explainable AI

- Permutation analysis: SHAP
- Gradient-based method: IG
- White-box LSTM networks
- Attention maps of transformer model

Interpretable LSTM



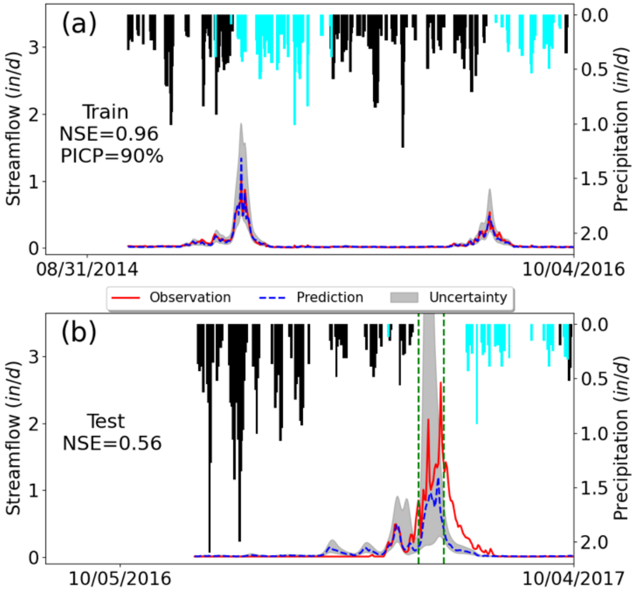


❖ Validate model decisions, ensure physical consistency and identify key drivers for prediction.

Lu, et al., *ICLR* 2022; Tayal, Lu, et al., *ICDM* 2024

Reliable AI

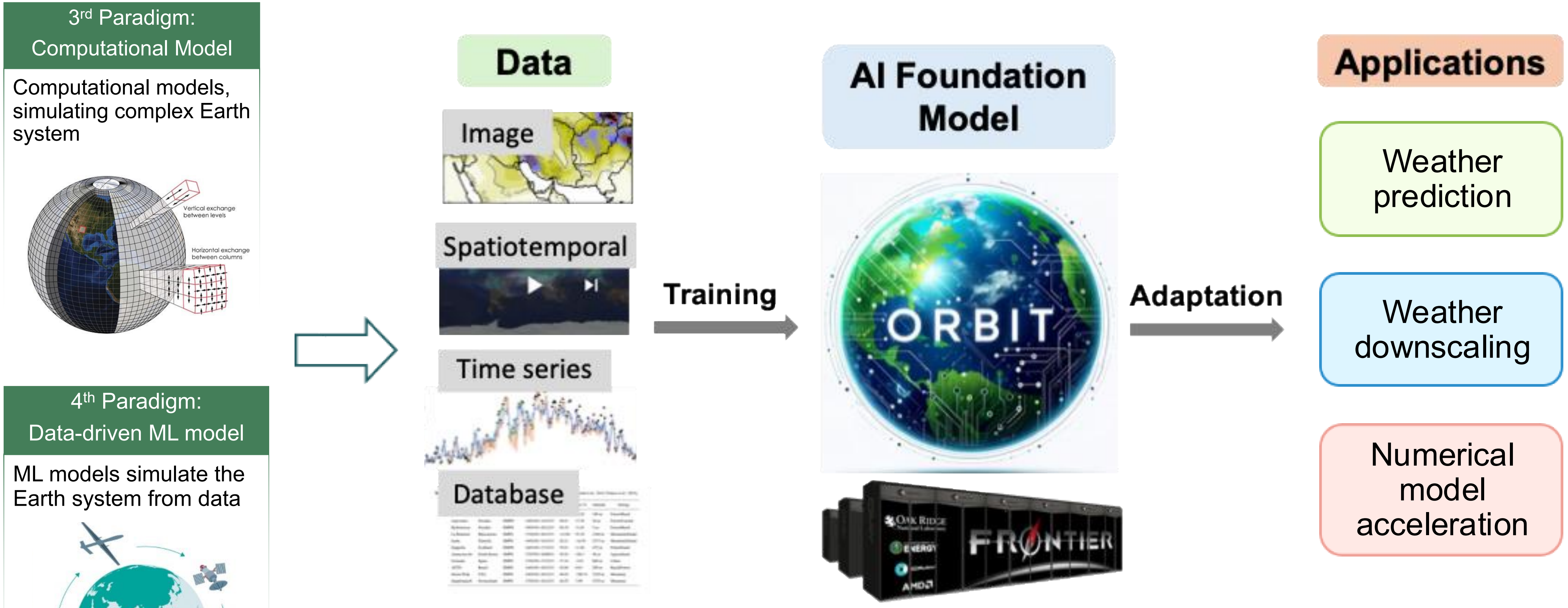
- Bayesian neural networks
- Gaussian processes
- Ensemble-based methods
- Prediction interval methods



❖ Quantify prediction uncertainty to ensure reliability under changing conditions.

Liu, Lu, et al., *ICLR* 2022; *Frontier in Water* 2023

From physics-based to data-driven, now to AI foundation models



3rd Paradigm: Computational Model

Computational models, simulating complex Earth system

4th Paradigm: Data-driven ML model

ML models simulate the Earth system from data

Data

Image

Spatiotemporal

Time series

Database

Training

AI Foundation Model

Adaptation

Applications

Weather prediction

Weather downscaling

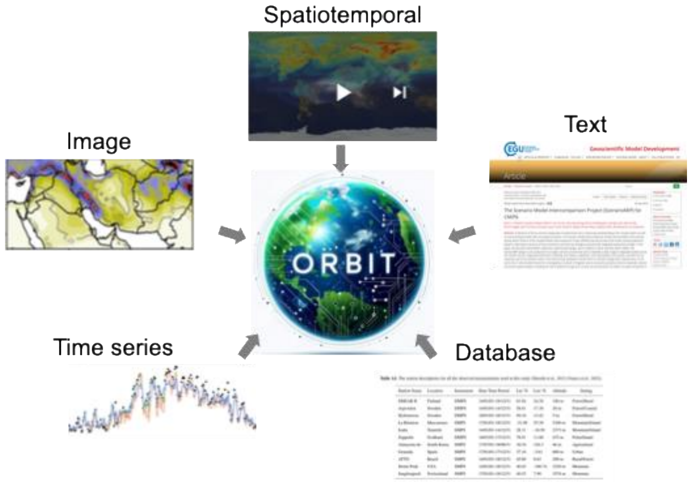
Numerical model acceleration

❖ AI foundation model is a large-scale neural network trained on extensive, diverse datasets and adaptable to a variety of modeling tasks.

