Bridging Prediction and Attribution: Identifying Forward and Backward Causal Influence Ranges Using Assimilative Causal Inference*

Marios Andreou† and Nan Chen†

Abstract. Causal inference identifies cause-and-effect relationships between variables. While traditional approaches rely on data to reveal causal links, a recently developed method, assimilative causal inference (ACI), integrates observations with dynamical models. It utilizes Bayesian data assimilation to trace causes back from observed effects by quantifying the reduction in uncertainty. ACI advances the detection of instantaneous causal relationships and the intermittent reversal of causal roles over time. Beyond identifying causal connections, an equally important challenge is determining the associated causal influence range (CIR), indicating when causal influences emerged and for how long they persist. In this paper, ACI is employed to develop mathematically rigorous formulations of both forward and backward CIRs at each time. The forward CIR quantifies the temporal impact of a cause, while the backward CIR traces the onset of triggers for an observed effect, thus characterizing causal predictability and attribution of outcomes at each transient phase, respectively. Objective and robust metrics for both CIRs are introduced, eliminating the need for empirical thresholds. Computationally efficient approximation algorithms to compute CIRs are developed, which facilitate the use of closed-form expressions for a broad class of nonlinear dynamical systems. Numerical simulations demonstrate how this forward and backward CIR framework provides new possibilities for probing complex dynamical systems. It advances the study of bifurcation-driven and noise-induced tipping points in Earth systems, investigates the impact from resolving the interfering variables when determining the influence ranges, and elucidates atmospheric blocking mechanisms in the equatorial region. These results have direct implications for science, policy, and decision-making.

Key words. causal inference, causal influence range, data assimilation, Bayesian inverse problems, uncertainty quantification

MSC codes. 62F15, 62D20, 62M20, 93E11, 93E14, 60H10

Significance Statement. Assimilative causal inference (ACI) is a recent method that identifies instantaneous cause-and-effect relationships. By examining the reduction in uncertainty when incorporating the observed potential effects, it effectively captures intermittent reversals of causality over time. However, determining the causal influence range (CIR), which involves understanding when an influence begins and for how long it lasts, remains a critical challenge. This paper introduces mathematically rigorous formulations for both forward and backward CIRs. The forward CIR forecasts the future impact of a cause, while the backward CIR traces the origins of an observed effect in the past. They characterize predictability and attribution, respectively. Importantly, we develop threshold-free metrics and computationally efficient algorithms for both CIRs, allowing for tractable analysis and advancing the study of complex dynamical systems. This framework provides a new avenue for forecasting the emergence and persistence of extreme events, with broad implications to science and disaster preparedness.

---

*Submitted to SIAM/ASA Journal on Uncertainty Quantification on October 21, 2025.

Funding: This work was funded by the Office of Naval Research (N00014-24-1-2244) and the Army Research Office (W911NF-23-1-0118).

†Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, 53706, WI, USA (mandreou@math.wisc.edu, https://mariosandreou.short.gy/Homepage & chennan@math.wisc.edu, https://people.math.wisc.edu/~nchen29/).

1. Introduction. Causal inference identifies cause-and-effect relationships between state variables [64, 62, 78]. It is widely used across various scientific and societal areas. Beyond improving the understanding of the underlying dynamics, causal inference additionally plays a crucial role with assisting in the discovery of the governing equations, evaluating interventions, and informing decision-making [72, 16, 8, 82, 71, 69, 92, 80, 94].

Most of the traditional causal inference methods utilize multivariate time series data. Methods falling into this category include direct time series analysis (Granger causality [32], transfer entropy [79, 9], causation entropy [84, 85]), state space reconstruction methods (topological causality [34], convergent cross mapping [83, 61]) and statistical learning of structural causal models (interventional data methods [81, 12], mixture causal discovery [91]). In these approaches, the identified average-in-time causal relationship provides an overall indication of the causal link. However, many complex dynamical systems exhibit strong intermittency and nonlinearity, which means the cause-and-effect relationships between state variables may cease or shift repeatedly over time. Therefore, identifying the instantaneous causal link is crucial for understanding transient behaviors such as extreme events, bifurcations, and regime switches. There exist model-based strategies that track time-dependent relationships, such as those utilizing functional operators (information flow [44], Koopman causality [73]) and linear response theory (fluctuation-dissipation theorem [26], Green's function [50]). However, these methods are usually computationally intractable even when validating the hypothesis of a single possible causal scenario, while some of them are only able to provide temporal results for a very short time.

Assimilative causal inference (ACI) [6], an approach which aims to address the time evolution of causal relationships, has recently been developed. It integrates data and models for causal inference through Bayesian data assimilation, formulating the study of causality as an inverse problem of uncertainty quantification, where causes are traced backward from observed effects. Key advantages of ACI include the online tracking of cause-and-effect roles which can intermittently swap, the accommodation of short and incomplete datasets that contain observations only from potential effects, and computational efficiency in high-dimensional settings by leveraging established techniques from data assimilation.

Although the discovery of instantaneous causal links is a critical first step, a comprehensive causal analysis must also quantify the associated causal influence range (CIR) that defines the temporal domain of significant causal impact. This concept has two distinct interpretations: A forward (predictive) CIR, describing the future period over which a current cause will remain influential, and a backward (attributable) CIR, describing the past period where causal progenitors responsible for a currently observed effect emerged:

- **(A)** The forward CIR quantifiably answers the operational question: "For how long will this causal effect persist?". It assesses the remaining duration of an ongoing event or, following an early warning signal, infers the likelihood of an extreme event's onset. By providing a forecast of causal persistence, the forward CIR facilitates proactive preparedness for preventing catastrophic events, such as droughts, tsunamis, or cyclones.
- **(B)** The backward CIR addresses the question: "When did the causal precursors for this observed event emerge?". It pinpoints the onset of a damaging event and reveals its underlying driving mechanisms. By supporting root-cause analysis through causal attribution, the backward CIR guides targeted strategies for disaster preparedness

and risk reduction. For instance, it can identify which factors most strongly influence long-term changes in Earth system processes, thereby helping to prioritize the most effective adaptation measures.

This work builds upon the ACI framework to develop mathematically rigorous metrics for forward and backward CIRs. These formulations provide the first general framework for quantifying the temporal reach of a causal link in turbulent systems, both in the predictive (future) and attributable (past) directions. A key advantage of this method is that it utilizes Bayesian data assimilation in ACI to yield computable expressions for CIRs. These expressions are applicable to complex dynamical systems, including those with diverse autocorrelation structures and chaotic dynamics. Critically, these CIR measures are objectively defined without relying on empirical thresholds. Furthermore, they admit computationally efficient algorithms with closed-form approximations for a broad class of turbulent systems, making asymptotic analysis and practical estimation possible even in high-dimensional applications.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of ACI fundamentals and develops the theory for forward and backward CIRs, including derivations of computationally efficient approximations. We also introduce a broad class of nonlinear dynamical systems that enables the analytical evaluation of CIR metrics without requiring ensemble approximations. Section 3 applies the CIR framework to: (a) Distinguishing bifurcation-driven from noise-induced tipping points in Earth system models, (b) clarifying prediction and attribution in multiscale atmospheric models by resolving interfering dynamical components, and (c) identifying atmospheric blocking mechanisms in an equatorial circulation model with flow-wave interactions and seasonal feedbacks. Concluding remarks and discussion are presented in Section 4, with additional mathematical derivations in the Appendix.

2. Methodology.

2.1. Notations. Throughout this work, boldface variables are used for multidimensional quantities. Lowercase variables denote column vectors, while uppercase ones denote matrices. The only exception is $\mathbf{W}$ (with some subscript), which represents a vector-valued Wiener process due to literary tradition. Furthermore, for simplicity, we follow the notational convention from physics and do not distinguish between random variables and their realizations.

Let $\mathscr{B} = (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be an augmented probability space for filtering on the time interval $[0, T]$, $T \in (0, +\infty)$. Denote by $\big(\mathbf{x}(t,\omega), \mathbf{y}(t,\omega)\big) \in \mathbb{R}^{k+l}$, for $t \in [0,T]$ and $\omega \in \Omega$, a $(k+l)$-dimensional partially-observable stochastic process on $\mathscr{B}$, where $\mathbf{x}$ is the k-dimensional observable component while $\mathbf{y}$ is the l-dimensional unobservable part. Without loss of generality, we consider real-valued processes, with $\mathbf{x}$ being continuously observed (an analogous framework can be developed for discrete-time observations). For notational simplicity, we henceforth drop the sample space (Ω) dependence, but it is always implied. We assume $(\mathbf{x}, \mathbf{y})$ is adapted to the filtration $\mathbb{F}$ and that $\mathbf{x}(s \le t)$ represents a time series of $\mathbf{x}$ over $[0, t]$, i.e., a realization for a fixed $\omega \in \Omega$. Finally, for explicitness, we hereafter write $\big(\cdot\,\big|\mathbf{x}(s \le t)\big)$ to indicate the fact that we are conditioning on the σ-algebra generated by $\{\mathbf{x}(s)\}_{s\le t}$; nevertheless, throughout this work we assume that the observed time series of $\mathbf{x}$ is free of observational noise, with noise-contaminated realizations (i.e., the effect of observational randomness on the conditional distributions) being left for consideration in future upcoming work.

2.2. Fundamentals of Assimilative Causal Inference (ACI). We provide a brief review of the recently developed ACI framework [6]. We assume that the evolution of the process $(\mathbf{x}, \mathbf{y})$ is governed by the following stochastic system of coupled Itô diffusions [70]:

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}^{\mathbf{x}}(t, \mathbf{x}, \mathbf{y}) + \mathbf{\Sigma}_1^{\mathbf{x}}(t, \mathbf{x})\dot{\mathbf{W}}_1(t) + \mathbf{\Sigma}_2^{\mathbf{x}}(t, \mathbf{x})\dot{\mathbf{W}}_2(t), \tag{2.1a}$$

$$\frac{\mathrm{d}\mathbf{y}(t)}{\mathrm{d}t} = \mathbf{f}^{\mathbf{y}}(t, \mathbf{x}, \mathbf{y}) + \mathbf{\Sigma}_1^{\mathbf{y}}(t, \mathbf{x}, \mathbf{y})\dot{\mathbf{W}}_1(t) + \mathbf{\Sigma}_2^{\mathbf{y}}(t, \mathbf{x}, \mathbf{y})\dot{\mathbf{W}}_2(t), \quad t \in [0, T], \tag{2.1b}$$

where $\dot{\mathbf{W}}_1$ and $\dot{\mathbf{W}}_2$ are independent real-valued Gaussian white noises, also independent of the initial distribution of $(\mathbf{x}(0), \mathbf{y}(0))$. Certain regularity and identifiability conditions are enforced such that the conditional distribution of $\mathbf{y}$ given the data $\mathbf{x}$ contains all available information about $\mathbf{y}$ over the considered observational time frame [70] (see [6] for full details).

The ACI framework differs from the classical notion of predictive causality as outlined by Granger [32, 33]. ACI solves an inverse problem of uncertainty quantification in statistical inference. Under the Bayesian data assimilation pipeline, candidate causal variables are traced backward from data of the possible subsequent observed effects. As such, identifying the cause-and-effect relationship is not based on running forward and extrapolating the system in (2.1) to obtain the informational response on the target variables' entropy, as in the usual setting [32, 93]. This bypasses the need to impose restrictive model structures, require the observation of all state variables, or demand sufficiently long time series for the temporal average.

The ACI framework applies Bayes' theorem to combine a prior distribution for the state $\mathbf{y}(t)$, obtained by forward integration of the system dynamics in (2.1) (the forecast step), with the likelihood from a single realization of $\mathbf{x}$. This yields a posterior distribution for $\mathbf{y}(t)$ (the analysis step). Building on this, we identify $\mathbf{y}(t)$ as an instantaneous cause of $\mathbf{x}$ at time $t \in [0, T]$ if incorporating future observations of $\mathbf{x}$ from $(t, T]$ reduces the uncertainty in estimating $\mathbf{y}(t)$ beyond the reduction achieved by using only its history, $\mathbf{x}(s \leq t)$. This excess uncertainty reduction is quantified using the relative entropy [40, 52]:

$$\mathcal{P}\big(p_t^{\mathrm{s}}(\mathbf{y}|\mathbf{x}), p_t^{\mathrm{f}}(\mathbf{y}|\mathbf{x})\big) := \int p_t^{\mathrm{s}}(\mathbf{y}|\mathbf{x}) \log\big(p_t^{\mathrm{s}}(\mathbf{y}|\mathbf{x})/p_t^{\mathrm{f}}(\mathbf{y}|\mathbf{x})\big)\mathrm{d}\mathbf{y}(t), \tag{2.2}$$

where $p_t^{\mathrm{s}}(\mathbf{y}|\mathbf{x}) := p\big(\mathbf{y}(t)|\mathbf{x}(s \leq T)\big)$ is the smoother and $p_t^{\mathrm{f}}(\mathbf{y}|\mathbf{x}) := p\big(\mathbf{y}(t)|\mathbf{x}(s \leq t)\big)$ is the filter posterior probability density function (PDF) of $\mathbf{y}(t)$. Note that this choice of statistical divergence is without loss of generality and chosen so purely on the basis of mathematical convenience [6]. Therefore, whenever

$$\mathcal{P}\big(p_t^{\mathrm{s}}(\mathbf{y}|\mathbf{x}), p_t^{\mathrm{f}}(\mathbf{y}|\mathbf{x})\big) > 0, \tag{2.3}$$

meaning there is information to be gained about $\mathbf{y}(t)$'s state from incorporating the future data of $\mathbf{x}$, we identify an assimilative causal relationship from $\mathbf{y}(t)$ to $\mathbf{x}$ [6], denoted as

$$\mathbf{y}(t) \to \mathbf{x}. \tag{2.4}$$

Essentially, ACI performs natural causal inference by assuming that only potential subsequent effects need to be observed and attributed to candidate past causes, thus removing the restriction of observing the whole state space. This is practically meaningful as the data of the causes to be identified may not always be available.

2.2.1. Conditional ACI: Causality in the Presence of Non-Target Variables. A more general problem is to determine whether $\mathbf{y}(t)$ causes a subset of the observed variables, $\mathbf{x}_\mathrm{A}$ where $\mathbf{x} = (\mathbf{x}_\mathrm{A}, \mathbf{x}_\mathrm{B})$, while in the presence of the remaining non-target variables, $\mathbf{x}_\mathrm{B}$, where $\mathbf{x}_\mathrm{B}$ can play the role of confounders, mediators, moderators, colliders, or instrumental variables in the underlying causal link (if it exists). This is sometimes referred to as a conditional causal relationship, where $\mathbf{y}(t)$ instantaneously causes $\mathbf{x}_\mathrm{A}$ beyond the contributions of $\mathbf{x}_\mathrm{B}$.

In [6], a conditional variation of ACI is developed. It assesses whether the conditional assimilative causal relationship

$$\mathbf{y}(t) \to \mathbf{x}_\mathrm{A} \big| \mathbf{x}_\mathrm{B}, \tag{2.5}$$

holds at $t \in [0, T]$ in an analogous manner to (2.3), where (2.5) is established if [6]:

$$\mathcal{P}\big(p_t^{\mathrm{s}|\mathbf{x}_\mathrm{B}}(\mathbf{y}|\mathbf{x}_\mathrm{A}), p_t^{\mathrm{f}|\mathbf{x}_\mathrm{B}}(\mathbf{y}|\mathbf{x}_\mathrm{A})\big) > 0, \tag{2.6}$$

where

$$p_t^{\mathrm{s}|\mathbf{x}_\mathrm{B}}(\mathbf{y}|\mathbf{x}_\mathrm{A}) := \lim_{\mathrm{Var}(\mathbf{x}_\mathrm{B})\to+\infty} p\big(\mathbf{y}(t)\big|\mathbf{x}_\mathrm{A}(s \le T), \mathbf{x}_\mathrm{B}(s \le T)\big) = \lim_{\mathrm{Var}(\mathbf{x}_\mathrm{B})\to+\infty} p_t^{\mathrm{s}}(\mathbf{y}|\mathbf{x}), \tag{2.7a}$$

$$p_t^{\mathrm{f}|\mathbf{x}_\mathrm{B}}(\mathbf{y}|\mathbf{x}_\mathrm{A}) := \lim_{\mathrm{Var}(\mathbf{x}_\mathrm{B})\to+\infty} p\big(\mathbf{y}(t)\big|\mathbf{x}_\mathrm{A}(s \le t), \mathbf{x}_\mathrm{B}(s \le t)\big) = \lim_{\mathrm{Var}(\mathbf{x}_\mathrm{B})\to+\infty} p_t^{\mathrm{f}}(\mathbf{y}|\mathbf{x}), \tag{2.7b}$$

are the smoother- and filter-based posterior-like PDFs of $\mathbf{y}(t)$, obtained by assigning infinite uncertainty to $\mathbf{x}_\mathrm{B}$'s marginal likelihood during the analysis stage of Bayesian data assimilation. This formal-limit modification of the posterior PDFs prohibits $\mathbf{x}_\mathrm{B}$ from affecting $\mathbf{y}(t)$'s state estimation during the Bayesian update. Unlike marginalizing or conditioning on $\mathbf{x}_\mathrm{B}$, this method ensures that uncertainty reduction stemming from the future data of $\mathbf{x}$ arises solely from the observational contributions of $\mathbf{x}_\mathrm{A}$ [6], thus allowing us to determine whether $\mathbf{y}(t)$ affects $\mathbf{x}_\mathrm{A}$ beyond the influence of $\mathbf{x}_\mathrm{B}$.

2.3. Causal Influence Range (CIR) Estimation in ACI. In what follows, for the sake of simplicity, we discuss the theory of unconditional CIR through the lens of ACI, where a similar setup can be built for the conditional CIR by substituting the appropriate equations ((2.3)–(2.4) in Section 2.2) with their conditional counterparts ((2.5)–(2.7) in Section 2.2.1).

