Tracing and Testing Information Pathways to Extreme Events: Causal Mechanisms and Model Interrogation

Nan Chen

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

Extreme and Rare Events: Modeling, Algorithms, and Applications
Sep 26–27, 2026, ICERM, Brown University

Introduction

- Extreme events are rare environmental conditions that exceed normal historical thresholds. They usually correspond to the tails of a probability density function.
- Similar events may arise from stochastic excitation, nonlinear instability, slowly varying feedbacks, or compensation among processes.
- It's important to understand **how the system reaches an extreme event** and **if there exist early-warning signals**.
- For prediction, attribution, and model credibility, the relevant object is therefore an **information pathway through the dynamics**.

To understand extreme events, we often ask ...

1. What pathway drives the event, and over what time range?
2. Does a model transmit that information through the right dynamics?

Note: Although the focus of the talk is about extreme events, the methods presented here are designed for general complex dynamical systems.

Outline

Part I: Trace pathways in the system

Assimilative causal inference (ACI) asks what drives the event and how long the influence persists.

Part II: Test pathways inside a model

Reconstruction-based model interrogation asks whether the model propagates information through the right dynamical pathways.

Part I

Tracing information pathways

Assimilative causal inference and causal influence range

What is Causal Inference?

- Many widely used causal inference methods in dynamical systems are time-series based:
  - Correlation or lagged correlation: correlation is not causality
  - Granger, transfer entropy, etc: time-averaged and predictive relationships, lacking state-dependent and instantaneous cause-effect roles
- Recent model-based methods (information transfer, statistical response, etc): often rely on explicit probability density functions and their derivatives, which become computationally intractable in high dimensions.

Most of these are predictive causal inference.

Assimilative Causal Inference (ACI)

ACI is a causal inference framework that leverages Bayesian data assimilation to reveal cause-and-effect relationships.

- ACI identifies dynamic causal interactions without requiring explicit observations or interventions on candidate causes, accommodates short datasets, and scales efficiently to high dimensions.
- It enables online tracking of causal roles, which may reverse intermittently, and provides a mathematically rigorous criterion for causal influence range.
- Because the diagnosis is state-dependent, ACI can distinguish **different mechanisms that lead to similar-looking extreme events**.

ACI: Use Future Effects to Infer a Present Cause

Let x be the observed potential effect and y a candidate (possibly unobserved) cause. Assume the underlying turbulent dynamical model and one observed time series of x is available. At time t,

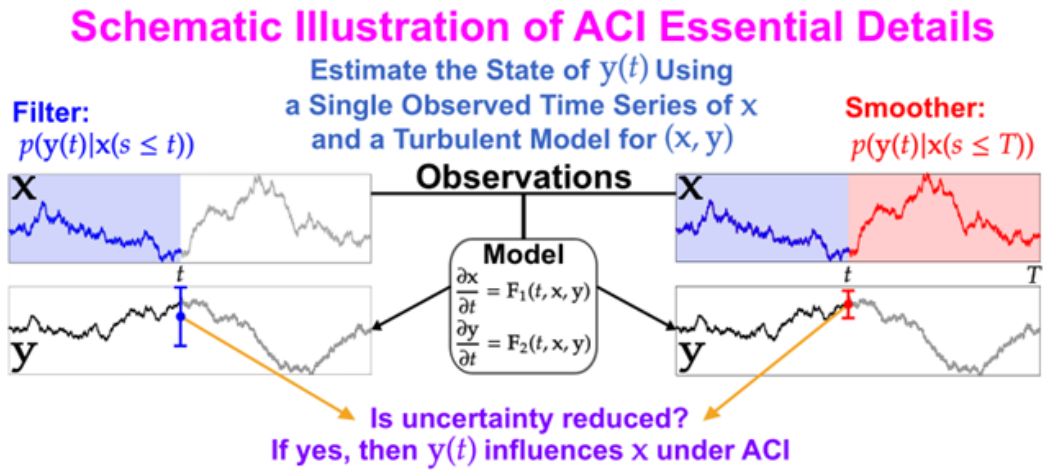
$$p_t^f(y \mid x) = p\big(y(t) \mid x(s \le t)\big), \qquad p_t^s(y \mid x) = p\big(y(t) \mid x(s \le T)\big), \quad T > t.$$