AI foundation model can advance Earth system modeling

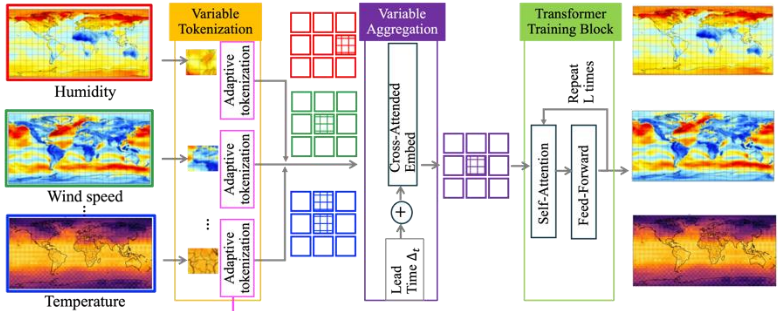
Heterogeneous Data

- From lab, field, satellite, and model simulations
- Multiple types, scales, and resolutions
- Cannot be fully integrated by numerical models or task-specific ML models



Scalable Model

- Vision Transformer model
- Integrate heterogenous data
- Performance scales with data size

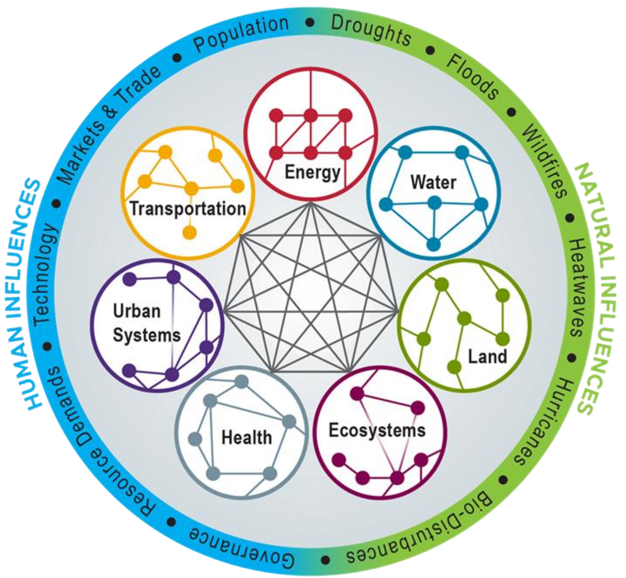


Foundation model:

- Integrate rich, multimodal data
- Reduce reliance on labeled data
- Ensure high versatility
- Improve accuracy, efficiency, and generalization

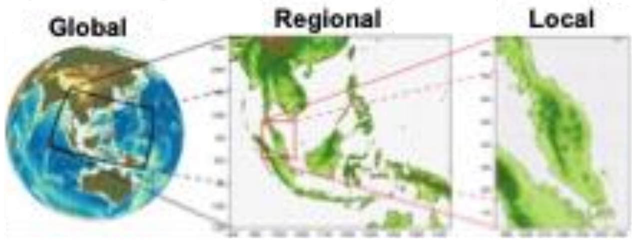
Various Applications

- ESM advances various scientific applications and impacts multiple sectors
- Foundation models save effort, cost, and energy



ORBIT: our AI foundation model for Earth system modeling

Downscaling



Generated high-res global weather data to support impact modeling

Weather Forecast

Forecasted weather 14 days ahead, demonstrating superior performance

ELM Acceleration

E3SM
Energy Exascale Earth System Model

Land ELM

Accelerated ELM simulation, advancing process insight and inform decision-making

Xiao et al., *SC24, SC25*;
DOE code: https://www.osti.gov/doecode/biblio/167711

ORBIT: weather forecasting

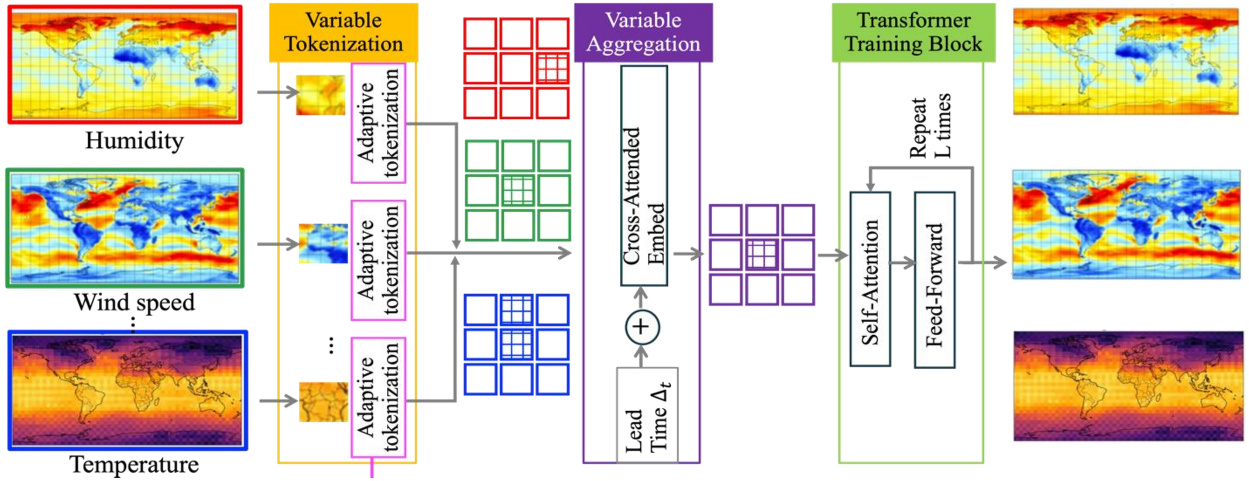
Pre-train on CMIP6 simulation dataset

- Simulation data from 10 CMIP6 models;
- Each provides 65-100 yrs of data at 6h interval;
- Consider 91 variables with spatial-res of 128*256;
- 1.2 million data point and 223.6 billion tokens.