While the uncertainty reduction in the smoother solution relative to the filter in (2.3) identifies the causal relationship (2.4), it does not reveal the associated temporal extent of $\mathbf{y}$'s causal impact on $\mathbf{x}$. This time span is exactly the CIR of a causal relationship, as discussed in Section 1. Depending on the adopted time directionality during the assessment of this temporal extent, there are two distinct formulations of CIR:

- **(A)** Forward CIR (Section 2.3.2): The time window after t that describes the future temporal extent of the causal state's, $\mathbf{y}(t)$, influence on $\mathbf{x}$. It defines the period for which $\mathbf{y}(t)$ remains a significant predictor for the effect variable $\mathbf{x}$. It characterizes the predictability and persistence of the causal variable.
- **(B)** Backward CIR (Section 2.3.4): The interval before T that describes the past temporal extent of the currently observed effect's, $\mathbf{x}(T)$, attribution to $\mathbf{y}$'s impact. It identifies the period during which the causal variable's, $\mathbf{y}$, past states acted as $\mathbf{x}(T)$'s precursors. It enables the attribution of an outcome to its historical causes, pinpointing its origin.

Note that, in the context of the backward CIR, causal attribution proceeds backward from a fixed observation time, which is nevertheless arbitrary. Therefore, we adopt the convention of denoting this reference time as $T>0$. Although the notation $\mathbf{x}(t)$ offers a more direct analogy to the forward CIR, the use of $\mathbf{x}(T)$ facilitates the discussion of the theory in the subsequent sections.

2.3.1. Framework for Assessing Forward and Backward CIRs. In complex turbulent dynamical systems, memory effects (e.g., autocorrelation) necessarily decay over time. Consequently, in Bayesian data assimilation, new observations of $\mathbf{x}$ typically influence the state estimation of $\mathbf{y}$ only over a finite time window [7]. Within the ACI framework, this fact implies that the causal variable $\mathbf{y}$ effectively and measurably influences the states of the effect variable $\mathbf{x}$ only for a limited duration. As a result, both the forward and backward CIRs are inherently finite in chaotic settings, enabling their rigorous mathematical treatment under a suitable formulation.

Let $T>0$ be the length of the total period of available observations, and let $t\in[0,T]$ be the current time. For every future time $T'\in[t,T]$, which we refer to as a lagged observational time (as it lags behind T), we define the following state estimates:

- Complete smoother PDF, $p\big(\mathbf{y}(t)\big|\mathbf{x}(s\leq T)\big)$: The optimal state estimate of $\mathbf{y}(t)$, conditional on all available data up to time T.
- Lagged smoother PDFs, $p\big(\mathbf{y}(t)\big|\mathbf{x}(s\leq T')\big)$: A suboptimal state estimate, conditional only on the data up to the lagged observational time T'.

The statistical discrepancy between these PDFs can be intrinsically quantified through the relative entropy, akin to (2.3):

$$0\leq \delta(t,T'):=\mathcal{P}\big(p(\mathbf{y}(t)|\mathbf{x}(s\leq T)),p(\mathbf{y}(t)|\mathbf{x}(s\leq T'))\big),\quad 0\leq t\leq T'\leq T. \tag{2.8}$$

We subsequently refer to $\delta(t,T')$ as the CIR metric, a function of both natural time t and lagged observational time T' (and of T, implicitly, but which we omit for notational simplicity). This quantity measures the lack of information in the state estimation of $\mathbf{y}(t)$ when omitting the observational data of $\mathbf{x}$ in $(T',T]$. This CIR metric is used to measure the actual CIR. Under ACI's inverse-problem-based formulation of causality, the finite-memory of the dynamics manifests itself in the temporal decay of:

(A) $\mathbf{y}(t)$'s influence on the future states of $\mathbf{x}(T')$ as T' increases from t towards T. This is because increasing data incorporation leads to the CIR metric in (2.8) to decrease to zero pointwise due to properties of the relative entropy [14]. In this setting, the CIR metric in (2.8) is treated as a function of T' with t being fixed, and the region for which it is significant in T' defines the forward CIR over $[t,T]$.

(B) The attribution of $\mathbf{x}(T)$ to the past states $\mathbf{y}(t)$ as t decreases from T when T' approaches T from below. In this backward-looking context, the CIR metric is treated as a function of t, with T fixed and T' sufficiently close to it such that the lagged smoother PDFs approximate the optimal complete one. This is intuitively understood in a temporal discretization for an observational rate of $\Delta t\ll 1$, where $T'=T-\Delta t$ (see Section D.2 in the Appendix). Here, the finite memory effect similarly causes the CIR metric in (2.8) to diminish as t decreases from T in this limiting regime: A causal relationship from $\mathbf{y}(t)$ to $\mathbf{x}(T)$ is discovered if the additional observation of $\mathbf{x}(T)$ provides significant uncertainty reduction in estimating $\mathbf{y}(t)$ beyond that stemming from the data $\mathbf{x}(s \leq T - \Delta t)$ as $\Delta t \to 0^+$ for t close to T. The interval in natural time t over which the CIR metric in (2.8) is significant in this limiting regime defines the backward CIR over $[0, T]$.

Therefore, by appropriately treating the CIR metric in (2.8) as either a function of T' or t, we can rigorously define ways to quantify the forward and backward CIR, respectively.

Panel (I) of Figure 1 provides a high-level overview of the forward and backward CIRs, in the context of a real-world scenario. Panel (II) provides further details through a schematic diagram that outlines the essential mathematical aspects of the ACI-based CIR framework developed in this work. Sections 2.3.2 and 2.3.4, and Figure 2 contain more technical details.

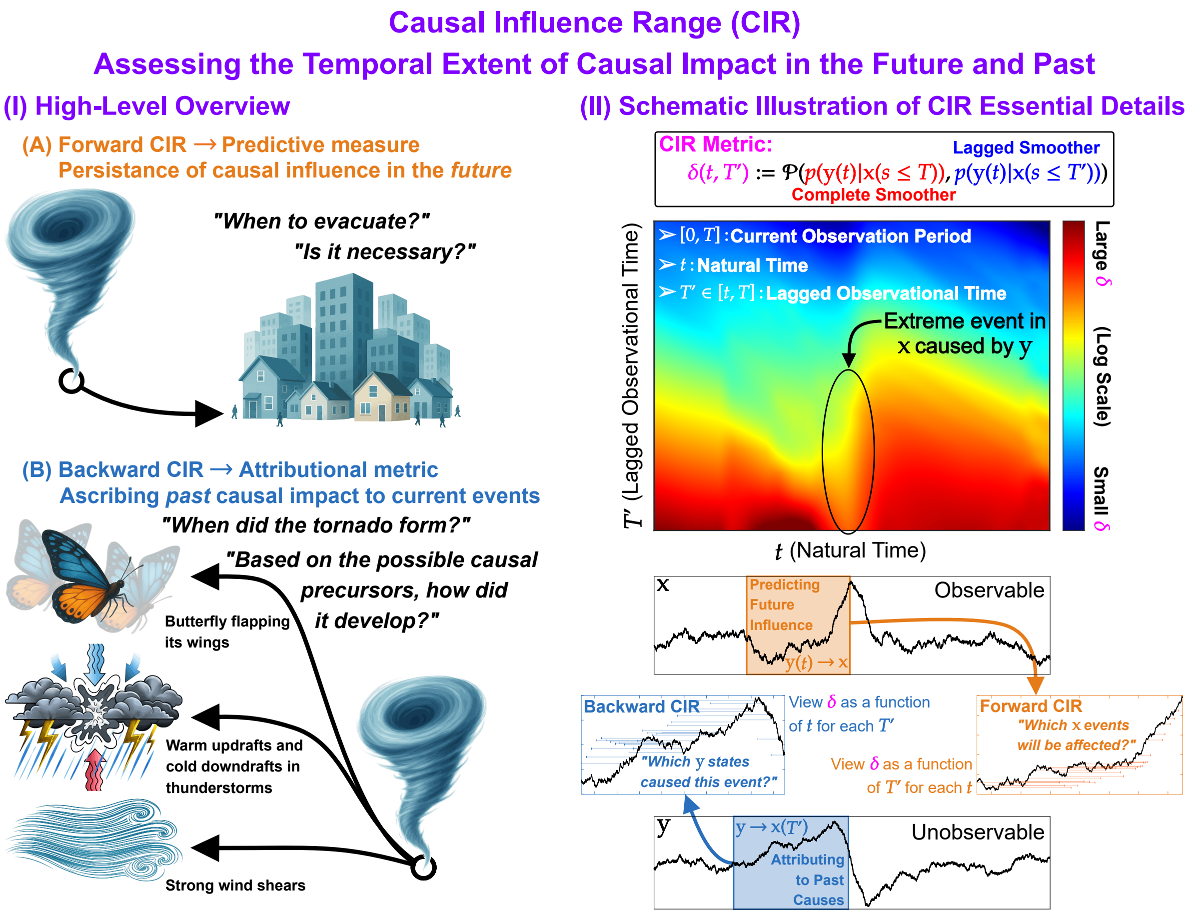


Figure 1. *The ACI-based formulations of the forward and backward CIR measures. Panel (I): A high-level overview of the developed methodology via a real-world scenario (tornado forecast and attribution). Panel (II): Schematic illustration of the framework from a more technical viewpoint.*

2.3.2. Inferring Causal Predictability: Forward CIR. The forward CIR quantifies how far into the future a causal state $\mathbf{y}(t)$ influences $\mathbf{x}$. To formalize this, we treat the CIR metric

δ from (2.8) as a function of the lagged observational time $T' \in [t, T]$ for each $t \in [0, T]$. For analysis, we normalize its observational time domain to $\mathrm{I} := [0, 1]$ via the transformation $T' = h(\tau) = t + \tau(T - t)$, defining the forward CIR metric:

$$\delta^{\mathrm{f}}(\tau; t) := \delta(t, t + \tau(T - t)), \quad \tau \in \mathrm{I}. \tag{2.9}$$

The maximum information deficit at time t is then:

$$M^{\mathrm{f}}(t) := \sup_{\tau \in \mathrm{I}} \big\{\delta^{\mathrm{f}}(\tau; t)\big\} > 0, \quad t \in [0, T], \tag{2.10}$$

which we assume exists and is nonzero, a condition satisfied in most practical settings.

Remark 2.1. Key properties of the forward CIR metric include:

- At $\tau = 0$, $\delta^{\mathrm{f}}(0; t) = \delta(t, t) = \mathcal{P}\big(p_t^{\mathrm{s}}(\mathbf{y}|\mathbf{x}), p_t^{\mathrm{f}}(\mathbf{y}|\mathbf{x})\big)$ equals the ACI metric assessing $\mathbf{y}(t) \to \mathbf{x}$ (see (2.3)–(2.4)). In chaotic systems, $M^{\mathrm{f}}(t)$ typically occurs at $\tau = 0$ or shortly after, reflecting the fact that causal influence is strongest immediately following the cause.
- For $\tau \in (0, 1)$, $\delta^{\mathrm{f}}(\tau; t)$ measures the information deficit from incorporating only observations up to $T' = h(\tau) \le T$ rather than T. Therefore, the subinterval in I over which $\delta^{\mathrm{f}}(\tau; t)$ is large indicates significant future causal influence, as substantial information gain comes from including observations in $(h(\tau), T]$.
- At $\tau = 1$, $\delta^{\mathrm{f}}(1; t) = 0$ by the positive-definiteness of relative entropy [14].

Since future observations of $\mathbf{x}$ affect $\mathbf{y}(t)$'s state estimate only over a finite interval, $\delta^{\mathrm{f}}(\tau; t)$ generally decreases from its maximum near $\tau = 0$ to zero at $\tau = 1$.

The subjective forward CIR length of $\mathbf{y}(t) \to \mathbf{x}$, dependent on a threshold ε, is:

$$\widetilde{\boldsymbol{\tau}}^{\mathrm{f}}(t, \varepsilon) := (T - t) \sup \big\{\tau \in \mathrm{I} : \delta^{\mathrm{f}}(\tau; t) > \varepsilon\big\}, \quad t \in [0, T], \tag{2.11}$$

with $\widetilde{\boldsymbol{\tau}}^{\mathrm{f}}(t, \varepsilon) = 0$ if $\varepsilon \ge M^{\mathrm{f}}(t)$. The corresponding subjective forward CIR interval is:

$$\widetilde{\mathrm{CIR}}^{\mathrm{f}}(t, \varepsilon) := \big[t, t + \widetilde{\boldsymbol{\tau}}^{\mathrm{f}}(t, \varepsilon)\big] \subseteq [t, T], \quad t \in [0, T]. \tag{2.12}$$

Remark 2.2. The subjective forward CIR length bounteously estimates the persistence of $\mathbf{y}(t)$'s causal influence on $\mathbf{x}$'s future states, since it satisfies

$$\widetilde{\boldsymbol{\tau}}^{\mathrm{f}}(t, \varepsilon) \ge (T - t)\lambda_{\mathrm{I}}\big(\big\{\tau \in \mathrm{I} : \delta^{\mathrm{f}}(\tau; t) > \varepsilon\big\}\big),$$

where λ_{I} is the Lebesgue measure on I.

To remove the ε-threshold dependence from (2.11)–(2.12), we define the objective forward CIR length by averaging over all ε:

$$\boldsymbol{\tau}^{\mathrm{f}}(t) := \frac{1}{M^{\mathrm{f}}(t)} \int_0^{M^{\mathrm{f}}(t)} \widetilde{\boldsymbol{\tau}}^{\mathrm{f}}(t, \varepsilon) \mathrm{d}\varepsilon, \quad t \in [0, T], \tag{2.13}$$

with the corresponding objective forward CIR interval given by:

$$\mathrm{CIR}^{\mathrm{f}}(t) := \big[t, t + \boldsymbol{\tau}^{\mathrm{f}}(t)\big] \subseteq [t, T], \quad t \in [0, T]. \tag{2.14}$$

This objective metric always lies within the maximal subjective interval, while always containing the minimal one. The subjective and objective CIRs can be regarded as analogs of the autocorrelation function and the decorrelation time, respectively, used to measure the memory of a stochastic system [6].

Remark 2.3. When $M^{\mathrm{f}}(t) \approx 0$ (weak evidence for $\mathbf{y}(t) \to \mathbf{x}$), (2.13) may yield an inflated CIR length if $\delta^{\mathrm{f}}(\tau; t)$ slowly increases toward its near-zero maximum $M^{\mathrm{f}}(t)$ as τ decreases. Such values typically coincide with small local maxima of the ACI metric in (2.3) and should be interpreted jointly with it as to avoid misrepresentation of temporal causal impact.

2.3.3. Efficient Computation of the Objective Forward CIR. Computationally, we can approximate the objective forward CIR length in (2.13) without quadrature over ε [6]:

$$\boldsymbol{\tau}^{\mathrm{f}}_{\mathrm{approx}}(t) := \frac{T-t}{M^{\mathrm{f}}(t)} \int_0^1 \delta^{\mathrm{f}}(\tau; t)\mathrm{d}\tau = \frac{1}{M^{\mathrm{f}}(t)} \int_t^T \delta(t, T')\mathrm{d}T' \leq \boldsymbol{\tau}^{\mathrm{f}}(t), \tag{2.15}$$

with equality if and only if $\delta^{\mathrm{f}}(\cdot; t)$ is nonincreasing on I. In this case, $M^{\mathrm{f}}(t) = \delta^{\mathrm{f}}(0; t)$ equals the ACI metric assessing $\mathbf{y}(t) \to \mathbf{x}$.

Remark 2.4. Since (2.15) is an underestimate, it yields shorter objective forward CIR intervals compared to the exact definition.

2.3.4. Assigning Causal Attribution: Backward CIR. This subsection parallels the development of the forward CIR in Section 2.3.2.

Whereas forward CIR quantifies future influence, backward CIR attributes observed effects at time T to past causes. We treat the CIR metric δ from (2.8) as a function of $t \in [0, T']$ for each $T' \in [0, T]$. As noted at the beginning of Section 2.3, T is fixed but arbitrary.

For analysis, we normalize its natural time domain to $\mathrm{I} = [0, 1]$ via $t = g(\tau) = \tau T'$ and center the result relative to $t = 0$, defining the backward CIR metric:

$$\delta^{\mathrm{b}}(\tau; T') := |\delta(\tau T', T') - \delta(0, T')|, \quad \tau \in \mathrm{I}. \tag{2.16}$$

The absolute value ensures well-posedness, though $\delta(t, T') - \delta(0, T')$ is typically positive for t sufficiently far from zero. As described in Section 2.3.1, we focus on the limit $T' \to T^-$ of complete information incorporation, assuming the complete backward CIR metric:

$$\lim_{T' \to T^-} \delta^{\mathrm{b}}(\tau; T'), \tag{2.17}$$

exists and is well-defined for each $\tau \in \mathrm{I}$.

Remark 2.5. Although $\delta^{\mathrm{b}}(\tau; T) = 0$ by relative entropy properties (see (2.8) and (2.16)) [14], $\lim_{T' \to T^-} \delta^{\mathrm{b}}(\tau; T')$ generally differs from zero. That is, the backward CIR metric is not necessarily left-continuous at T as a function of T'. This discontinuity becomes pronounced when approximating the limit with discrete-time observations at small observation rates (see Section D.2 in the Appendix). Notably, in this context, the uncertainty reduction in $\mathbf{y}(t)$ from $\mathbf{x}(s \leq T)$ beyond $\mathbf{x}(s \leq T - \Delta t)$, with $\Delta t \ll 1$ a discrete time step, indicates precisely the role of the effect at T when tracing it back to causes earlier in time.

The maximum information deficit under the complete information limit is then:

$$M^{\mathrm{b}}(T) := \sup_{\tau\in \mathrm{I}} \left\{ \lim_{T'\to T^-} \delta^{\mathrm{b}}(\tau; T') \right\} > 0, \quad T > 0, \tag{2.18}$$

which we assume exists and is nonzero, both valid assumptions in most applications.

Remark 2.6. Key properties of the (complete) backward CIR metric include:

- At $\tau = 1$, $\delta(T', T')$ equals the ACI metric assessing $\mathbf{y}(T') \to \mathbf{x}$ (see (2.3)–(2.4)), which is offset by the bias term $\delta(0, T')$ in (2.16) that represents the information deficit from excluding the data in $(T', T]$ when estimating $\mathbf{y}(0)$. Typically, $M^{\mathrm{b}}(T)$ occurs at or near $\tau = 1$, indicating that effects at T are primarily attributed to recent causes.
- For $\tau \in (0, 1)$, $\lim_{T'\to T^-} \delta^{\mathrm{b}}(\tau; T')$ measures the absolute difference in information gain when estimating $\mathbf{y}(t) = \mathbf{y}(g(\tau)) = \mathbf{y}(\tau T')$ versus $\mathbf{y}(0)$ from incorporating nearly all observations of $\mathbf{x}$. Therefore, the subinterval in I over which the complete backward CIR metric is large indicates significant causal attribution into the past, as substantial past information gain is ascribed to the currently observed effect $\mathbf{x}(T)$.
- At $\tau = 0$, $\delta^{\mathrm{b}}(0; T') = 0$ by construction.

