AI-Enabled Approaches for Earth System Models: Bridging Model Hierarchies and Diagnosing Mechanistic Deficiencies

Nan Chen

Department of Mathematics
Institute for Foundations of Data Science
University of Wisconsin-Madison

Unpacking AI Weather Emulators: What and How Do They Learn?
September 28 – October 2, 2026, IMSI, Chicago

Beyond Forecast Skill: Using AI and Data Assimilation to Understand and Evaluate Models

AI weather emulators are often judged by forecast skill. But a recurring question at this workshop is:

Can we get the right answer for the right reason?

Good forecast skill alone does not tell us whether an AI emulator has learned credible dynamics.

Can **AI and data assimilation** help us use models and observations together to **understand and evaluate complex dynamical systems**?

1. Bridging model hierarchies

- Connect models at very different levels of complexity.
- Use AI to combine physical insight with high-dimensional Earth-system representations, rather than simply replace physics-based models.

2. Diagnosing model errors and mechanistic deficiencies

- Go beyond “does the output look right?”
- Selectively constrain one part of the system and trace how the response propagates along its dynamical pathways.

Data Assimilation as a Common Information Interface

Data assimilation (DA) combines a dynamical model with **partial and noisy information** to update the system state and its uncertainty. Importantly here, the model dynamics propagate the injected information to variables and regions that are not directly observed.

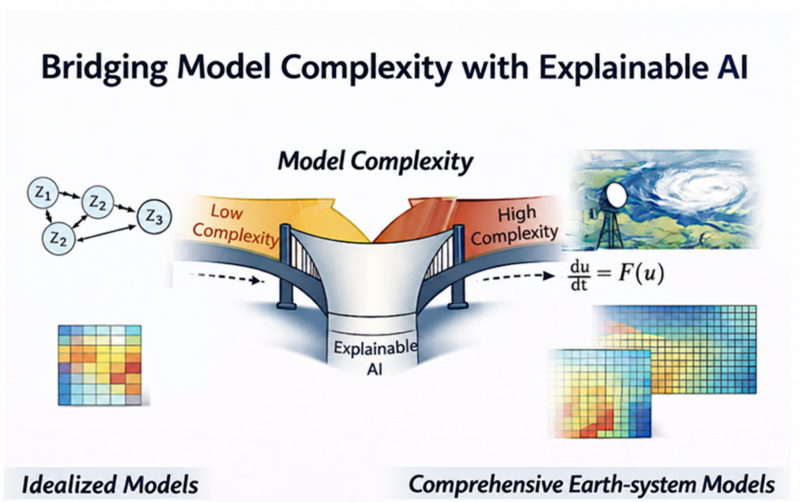
In this talk, DA is not mainly used for forecast initialization. We use it as a controlled way to **inject selected information** and study how a model **propagates that information**.

For high-dimensional Earth-system models, this update can be expensive. **Latent DA** performs the update in a reduced space while still constraining the high-dimensional state.

Key idea: design the latent space for **assimilation and information propagation, not only compression**.

Behnoudfar, Moser, Bocquet, Cheng, Chen, *npj Climate and Atmospheric Science* 9:65 (2026).

1. Bridging Model Hierarchies



Idealized / conceptual models

- inexpensive, interpretable, and useful for long simulations;
- can capture selected mechanisms and important statistics, including extremes;
- but contain only a subset of variables and scales.

Comprehensive Earth-system models

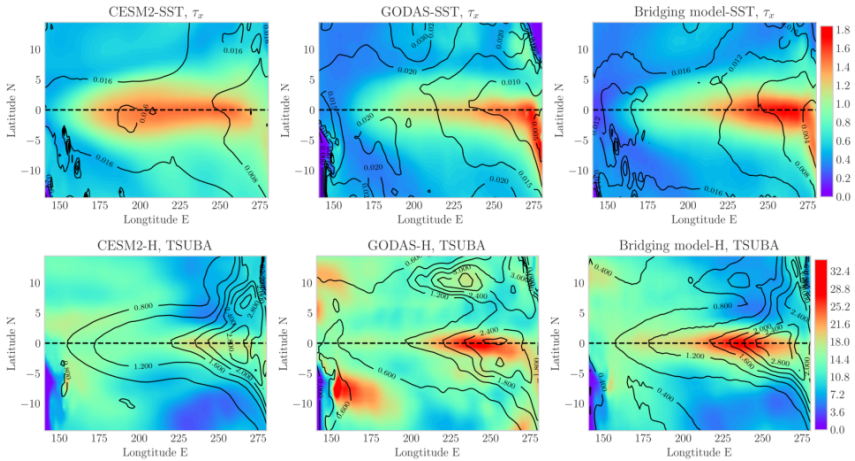
- resolve many variables and spatial structures;
- but remain expensive and can have systematic dynamical and statistical biases.