Fine-tune on ERA5 reanalysis data

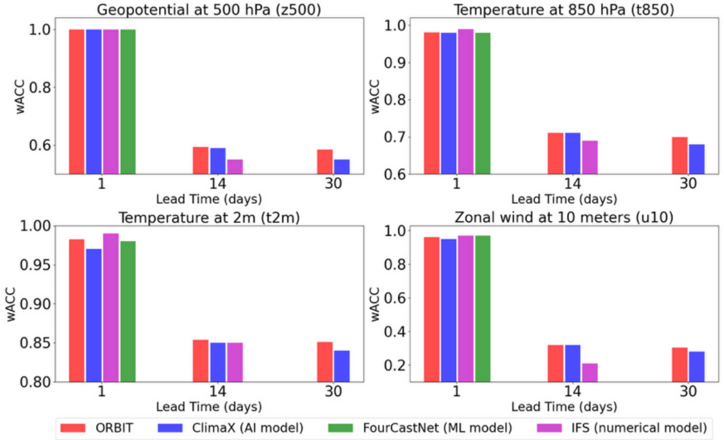
- Hourly data in 1980-2020
- 91 variables, 1.4 degree
- 4 TB (68 billion tokens)

Develop ViT models for effective and efficient learning



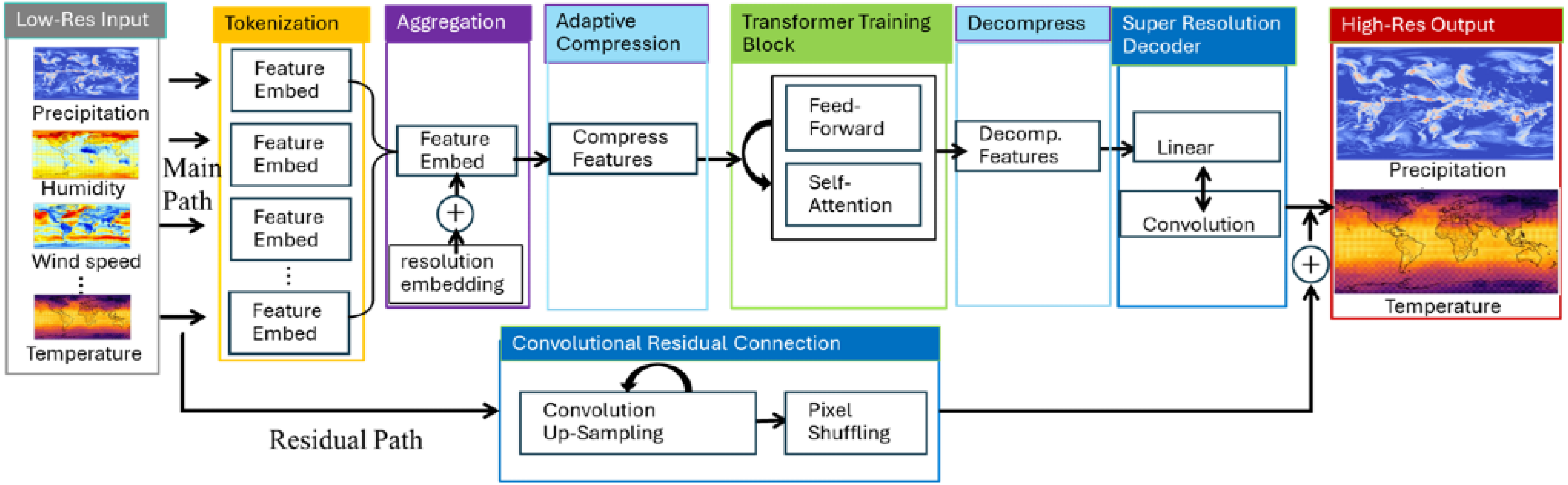
- Four model sizes with 115M, 1B, 10B, and 113B par.
- Largest AI model for ESM.

ORBIT enables fast and accurate weather forecasts



- ❖ ORBIT achieves competitive weather forecasting performance compared to state-of-the-art numerical, ML, and AI models.
- ❖ Single-GPU inference time under a second.

ORBIT: downscaling at national and global scale



- Multi-task Vision Transformer downscales multiple variables simultaneously;
- Resolution embedding and Bayesian loss improve accuracy, generalizability.