Due to the finite-memory dynamics [7], $\lim_{T'\to T^-} \delta^{\mathrm{b}}(\tau; T')$ generally decreases from its maximum near $\tau = 1$ to zero at $\tau = 0$.

The subjective backward CIR length of $\mathbf{y} \to \mathbf{x}(T)$, dependent on a threshold ε, is:

$$\widetilde{\boldsymbol{\tau}}^{\mathrm{b}}(T, \varepsilon) := T\left(1 - \sup\left\{\tau \in \mathrm{I} : \lim_{T'\to T^-} \delta^{\mathrm{b}}(\tau; T') \le \varepsilon\right\}\right), \quad T > 0, \tag{2.19}$$

with $\widetilde{\boldsymbol{\tau}}^{\mathrm{b}}(T, \varepsilon) = 0$ when $\varepsilon < 0$. The corresponding subjective backward CIR interval is:

$$\widetilde{\mathrm{CIR}}^{\mathrm{b}}(T, \varepsilon) := \left[T - \widetilde{\boldsymbol{\tau}}^{\mathrm{b}}(T, \varepsilon), T\right] \subseteq [0, T], \quad T > 0. \tag{2.20}$$

Remark 2.7. The bias term $\delta(0, T')$ ensures the supremum in (2.19) exists since 0 always belongs to the set. Furthermore, this also ensures that the subjective backward CIR length conservatively estimates the temporal attribution of $\mathbf{x}(T)$ to $\mathbf{y}$'s past states, as:

$$\widetilde{\boldsymbol{\tau}}^{\mathrm{b}}(T, \varepsilon) \le T\left(1 - \lambda_{\mathrm{I}}\left(\left\{\tau \in \mathrm{I} : \lim_{T'\to T^-} \delta^{\mathrm{b}}(\tau; T') \le \varepsilon\right\}\right)\right),$$

with λ_{I} denoting the Lebesgue measure on I.

Averaging over all possible ε thresholds gives the objective backward CIR length:

$$\boldsymbol{\tau}^{\mathrm{b}}(T) := \frac{1}{M^{\mathrm{b}}(T)} \int_0^{M^{\mathrm{b}}(T)} \widetilde{\boldsymbol{\tau}}^{\mathrm{b}}(T, \varepsilon) \mathrm{d}\varepsilon, \quad T > 0, \tag{2.21}$$

with the corresponding objective backward CIR interval defined as:

$$\mathrm{CIR}^{\mathrm{b}}(T) := \left[T - \boldsymbol{\tau}^{\mathrm{b}}(T), T\right] \subseteq [0, T], \quad T > 0. \tag{2.22}$$

As with the forward CIR, this objective interval lies within the extreme subjective intervals.

Remark 2.8. When $M^{\mathrm{b}}(T) \approx 0$ (weak evidence for $\mathbf{y} \to \mathbf{x}(T)$ causality), (2.21) may yield an inflated CIR length if the complete backward CIR metric gradually increases toward its near-zero maximum $M^{\mathrm{b}}(T)$ as τ increases. Like in Remark 2.3, such values typically coincide with small ACI maxima and should be construed together to avoid causal misidentification.

2.3.5. Efficient Computation of the Objective Backward CIR. Direct computation of the objective backward CIR length in (2.21) is often prohibitive. Its subjective counterpart in (2.19) rarely admits analytical solutions, requires repeated smoother computations with each new observation of $\mathbf{x}$, and involves quadrature that scales at best quadratically with the number of discretization points.

Mirroring (2.15), we develop a computationally efficient upper-bound approximation to (2.21) that becomes exact when the complete backward CIR metric is nondecreasing, usually the case in practice. The theorem below requires only measurability of $\lim_{T'\to T^-}\delta^{\mathrm{b}}(\cdot;T')$ over I (see Section A.2), with the proof given in Section B of the Appendix.

Theorem 2.9 (Computationally Efficient Approximation of the Objective Backward CIR). *Assume $M^{\mathrm{b}}(T)=\sup_{\tau\in\mathrm{I}}\{\lim_{T'\to T^-}\delta^{\mathrm{b}}(\tau;T')\}>0$ exists for each $T>0$. Then:*

$$\begin{aligned}\boldsymbol{\tau}^{\mathrm{b}}_{\mathrm{approx}}(T) &:= \frac{T}{M^{\mathrm{b}}(T)}\int_0^1 \lim_{T'\to T^-}\delta^{\mathrm{b}}(\tau;T')\mathrm{d}\tau\\ &= \frac{T}{M^{\mathrm{b}}(T)}\int_0^{M^{\mathrm{b}}(T)}\Big(1-\lambda_{\mathrm{I}}\Big(\Big\{\tau\in\mathrm{I}:\lim_{T'\to T^-}\delta^{\mathrm{b}}(\tau;T')\le\varepsilon\Big\}\Big)\Big)\mathrm{d}\varepsilon \ge \boldsymbol{\tau}^{\mathrm{b}}(T),\end{aligned}\tag{2.23}$$

with equality if and only if $\lim_{T'\to T^-}\delta^{\mathrm{b}}(\cdot;T')$ is nondecreasing on I. In this case, the maximum occurs at $\tau=1$:

$$\begin{aligned}M^{\mathrm{b}}(T) &= \lim_{T'\to T^-}|\delta(T',T')-\delta(0,T')|\\ &= \lim_{T'\to T^-}\big|\mathcal{P}\big(p^{\mathrm{s}}_{T'}(\mathbf{y}|\mathbf{x}),p^{\mathrm{f}}_{T'}(\mathbf{y}|\mathbf{x})\big)-\mathcal{P}\big(p(\mathbf{y}(0)|\mathbf{x}(s\le T)),p(\mathbf{y}(0)|\mathbf{x}(s\le T'))\big)\big|,\end{aligned}\tag{2.24}$$

where the first term inside the absolute value quantifies the causal intensity of $\mathbf{y}(T')\to\mathbf{x}$, while the second measures the initial-time information deficit from limited data assimilation.

Thus, we can approximate the ε-average in (2.21) with a time integral of the complete backward CIR metric, which becomes exact under monotonicity conditions.

Remark 2.10. As (2.23) is an overestimate, it yields larger objective backward CIR intervals compared to the exact definition.

Remark 2.11. Dimensional analysis confirms $\boldsymbol{\tau}^{\mathrm{b}}_{\mathrm{approx}}(T)$ maintains consistent time units, ensuring physical validity.

2.3.6. Temporal Behavior of the Objective Backward CIR in Conditionally Linear Systems. Analytical study of the forward CIR metric $\delta^{\mathrm{f}}(\tau;t)$ in (2.9) is challenging, as the CIR metric can exhibit erratic behavior when viewed as a function of lagged observational time T'. This reflects the inherent difficulty of extrapolating temporal causal impact into the future. In contrast, the backward CIR metric $\delta^{\mathrm{b}}(\tau;T')$ in (2.16) and its complete variant in (2.17) are more analytically tractable. By focusing on causal attribution rather than prediction and leveraging the efficient approximation in Theorem 2.9, we can establish rigorous asymptotic results ($T>0$ sufficiently large) for specific system classes.

Consider conditionally linear systems with state-independent feedbacks:

$$\begin{aligned}\mathrm{d}x &= (\lambda^x y + f^x(t,x))\mathrm{d}t+\sigma^x\mathrm{d}W_1,\\ \mathrm{d}y &= (\lambda^y y + f^y(t,x))\mathrm{d}t+\sigma^y\mathrm{d}W_2,\end{aligned}\tag{2.25}$$

where again x is observed while y is unobserved. Despite the conditionally linear structure of (2.25), the system remains nonlinear (in x) and captures many complex dynamical regimes (see Section C in the Appendix for concrete examples).

Theorem 2.12 (Temporal Behavior of the Objective Backward CIR Length for Conditionally Linear Systems). *Consider (2.25) under the assumptions of Theorem 2.9, including existence of filter and smoother equilibrium distributions. For large T (which need not be that large in practice due to conditional linearity), the approximate objective backward CIR satisfies:*

$$\tau^{\mathrm{b}}_{approx}(T) = \Theta(1) \Leftrightarrow A_1 \le \tau^{\mathrm{b}}_{approx}(T) \le A_2, \tag{2.26}$$

where A_1 and A_2 are T-independent constants.

Remark 2.13. Under the assumptions of Theorem 2.12, in practice $\tau^{\mathrm{b}}_{\mathrm{approx}}(T)$ oscillates intermittently between the T-constant bounds in (2.26) due to the inherent dynamical noise in (2.25). See Section 3.1, Figures 4(c) and 6(c) for numerical validation.

2.3.7. Schematic Illustration and Physical Interpretation of the Forward and Backward CIRs and Their Relationship. Panels (A) and (B) of Figure 2 illustrate the underlying mechanisms of the forward and backward CIRs, respectively, through an example simulation. Both the objective and subjective formulations are depicted. For the forward CIR, the analysis focuses on a time t corresponding to the onset of an extreme event in $\mathbf{x}$, triggered by $\mathbf{y}(t)$. In contrast, the backward CIR results pertain to time T at the peak of that extreme event.

To interpret these results, note that the subjective CIR lengths in (2.11) and (2.19) are affine transformations of the generalized inverses of the respective CIR metrics, given in (2.9) and (2.16) respectively [27, 24, 38]. Here, the forward CIR metric fails to be a nonincreasing function in lagged observational time T'. Consequently, its generalized inverse and by extension the subjective forward CIR length are multivalued functions. This demonstrates the strong need for objective CIR measurements. It is also validated that the computationally efficient approximation in (2.15) is a strict underestimate of the objective forward CIR: $\tau^{\mathrm{f}}_{\mathrm{approx}}(t) < \tau^{\mathrm{f}}(t)$. Conversely, the complete backward CIR metric is an increasing and continuous function of natural time. Hence, a one-to-one correspondence exists between the subjective backward CIR length and its inverse, while the computationally efficient approximation in (2.16) precisely recovers the associated objective CIR: $\tau^{\mathrm{b}}_{\mathrm{approx}}(T) = \tau^{\mathrm{b}}(T)$.

Intuitively, the forward and backward CIRs provide complementary temporal perspectives to state-space causality, assigning temporal extents in opposite directions:

- **(A)** Forward CIR characterizes the future persistence of $\mathbf{y}(t)$'s causal influence on $\mathbf{x}$, contingent on evidence for $\mathbf{y}(t) \to \mathbf{x}$ at time $t \in [0, T]$.
- **(B)** Backward CIR locates the past onset of causal precursors for the observed effect $\mathbf{x}(T)$, conditional on significant evidence towards $\mathbf{y} \to \mathbf{x}(T)$ at time $T > 0$.

This complementary nature between the two CIRs is evident in the parallel properties developed in Sections 2.3.2 and 2.3.4 (compare Remarks 2.1–2.4 with Remarks 2.6–2.10). Together, they unify causal prediction and attribution within the ACI framework.

Their relationship naturally mirrors fundamental concepts from physics. In relativistic spacetime, the forward CIR defines the future light cone of $\mathbf{y}(t)$ in the $\mathbf{x}$-marginal state space, while the backward CIR defines the past light cone of $\mathbf{x}(T)$ in the $\mathbf{y}$-marginal state space

[75, 65]. More concretely, hyperbolic PDEs provide a direct analogy through their finite information propagation rate [25]. The forward CIR mirrors the range of influence, delimiting how $\mathbf{y}(t)$'s effects propagate forward, while the backward CIR corresponds to the domain of dependence, identifying which past $\mathbf{y}$ values caused $\mathbf{x}(T)$. However, unlike fixed characteristic cones, CIRs are dynamical regions that evolve with the causal flow between $\mathbf{y}$ and $\mathbf{x}$ subspaces.

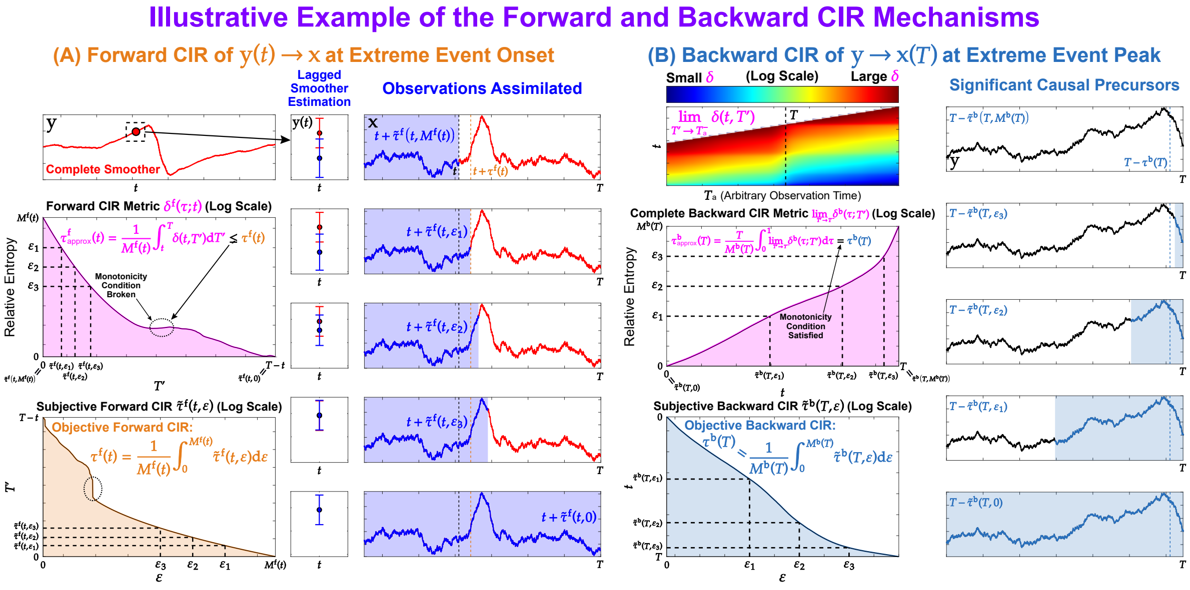


Figure 2. *Mathematical details of the forward and backward CIRs during an illustrative extreme event scenario. Panel (A): Mechanisms of the forward CIR metric and lengths, subjective and objective, at the onset of an extreme event. Panel (B): Similar to Panel (A), but for the complete backward CIR metric and associated lengths, subjective and objective, at the peak of the extreme event.*

2.4. Computational Tool for Exact CIR Inference: Conditional Gaussian Nonlinear Systems. Computing objective CIR lengths requires evaluating the CIR metrics $\delta^{\mathrm{f}}(\tau;t)$ and $\delta^{\mathrm{b}}(\tau;T')$, which depend on the lagged and complete smoother PDFs. While for the general nonlinear system in (2.1) ensemble methods have to be utilized, a wide range of nonlinear systems, called conditional Gaussian nonlinear systems (CGNSs), admit closed-form solutions for both filter and smoother posterior distributions. These systems take the form [47, 19, 18]:

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \big(\mathbf{\Lambda}^{\mathbf{x}}(t,\mathbf{x})\mathbf{y}(t) + \mathbf{f}^{\mathbf{x}}(t,\mathbf{x})\big) + \mathbf{\Sigma}_1^{\mathbf{x}}(t,\mathbf{x})\dot{\mathbf{W}}_1(t) + \mathbf{\Sigma}_2^{\mathbf{x}}(t,\mathbf{x})\dot{\mathbf{W}}_2(t), \tag{2.27a}$$

$$\frac{\mathrm{d}\mathbf{y}(t)}{\mathrm{d}t} = \big(\mathbf{\Lambda}^{\mathbf{y}}(t,\mathbf{x})\mathbf{y}(t) + \mathbf{f}^{\mathbf{y}}(t,\mathbf{x})\big) + \mathbf{\Sigma}_1^{\mathbf{y}}(t,\mathbf{x})\dot{\mathbf{W}}_1(t) + \mathbf{\Sigma}_2^{\mathbf{y}}(t,\mathbf{x})\dot{\mathbf{W}}_2(t), \tag{2.27b}$$

where $\mathbf{x}$ appears nonlinearly while $\mathbf{y}$ exclusively linearly. Nevertheless, the marginal distributions of both $\mathbf{x}$ and $\mathbf{y}$ as well as their joint distribution can be highly non-Gaussian. Only the conditional distribution of $\mathbf{y}$ given a realization of $\mathbf{x}$ is Gaussian. CGNSs are widely applicable in neuroscience, ecology, and geophysics due to their efficacy in continuous data assimilation.

We employ an adaptive-lag online smoother for CGNSs [7] that enables exact computation of both CIR metrics. This algorithm efficiently computes lagged smoother PDFs in real-time by adaptively performing only the most influential updates, capturing essential information from new observations while minimizing computational cost. Section D in the Appendix details this method and its application to CIR metric computation.

3. CIRs for Probing Complex Dynamical Systems. In this section, ACI and its forward and backward CIR formulations are utilized to address three key scientific challenges: (a) Understanding bifurcation-driven and noise-induced tipping points in a conceptual Earth system model, (b) resolving interfering observed effects to improve prediction and attribution in low-frequency atmospheric variability models, and (c) investigating atmospheric blocking and unblocking mechanisms using a conceptual equatorial circulation model with flow-wave interactions. All objective forward and backward CIR lengths in the following figures are computed using their approximations in (2.15) and (2.23), respectively. Implementation details for these metrics appear in Sections D.1 and D.2 of the Appendix.

3.1. Studying Bifurcation-Driven and Noise-Induced Tipping Points in a Model with Intermittent Instabilities. Tipping points in Earth system models explain dynamical transitions between physical regimes, but distinguishing whether these are driven by external forcings (e.g., Milankovitch cycles [10]) or internal variability remains a challenging task.