ACI causal criterion

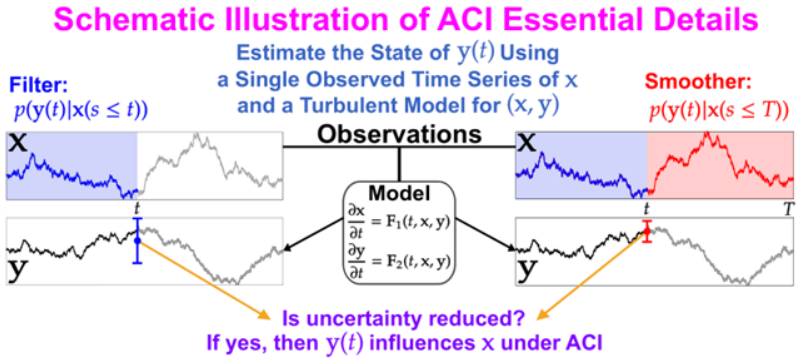
Consider the information distance between the two distributions. If

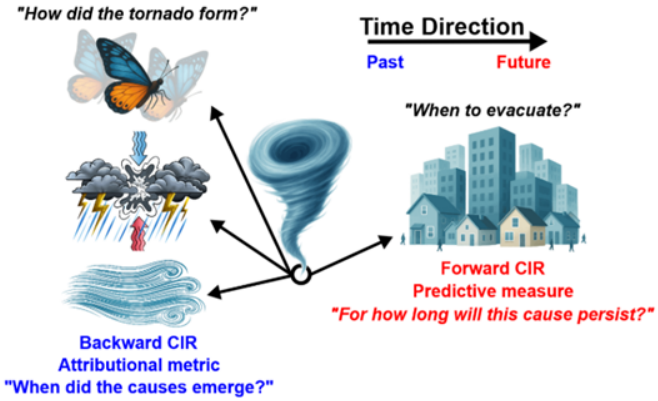
$$\mathcal{P}\left(p_t^s, p_t^f\right) = \int p_t^s \log \frac{p_t^s}{p_t^f} \, \mathrm{d}y > 0,$$

then future observations of x contain additional information about $y(t)$; ACI identifies $y(t) \to x$ at that time.



Causal Influence Range (CIR) in the ACI framework





CIR turns the ACI signal into a **time scale**, but there are two time directions.

Forward CIR (prediction).

- Starting from a causal influence active now: **how long will it remain relevant?**
- Quantifies the remaining influence range of an ongoing event and supports early warning.

Backward CIR (diagnosis).

- Looking back from an event: **how far into the past can a relevant precursor be traced?**
- Quantifies precursor range and helps identify event onset and driving mechanisms.

Subjective CIR. For a tolerance ϵ, let T'_ϵ be the shortest future window for which the truncated smoother is within ϵ of the full smoother:

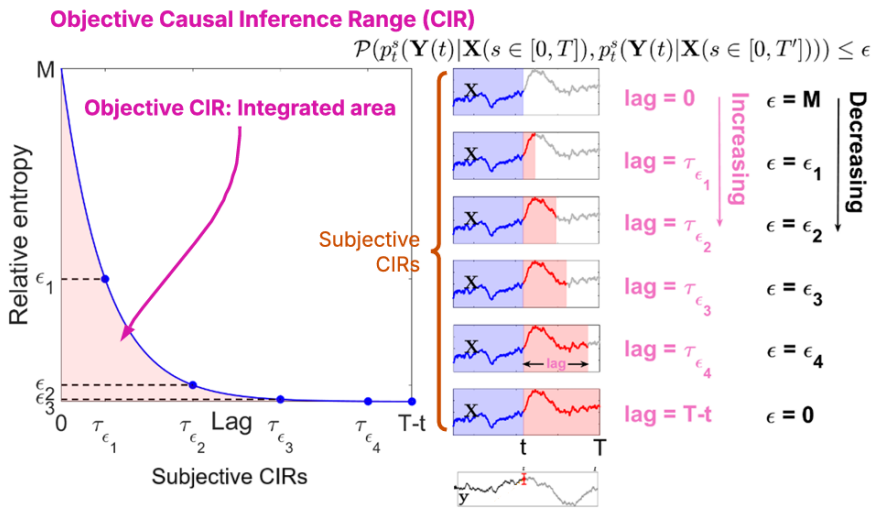
$$\mathcal{P}\Big(p(\mathbf{Y}(t) \mid \mathbf{X}(s \in [0, T])),\ p(\mathbf{Y}(t) \mid \mathbf{X}(s \in [0, T'_\epsilon]))\Big) \leq \epsilon, \qquad \tau_\epsilon = T'_\epsilon - t.$$