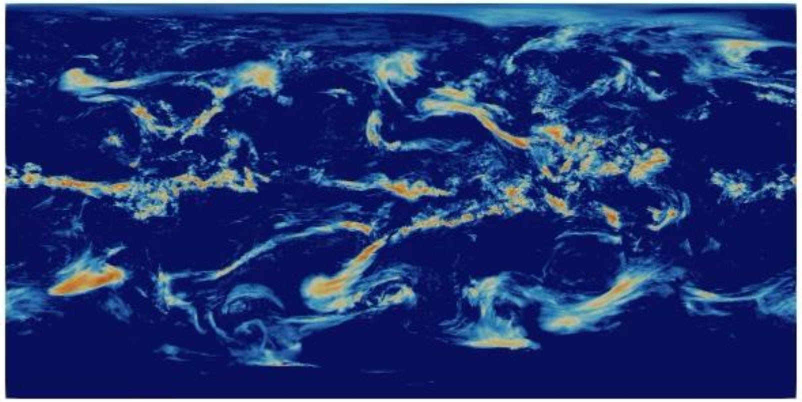
Dataset name	Region	Resolution (km)	Input Vars	Output Vars	Sample Size (in/out)	# Sample Pairs	Size (GB)
Pretraining							
ERA5 → ERA5	Global	622 → 156	23	3	[32, 64, 23] → [128, 256, 3]	367,920	200
ERA5 → ERA5	Global	112 → 28	23	3	[180, 360, 23] → [720, 1440, 3]	367,920	6,328
PRISM → PRISM	US	16 → 4	7	3	[180, 360, 7] → [720, 1440, 3]	14,235	189
DAYMET → DAYMET	US	16 → 4	7	3	[180, 360, 7] → [720, 1440, 3]	14,946	200
Fine-Tuning							
[ERA5, DAYMET] → DAYMET	US	28 → 7	23	3	[120, 240, 23] → [480, 960, 3]	14,946	113
Model Inference Evaluation							
ERA5 → IMERG	Global	28 → 7	23	3	[720, 1440, 23] → [2880, 5760, 3]	1,488	132

❖ Scales to 10B-parameter model across 65,536 GPUs, achieving 4.1 ExaFLOPS sustained throughput and 98% strong scaling efficiency on 8192 Frontier nodes.

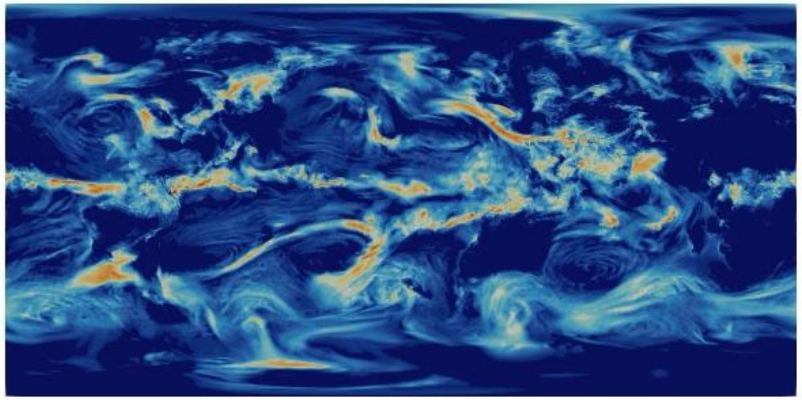
ORBIT accurately generated high-res global *precipitation* data

July 1 2020

Prediction:
Downscaled 7km global precipitation

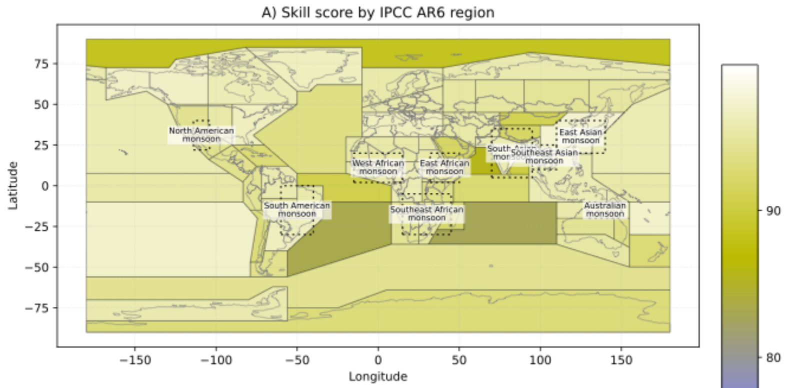


Input:
28km ERA5 data

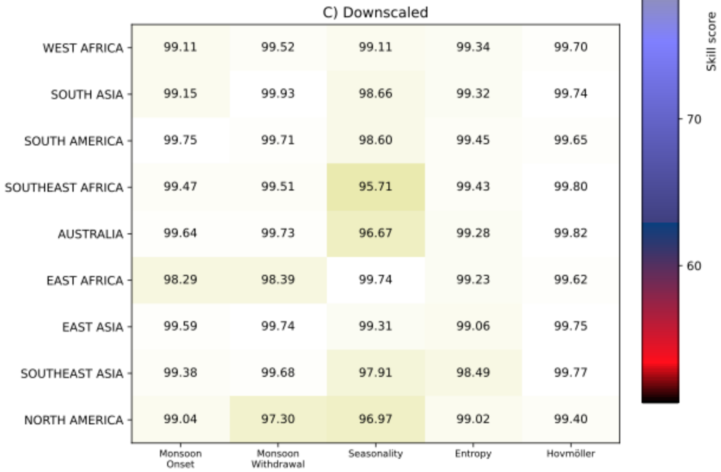


High Downscaling Skill Across Global Monsoon Regions

- Achieved high downscaling skill across 58 IPCC AR6 regions.



- Achieved high skill in 9 monsoon regions, where precipitation is governed by strong seasonality and complex land, ocean, atmosphere interactions.