Mathematically, tipping points are analyzed through bifurcation structures where quasi-static attractors lose stability, potentially leading to new attractors, transient excursions, or equilibrium departure [86]. Understanding these mechanisms is crucial for assessing responses in Earth processes under perturbations [2]. We focus on two types of tipping points. First, on bifurcation-driven tipping, which occurs when control parameters cross critical thresholds due to smooth perturbations [42]. Early warning signals like rising autocorrelation and variance often precede tipping but cannot reliably distinguish genuine precursors from internal variability artifacts [13]. Furthermore, in multiscale systems, bifurcations in unobserved fast components may escape detection, where rapid exogenous changes can outpace early warning signals in inertial systems [41]. ACI-based CIR metrics overcome these limitations: Forward CIR enables real-time inference of impending tipping without solely relying on observations or low-order statistical moments, while backward CIR supports post-tipping attribution to triggering precursors. Second, on noise-induced tipping, which is driven by random fluctuations or internal variability. These transitions lack clear early warning signals since noise can trigger regime shifts without potential changes. In Earth models, explained via stochastic dynamical systems [31], weather modes can act as stochastic perturbations that excite atmospheric processes through stochastic resonance [29]. Forward and backward CIR measures can jointly enable efficient detection, anticipation, and attribution of noise-induced regime transitions.

These types of tipping points are studied using the following three-dimensional system:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x - d_x x^3 - \alpha y + \sigma_x \dot{W}_x, \tag{3.1a}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \gamma x + \beta - \frac{d_y}{\varepsilon} y + \frac{\sigma_y}{\sqrt{\varepsilon}} \dot{W}_y, \tag{3.1b}$$

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = -d_\gamma(\gamma - \bar{\gamma})\mathrm{d}t + \sigma_\gamma \dot{W}_\gamma. \tag{3.1c}$$

This model can mimic essential structural features of low-frequency variability present in general circulation models, which cannot be captured by linear Gaussian approximations [55]. The small parameter $0 < \varepsilon < 1$ separates the fast weather modes (y) from the slow (x) timescales, while γ provides stochastic parameterization with multiplicative noise. We apply our framework to study:

- Bifurcation-driven tipping: In (3.1a), y acts as a random forcing on x, intermittently triggering instability or regime shifts by pushing the system beyond its global stability threshold. Forward CIR analysis of $y \to x|\gamma$ assesses future tipping risk, while backward CIR attributes tips to y's past behavior as a bifurcation parameter.
- Noise-induced tipping: The colored noise introduced by γ in (3.1c) reduces system memory relative to additive-noise alternatives, due to the turbulent dynamics generated [46, 15]. Simultaneous study of the forward and backward CIRs for $\gamma \to y|x$ identifies regime shifts linked to low-frequency variability.

The following parameter values are used in this study: $\varepsilon \in \{0.01, 0.1\}$, $d_x = 1/3$, $\alpha = 4$, $\sigma_x = 0.2$, $d_y = 0.2$, $\beta = -0.8$, $\sigma_y = 0.3$, $d_\gamma = 0.5$, $\bar{\gamma} = 1$, $\sigma_\gamma = 2$.

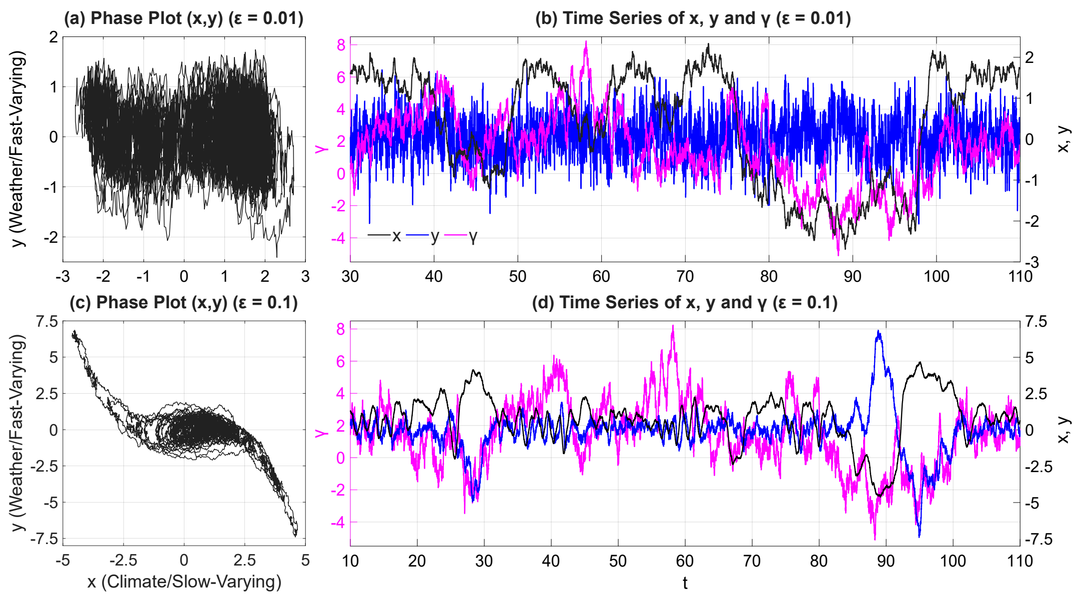


Figure 3. *Panel (a): Phase plot of $(x(t), y(t))$ for $\varepsilon = 0.01$. Panel (b): Time series for $\varepsilon = 0.01$. Panel (c)–(d): Same for $\varepsilon = 0.1$. Time windows: $t \in [30, 110]$ ($\varepsilon = 0.01$), $t \in [10, 110]$ ($\varepsilon = 0.1$). The random number seeds are fixed, therefore the time series of γ in these two simulations is the same.*

Figure 3 reveals distinct behaviors in the two regimes with different ε values. When $\varepsilon = 0.01$, a bistable slow mode x is observed, with transitions at $t \in [76, 82]$ ($x \approx 6 \to -2.5$) and $t \in [97.5, 100]$ (reverse). A further near-tip instability occurs around $t = 40$. Transitions coincide with $\gamma < 0$, driving anti-correlated y amplifications that trigger bifurcations in x via the $-\alpha y$ forcing. As such, tips are partly initiated due to the internal variability in (3.1b). On the other hand, when $\varepsilon = 0.1$, intermittent oscillations and unstable excursions in x occur with closer x-y timescales, which is evident in near-synchronous oscillations during stable periods.

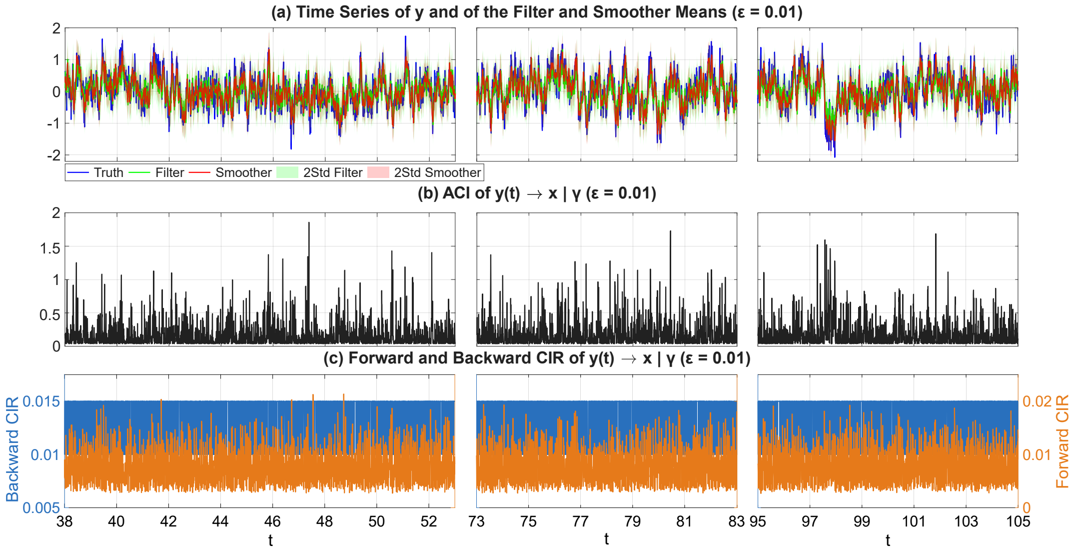


Figure 4. *ACI and CIR analysis of* $y \to x|\gamma$ *for* $\varepsilon = 0.01$. *Panel (a): Time series with filter/smoother distributions. Panel (b): ACI metric. Panel (c): Forward (orange) and backward (blue) CIR. Columns show* $t \in [38, 53]$, $t \in [73, 83]$, *and* $t \in [95, 105]$, *respectively.*

Figures 4 and 5 present ACI and CIR analyses for $\varepsilon = 0.01$, validating both the physical intuition behind (3.1) and our theoretical results while clarifying causal mechanisms during regime transitions. For $y \to x|\gamma$, ACI peaks correlate with tip generation and duration, particularly during the build-up phase of the second transition at $t \in [97.5, 100]$, with additional local maxima during the near-tip instability around $t \in [40, 50]$. This temporal pattern reflects y's role as a constant-feedback forcing, where peaks coincide with strong y values, while rapid fluctuations stem from slow-fast scale separation and internal variability.

The ACI for $\gamma \to y|x$ shows local maxima preceding both tips and the near-tip instability. Notably, it peaks immediately after regime establishment, indicating that strong negative y values help terminate the near-tip instability. However, during intermediate transition phases, this ACI metric drops to near zero, suggesting y drives tip maintenance and potential nullification while γ's influence recedes temporarily.

Both forward and backward CIRs for $y \to x|\gamma$ remain short due to the fast weather timescale, indicating instantaneous "short-memory" causal impact. The backward CIR's uniformity in observational time, oscillating between 0.01 and 0.015, confirms Theorem 2.12, with equilibrium statistics reached by $T = 38$ given the CGNS structure. In contrast, $\gamma \to y|x$

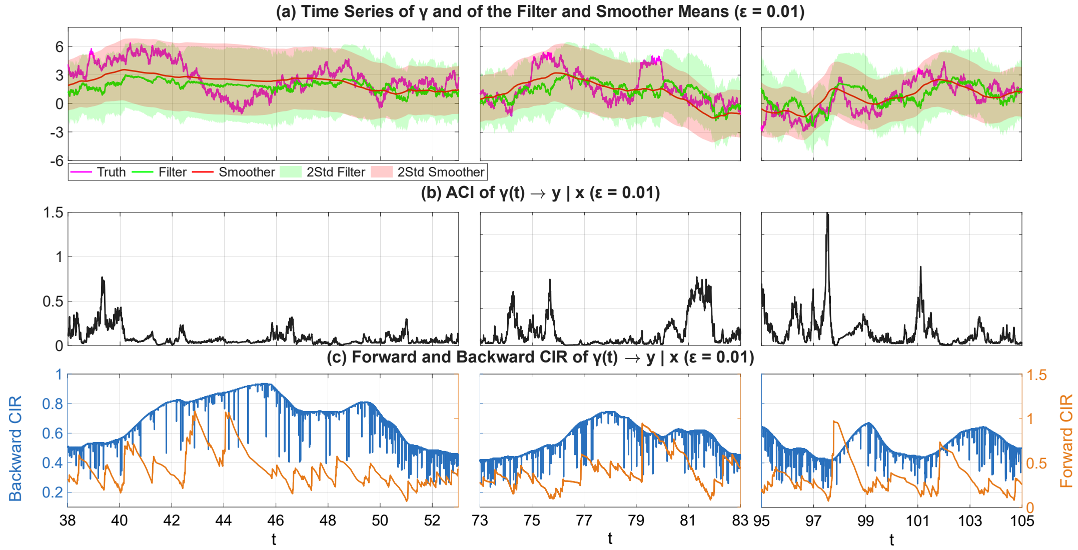


Figure 5. *Same as Figure 4, but for $\gamma \to y|x$.*

exhibits smaller ACI values but consistently larger CIRs. Forward CIR peaks align with gradual ACI increases, with slow decay reflecting extended causal horizons, while backward CIR effectively attributes tips to past internal variability, though occasional operationally inflated values occur during transitional periods dominated by weather fluctuations.

These analyses reveal distinct causal patterns for each event. The near-tip instability ($t \in [40, 50]$) is jointly initiated by bifurcation and noise mechanisms, with comparable pre-excursion ACI values for both relationships. During the instability itself, the ACI metric of $\gamma \to y|x$ drops to near zero while that of $y \to x|\gamma$ maintains multiple local maxima, indicating resolution primarily through y's bifurcation-parameter behavior. The first tip ($t \in [76, 82]$) is predominantly noise-induced with bifurcation contributions, as both ACI metrics peak beforehand, though y dominates the transition dynamics. Post-stabilization, $\gamma \to y|x$ regains significance with long CIRs, consistent with $\gamma < 0$ maintaining the negative attracting state of x. The second, faster tip ($t \in [97.5, 100]$) is primarily bifurcation-driven, evidenced by the brief $\gamma \to y|x$ peak and short CIRs, while y's rapid variability triggers the transition following $\gamma > 0$, with subsequent sign change of γ enabling positive-state stabilization for x.

Figures 6 and 7 present the case with $\varepsilon = 0.1$. In this regime, x exhibits regular oscillations when $\gamma > 0$ but becomes more irregular and unstable when $\gamma < 0$. Correspondingly, y behaves similarly but with a slight delay due to the reduced timescale separation. Notably, x and y become positively correlated for $\gamma > 0$ but anti-correlated for $\gamma < 0$, as dictated by the dynamics in (3.1a)–(3.1b).

Both ACI metrics show enhanced amplitude for $\varepsilon = 0.1$, with $y \to x|\gamma$ exhibiting particularly larger CIR values that reflect the reduced timescale separation. When γ enters strongly negative regimes, it becomes the dominant causal driver, inducing unstable oscillations in both variables. However, the corresponding short CIRs of $\gamma \to y|x$ during these unstable periods

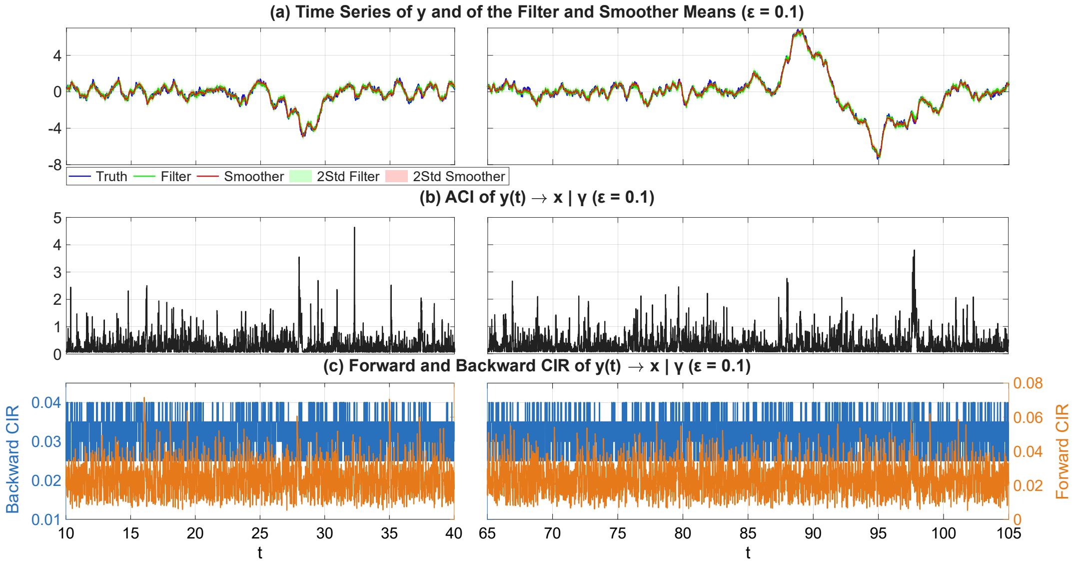


Figure 6. *Same as Figure 4, but for* $\varepsilon = 0.1$. *Time windows:* $t \in [10, 40]$ *and* $[65, 105]$.

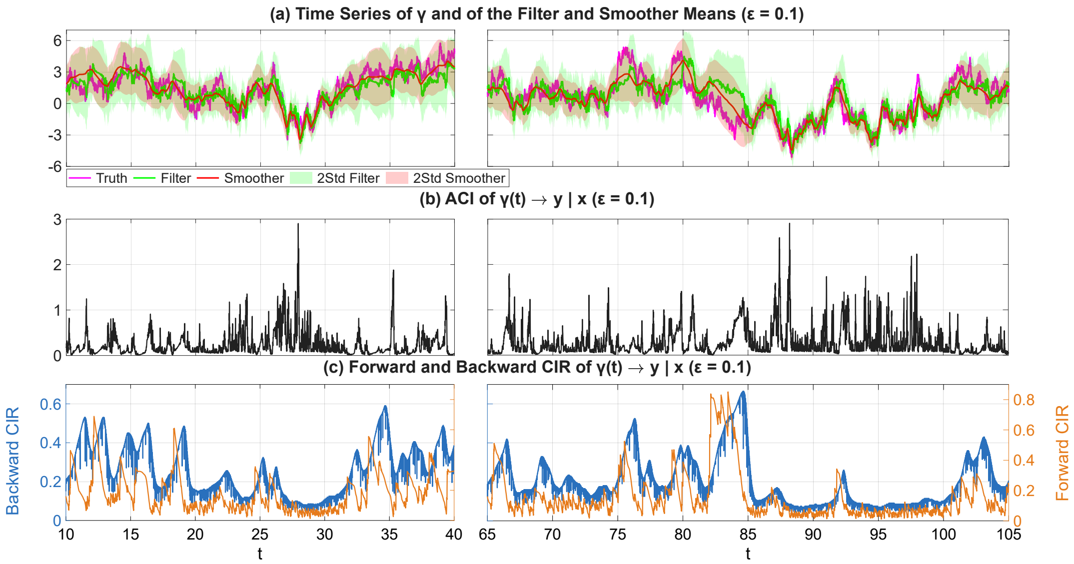


Figure 7. *Same as Figure 6, but for* $\gamma \to y|x$.

indicate that γ's causal influence remains temporally confined, affecting instantaneous behavior without extended temporal impact. Prior to the return to stable oscillations, particularly around $t = 27.5$ and $t = 97.5$, the ACI for $y \to x|\gamma$ displays concurrent strong local ACI peaks with significant dispersion components (see signal-dispersion decomposition in Sections D.1 and D.2 of the Appendix). These peaks serve as operational precursors, confirming that internal variability, rather than mean bifurcation dynamics, governs the onset and demise of unstable phases. Once x stabilizes around $x \approx 2$ with regular oscillations, the system becomes predominantly bifurcation-driven, exhibiting correlated but delayed responses through the $-\alpha y$ and γx feedback loop.