The resulting subjective CIR τ_ϵ depends on the chosen tolerance ϵ and may be non-unique.

Objective CIR. Let $M = \mathcal{P}(p_t^s, p_t^f)$ denote the total information gain from smoothing. The objective CIR is defined as

$$\text{CIR} = \int_0^M \tau_\epsilon \, d\epsilon,$$

which aggregates influence across all resolution levels and provides a robust, threshold-free measure.



The subjective and objective CIRs are analogous to the autocorrelation function and decorrelation time.

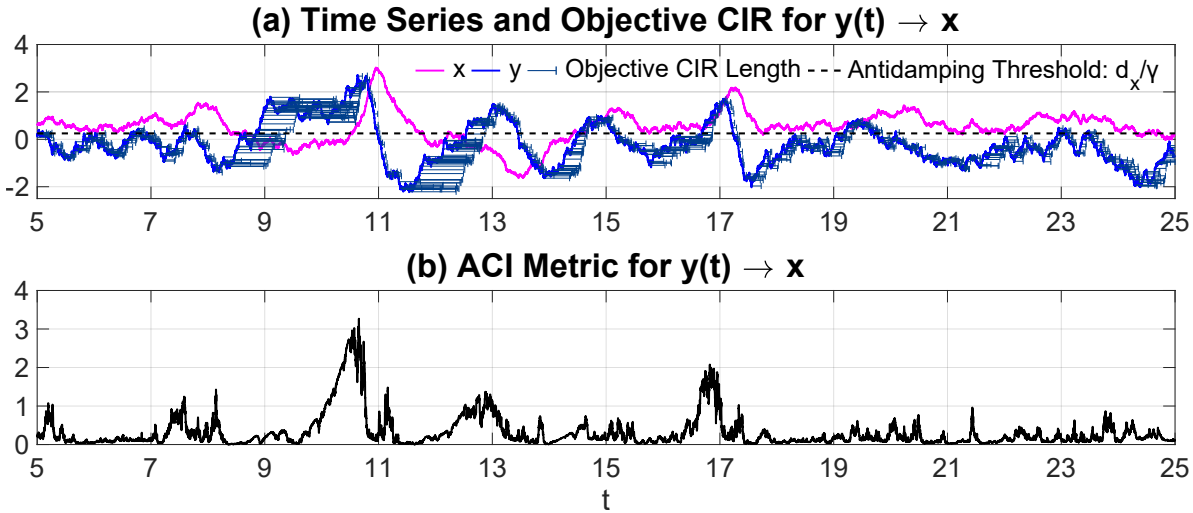
An efficient algorithm with rigorous error analysis has been developed to compute the CIR.

Example: Hidden Effect for Extreme Events and Influence Range

A prototype stochastic model:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -d_x x + \gamma x y + f_x + \sigma_x \dot{W}_x$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -d_y y - \gamma x^2 + f_y + \sigma_y \dot{W}_y$$



- ACI reveals that strong causal influence from $y \to x$ aligns with the onset and peak of extreme events in x.
- After the peak, the ACI value drops sharply, reflecting negative feedback that suppresses the event.
- The CIR shows sustained influence during the triggering phase and minimal influence during decay.
- CIR further indicates that extreme events develop gradually, with triggering conditions established well in advance.

Part II

Testing information pathways

Model interrogation through controlled reconstruction

From Understanding a Pathway to Testing a Model

Part I: What pathway matters?