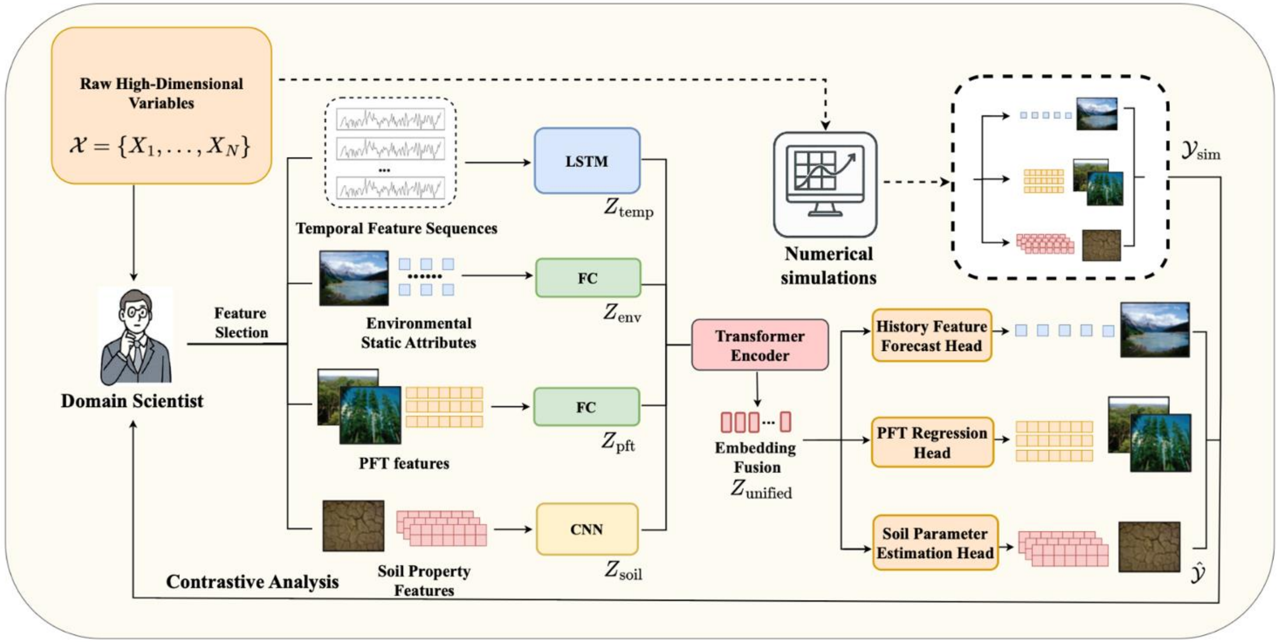
C) Downscaled

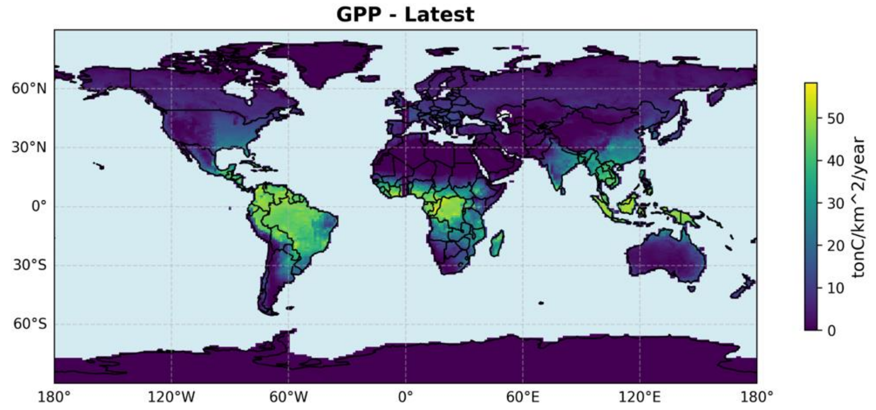
	Monsoon Onset	Monsoon Withdrawal	Seasonality	Entropy	Hovmöller
WEST AFRICA	99.11	99.52	99.11	99.34	99.70
SOUTH ASIA	99.15	99.93	98.66	99.32	99.74
SOUTH AMERICA	99.75	99.71	98.60	99.45	99.65
SOUTHEAST AFRICA	99.47	99.51	95.71	99.43	99.80
AUSTRALIA	99.64	99.73	96.67	99.28	99.82
EAST AFRICA	98.29	98.39	99.74	99.23	99.62
EAST ASIA	99.59	99.74	99.31	99.06	99.75
SOUTHEAST ASIA	99.38	99.68	97.91	98.49	99.77
NORTH AMERICA	99.04	97.30	96.97	99.02	99.40

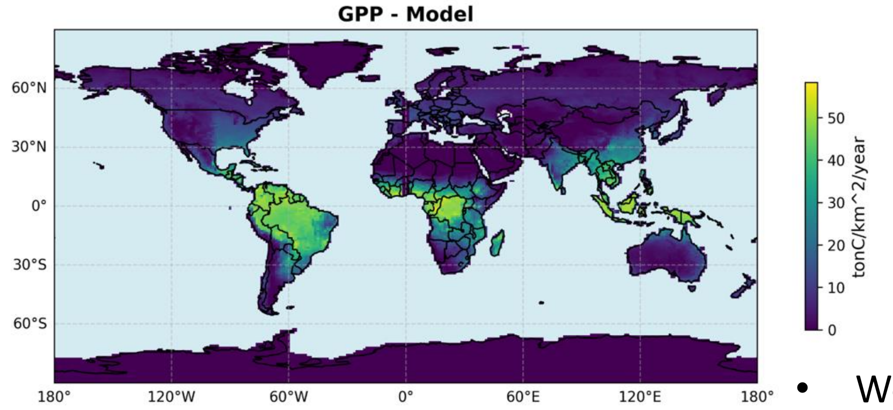
- Generated global 7km-res precipitation fields validated against IMERG satellite data.
- Accurately reproduced regional precipitation patterns and monsoon characteristics.

ORBIT: accelerating ELM simulation

- We leverage the ORBIT architecture to build a fast surrogate of ELM.
- Initial focus is on accelerating the BGC spin-up.




GPP - Latest


GPP - Model

- Achieved 10x speedup while maintaining high predictive accuracy ($R^2 > 0.97$), validated by ELM simulations, domain knowledge, and observations.
- Extending the approach to accelerate the entire ELM and additional E3SM components.