The ACI framework proves particularly effective for identifying noise-induced instabilities even in the absence of clear early warning signals. The instability emerging around $t = 85$ demonstrates this capability, with the forward CIR peak at $t = 82.5$ providing advance prediction and strong backward CIRs during the initial stages of the unstable phase correctly attributing the event to γ's evolution toward negative values. This behavior, while present for $\varepsilon = 0.01$, becomes more pronounced with the larger ε as it better resolves the underlying causal structure. The CIR for $y \to x|\gamma$ exhibits more intricate temporal structure compared to the $\varepsilon = 0.01$ case, consistent with the closer x-y timescales, while Theorem 2.12 remains valid with backward CIR oscillations between 0.025 and 0.04 through intermediate transitions.

3.2. Impact of Non-Target Variable Resolution on the CIR Metrics in a Multiscale Model for Atmospheric Variability. Consider a system that includes quadratic terms which preserve the system's energy through the nonlinear operator, multiscale spatiotemporal features, and correlated additive and multiplicative noise. These features enable an efficient search of the non-Gaussian parameter space with far fewer degrees of freedom [53, 54]:

$$\begin{aligned}\frac{\mathrm{d}x_1}{\mathrm{d}t} &= a_1x_1 - c_1x_1^3 - x_2(M + M_1x_1 + M_2x_2) + I_{11}x_1y_1 + I_{12}x_1y_2 + L_{11}y_1 + L_{12}y_2 \\ &\quad + f_1^x(t) + \sigma_{x_1}\dot{W}_{x_1} + \frac{\sigma_{y_1}}{\gamma_1}(L_{11} - I_{11}x_1)\dot{W}_{y_1} + \frac{\sigma_{y_2}}{\gamma_2}(L_{12} - I_{12}x_1)\dot{W}_{y_2},\end{aligned} \tag{3.2a}$$

$$\begin{aligned}\frac{\mathrm{d}x_2}{\mathrm{d}t} &= -c_2x_2 + x_1(M + M_1x_1 + M_2x_2) + I_{21}x_2y_1 + I_{22}x_2y_2 + L_{21}y_1 + L_{22}y_2 \\ &\quad + f_2^x(t) + \sigma_{x_2}\mathrm{d}\dot{W}_{x_2} + \frac{\sigma_{y_1}}{\gamma_1}(L_{21} - I_{21}x_2)\dot{W}_{y_1} + \frac{\sigma_{y_2}}{\gamma_2}(L_{22} - I_{22}x_2)\dot{W}_{y_2},\end{aligned} \tag{3.2b}$$

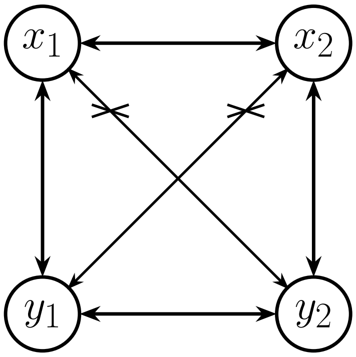
$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = -\frac{\gamma_1}{\varepsilon}y_1 - L_{11}x_1 - L_{21}x_2 + Ny_2 - I_{11}x_1^2 - I_{21}x_2^2 + f_1^y(t) + \frac{\sigma_{y_1}}{\sqrt{\varepsilon}}\dot{W}_{y_1}, \tag{3.2c}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = -\frac{\gamma_2}{\varepsilon}y_2 - L_{12}x_1 - L_{22}x_2 - Ny_1 - I_{12}x_1^2 - I_{22}x_2^2 + f_2^y(t) + \frac{\sigma_{y_2}}{\sqrt{\varepsilon}}\dot{W}_{y_2}. \tag{3.2d}$$

The model parameters used in this study are: $a_1 = 1$, $c_1 = 1/3$, $M = 0.5$, $M_1 = 0.5$, $I_{11} = 0.6$, $I_{12} = 0$, $L_{11} = 1$, $L_{12} = 0$, $f_1^x(t) = 0$, $\sigma_{x_1} = 0.15$, $c_2 = 0.4$, $M_2 = -1.5$, $I_{21} = 0$, $I_{22} = 2.5$, $L_{21} = 0$, $L_{22} = 1.5$, $f_2^x(t) = 4 + 2\sin(2\pi t/18)$, $\sigma_{x_2} = 0.3$, $\gamma_1 = 0.5$, $\varepsilon = 0.1$, $N = 4$, $f_1^y(t) = 1$, $\sigma_{y_1} = 1$, $\gamma_2 = 1.2$, $f_2^y(t) = -1$, and $\sigma_{y_2} = 2$. Under these model parameter values, the causal network induced by this four-dimensional system is:

The system in (3.2) is a reduced-order model for large-scale geophysical flow [51, 31], often derived via a stochastic-mode reduction on general circulation models [56, 57]. It distinguishes between the resolved slow atmospheric or environmental variables $\mathbf{x} = (x_1, x_2)^\mathtt{T}$ and fast weather components $\mathbf{y} = (y_1, y_2)^\mathtt{T}$ through the timescale separation governed by $0 < \varepsilon < 1$, which introduces variations in system memory and autocorrelation.

The model captures key physical processes. Quadratic nonlinearities enforce energy conservation while inducing intermittent instabilities in the atmospheric components via anti-



damping. The linear operator generates multiscale structures typical of turbulent flows, decomposable into a skew-symmetric part mimicking Coriolis effects and topographic Rossby wave propagation, and a negative-definite component providing stochastic stability through dissipative processes like surface drag. In the weather dynamics, oscillatory structures arise from terms like $+Ny_2$ and $-Ny_1$, which dominate for large N while atmospheric variables remain slow if weather feedbacks are rapidly damped. Forcings are scale-dependent: Large-scale terms f_1^x, f_2^x represent slow external inputs such as decadal oscillations [88, 58], while small-scale terms f_1^y, f_2^y capture faster intraseasonal variability like Madden–Julian oscillation effects [87].

The goal of this study is to investigate how resolving ancillary variables clarifies the underlying causal structure. By conditioning on (x_2, y_2) when assessing $y_1 \to x_1$ and on (x_1, y_1) for $y_2 \to x_2$ through conditional ACI, we decompose the joint relationship $(y_1, y_2) \to (x_1, x_2)$ to identify dominant causal drivers and their temporal influence.

The time series in Figure 8(a–d) reveal distinct dynamical behaviors. The variable x_1 exhibits intermittent excursions towards a weaker but unstable state, showing negative skewness and bimodality due to cubic dissipation, multiplicative noise, and weak weather-mode feedbacks [53, 17]. In contrast, x_2 remains strictly positive with significant positive skewness and intermittent instabilities due to energy conservation restrictions. The atmosphere-to-weather influence components are asymmetric: x_2 exerts stronger control on y_2 than x_1 does on y_1, consistent with $I_{11} < I_{22}$ and $L_{11} < L_{22}$. Consequently, y_1 follows a sub-Gaussian distribution while y_2 shows slight negative skewness, with both weather variables evolving faster than their atmospheric counterparts ($\varepsilon = 0.1$). All marginals and joint distributions display strong non-Gaussianity from intermittent instabilities, regime switches, and multiscale dynamics.

Panels (e–j) demonstrate how resolving ancillary variables affects ACI metrics and CIRs. Conditioning increases CIR lengths relative to the joint unconditional case (Panels (e–f)), revealing the true causal structure by showing when large joint ACI values stem from one marginal relationship while the other remains negligible. Specifically, accounting for (x_2, y_2) in $y_1 \to x_1$ (Panels (g–h)) shows that weather-to-atmosphere feedback from y_1 to x_1 drives the dispersion component of the joint ACI metric (see signal-dispersion decomposition in Sections D.1 and D.2 of the Appendix), particularly during intervals like $t \in [50, 55]$ where y_1's rapid fluctuations generate x_1's bimodal tendencies. Conversely, conditioning on (x_1, y_1) in $y_2 \to x_2$ (Panels (i–j)) reveals longer CIRs attributed to y_2's weather effects, with significant CIRs at around $t = 57.5$, $t = 65$, and other instants where $y_1 \to x_1$ ACI drops to zero. These periods correspond to x_2's intermittent instabilities and positive bursts, where energy conservation forces x_1 into near-tip unstable excursions. The conditional $y_1 \to x_1$ CIRs show strong causal influence with extended temporal influence during x_1's deviations above its mean state at around $t \in [50, 55]$ and $t = 82.5$, enabling prediction 0.2 units ahead and attribution 0.3 units into the past to strong y_1 amplitudes. Notably, the conditional ACI for $y_2 \to x_2$ never collapses to zero, indicating persistent uncertainty reduction from smoother solutions that is consistent with longer CIRs. In contrast, the ACI metric of $y_1 \to x_1$ becomes negligible during x_1 excursions driven by x_2's instabilities, as these deterministic events are well-captured by filter solutions alone.

3.3. Atmospheric Blocking Mechanisms in an Equatorial Circulation Model with Seasonality and Flow-Wave Interactions. The noisy Lorenz-84 model is a conceptual system describing coarse-grained equatorial atmospheric circulation [88, 74]. Its deterministic component is derived from a Galerkin truncation of the quasi-geostrophic potential vorticity equations [19, 90]. The model captures key features of the Hadley cell and can be extended to include atmosphere-ocean coupling for studying oceanic variability and seasonality [19, 68, 28]. The model reads [48]:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -y^2 - z^2 - a\big(x - f(t)\big) + \sigma_x \dot{W}_x, \tag{3.3a}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -bxz + xy - y + g + \sigma_y \dot{W}_y, \tag{3.3b}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = bxy + xz - z + \sigma_z \dot{W}_z. \tag{3.3c}$$

The model parameters used for this case study are: $a = 1/4$, $f(t) = 8 + 3\cos(2\pi t/73)$, $\sigma_x = 0.2$, $b = 4$, $g = 1$, and $\sigma_y = \sigma_z = 0.2$, which follow those originally proposed by Lorenz and define a characteristic time scale of about five days [48, 49]. The Lorenz-84 model (3.3) describes atmospheric wave-mean flow interactions where the unobserved variable x represents the amplitude of the westerly zonal flow, while the observed variables y and z are the cosine and sine phases of a large-scale baroclinic wave. Under a thermal wind balance formulation, x can describe the zonally-averaged meridional temperature gradient [88, 51]. The quadratic nonlinearities conserve energy and capture wave amplification at the expense of the zonal flow, with wave displacement governed by the advection strength b. The Prandtl number $a < 1$ allows the zonal flow to dampen slower than waves. External forcings include the seasonal thermal contrast $f(t) = 8 + 3\cos(2\pi t/73)$, which drives the zonal flow and incorporates a one-year period, and the constant asymmetric forcing $g = 1$ that represents topographic influences. The model exhibits bistability between two stable equilibria, mimicking transitions between zonal (strong westerly jets) and blocked (amplified waves) atmospheric regimes. This provides a framework for using ACI to study how the jet x influences waves y and z during blocking events.

Below, we employ forward and backward CIRs to quantify the temporal causal extent of the zonal wind current (x) on the large-scale vortices (y,z), mirroring teleconnections where regional Earth process anomalies are related via atmospheric pathways. A weak jet stream during blocking may be traced to past wave amplification that drives poleward heat transport, while an increasing meridional temperature gradient can precede future rapidly oscillating waves.

Figure 9(a–c) reveals chaotic teleconnection patterns: x evolves more slowly than the wave components y and z. When x oscillates away from zero, it induces regular wave oscillations ($t \in [15, 17.5]$, $t \in [30, 50]$, $t \in [110, 125]$), whereas near-zero x values (atmospheric blocking) cause protracted wave oscillations ($t \approx 50$, $t \in [95, 100]$, $t \in [102.5, 110]$). During blocking, phase-shifted y and z reach opposite peaks before decaying irregularly due to energy conservation, consistent with poleward heat transport weakening the jet through the vortical term $-y^2 - z^2$. Wave growth is in time suppressed as weak x enforces linear stability in (3.3b)–(3.3c), eventually leading to zonal flow resurgence.

Analysis of the conditional relationships $x \to y|z$ (Panels (d–e)) and $x \to z|y$ (Panels (f–g)) shows x predominantly drives y and z during stable oscillations away from zero, particularly when approaching and then exceeding their antidamping threshold of $x = 1$ ($t = 45$, $t \in [60, 70]$, $t = 80$, $t = 115$, $t = 125$). The induced rapid y and z oscillations show significant ACI but short CIRs, indicating x initiates but doesn't maintain instabilities due to energy conservation. The strongest ACI values occur around $t = 80$, where transitional phases between damped and rapid oscillations confirm the model's core mechanism: Zonal jet damping by significant vortical amplitudes versus wave instability from increasing meridional temperature gradients.

During x's slow irregular oscillations away from zero ($t \approx 5$, $t \in [15, 17.5]$, $t \in [30, 47.5]$, $t \in [110, 125]$), both ACI and CIR show strong peaks, suggesting zonal displacement dominates amplification with persistent causal influence. Similar patterns appear for $x \to z|y$, though phase differences from the asymmetric topographic forcing g create discrepancies. During blocking events ($t \approx 50$, $t \in [95, 100]$, $t \in [102.5, 110]$), ACI drops sharply as waves become primary drivers, while unblocking is attributed to weak x via the backward CIR, as it generates linear stability. The extended blocking during $t \in [100, 110]$ shows distinct patterns, where diminished zonal gradients result from heat transport without full blocking, a nuance often missed in physical studies. Throughout, backward CIR attributes weak jets to prior wave amplification, while forward CIR warns of future rapid oscillations during strong jets.

4. Discussion and Conclusion. In this paper, we developed mathematically rigorous formulations for quantifying the temporal extent of causal impact within the ACI framework. We introduced CIRs that measure the future period in which a current cause will remain influential (forward CIR) and the past period responsible for a currently observed effect (backward CIR). Theoretically grounded approximations enable efficient computation of these temporal measures and facilitate asymptotic analysis in tractable systems. This work advances causal inference beyond static relationships, bridging prediction and attribution of intermittent extreme events and regime transitions. Applications reveal the effectiveness of the framework in identifying causal drivers of tipping points, particularly for those triggered by internal variability that lack conventional early warning signals. The conditional formulation further enables the identification of overlapping causal influences in high-dimensional multiscale systems by resolving interfering non-target effects. The CIRs also advance the investigation of atmospheric blocking mechanisms through a conceptual equatorial circulation model with seasonal feedbacks and flow-wave interactions.

Future research directions include the following. First, developing methods to identify significant marginal causes in multivariate settings would enhance causal discovery. Second, deriving analytical expressions for conditional ACI with cross-interacting noise feedbacks would strengthen theoretical foundations. Third, investigating ACI under noise-contaminated observations would improve practical applicability. Finally, assessing the robustness of the framework against model error remains crucial. Such studies could also open new avenues for constructing reduced-order parsimonious models.

Acknowledgments. The research of N.C. is funded by the Office of Naval Research N00014-24-1-2244 and the Army Research Office W911NF-23-1-0118. M.A. is partially supported as a research assistant under the second grant.

The Appendix follows.

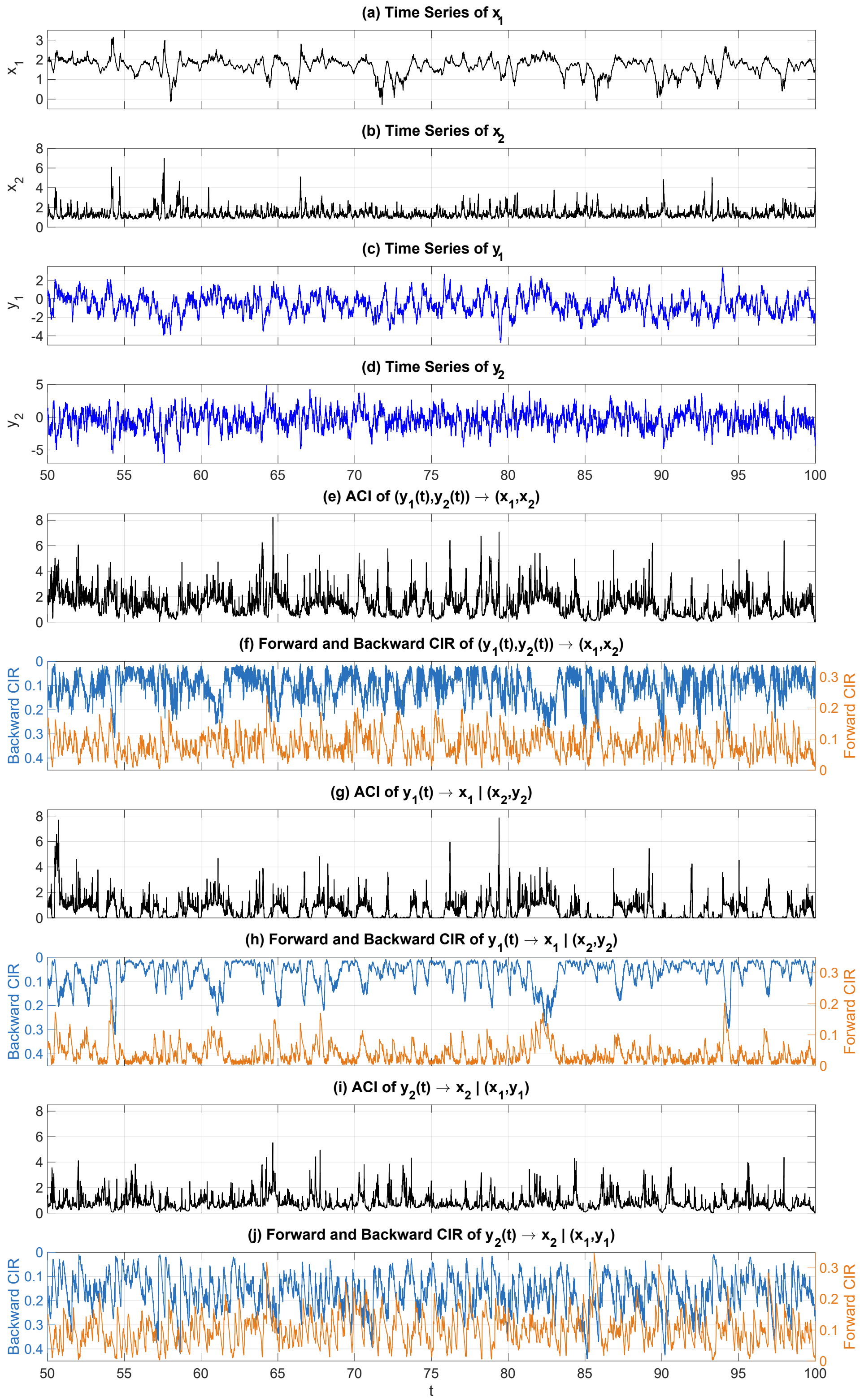


Figure 8. *Panels (a–d): Time series of x_1, x_2, y_1, and y_2. Panel (e): ACI metric used to assess the causal relationship $(y_1, y_2) \to (x_1, x_2)$ (see (2.3)–(2.4)). Panel (f): (Objective) Forward (orange) and backward (blue; reverse y-axis) CIR of $(y_1, y_2) \to (x_1, x_2)$. Panels (g–h): Same as Panels (e–f), but for the conditional causal relationship $y_1 \to x_1\big|(x_2, y_2)$ (see (2.5)–(2.7)). Panels (i–j): Same as Panels (e–f), but for the conditional causal relationship $y_2 \to x_2\big|(x_1, y_1)$ (see (2.5)–(2.7)). Time window: $t \in [50, 100]$.*