ACI and CIR combine **model dynamics and a specific observed realization** to ask:

- Which variables carry information to the target?
- When does that influence become important?
- How long does it persist?

The result is a causal pathway **within the assumed dynamical model**, constrained by observations.

But can we trust the model pathway?

The interpretation depends on how faithfully the model represents the real dynamics.

- structural or parameterization errors can distort information pathways;
- compensating errors can still produce a plausible output;
- a pathway can therefore look convincing inside the model but be dynamically wrong.

Therefore, **before trusting a diagnosed mechanism, we need to test the dynamics that produced it.**

Part II: Test the model itself

Use observations as **controlled probes** to ask whether the model transmits information through the right pathways.

Why Direct Inspection is not Enough?

For a simple model, we may be able to inspect a few equations and use physical reasoning to identify a questionable mechanism.

However, for high-dimensional complex turbulent systems, especially a modern Earth-system model, this becomes much harder:

- many variables and processes interact across scales;
- parameterizations, numerics, and learned components act simultaneously;
- compensating errors can hide an incorrect pathway;
- the full information flow may not have a compact, analytically tractable representation.

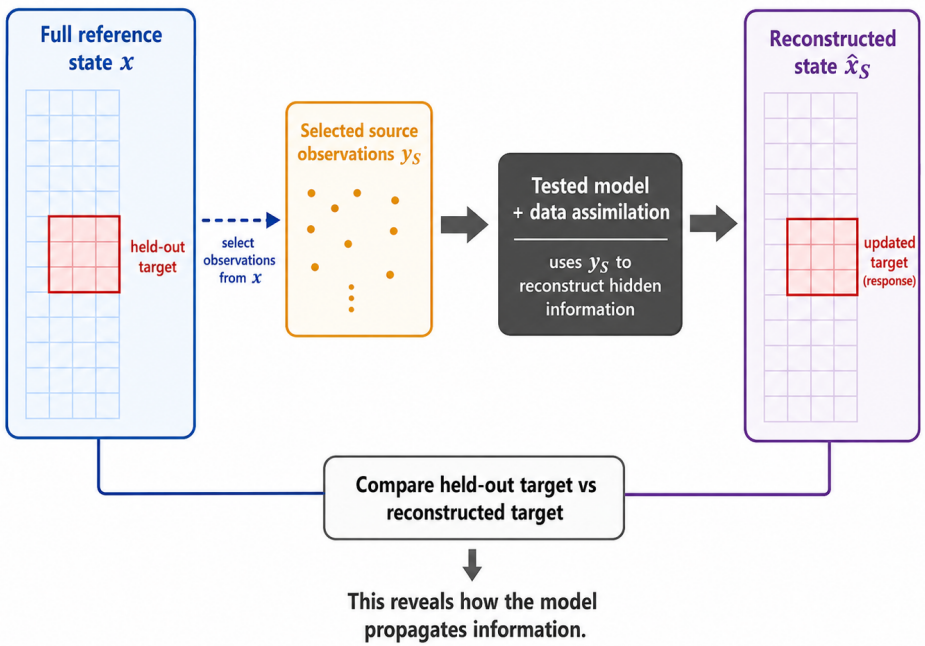
How We Probe a Model

1. **Define the target.** Choose a held-out state, event, parameterization scheme, or performance metric whose underlying dynamics we want to test.
2. **Add one information source.** Assimilate observations from a candidate variable, region, physical subsystem, or sensor class.
3. **Measure the model response.** Compare the reconstruction with the reference and repeat across candidate sources to map how information propagates through the model.

Core theoretical insight:

Under an ideal conditional reconstruction, adding correct information cannot increase the target MSE.

A harmful or misdirected response therefore exposes an inconsistent information pathway.



Bayesian data assimilation provides the theoretical foundation; scientific machine learning provides latent-space scalability.

Information-Gain Benchmark: What Should Added Info Do?

Hold out a target z, reconstruct it from information S, then add one new source j.

The controlled reconstruction experiment

Using the current information S,

$\widehat{\mathbf{z}}_S =$ reconstruction of the held-out target.