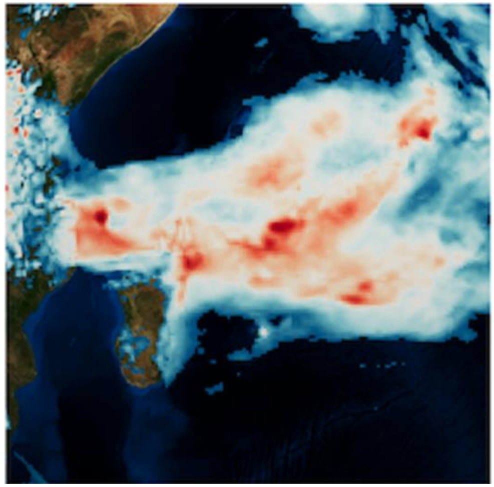
- Work is led by Dali Wang, Dan Ricciuto, Xiaojuan Yang, Xiaoying Shi, and Peter Thornton

ORBIT: AI foundation model advances Earth and energy systems

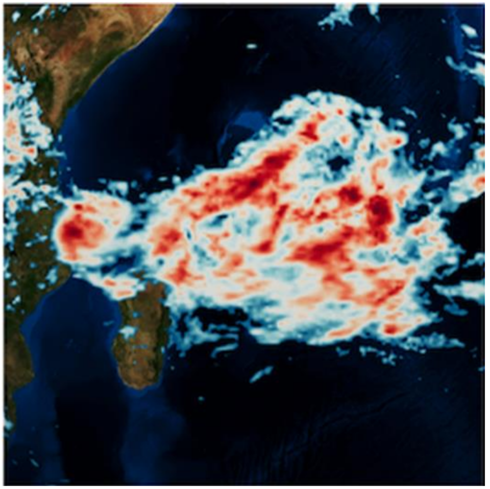
- **Weather forecasting**:
  - Top Supercomputing Achievement Award
  - Gordon Bell Prize Finalist in 2024

- **Downscaling**:
  - Nominated SC25 Best Paper Award
  - Gordon Bell Prize Finalist in 2025