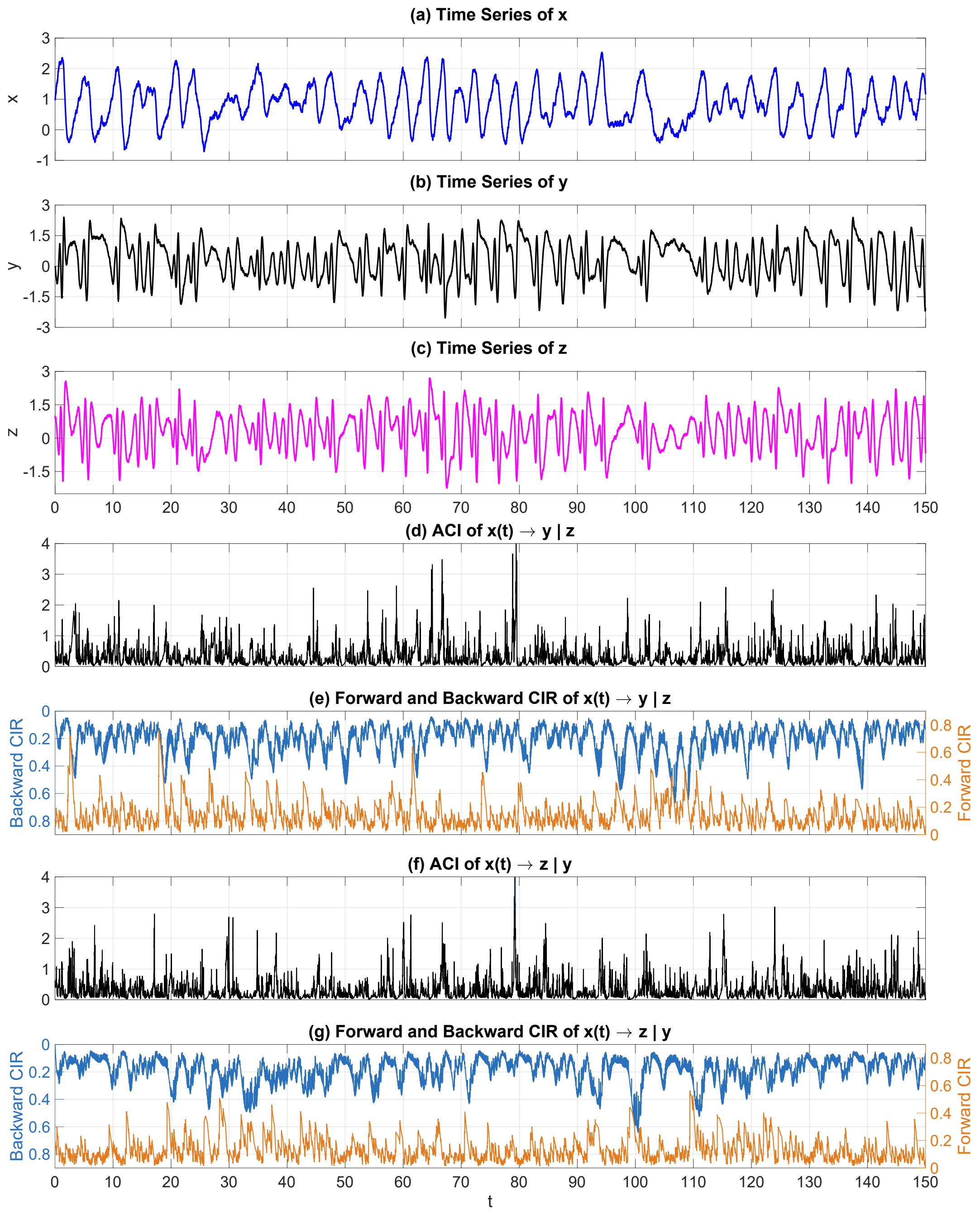


Figure 9. *Same as Figure 8, but for the noisy Lorenz-84 model in (3.3) and the conditional assimilative causal relationships $x \to y|z$ (Panels (d–e)) and $x \to z|y$ (Panels (f–g)). Time window: $t \in [0, 150]$.*

Appendix A. Required Regularity Conditions for the CIR Metrics.

In this section, we present results that concern the measurability of the forward and backward variants of the CIR metric $\delta(t,T')$ in (2.9) and (2.16), respectively. These are essential for establishing and justifying the theoretical results presented in this work (see (2.15) and Theorem 2.9). The following remarks and assumptions are needed for the subsequent subsections:

- The relative entropy in (2.2), is part of the family of f-divergences [67, 3]:

$$\mathcal{D}_f(\mathbb{P},\mathbb{Q}) := \int f(\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q})\mathrm{d}\mathbb{Q}.$$

  In particular, it is induced by the $f(v)=v\log v$ generator function. It is known that an f-divergence induced by a lower semi-continuous (LSC) f is itself LSC in the weak topology of the restricted product space of Borel probability measures defined over a Hausdorff topological state space with $\mathbb{P}\ll\mathbb{Q}$; this product space is equivalent to the direct product of the weak topologies [11]. In other words, if both $\mathbb{P}$ and $\mathbb{Q}$ are dominated by the appropriate Lebesgue measure and enjoy PDFs p and q, respectively, then the f-divergence is weakly LSC in both the primary density p as well as the reference PDF q [4, 45].
- We assume that all backward interpolation (lagged and complete smoother) PDFs,

$$p\big(\mathbf{y}(t)\big|\mathbf{x}(s\leq T')\big),$$

  are weakly continuous in $T'\in[t,T]$ for each fixed $t\in[0,T]$, as well in $t\in[0,T']$ for each fixed $T'\in[0,T]$. These conditions are necessary because weak continuity, not merely weak LSC, is required for the subsequent arguments, specifically the fact that the composition of LSC functions is not necessarily LSC. Note here that this weak continuity condition is equivalent to weak sequential continuity, since we consider at least a reflexive Polish state space (e.g., separable Hilbert space). These weak-continuity properties can be analytically verified for the backward interpolation distributions arising from the class of continuous-time nonlinear dynamical systems considered in (2.1) [70], including the CGNSs introduced in (2.27) of Section 2.4 [47], under mild regularity assumptions.
- A Borel measurable function is always Lebesgue measurable in our adopted setup.

A.1. Measurability of the Forward CIR Metric. Consider the forward CIR metric from (2.9):

$$\delta^{\mathrm{f}}(\tau;t) := \delta(t,t+\tau(T-t)) = \mathcal{P}\left(p\big(\mathbf{y}(t)|\mathbf{x}(s\leq T)\big), p\big(\mathbf{y}(t)|\mathbf{x}(s\leq t+\tau(T-t))\big)\right), \quad \tau\in[0,1],$$

obtained from δ as a function of lagged observational time $T'=h(\tau)=t+\tau(T-t)\in[t,T]$ for arbitrary $t\in[0,T]$. Then, as the composition between a weakly LSC function (relative entropy) and weakly continuous functions (backward interpolation PDFs and $h(\tau)=t+\tau(T-t)$), $\delta^{\mathrm{f}}(\tau;t)$ is a weakly LSC function on $[0,1]$ for each $t\in[0,T]$. This weak lower semi-continuity implies that $\delta^{\mathrm{f}}(\tau;t)$ is also weakly measurable on $[0,1]$ for each $t\in[0,T]$ [23, 37]. Although one might consider attempting to establish that it is also almost surely separably valued to apply Pettis' measurability theorem and conclude strong measurability [66], this step is unnecessary in our setting. For separable spaces, weak and strong measurability coincide, and the space of Borel probability measures defined over a Polish state space, such as the one considered in this work, is itself Polish [35, 36], as established by Prokhorov's theorem [89, 39, 63]. As such, the forward CIR metric defined in (2.9) is a measurable function of $\tau \in \mathrm{I}$ for each $t \in [0, T]$.

A.2. Measurability of the Backward CIR Metric. Consider the backward CIR metric from (2.16):

$$\begin{aligned}\delta^{\mathrm{b}}(\tau; T') &= |\delta(\tau T', T') - \delta(0, T')| \\ &= \big|\mathcal{P}\big(p(\mathbf{y}(\tau T')|\mathbf{x}(s \leq T)), p(\mathbf{y}(\tau T')|\mathbf{x}(s \leq T'))\big) \\ &\qquad\qquad - \mathcal{P}\big(p(\mathbf{y}(0)|\mathbf{x}(s \leq T)), p(\mathbf{y}(0)|\mathbf{x}(s \leq T'))\big)\big|,\end{aligned}$$

for $\tau \in [0, 1]$, obtained from δ as a function of natural time $t = g(\tau) = \tau T' \in [0, T']$ for arbitrary $T' \in [0, T]$ (of interest is when T' is arbitrarily close to T, as to simulate the limit of complete information incorporation). Then, as an affine transformation of the composition between a weakly LSC function (relative entropy) and weakly continuous functions (backward interpolation PDFs and $g(\tau) = \tau T'$), $\delta^{\mathrm{b}}(\tau; T')$ is a weakly LSC function on $[0, 1]$ for each $T' \in [0, T]$. This weak lower semi-continuity implies that $\delta^{\mathrm{b}}(\tau; T')$ is also weakly measurable on $[0, 1]$ for each $T' \in [0, T]$ [23, 37]. Although one might consider attempting to establish that it is also almost surely separably valued in order to apply Pettis' measurability theorem and conclude strong measurability [66], this step is unnecessary in our setting. For separable spaces, weak and strong measurability coincide, and the space of Borel probability measures defined over a Polish state space, such as the one considered in this work, is itself Polish [35, 36], as established by Prokhorov's theorem [89, 39, 63]. As such, the backward CIR metric defined in (2.16) is a measurable function of $\tau \in \mathrm{I}$ for each $T' \in [0, T]$. As the pointwise limit of measurable functions, so is the backward CIR metric after the limit of complete information incorporation, i.e., the complete backward CIR metric $\lim_{T' \to T^-} \delta^{\mathrm{b}}(\tau; T')$ is a measurable function of $\tau \in \mathrm{I}$ for each $T > 0$.

Appendix B. Proof of Theorem 2.9.

Here we provide the proof of Theorem 2.9.

Proof. First, observe that $\lambda_{\mathrm{I}}(\cdot)$ defines a probability measure on $\mathrm{I} = [0, 1]$. Under the (Lebesgue) measurability assumption for $\lim_{T' \to T^-} \delta^{\mathrm{b}}(\cdot; T')$, we can then interpret it as a continuous random variable on the probability space $(\mathrm{I}, \mathcal{L}_{\mathrm{I}}, \lambda_{\mathrm{I}})$, where $\mathcal{L}_{\mathrm{I}}$ is the σ-algebra of Lebesgue measurable subsets of I, with its support being $[0, M^{\mathrm{b}}(T)]$.

Its survival function,

$$1 - \lambda_{\mathrm{I}}\Big(\Big\{\tau \in \mathrm{I} : \lim_{T' \to T^-} \delta^{\mathrm{b}}(\tau; T') \leq \varepsilon\Big\}\Big),$$

is well-defined because (a) it is nonincreasing and right-continuous with left limits, (b) it equals 1 for $\varepsilon \leq 0$, and (c) it equals zero for $\varepsilon \geq M^{\mathrm{b}}(T)$.

Applying the expected survival time formula to this nonnegative, compactly supported random variable yields:

$$\begin{aligned}\mathbb{E}\Big[\lim_{T'\to T^-}\delta^{\mathrm{b}}(\tau;T')\Big] &:= \int_0^1 \lim_{T'\to T^-}\delta^{\mathrm{b}}(\tau;T')\mathrm{d}\tau\\ &= \int_0^{M^{\mathrm{b}}(T)}\Big(1-\lambda_{\mathrm{I}}\Big(\Big\{\tau\in\mathrm{I}:\lim_{T'\to T^-}\delta^{\mathrm{b}}(\tau;T')\le\varepsilon\Big\}\Big)\Big)\mathrm{d}\varepsilon.\end{aligned}$$

The theorem's main result follows by dividing through by $M^{\mathrm{b}}(T)>0$, multiplying by T, and applying the inequality:

$$\widetilde{\boldsymbol{\tau}}^{\mathrm{b}}(T,\varepsilon)\le T\Big(1-\lambda_{\mathrm{I}}\Big(\Big\{\tau\in\mathrm{I}:\lim_{T'\to T^-}\delta^{\mathrm{b}}(\tau;T')\le\varepsilon\Big\}\Big)\Big).$$

The equality condition follows immediately from this, since equality in the above holds if and only if $\lim_{T'\to T^-}\delta^{\mathrm{b}}(\cdot;T')$ is nondecreasing in I. ■

Appendix C. Conditionally Linear Dynamics with State-Independent Feedbacks as Representatives of Complex Systems.

While the system (2.25) in Section C might appear overly simplistic, many complex stochastic nonlinear dynamical systems with multivariate components $\mathbf{x}$ and $\mathbf{y}$ can be put in such a reduced-order canonical form. In this section, we provide a brief overview of how this is possible through some practical generalizations:

- In (2.25), the noise feedbacks are purely additive or state-independent. This is without loss of generality, as transitioning from state-dependent noise feedbacks to fully additive ones is possible via the use of a suitable Lamperti transform (which adjusts the system's potential to account for the transition from multiplicative to additive noise) [60].
- In (2.25), nonexistent noise cross-interactions are assumed, i.e., W_1 does not appear in y's evolution nor does W_2 in x's equation. The case of correlated noise forcings can be easily reduced to the uncorrelated one via a whitening transformation due to the linear mean dynamics and state-independent noise forcings [70, 47].
    - ► We note here that the proof of Theorem 2.12 (see Section E) can be adapted to provide an analogous result even in the case where there are cross-interacting or colored noise feedbacks, by accounting for their correlations in the auxiliary components of the online smoother formulae (see Theorem D.1) [7].
- In (2.25), we assume x and y are one-dimensional. This is without loss of generality, as it is known that a diagonal approximation of the filter and smoother posterior covariance matrices is possible for multidimensional variants [20].
- In (2.25), y's feedbacks, both in x and in y, do not vary in time. For time-varying coefficients (that also possibly depend on x), which satisfy sufficient conditions for well-posedness, stochastic linear stability, and the existence of equilibrium filter and smoother distributions (posterior statistical attractors), then a simple linearization procedure can be utilized to retrieve similar results to those established in Theorem 2.12.

Appendix D. An Adaptive-Lag Online Smoother for CGNSs and its Use in Deriving Closed-Form Expressions for the CIR Lengths.

In this section, we briefly review the pertinent results from [7], which developed an adaptive-lag online smoother for CGNSs. In the subsequent subsections, we outline how the main theorem cited here can be applied to compute the complete and lagged smoother PDFs from Section 2.3.1 in the case of a CGNS. These distributions are required for computing the forward and backward CIR metrics from Sections 2.3.2 and 2.3.4, which, in turn, are needed for evaluating the corresponding lower- and upper-bound approximations of the objective forward and backward CIR lengths, respectively.