Measure its target error by

$$R(S) = \frac{1}{T}\int_0^T \left\|\mathbf{z}(t) - \widehat{\mathbf{z}}_S(t)\right\|^2 \, \mathrm{d}t.$$

Now add only source j:

$$S \quad \longrightarrow \quad S \cup \{j\}.$$

Define the resulting gain

$$\Delta_j(S) = R(S) - R(S \cup \{j\}).$$

$\Delta_j > 0$: target error decreases,
$\Delta_j = 0$: no net improvement,
$\Delta_j < 0$: target error increases.

The theoretical benchmark

Suppose the reconstruction is the **true conditional mean**:

$$\widehat{\mathbf{z}}_S^{\,\mathrm{opt}} = \mathbb{E}[\mathbf{z} \mid \mathcal{Y}_S],$$

where $\mathcal{Y}_S$ denotes all information supplied by S.
Adding source j gives a larger information set,

$$\mathcal{Y}_S \subseteq \mathcal{Y}_{S\cup\{j\}}.$$

Conditional projection then gives

$$\begin{aligned}\Delta_j^{\mathrm{opt}}(S) &= R_{\mathrm{opt}}(S) - R_{\mathrm{opt}}(S \cup \{j\}) \\ &= \mathbb{E}\left\|\widehat{\mathbf{z}}_{S\cup\{j\}}^{\,\mathrm{opt}} - \widehat{\mathbf{z}}_S^{\,\mathrm{opt}}\right\|^2 \geq 0.\end{aligned}$$

More correct information cannot worsen the optimal MSE.

No linearity or Gaussianity is required.

Update Calibration: How Do We Diagnose the Model Response?

Adding source j changes the reconstruction of the target z:

$$\widehat{\mathbf{z}}_S \longrightarrow \widehat{\mathbf{z}}_{S\cup\{j\}}, \qquad \underbrace{\mathbf{e}_S = \mathbf{z} - \widehat{\mathbf{z}}_S}_{\text{error before adding } j}, \qquad \underbrace{\mathbf{u}_j = \widehat{\mathbf{z}}_{S\cup\{j\}} - \widehat{\mathbf{z}}_S}_{\text{update produced by source } j}.$$

How well does the update correct the target error?

Adding source j changes the reconstruction by $\mathbf{u}_j$, so the target error changes from e_S to $\mathbf{e}_{S\cup\{j\}} = \mathbf{e}_S - \mathbf{u}_j$. Define

$$\boxed{A_j = \frac{\langle \mathbf{e}_S, \mathbf{u}_j \rangle_T}{\|\mathbf{u}_j\|_T^2}}$$

Then $A_j u_j$ is the component of the **existing target error along the update direction**:

$$\mathbf{e}_S = A_j \mathbf{u}_j + \mathbf{e}_{S,j}^{\perp}.$$

After applying the update,

$$\mathbf{e}_{S\cup\{j\}} = (A_j - 1)\mathbf{u}_j + \mathbf{e}_{S,j}^{\perp}.$$

Therefore, $A_j = 1$ means that the error along the model-produced update direction is **exactly removed**.

How A_j explains the realized gain

The gain can be written exactly as

$$\Delta_j = \|\mathbf{e}_S\|_T^2 - \|\mathbf{e}_S - \mathbf{u}_j\|_T^2 = \boxed{(2A_j - 1)\|\mathbf{u}_j\|_T^2}.$$

Therefore:

$A_j < 1/2$	$\Delta_j < 0$: harmful
$A_j = 1/2$	$\Delta_j = 0$
$1/2 < A_j < 1$	helpful, but too large
$A_j = 1$	calibrated
$A_j > 1$	helpful, but too small

Δ_j: Does source j help?
A_j: Does the update produced by source j actually correct the error that was present before the source was added?

Example: Identical Output Statistics, Different Dynamics

Reference

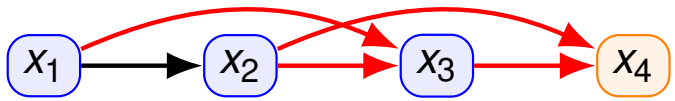
$\mathbf{x}_{n+1} = A\mathbf{x}_n + \xi_n, \qquad x_1 \to x_2 \to x_3 \to x_4.$

Hold out x_4 and progressively observe x_3, x_2, x_1.