E3SM simulated weather pattern

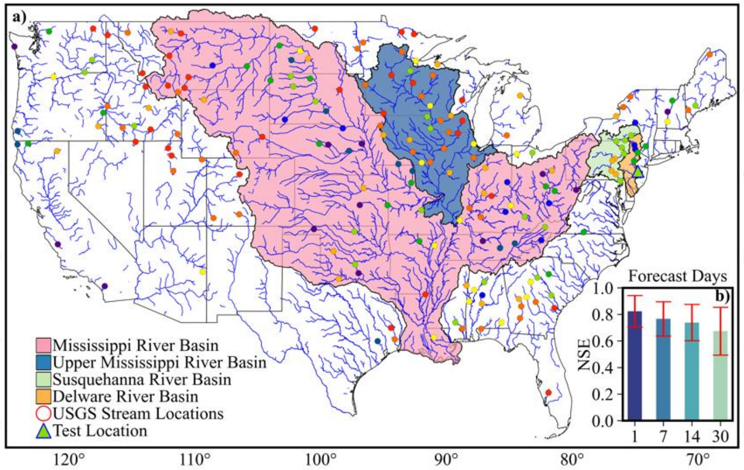


AI downscaled high-res weather data



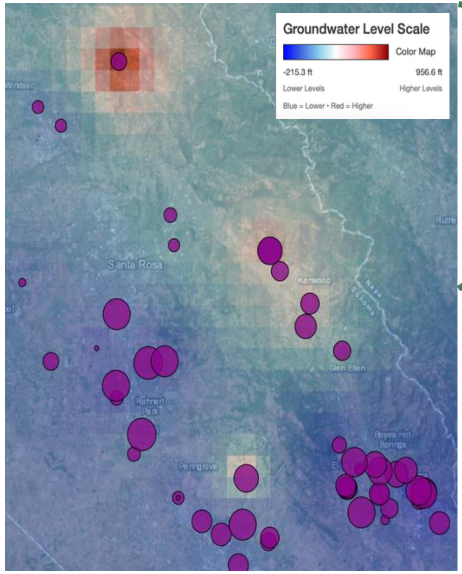
AI accelerated ELM

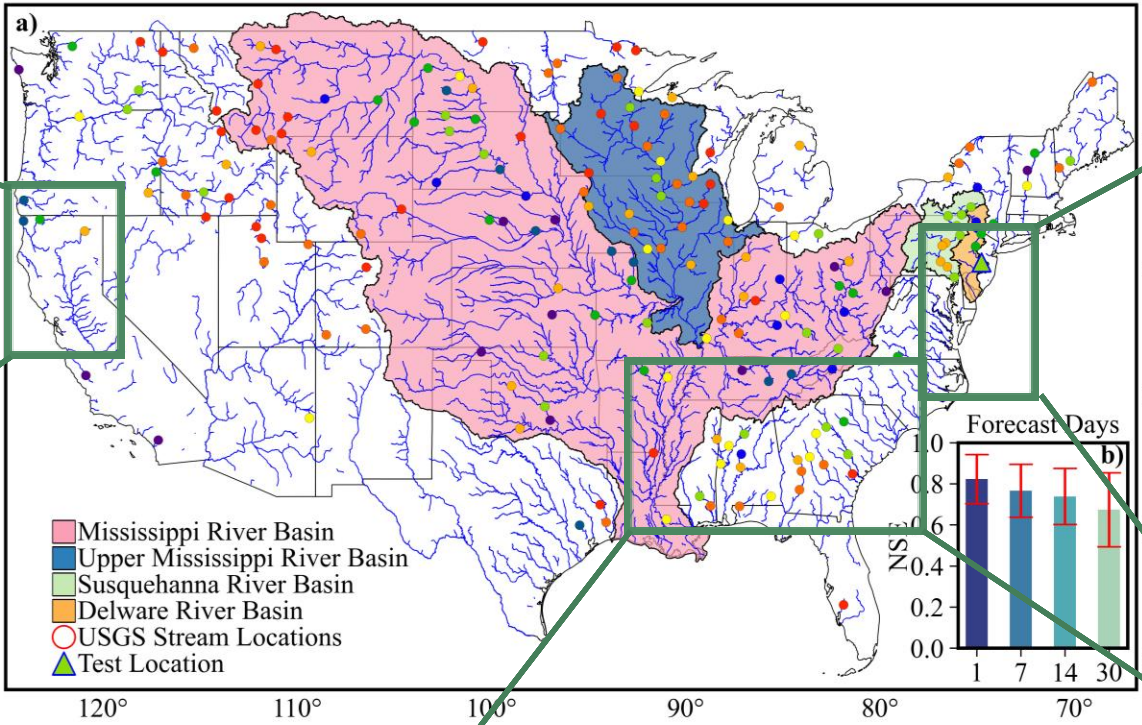
AI-advanced water-energy system



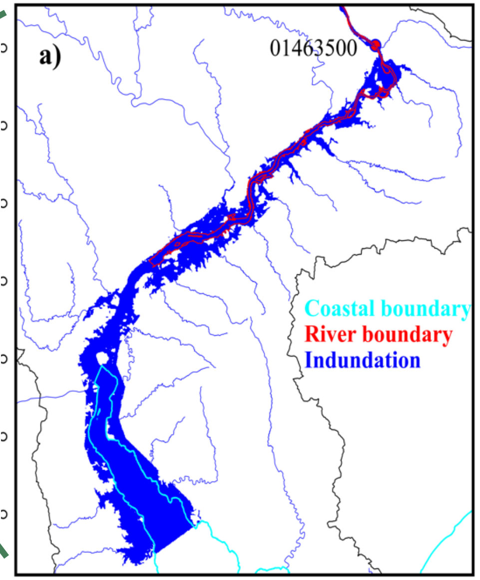
AI-enhanced water-energy system modeling capabilities

Pacific west region: Groundwater mapping

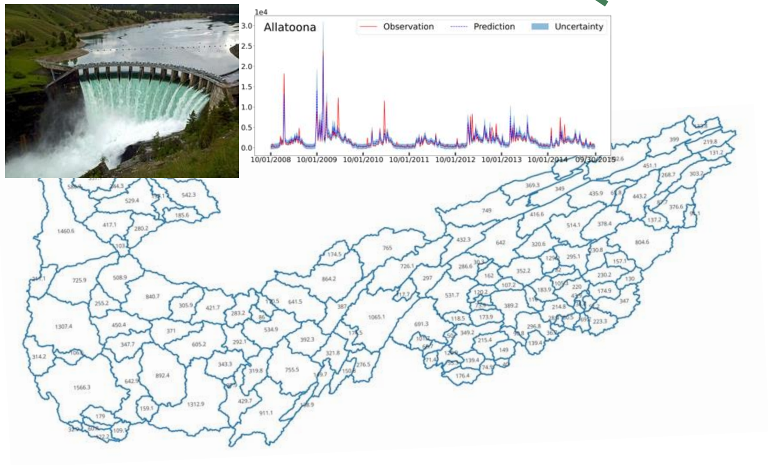




Mid-Atlantic region: Flood prediction



Southeast US: Reservoir inflow forecasting

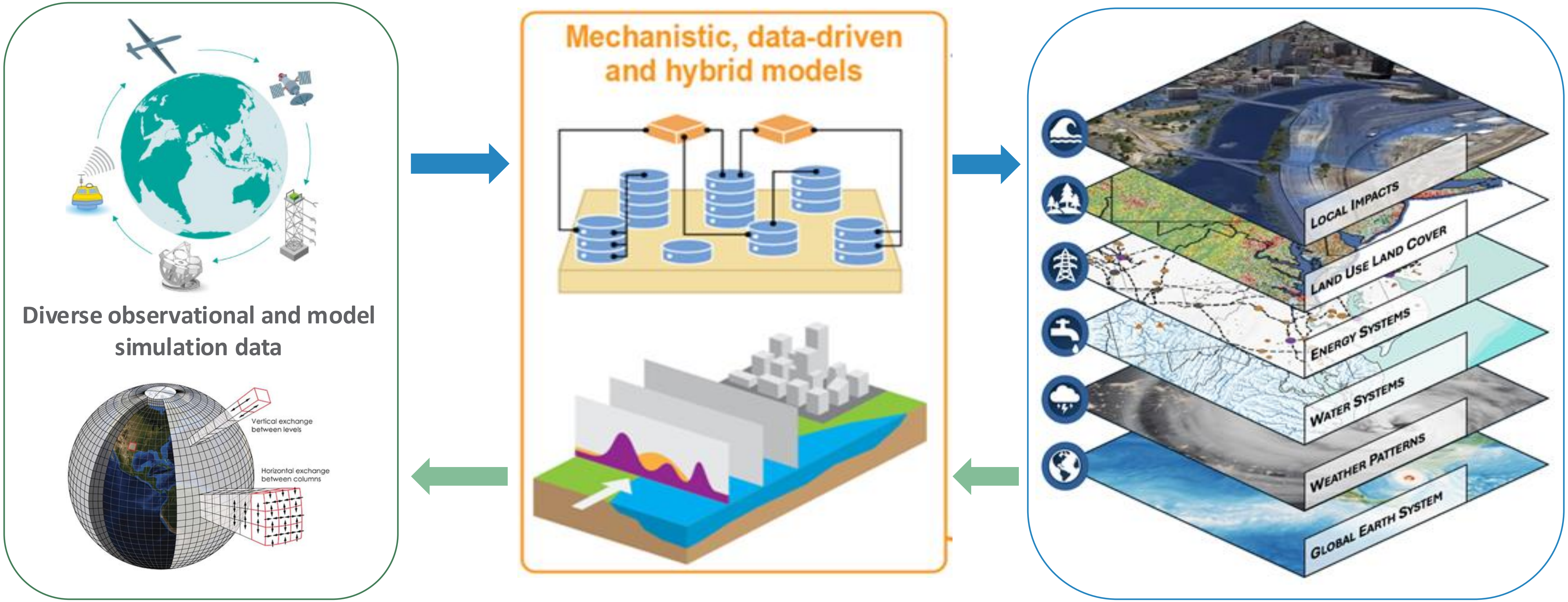


- Fan, Lu, et al., *Frontiers in Water*, 2023.
- Fan, Liu, and Lu, *JH,* 2023.
- Fan, Lu, et al., *EMS*, 2023.
- Ambika, Lu, et al., *GRL*, 2025.