We first adopt a uniform discretization of the interval $[0,T]$ to transition to discrete time, with $0 = t_0 < t_1 < \cdots < t_N = T$ and $t_j = j\Delta t$, where $j \in [\![N]\!] := \{0,1,\ldots,N\}$ and $N \in \mathbb{N}$, with N large enough so that the partition norm $\Delta t = T/N$ satisfies $0 < \Delta t \ll 1$. In what follows, T is assumed to be arbitrary and allowed to increase freely in $(0,+\infty)$, modeling the successive observation of $\mathbf{x}$, with $\Delta t \ll 1$ predetermined from either the model simulation's integration time-step or observation rate. (The assumption of uniform Δt is without loss of generality [7].) Therefore, $N := \lceil T/\Delta t \rceil$ with T being sequentially incremented by Δt. Under this time discretization, both the natural (t) and lagged observational (T') time variables of δ in (2.8) take values in the discrete set $\{t_j\}_{j\in[\![N]\!]}$. We use the index j to denote discretization along natural time and n over observational time, while the superscript notation $\cdot^j$ is used to denote the discrete-time approximation of the corresponding continuous vector- or matrix-valued function evaluated at t_j; for example, $\mathbf{x}^j := \mathbf{x}(t_j)$ and $\boldsymbol{\Lambda}^{\mathbf{x},j} := \boldsymbol{\Lambda}^{\mathbf{x}}(t_j,\mathbf{x}(t_j))$. Additionally, we write $\big(\cdot\big|\mathbf{x}(s\leq n)\big)$ to indicate that we condition on the observations $\{\mathbf{x}^j\}_{j\in[\![n]\!]}$ for $n \in [\![N]\!]$. For arbitrary j, we define the following auxiliary matrices:

$$
\begin{aligned}
\mathbf{E}^j &:= \mathbf{I}_{l\times l} + \Big[(\boldsymbol{\Sigma}^{\mathbf{y}} \circ \boldsymbol{\Sigma}^{\mathbf{x}})^j\big((\boldsymbol{\Sigma}^{\mathbf{x}} \circ \boldsymbol{\Sigma}^{\mathbf{x}})^j\big)^{-1}\mathbf{G}^{\mathbf{x},j} - \mathbf{G}^{\mathbf{y},j}\Big]\Delta t \in \mathbb{R}^{l\times l},\\
\mathbf{F}^j &:= -\mathbf{R}_{\mathrm{f}}^j\Big[(\mathbf{K}^j)^{\mathtt{T}} + \Big((\mathbf{G}^{\mathbf{x},j})^{\mathtt{T}}\mathbf{K}^j\mathbf{R}_{\mathrm{f}}^j(\mathbf{K}^j)^{\mathtt{T}} - (\mathbf{R}_{\mathrm{f}}^j)^{-1}(\mathbf{H}^j)^{\mathtt{T}}\mathbf{R}_{\mathrm{f}}^j(\mathbf{K}^j)^{\mathtt{T}} + (\boldsymbol{\Lambda}^{\mathbf{y},j})^{\mathtt{T}}(\mathbf{K}^j)^{\mathtt{T}}\Big)\Delta t\\
&\qquad\qquad - (\boldsymbol{\Lambda}^{\mathbf{x},j})^{\mathtt{T}}\Big(\big((\boldsymbol{\Sigma}^{\mathbf{x}} \circ \boldsymbol{\Sigma}^{\mathbf{x}})^j\big)^{-1} + \mathbf{K}^j\mathbf{R}_{\mathrm{f}}^j(\mathbf{K}^j)^{\mathtt{T}}\Delta t\Big)\Big] \in \mathbb{R}^{l\times k},
\end{aligned}
$$

where

$$
\begin{gathered}
\mathbf{G}^{\mathbf{x},j} := \boldsymbol{\Lambda}^{\mathbf{x},j} + (\boldsymbol{\Sigma}^{\mathbf{x}} \circ \boldsymbol{\Sigma}^{\mathbf{y}})^j(\mathbf{R}_{\mathrm{f}}^j)^{-1} \in \mathbb{R}^{k\times l}, \quad \mathbf{G}^{\mathbf{y},j} := \boldsymbol{\Lambda}^{\mathbf{y},j} + (\boldsymbol{\Sigma}^{\mathbf{y}} \circ \boldsymbol{\Sigma}^{\mathbf{y}})^j(\mathbf{R}_{\mathrm{f}}^j)^{-1} \in \mathbb{R}^{l\times l},\\
\mathbf{H}^j := (\mathbf{R}_{\mathrm{f}})^{-1}\Big(\boldsymbol{\Lambda}^{\mathbf{y},j}\mathbf{R}_{\mathrm{f}}^j + \mathbf{R}_{\mathrm{f}}^j(\boldsymbol{\Lambda}^{\mathbf{y},j})^{\mathtt{T}} + (\boldsymbol{\Sigma}^{\mathbf{y}} \circ \boldsymbol{\Sigma}^{\mathbf{y}})^j\Big) \in \mathbb{R}^{l\times l},\\
\mathbf{K}^j := \big((\boldsymbol{\Sigma}^{\mathbf{x}} \circ \boldsymbol{\Sigma}^{\mathbf{x}})^j\big)^{-1}\mathbf{G}^{\mathbf{x},j} \in \mathbb{R}^{k\times l}.
\end{gathered}
$$

Under the preceding exposition, the following theorem holds, characterizing the online smoother distributions of the CGNS framework in discrete-time settings [7]. Specifically, it outlines the procedure for online estimation of the smoother PDF in a CGNS, i.e., its iterative calculation at each time instant when considering a newly obtained observation from $\mathbf{x}$.

Theorem D.1 (Optimal Online Forward-In-Time Discrete Smoother). *Let $\mathbf{x}(t)$ and $\mathbf{y}(t)$ satisfy (2.27a)–(2.27b). Then, under suitable regularity conditions, the discrete smoother (backward interpolation) distributions are Gaussian,*

$$
\mathbb{P}\big(\mathbf{y}^j\big|\mathbf{x}(s\leq n)\big) \overset{(\mathrm{d})}{\sim} \mathcal{N}_l(\boldsymbol{\mu}_s^{j,n}, \mathbf{R}_s^{j,n}),
$$

for $j \in [\![n]\!]$, $n \in [\![N]\!]$, *where*

$$\boldsymbol{\mu}_{\mathrm{s}}^{j,n} := \mathbb{E}\left[\mathbf{y}^j\middle|\mathbf{x}(s \leq n)\right], \quad \mathbf{R}_{\mathrm{s}}^{j,n} := \mathbb{E}\left[(\mathbf{y}^j - \boldsymbol{\mu}_{\mathrm{s}}^{j,n})(\mathbf{y}^j - \boldsymbol{\mu}_{\mathrm{s}}^{j,n})^{\mathtt{T}}\middle|\mathbf{x}(s \leq n)\right],$$

satisfy the following recursive backward difference equations for $j \in [\![n-2]\!]$ *and* $n \in \{2, \dots, N\}$*:*

$$\boldsymbol{\mu}_{\mathrm{s}}^{j,n} = \boldsymbol{\mu}_{\mathrm{s}}^{j,n-1} + \mathbf{D}^{j,n-2}\left(\boldsymbol{\mu}_{\mathrm{s}}^{n-1,n} - \boldsymbol{\mu}_{\mathrm{f}}^{n-1}\right), \tag{D.1a}$$

$$\mathbf{R}_{\mathrm{s}}^{j,n} = \mathbf{R}_{\mathrm{s}}^{j,n-1} + \mathbf{D}^{j,n-2}\left(\mathbf{R}_{\mathrm{s}}^{n-1,n} - \mathbf{R}_{\mathrm{f}}^{n-1}\right)(\mathbf{D}^{j,n-1})^{\mathtt{T}}, \tag{D.1b}$$

where the update matrices $\mathbf{D}^{j,n-2}$ *are defined by*

$$\mathbf{D}^{n-1,n-2} := \mathbf{I}_{l\times l} \quad \& \quad \mathbf{D}^{j,n-2} := \overset{\curvearrowright}{\prod_{i=j}^{n-2}} \mathbf{E}^i := \mathbf{E}^j\mathbf{E}^{j+1}\cdots\mathbf{E}^{n-2}, \tag{D.2}$$

while for $j = n-1$ *and* $n \in \{1, \dots, N\}$ *we have:*

$$\begin{aligned}\boldsymbol{\mu}_{\mathrm{s}}^{n-1,n} &= \mathbf{E}^{n-1}\boldsymbol{\mu}_{\mathrm{f}}^{n} + \mathbf{b}^{n-1},\\ \mathbf{R}_{\mathrm{s}}^{n-1,n} &= \mathbf{E}^{n-1}\mathbf{R}_{\mathrm{f}}^{n}(\mathbf{E}^{n-1})^{\mathtt{T}} + \mathbf{P}_{n}^{n-1},\end{aligned}$$

with the $\mathbf{b}^{n-1}$ *and* $\mathbf{P}_n^{n-1}$ *auxiliary residual terms being given as*

$$\begin{aligned}\mathbf{b}^{n-1} &:= \boldsymbol{\mu}_{\mathrm{f}}^{n-1} - \mathbf{E}^{n-1}\left((\mathbf{I}_{l\times l} + \boldsymbol{\Lambda}^{\mathbf{y},n-1}\Delta t)\boldsymbol{\mu}_{\mathrm{f}}^{n-1} + \mathbf{f}^{\mathbf{y},n-1}\Delta t\right)\\ &\qquad + \mathbf{F}^{n-1}\left(\mathbf{x}^n - \mathbf{x}^{n-1} - (\boldsymbol{\Lambda}^{\mathbf{x},n-1}\boldsymbol{\mu}_{\mathrm{f}}^{n-1} + \mathbf{f}^{\mathbf{x},n-1})\Delta t\right),\\ \mathbf{P}_n^{n-1} &:= \mathbf{R}_{\mathrm{f}}^{n-1} - \mathbf{E}^{n-1}(\mathbf{I}_{l\times l} + \boldsymbol{\Lambda}^{\mathbf{y},n-1}\Delta t)\mathbf{R}_{\mathrm{f}}^{n-1} - \mathbf{F}^{n-1}\boldsymbol{\Lambda}^{\mathbf{x},n-1}\mathbf{R}_{\mathrm{f}}^{n-1}\Delta t,\end{aligned}$$

and finally for $j = n$ *and* $n \in [\![N]\!]$ *we have* $\boldsymbol{\mu}_{\mathrm{s}}^{n,n} = \boldsymbol{\mu}_{\mathrm{f}}^{n}$ *and* $\mathbf{R}_{\mathrm{s}}^{n,n} = \mathbf{R}_{\mathrm{f}}^{n}$*, since the smoother and filter posterior Gaussian statistics coincide at the end point by definition.*

The Gaussianity of the discrete-time posterior distributions of a CGNS is proven in Chapter 13 of Liptser & Shiryaev [47] (see also Lemma A6 in [5] for the discrete-time optimal nonlinear smoother distribution proof specifically), while the proof of Theorem D.1 is given in Theorem 3.1 of [7].

In the sequel, we present the details for explicitly computing the approximations of the objective forward and backward CIR lengths in the unconditional setting (as given in (2.15) and (2.23), respectively) via analytical means. The same methodology extends naturally to conditional assimilative causal links, with the appropriate modifications.

D.1. Computation of the Forward CIR in CGNSs. From Theorem D.1, by adopting the following notation for $j \in [\![N]\!]$, $n \in \{j, j+1, \dots, N\}$:

$$p_n(\mathbf{y}^j|\mathbf{x}) \text{ corresponding to } \mathcal{N}_l(\boldsymbol{\mu}_{\mathrm{s}}^{j,n}, \mathbf{R}_{\mathrm{s}}^{j,n}), \tag{D.3}$$

we have for

$$\mathcal{P}_n^j := \mathcal{P}\big(p_N(\mathbf{y}^j|\mathbf{x}), p_n(\mathbf{y}^j|\mathbf{x})\big),$$

that:

$$\text{(D.4)}\qquad \begin{aligned}\mathcal{P}^j_n &= \frac{1}{2}\left(\boldsymbol{\mu}_{\mathrm{s}}^{j,N}-\boldsymbol{\mu}_{\mathrm{s}}^{j,n}\right)^{\mathtt{T}}\left(\mathbf{R}_{\mathrm{s}}^{j,n}\right)^{-1}\left(\boldsymbol{\mu}_{\mathrm{s}}^{j,N}-\boldsymbol{\mu}_{\mathrm{s}}^{j,n}\right)\\ &\quad+\frac{1}{2}\left(\mathrm{tr}\big(\mathbf{R}_{\mathrm{s}}^{j,N}(\mathbf{R}_{\mathrm{s}}^{j,n})^{-1}\big)-l-\log\left(\det\big(\mathbf{R}_{\mathrm{s}}^{j,N}(\mathbf{R}_{\mathrm{s}}^{j,n})^{-1}\big)\right)\right),\end{aligned}$$

due to the signal-dispersion decomposition of the relative entropy for Gaussian variables [6, 14]: The quadratic-form term in (D.4) is the signal part, while the term equal to half the Burg or log-det divergence from $\mathbf{R}_{\mathrm{s}}^{j,N}$ to $\mathbf{R}_{\mathrm{s}}^{j,n}$, is the dispersion part. Then, by the theory in Section 2.3.2 (see also SI Section B.3 of [6]), we have from (2.9) the following exact equality:

$$\text{(D.5)}\qquad \delta(t_j,t_n)=\mathcal{P}^j_n,\quad j\in[\![N]\!],\ n\in\{j,j+1,\dots,N\},$$

As such, in the case of a CGNS, we have that the discrete-time subjective forward CIR length of $\mathbf{y}(t_j)\to\mathbf{x}$ for the threshold parameter ε is given by (see (2.11)):

$$\text{(D.6)}\qquad \widetilde{\boldsymbol{\tau}}^{\mathrm{f}}(t_j,\varepsilon)=\max_{n\in\{j,\dots,N\}}\left\{\mathcal{P}^j_n>\varepsilon\right\}\Delta t-t_j\in[0,T-t_j],$$

which subsequently defines its objective discrete-time counterpart via (2.13), by averaging (D.6) over ε. In addition, we have that the lower-bound approximation of the associated objective forward CIR length in the case of discrete time for a CGNS, is given by (see (2.15)):

$$\text{(D.7)}\qquad [0,T-t_j]\ni\frac{\big\|\mathcal{P}^j_{\cdot}\big\|_{L^1(\{j,\dots,N\})}}{\big\|\mathcal{P}^j_{\cdot}\big\|_{L^\infty(\{j,\dots,N\})}}=\boldsymbol{\tau}^{\mathrm{f}}_{\mathrm{approx}}(t_j)\le\boldsymbol{\tau}^{\mathrm{f}}(t_j),$$

where for $j\in[\![N]\!]$ and a bounded grid function $g_n=g(t_n)$ defined on the equidistant grid points $\{t_n\}_{n\in\{j,j+1,\dots,N\}}$ (uniform partition of $[0,T]$ with mesh size of Δt), its L^1 and L^∞ norms on $\{j,j+1,\dots,N\}$ are defined as [43]:

$$\begin{aligned}\|g\|_{L^1(\{j,\dots,N\})}&=\Delta t\sum_{n=j}^{N}|g(t_n)|,\\ \|g\|_{L^\infty(\{j,\dots,N\})}&=\max_{n\in\{j,\dots,N\}}\{|g(t_n)|\},\end{aligned}$$

essentially approximating the analytic Riemann integral with its associated right Riemann sum.

D.2. Computation of the Backward CIR in CGNSs. From Theorem D.1, for Δt sufficiently small ($\Delta t\ll 1$), by adopting the following notation for $j\in[\![N]\!]$ (see also definitions of complete and lagged smoother PDFs in Section 2.3.1):

$$\text{(D.8)}\qquad \begin{aligned}p_T^{\mathrm{upd}}(\mathbf{y}^j|\mathbf{x})&\text{ corresponding to }\mathcal{N}_l(\boldsymbol{\mu}_{\mathrm{s}}^{j,N},\mathbf{R}_{\mathrm{s}}^{j,N})\ (n=N;\ \text{complete smoother}),\\ p_T^{\mathrm{lag}}(\mathbf{y}^j|\mathbf{x})&\text{ corresponding to }\mathcal{N}_l(\boldsymbol{\mu}_{\mathrm{s}}^{j,N-1},\mathbf{R}_{\mathrm{s}}^{j,N-1})\ (n=N-1;\ \text{lagged smoother}),\end{aligned}$$

where we have used $T' = T - \Delta t$ for the lagged smoother, we have for

$$\mathcal{P}_T^j := \mathcal{P}\big(p_T^{\text{upd}}(\mathbf{y}^j|\mathbf{x}), p_T^{\text{lag}}(\mathbf{y}^j|\mathbf{x})\big), \quad T > 0,$$

that [7]:

$$\begin{aligned}\mathcal{P}_T^j = {}& \frac{1}{2}\left(\boldsymbol{\mu}_{\text{s}}^{N-1,N} - \boldsymbol{\mu}_{\text{f}}^{N-1}\right)^{\mathtt{T}}(\mathbf{D}^{j,N-2})^{\mathtt{T}}(\mathbf{R}_{\text{s}}^{j,N-1})^{-1}\mathbf{D}^{j,N-2}\left(\boldsymbol{\mu}_{\text{s}}^{N-1,N} - \boldsymbol{\mu}_{\text{f}}^{N-1}\right) \\ & + \frac{1}{2}\left(\operatorname{tr}(\mathbf{Q}^{j,N}) - l - \log(\det(\mathbf{Q}^{j,N}))\right),\end{aligned} \tag{D.9}$$

due to the signal-dispersion decomposition of the relative entropy for Gaussian variables [6, 14], where $\mathbf{Q}^{j,N}$ is defined as the covariance ratio matrix [7]:

$$\begin{aligned}\mathbf{Q}^{j,N} &:= \left(\mathbf{R}_{\text{s}}^{j,N-1} + \mathbf{D}^{j,N-2}\left(\mathbf{R}_{\text{s}}^{N-1,N} - \mathbf{R}_{\text{f}}^{N-1}\right)(\mathbf{D}^{j,N-2})^{\mathtt{T}}\right)(\mathbf{R}_{\text{s}}^{j,N-1})^{-1} \\ &= \mathbf{I}_{l\times l} + \left(\mathbf{D}^{j,N-2}\left(\mathbf{R}_{\text{s}}^{N-1,N} - \mathbf{R}_{\text{f}}^{N-1}\right)(\mathbf{D}^{j,N-2})^{\mathtt{T}}\right)(\mathbf{R}_{\text{s}}^{j,N-1})^{-1}.\end{aligned} \tag{D.10}$$

(The dependence of $\mathcal{P}_T^j$ on T is explicitly noted here for clarity.) Then, by the theory in Section 2.3.4, we have from (2.16) the following approximation:

$$\lim_{T'\to T^-} \delta(t_j, T') \approx \mathcal{P}_T^j, \quad j \in [\![N]\!]. \tag{D.11}$$

Note that this is an approximation, unlike the equality in (D.5), as this result is contingent on $t_N - t_{N-1} = \Delta t = T/N \ll 1$ being sufficiently small so that the limit of complete information incorporation, $T' \to T^-$, is sufficiently well approximated in discrete time. As such, in the case of a CGNS, we have that the discrete-time subjective backward CIR length of $\mathbf{y} \to \mathbf{x}(T)$ for the threshold parameter $\varepsilon \geq 0$ can be approximated as (see (2.19)):

$$\tilde{\boldsymbol{\tau}}^{\text{b}}(T, \varepsilon) \approx T - \sup_{j\in[\![N]\!]}\left\{\left|\mathcal{P}_T^j - \mathcal{P}_T^0\right| \leq \varepsilon\right\}\Delta t \in [0, T], \tag{D.12}$$

which subsequently approximates its objective discrete-time counterpart via (2.21), by averaging (D.12) over ε. Similarly, we have that the upper-bound approximation of the associated objective backward CIR length in the case of discrete time for a CGNS, is in turn approximated as (see (2.23) and Theorem 2.9):

$$\boldsymbol{\tau}^{\text{b}}(T) \leq \boldsymbol{\tau}^{\text{b}}_{\text{approx}}(T) = \frac{\left\|\mathcal{P}_T^{\cdot} - \mathcal{P}_T^0\right\|_{L^1([\![N]\!])}}{\left\|\mathcal{P}_T^{\cdot} - \mathcal{P}_T^0\right\|_{L^\infty([\![N]\!])}} \in [0, T], \tag{D.13}$$

where for $j \in [\![N]\!]$ and a bounded grid function $g^j = g(t_j)$ defined on the equidistant grid points $\{t_j\}_{j\in[\![N]\!]}$ (uniform partition of $[0, T]$ with mesh size of Δt), its L^1 and L^∞ norms on $[\![N]\!]$ are defined as [43]:

$$\begin{aligned}\|g\|_{L^1([\![N]\!])} &= \Delta t \sum_{j=0}^{N} |g(t_j)|, \\ \|g\|_{L^\infty([\![N]\!])} &= \max_{j\in[\![N]\!]}\{|g(t_j)|\},\end{aligned}$$

essentially approximating the analytic Riemann integral with its associated right Riemann sum.