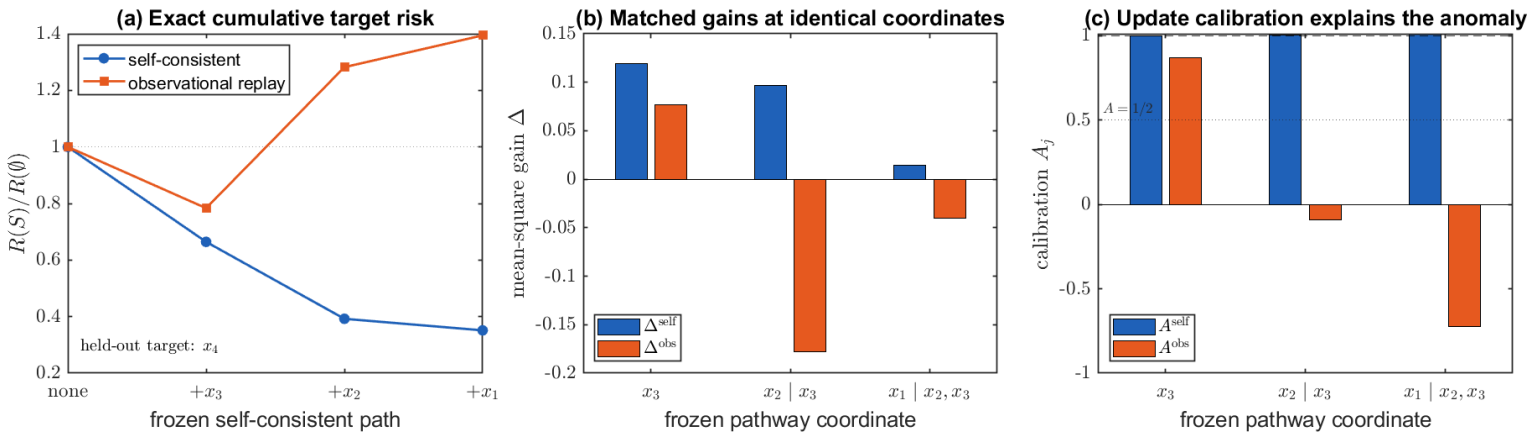
reference

Imperfect model

It has the same stationary Gaussian law as the reference but wrong temporal pathways (shortcut links and sign reversals).



imperfect model



- Model-generated data select the frozen route $x_3 \to x_2 \to x_1$; self-consistent risk decreases monotonically.
- Replaying the *same coordinates* with reference-system data makes the later steps harmful; $A_j < 1/2$ explains the negative gains.

ENSO Example: What Is Recharge-Discharge Timing?

ENSO is a coupled ocean-atmosphere phenomenon in the tropical Pacific. A useful way to describe its evolution is through the build-up and release of upper-ocean heat.

Recharge
upper-ocean heat content builds

→

El Niño
SST anomaly peaks

→

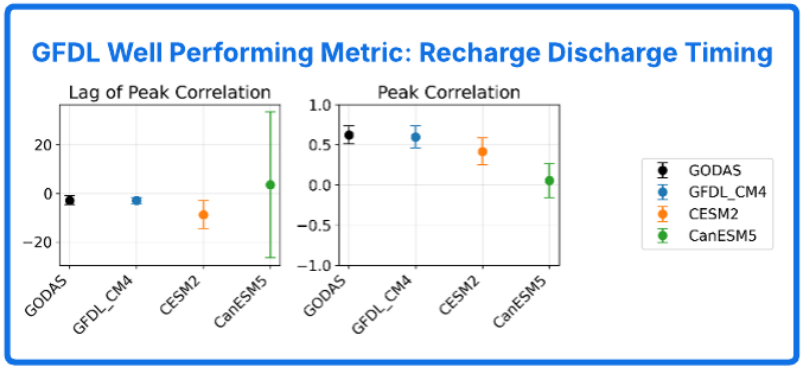
Discharge
upper-ocean heat content decreases

What do we measure?