Bridging scales through a two-way modeling framework

AI models integrate large-scale data to enhance local accuracy →

Large-scale data | **AI foundation model** | **Energy-related applications**

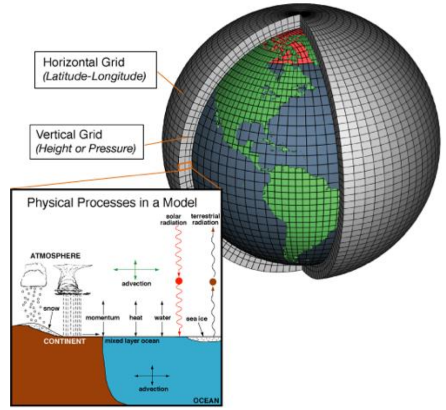


Diverse observational and model simulation data

← **Local testbed insights upscale science for broader impact**

Advancing Earth system modeling using AI/ML

Numerical Model



Challenges:

- High computational costs
- Large, multiple uncertainties
- Cannot integrate diverse data

Our AI/ML solutions:

- Surrogate modeling
- Inversion-free prediction
- Generative AI
- UQ; data assimilation

Data-Driven ML Model

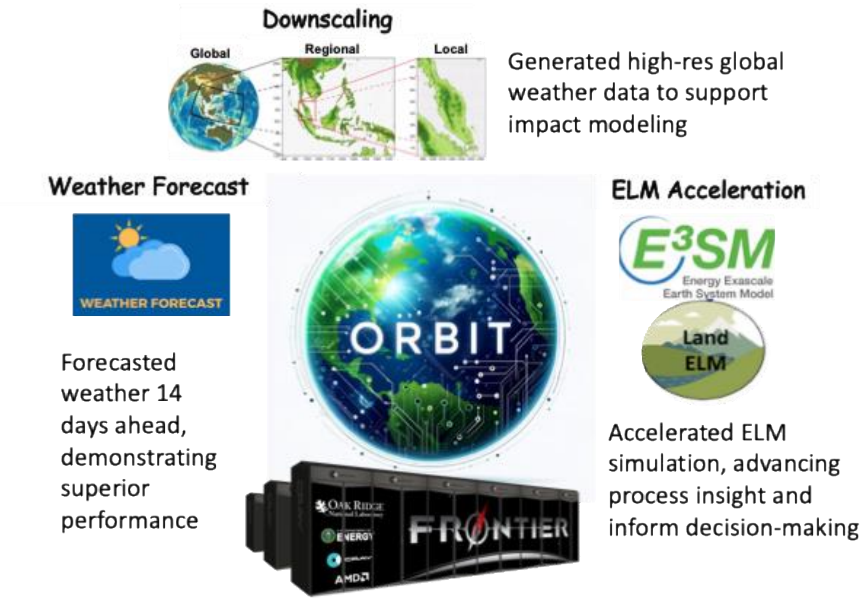
Challenges:

- Lack of explainability
- Poor generalizability
- Need trustworthiness

Our AI/ML solutions :

- Interpretable ML
- Physics-informed ML
- Physics-ML hybrid modeling
- UQ for AI/ML predictions

AI Foundation Model



Generated high-res global weather data to support impact modeling

Forecasted weather 14 days ahead, demonstrating superior performance

Accelerated ELM simulation, advancing process insight and inform decision-making

Our study: AI foundation model for ESM

- Integrate ”big data” and knowledge
- Use for a wide range of modeling tasks
- Save cost, effort, and energy
- Improve prediction, understanding, and generalizability

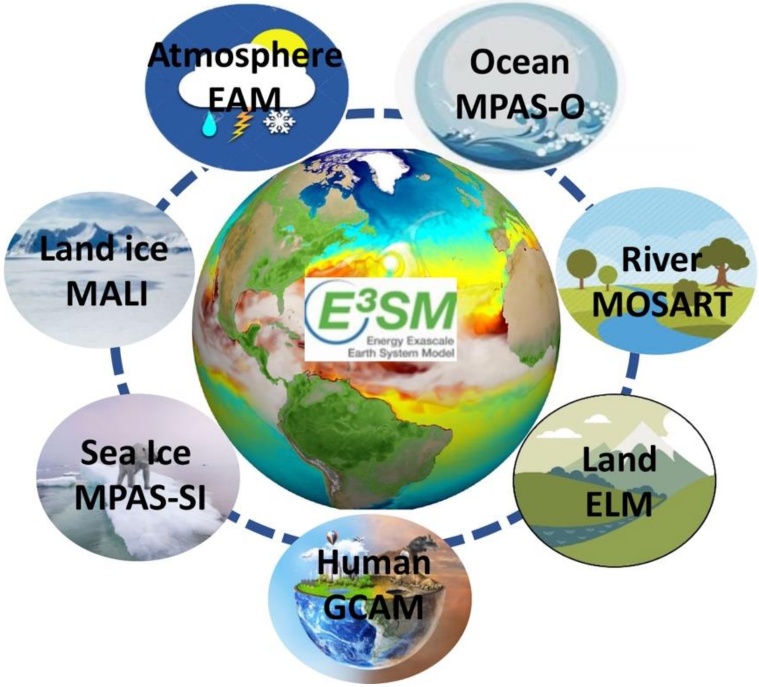
Our AI/ML efforts in advancing E3SM modeling capabilities

❖ **Emulators**:

- Build an accurate emulator of ELM with limited ensembles
- Improve comp. efficiency of ensemble generation and UQ

❖ **Downscaling**:

- Refine coarse-res ensemble outputs into high-res predictions
- Improve flood simulations
- Inform hydropower operations



❖ **Hybrid ML-process modeling**:

- Use explainable AI to guide ELM development
- Build a hybrid ML-process river model

❖ **Model calibration & UQ**:

- Use generative AI to automate parameter tuning and UQ
- Combine emulator and Bayesian inference for model calibration