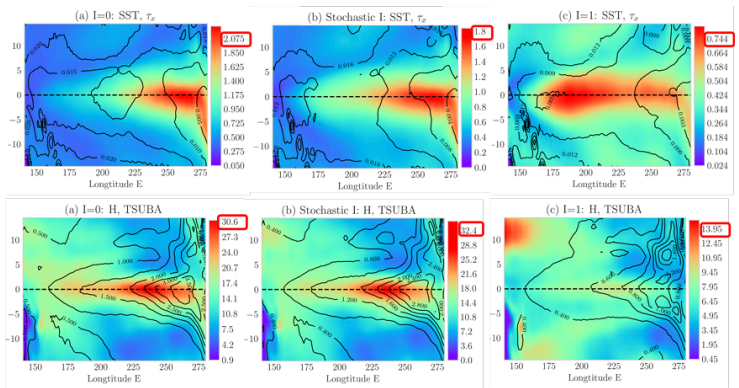
Motivation: Finite observational and reanalysis records cannot densely sample all relevant regimes, especially rare events.

Bridging model: a surrogate of the comprehensive Earth-system model that is **constrained by structured information from a simpler, physically interpretable model through latent DA**. The simpler-model output acts as “pseudo-observations”, allowing low-dimensional physical information to constrain a much richer spatial state.

ENSO: From Interpretable Dynamics to High-Resolution Fields

- **ENSO diversity** is a demanding test: capturing spatial patterns, timing, non-Gaussian statistics, and extremes remains challenging.
- A recently developed intermediate-coupled model can capture important ENSO statistics and mechanisms, but it is defined only along the equatorial Pacific and does not represent the full spatial state.





- The bridging model transfers that information into the high-dimensional Earth-system representation (CESM2).
- In the quantities shown here, the bridge moves CESM2 toward reanalysis **both along and away from the equator**.

Because the idealized model is cheap, we can also run efficient physics-guided "what-if" experiments.

- e.g., persistently weakened or strengthened Walker circulation;
- propagate the scenario into high-resolution fields through the bridge.

What the Bridging Framework Can Enable

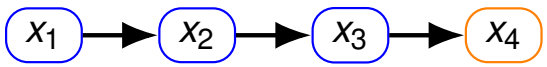
- **Long, high-dimensional synthetic records.** The bridge can augment the limited set of expensive comprehensive-model simulations with additional high-dimensional synthetic data for **AI emulators** and uncertainty studies.
- **A route to rare events and extremes.** Long, inexpensive simulations can be generated in the lower-dimensional model and then translated into much richer spatial fields through the bridge.
- **Physics-guided "what-if" experiments.** Perturb a physically interpretable component in the lower-dimensional model, then propagate its response into the high-dimensional Earth-system representation.

The bridge combines interpretability and computational efficiency with high-dimensional spatial structure.

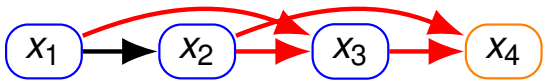
2. Data Assimilation as a Probe of Model Dynamics

Why output agreement is not enough

Different internal pathways can produce similar historical output. **Compensating errors** may hide a wrong mechanism **until the forcing or background state changes**.



reference pathway



same output statistics, distorted pathways

Diagnostic idea:

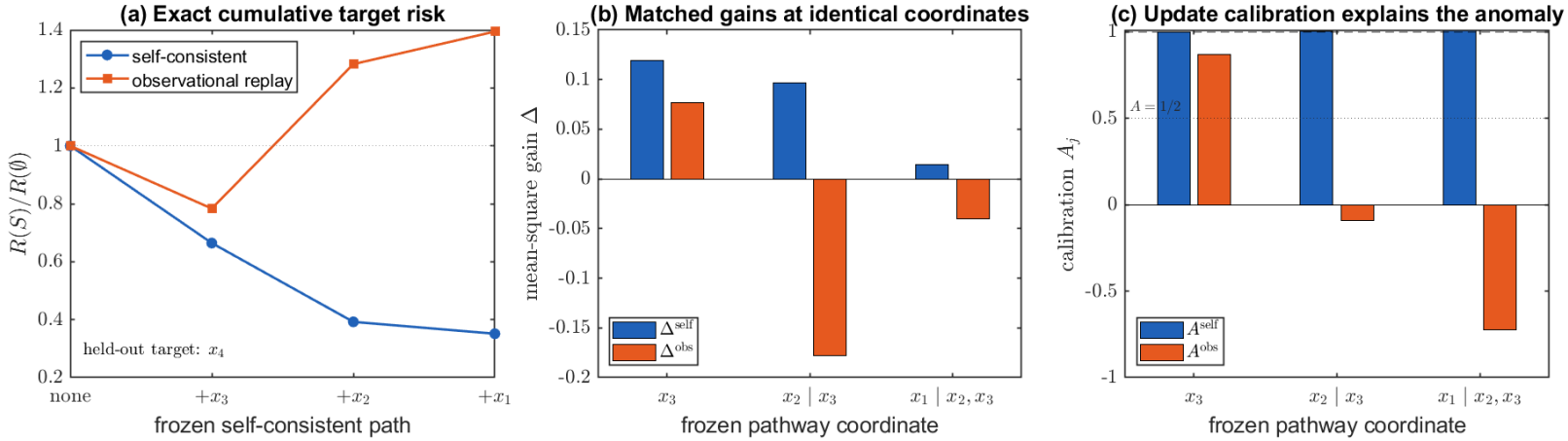
1. Hold out a target state, event, or metric.
2. Add one selected source of information.
3. Measure the induced reconstruction update.

Data assimilation injects targeted information and lets the model dynamics propagate it.

The reconstruction update is therefore a probe of how the model transmits information.

See Charlotte Moser's poster tomorrow for more details on the model-interrogation framework.

From Information Gain to Pathway Diagnosis



Mathematical benchmark

$$\Delta_j^{\text{opt}}(S) \geq 0 :$$

More information cannot increase the conditional-optimal MSE.

- Δ_j tells us **whether adding this pathway reduces the reconstruction error**.
- A_j tells us **whether the correction generated along that pathway has the right direction and magnitude.**

  Think of A_j as the factor that would make the update best match the pre-existing error along that update direction:

  $A_j = 1$: right size; $A_j > 1$: update too weak;

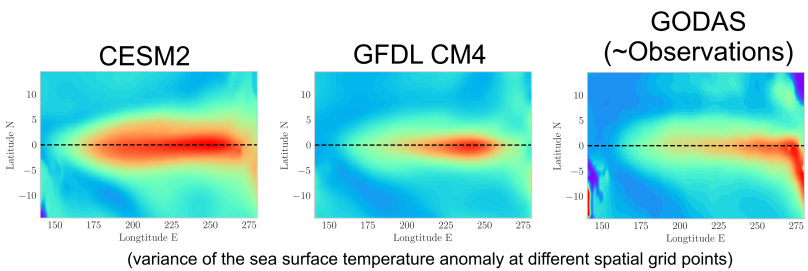
  $0 < A_j < 1$: update too strong; $A_j < 0$: wrong direction.

They localize **dynamically inconsistent pathways** and indicate where the model response should be examined or recalibrated.

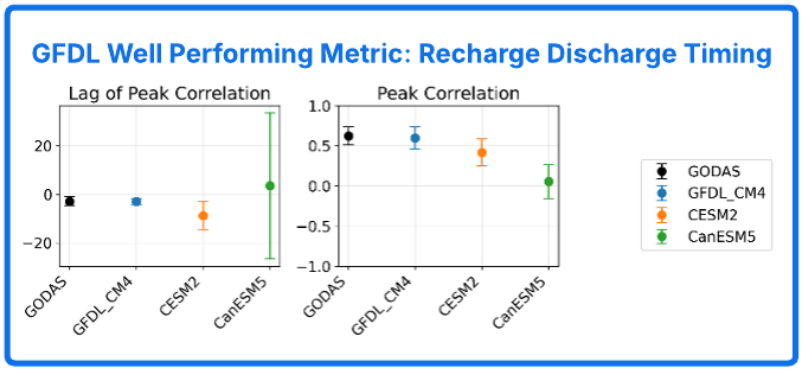
Moser, Behnoudfar, Chen, *Beyond Model Outputs: Interrogating Dynamical Models through Information Propagation*, in preparation.