The **recharge-discharge timing** measures the lag between the upper ocean heat content (or thermocline) signal and the ENSO sea surface temperature (SST) signal.

This timing is one signature of whether the model captures the recharge-discharge dynamics.

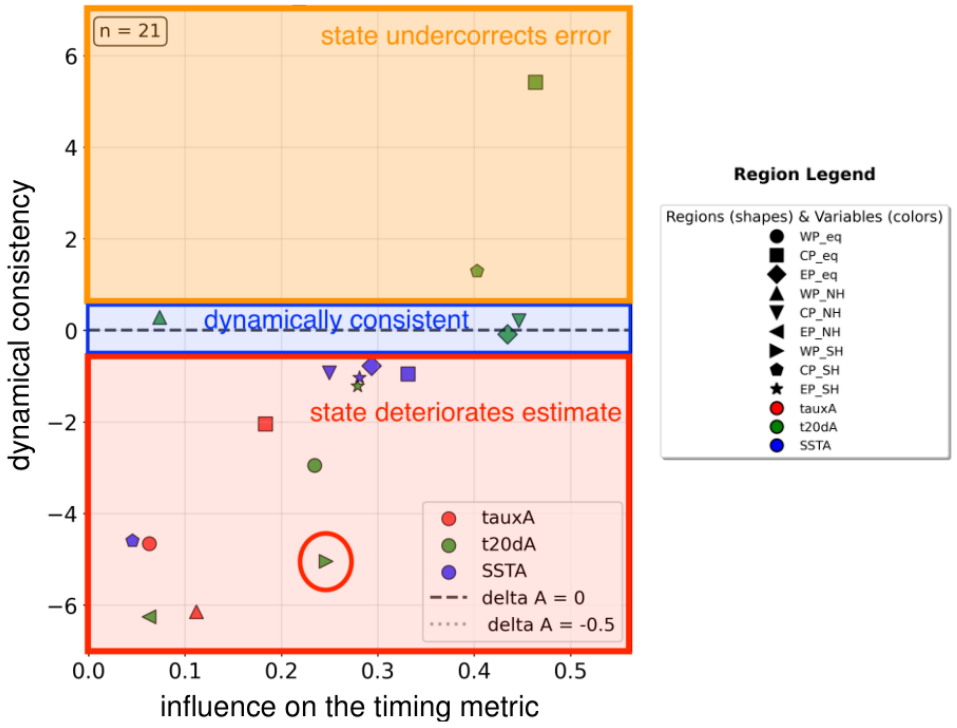
GFDL-CM4 reproduces this timing metric reasonably well.



ENSO Example: Does the Model Get the Pathway Right?

Controlled test

- **Target:**
  recharge-discharge timing.
- **Candidate sources:**
  21 region-variable combinations across the tropical Pacific.
- **Variables:**
  zonal wind stress, thermocline depth, and SST.



Each point asks:

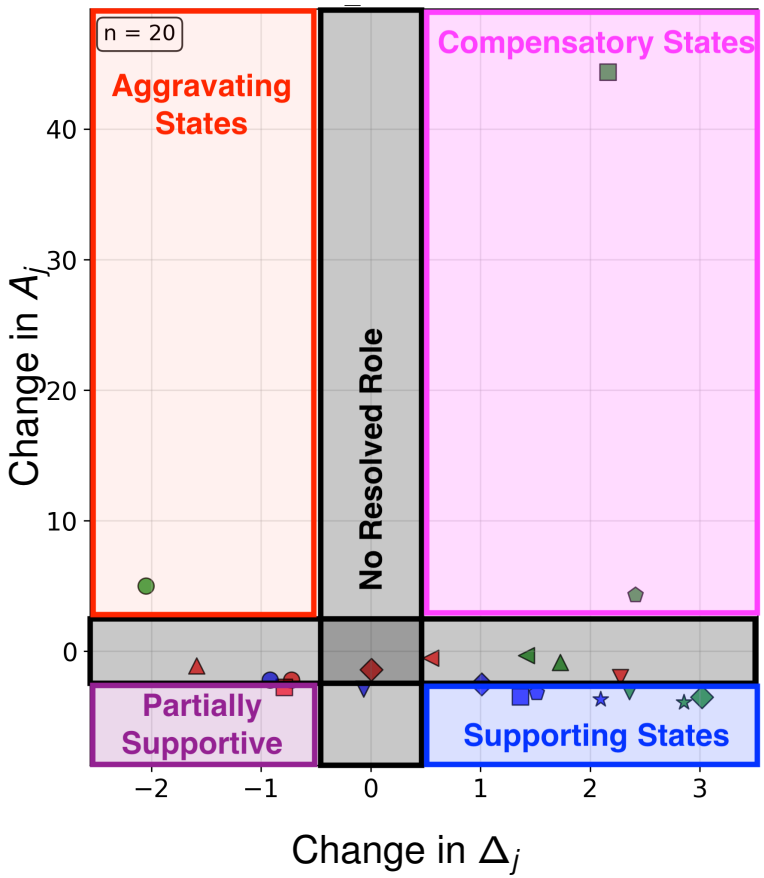
Each point is one controlled reconstruction: add one source and see how the target responds.

The circled thermocline-depth source is **influential**, but adding its information makes the reconstruction worse. We therefore use it as the seed for the next test.

A good output metric does not guarantee that the underlying pathway is right.

Tracing What Supports or Compensates for the Harmful Pathway

Seed pathway: `WP_SH t20dA`. Now add one more observed state and ask whether it repairs, aggravates, or bypasses the diagnosed response.



Interpretation

- **Support:** improves the target and the seed response.
- **Aggravation:** makes the seed response worse.
- **Compensation:** improves the output by another route.
- **No role:** no resolved effect in this context.

From Pathway Diagnosis to Better Extreme-Event Models

Extend the same framework to extreme events

The target and error metric can be defined around a specific event or event property, such as its **amplitude, location, duration, or timing**. The same pathway test can then reveal which model mechanisms succeed or fail during extremes.

From diagnosis to model improvement

Once a problematic pathway is localized, the diagnosis can guide **model recalibration, process correction, or residual-based closure**, with the goal of improving both ordinary behavior and extreme events.

Why this matters especially for extremes

A model may match ordinary statistics through compensating errors. An extreme event can move the system into a different part of state space, where those compensations need not persist. **Reliable extreme-event prediction therefore requires credible pathways, not only credible outputs.** This also provides a stronger basis for reduced-order models that retain the essential dynamics.

Summary

Assimilative causal inference: trace the system

ACI tells us **what is influencing an event at a given time** and **over what time range that influence matters**. The backward CIR points to when relevant precursors emerge, while the forward CIR tells us how long their influence is likely to persist.

Model interrogation: test the model

Getting the output right is not enough. We add observations one source at a time and see where the reconstruction gets worse. This can expose a bad pathway even when another part of the model is compensating for it.

The same ideas extend beyond extremes to regime transitions, model comparison, observation design, reduced models, scientific machine learning, and digital twins.


M. Andreou, N. Chen, and E. Bollt. “Assimilative Causal Inference.” *Nature Communications* 17, 1854 (2026).


M. Andreou and N. Chen. “Bridging Prediction and Attribution: Identifying Forward and Backward Causal Influence Ranges Using Assimilative Causal Inference.” *Physica D*, in press, 2026.


C. Moser, P. Behnoudfar, and N. Chen. “Beyond Model Outputs: Interrogating Dynamical Models through Information Propagation.” In preparation (2026).


C. Moser, N. Chen, and M. Andreou. (2026). “Mechanisms and Pathways of Extreme Events in Partially-Observed Stochastic Dynamical Systems.” arXiv preprint arXiv:2605.22692.


P. Behnoudfar, C. Moser, M. Bocquet, S. Cheng, and N. Chen. “Bridging Idealized and Operational Models: An Explainable AI Framework for Earth System Emulators.” *npj Climate and Atmospheric Science* 9:65 (2026).

Thank you!

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