Testing the Pathways Behind a Successful ENSO Metric

GFDL-CM4 already reproduces an important ENSO timing relationship between **upper-ocean heat content** and **SST** reasonably well.



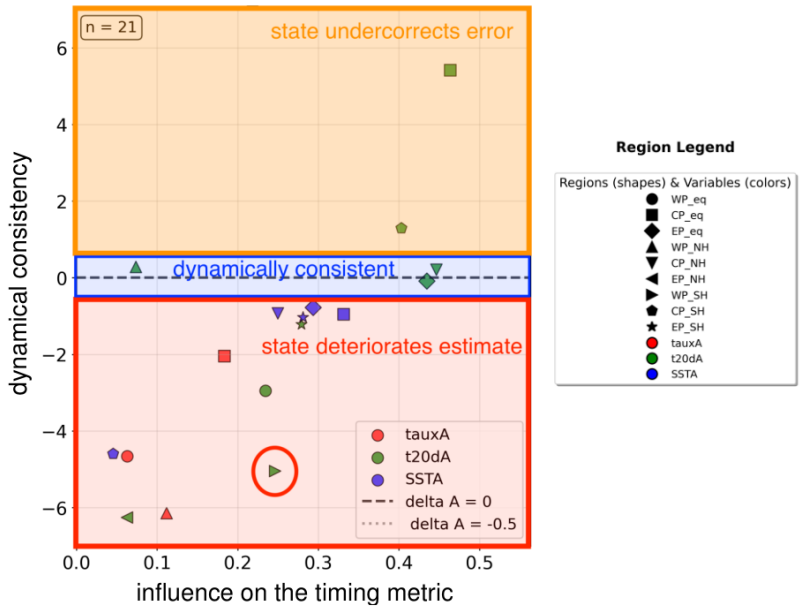
(variance of the sea surface temperature anomaly at different spatial grid points)



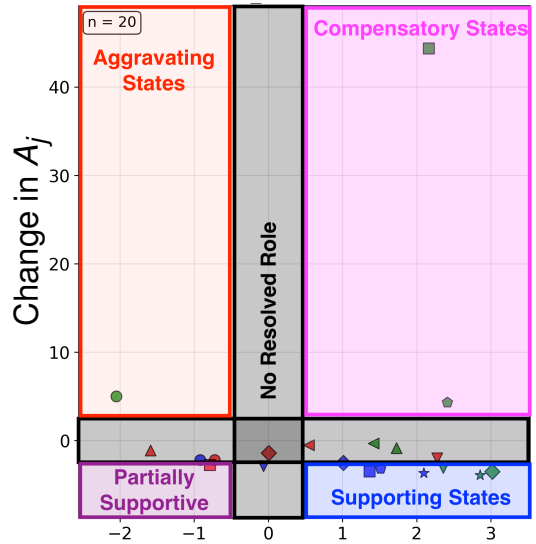
We therefore use it as a deliberately hard test:

Is the successful output supported by consistent information pathways?

Testing the Pathways Behind a Successful ENSO Metric



- Probe 21 Pacific region-variable pathways involving wind stress, thermocline depth, and SST.
- For one influential thermocline-depth source, the observation-driven response departs strongly from the self-consistent benchmark.



- Add further states and ask whether they **support**, **aggravate**, or **compensate for** the diagnosed mismatch.
- This localizes the inconsistency in information propagation.

A good output metric does not guarantee consistency of the underlying information pathways.

A Common Thread: Using Information to Understand Models

Latent data assimilation provides a common information interface.

1. Bridge model hierarchies

Use structured information from simpler, interpretable models to constrain much richer Earth-system representations.

2. Interrogate model dynamics

Use targeted observations to test how a model propagates information and to localize dynamical inconsistencies.

Potential uses: targeted calibration, emulator / closure evaluation, and observation design.

Related direction: Assimilative Causal Inference

The same information-based viewpoint can be extended to **state-dependent causal inference**. Combined with scalable latent DA, this provides a route to studying **precursors and influence ranges of extreme events** in higher-dimensional systems. (Andreou, Chen, Bollt, *Nature Communications* 17.1 (2026): 1854.)

AI and data assimilation can provide new ways to use information to understand and evaluate complex dynamical models.

Thank you!

(Please feel free to contact me at chennan@math.wisc.edu)