Appendix E. Proof of Theorem 2.12.

Here we provide the proof of Theorem 2.12, utilizing the results from Section D.2.

Proof. Under the conditionally linear structure with additive noise feedbacks in (2.25), and assumptions which establish linear stochastic stability, the posterior distributions of filter and smoother enjoy a statistical attractor, where the filter and smoother equilibrium distributions are necessarily Gaussian [20, 77, 21, 30, 76]. The steady-state algebraic equations for the equilibrium filter (algebraic Riccati equation) and smoother (algebraic symmetric Sylvester equation) covariance matrices [47, 1], denoted by R_{f}^{∞} and R_{s}^{∞} respectively, yield [7, 20, 77, 21, 30, 76]:

$$\begin{aligned} R_{\mathrm{f}}^{\infty} &= \frac{\lambda^y(\sigma^x)^2 + \sigma^x\sqrt{\Psi}}{(\lambda^x)^2}, \\ R_{\mathrm{s}}^{\infty} &= \frac{(\sigma^y)^2}{2\big((\lambda^y)^2 + (\sigma^y)^2/R_{\mathrm{f}}^{\infty}\big)}, \end{aligned}$$

where $\Psi := (\lambda^y\sigma^x)^2 + (\lambda^x\sigma^y)^2$. (For one-dimensional x and y these are scalar quantities.) As such, as T grows unboundedly (and by extension N, the current number of observations for x, per our setup from the beginning of Section D), we have that R_{f}^{N}, $R_{\mathrm{s}}^{N-1,N} = O(1)$. This extends to the filter and smoother means, with μ_{f}^{N}, $\mu_{\mathrm{s}}^{N-1,N} = \Theta(1)$ if f^x or f^y are periodic and bounded, which does not alter the assertion of this theorem.

Note that the system in (2.25) is a CGNS (see (2.27)), therefore the forward and backward variants of the CIR metric, (2.9) and (2.16) respectively, enjoy a Gaussian signal-dispersion decomposition [6, 14], as illustrated in Sections D.1 and D.2. Therefore, we can then make the following assertion for the backward one (recall the results from Section D.2, specifically (D.9)–(D.11)):

(E.1)
$$\begin{aligned} \lim_{T'\to T^-}\delta(t_j,T') &\approx \mathcal{P}_T^j = \mathcal{P}\big(p_T^{\mathrm{upd}}(\mathbf{y}^j|\mathbf{x}), p_T^{\mathrm{lag}}(\mathbf{y}^j|\mathbf{x})\big) \\ &\approx \frac{1}{2}\big(\mu_{\mathrm{s}}^{N-1,N} - \mu_{\mathrm{f}}^{N-1}\big)^2\big(D^{j,N-2}\big)^2\big(R_{\mathrm{s}}^{j,N-1}\big)^{-1} + \frac{1}{2}\big(Q^{j,N} - 1 - \log(Q^{j,N})\big), \end{aligned}$$

for each $j \in [\![N]\!]$, where for $T \gg 1$ (and $N \gg 1$) from (D.10) we have

$$Q^{j,N} \approx 1 + \frac{R_{\mathrm{s}}^{\infty} - R_{\mathrm{f}}^{\infty}}{R_{\mathrm{s}}^{j,\infty}}(D^{j,\infty})^2,$$

since in this scenario we have $R_{\mathrm{f}}^{N-1} \to R_{\mathrm{f}}^{\infty}$ and $R_{\mathrm{s}}^{N-1,N} \to R_{\mathrm{s}}^{\infty}$, which exist and are finite due to the existence of the filter and smoother posterior attractors, respectively. Therefore, it suffices to asymptotically analyze the update operators $D^{j,N-2}$ for $j \in [\![N]\!]$ as $T, N \gg 1$. For that, as already stated at the beginning of Section D, we assume Δt, defined either by the model simulation's numerical integration time step or the temporal rate at which observations of x are obtained, is sufficiently small ($\Delta t \ll 1$). For a continuously observed x, this is certainly true through the formal limit $\Delta t \to 0^+$.

From Theorem D.1, we observe the following for each $j \in [\![N]\!]$ when T, $N \gg 1$:

$$D^{j,N-2} = \prod_{i=j}^{N-2}(1 - G^{y,i}\Delta t) = O(1)(1 - G^{y,\infty}\Delta t)^{N-j}, \tag{E.2}$$

where [7]:

$$G^{y,\infty} = \frac{\sqrt{\Psi}}{(\lambda^x)^2 R_{\mathrm{f}}^{\infty}}(\sqrt{\Psi} + \lambda^y\sigma^x) > 0.$$

Recalling that we assume $\Delta t \ll 1$, Taylor expansion manipulations reveal that for large N we have

$$(D^{j,N-2})^2 = \Theta\big(e^{2G^{y,\infty}(j-N)\Delta t}\big) = \Theta\big(e^{2G^{y,\infty}(j\Delta t - T)}\big).$$

Furthermore, due to the existence of the filter and smoother equilibrium Gaussian statistics, it is immediate that

$$\big(O(1) =\big)\, \mu_{\mathrm{s}}^{N-1,N} - \mu_{\mathrm{f}}^{N-1} \to \mu_{\mathrm{s}}^{\infty} - \mu_{\mathrm{f}}^{\infty},$$

as $\mu_{\mathrm{s}}^{\infty} - \mu_{\mathrm{f}}^{\infty}$ exists and is finite (see Theorem D.1). As such, for each $j \in [\![N]\!]$ when T, $N \gg 1$, combining the previous results and specifically plugging (E.2) in (E.1) yields:

$$\lim_{T' \to T^-} \delta(t_j, T') \approx \mathcal{P}_T^j = O(1)\left(\Theta\big(e^{2G^{y,\infty}(j\Delta t - T)}\big)R_{\mathrm{s}}^{j,\infty} + Q^{j,N} - 1 - \log(Q^{j,N})\right), \tag{E.3}$$

with

$$Q^{j,N} = 1 + \frac{O(1)}{R_{\mathrm{s}}^{j,\infty}}\Theta\big(e^{2G^{y,\infty}(j\Delta t - T)}\big). \tag{E.4}$$

Since we consider $T \gg 1$, it is immediate from (E.4) that the information gain in (E.3) is predominantly signal-driven, i.e., the dispersion component is essentially negligible; see the case study in Section 3.1 for a numerical validation of this statement. Therefore, it follows that,

$$\mathcal{P}_T^j - \mathcal{P}_T^0 \approx O(1)\Theta\big(e^{2G^{y,\infty}(j\Delta t - T)}\big)R_{\mathrm{s}}^{j,\infty}. \tag{E.5}$$

Recalling the assumption that (2.25) enjoys linear stochastic stability, meaning its linear dynamics ensure geometric ergodicity [59, 22], the online smoother variance quickly relaxes towards the smoother equilibrium variance in j: $R_{\mathrm{s}}^{j,\infty} \to R_{\mathrm{s}}^{\infty}$. As a result, from this, (E.5), (D.13), and

$$0 < \frac{\displaystyle\int_0^T e^{r(t-T)}\mathrm{d}t}{\displaystyle\max_{t\in[0,T]}\big\{e^{r(t-T)}\big\}} = \frac{1}{r} - \frac{e^{-Tr}}{r},$$

for $r > 0$ arbitrary, then for T large enough we finally recover that

$$\frac{1}{r_1} \leq \tau^{\mathrm{b}}_{\mathrm{approx}}(T) \leq \frac{1}{r_2}, \quad T \gg 1,$$

for some r_1, $r_2 > 0$ which do not depend on T. ■

REFERENCES

[1] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank, *Matrix Riccati Equations in Control and Systems Theory*, Birkhäuser Basel, 2003, https://doi.org/10.1007/978-3-0348-8081-7, http://dx.doi.org/10.1007/978-3-0348-8081-7.

[2] R. V. Abramov and A. J. Majda, *A New Algorithm for Low-Frequency Climate Response*, Journal of the Atmospheric Sciences, 66 (2009), pp. 286–309.

[3] S. M. Ali and S. D. Silvey, *A general class of coefficients of divergence of one distribution from another*, Journal of the Royal Statistical Society: Series B (Methodological), 28 (1966), pp. 131–142.

[4] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford university press, 2000.

[5] M. Andreou and N. Chen, *A Martingale-Free Introduction to Conditional Gaussian Nonlinear Systems*, Entropy, 27 (2024), p. 2.

[6] M. Andreou, N. Chen, and E. Bollt, *Assimilative Causal Inference*, 2025, https://arxiv.org/abs/2505.14825, https://arxiv.org/abs/2505.14825.

[7] M. Andreou, N. Chen, and Y. Li, *An Adaptive Online Smoother with Closed-Form Solutions and Information-Theoretic Lag Selection for Conditional Gaussian Nonlinear Systems*, arXiv preprint arXiv:2411.05870, (2024).

[8] J. D. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, NJ, Dec. 2008.

[9] L. Barnett, A. B. Barrett, and A. K. Seth, *Granger causality and transfer entropy are equivalent for Gaussian variables*, Physical review letters, 103 (2009), p. 238701.

[10] A. Berger, *Milankovitch Theory and climate*, Reviews of Geophysics, 26 (1988), p. 624–657, https://doi.org/10.1029/rg026i004p00624, http://dx.doi.org/10.1029/RG026i004p00624.

[11] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer New York, 2011, https://doi.org/10.1007/978-0-387-70914-7, http://dx.doi.org/10.1007/978-0-387-70914-7.

[12] P. Brouillard, S. Lachapelle, A. Lacoste, S. Lacoste-Julien, and A. Drouin, *Differentiable Causal Discovery from Interventional Data*, Advances in Neural Information Processing Systems, 33 (2020), pp. 21865–21877.

[13] V. Brovkin, E. Brook, J. W. Williams, S. Bathiany, T. M. Lenton, M. Barton, R. M. DeConto, J. F. Donges, A. Ganopolski, J. McManus, et al., *Past abrupt changes, tipping points and cascading impacts in the Earth system*, Nature Geoscience, 14 (2021), pp. 550–558.

[14] D. Cai, R. Kleeman, and A. Majda, *A Mathematical Framework for Quantifying Predictability Through Relative Entropy*, Methods and Applications of Analysis, 9 (2002), p. 425–444, https://doi.org/10.4310/maa.2002.v9.n3.a8, http://dx.doi.org/10.4310/MAA.2002.v9.n3.a8.

[15] J. Casado, *Noise-induced coherence in an excitable system*, Physics Letters A, 235 (1997), pp. 489–492.

[16] A. Charakopoulos, G. Katsouli, and T. Karakasidis, *Dynamics and causalities of atmospheric and oceanic data identified by complex networks and Granger causality analysis*, Physica A: Statistical Mechanics and its Applications, 495 (2018), pp. 436–453.

[17] N. Chen, *Stochastic Methods for Modeling and Predicting Complex Dynamical Systems: Uncertainty Quantification, State Estimation, and Reduced-Order Models*, Springer Nature Switzerland, 2025, https://doi.org/10.1007/978-3-031-81924-7, http://dx.doi.org/10.1007/978-3-031-81924-7.

[18] N. Chen, Y. Li, and H. Liu, *Conditional Gaussian nonlinear system: A fast preconditioner and a cheap surrogate model for complex nonlinear systems*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 32 (2022).

[19] N. Chen and A. J. Majda, *Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification*, Entropy, 20 (2018), p. 509.

[20] N. Chen, A. J. Majda, and X. T. Tong, *Information barriers for noisy Lagrangian tracers in filtering random incompressible flows*, Nonlinearity, 27 (2014), p. 2133.

[21] N. Chen, A. J. Majda, and X. T. Tong, *Noisy Lagrangian Tracers for Filtering Random Rotating Compressible Flows*, Journal of Nonlinear Science, 25 (2015), pp. 451–488, https://doi.org/10.1007/s00332-014-9226-5, https://doi.org/10.1007/s00332-014-9226-5.

[22] N. Chen, A. J. Majda, and X. T. Tong, *Rigorous Analysis for Efficient Statistically Accurate Algorithms for Solving Fokker–Planck Equations in Large Dimensions*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 1198–1223.

[23] J. Diestel and J. Uhl, *Vector Measures*, Mathematical surveys and monographs, American Mathematical Society, 1977, https://books.google.com/books?id=EQFjD90fXWAC.

[24] P. Embrechts and M. Hofert, *A note on generalized inverses*, Mathematical Methods of Operations Research, 77 (2013), pp. 423–432.

[25] L. Evans, *Partial Differential Equations*, American Mathematical Society, Mar. 2010, https://doi.org/10.1090/gsm/019, http://dx.doi.org/10.1090/gsm/019.

[26] F. Falasca, P. Perezhogin, and L. Zanna, *Data-driven dimensionality reduction and causal inference for spatiotemporal climate fields*, Physical Review E, 109 (2024), p. 044202.

[27] C. Feng, H. Wang, X. M. Tu, and J. Kowalski, *A Note on Generalized Inverses of Distribution Function and Quantile Transformation*, Applied Mathematics, 3 (2012), pp. 2098–2100.

[28] R. Ferrari and P. Cessi, *Seasonal synchronization in a chaotic ocean–atmosphere model*, Journal of climate, 16 (2003), pp. 875–881.

[29] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, *Stochastic resonance*, Reviews of modern physics, 70 (1998), p. 223.

[30] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics, Springer Berlin Heidelberg, 2009, https://books.google.com/books?id=epf8sgEACAAJ.

[31] M. Ghil and S. Childress, *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics*, vol. 60, Springer Science & Business Media, 2012.

[32] C. W. Granger, *Investigating causal relations by econometric models and cross-spectral methods*, Econometrica: journal of the Econometric Society, (1969), pp. 424–438.

[33] C. W. Granger, *Testing for causality: A personal viewpoint*, Journal of Economic Dynamics and control, 2 (1980), pp. 329–352.

[34] D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik, *Topological Causality in Dynamical Systems*, Physical review letters, 119 (2017), p. 098301.

[35] T. (https://mathoverflow.net/users/5268/thanh), *Polish spaces in probability*. MathOverflow, https://mathoverflow.net/q/20919, https://arxiv.org/abs/https://mathoverflow.net/q/20919. URL:https://mathoverflow.net/q/20919 (version: 2019-02-22).

[36] M. S. (https://math.stackexchange.com/users/8297/martin sleziak), *Why do we study polish spaces?* Mathematics Stack Exchange, https://math.stackexchange.com/q/776638, https://arxiv.org/abs/https://math.stackexchange.com/q/776638. URL:https://math.stackexchange.com/q/776638 (version: 2014-05-01).

[37] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces*, Springer International Publishing, 2016, https://doi.org/10.1007/978-3-319-48520-1, http://dx.doi.org/10.1007/978-3-319-48520-1.

[38] T. Kämpke and F. J. Radermacher, *The Generalized Inverse of Distribution Functions*, in Income Modeling and Balancing: A Rigorous Treatment of Distribution Patterns, Springer, 2014, pp. 9–28.

[39] A. S. Kechris, *Classical Descriptive Set Theory*, Springer New York, 1995, https://doi.org/10.1007/978-1-4612-4190-4, http://dx.doi.org/10.1007/978-1-4612-4190-4.

[40] R. Kleeman, *Information Theory and Dynamical System Predictability*, Entropy, 13 (2011), p. 612–649, https://doi.org/10.3390/e13030612, http://dx.doi.org/10.3390/e13030612.

[41] C. Kuehn, *A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics*, Physica D: Nonlinear Phenomena, 240 (2011), pp. 1020–1035.

[42] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, *Tipping elements in the Earth's climate system*, Proceedings of the national Academy

of Sciences, 105 (2008), pp. 1786–1793.
[43] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, Society for Industrial and Applied Mathematics, Jan. 2007, https://doi.org/10.1137/1.9780898717839, http://dx.doi.org/10.1137/1.9780898717839.
[44] X. S. Liang and R. Kleeman, *Information Transfer between Dynamical System Components*, Physical review letters, 95 (2005), p. 244101.
[45] M. Liero, A. Mielke, and G. Savaré, *Optimal Entropy-Transport problems and a new Hellinger–Kantorovich distance between positive measures*, Inventiones mathematicae, 211 (2018), pp. 969–1117.
[46] B. Lindner, J. Garcia-Ojalvo, A. Neiman, and L. Schimansky-Geier, *Effects of noise in excitable systems*, Physics reports, 392 (2004), pp. 321–424.
[47] R. S. Liptser and A. N. Shiryaev, *Statistics of random processes I: General theory & Statistics of Random Processes II: Applications*, Springer Science & Business Media, 2001.
[48] E. N. Lorenz, *Irregularity: A fundamental property of the atmosphere*, Tellus A, 36 (1984), pp. 98–110.
[49] E. N. Lorenz, *Can chaos and intransitivity lead to interannual variability?*, Tellus a, 42 (1990), pp. 378–389.
[50] V. Lucarini and M. D. Chekroun, *Detecting and Attributing Change in Climate and Complex Systems: Foundations, Green's Functions, and Nonlinear Fingerprints*, Physical Review Letters, 133 (2024), p. 244201.
[51] A. Majda and X. Wang, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*, Cambridge University Press, 2006.
[52] A. J. Majda and N. Chen, *Model error, information barriers, state estimation and prediction in complex multiscale systems*, Entropy, 20 (2018), p. 644.
[53] A. J. Majda, C. Franzke, and D. Crommelin, *Normal forms for reduced stochastic climate models*, Proceedings of the National Academy of Sciences, 106 (2009), pp. 3649–3653.
[54] A. J. Majda, C. Franzke, and B. Khouider, *An Applied Mathematics Perspective on Stochastic Modelling for Climate*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366 (2008), pp. 2427–2453.
[55] A. J. Majda, B. Gershgorin, and Y. Yuan, *Low-frequency climate response and fluctuation–dissipation theorems: Theory and practice*, Journal of the Atmospheric Sciences, 67 (2010), pp. 1186–1201.
[56] A. J. Majda, I. Timofeyev, and E. Vanden Eijnden, *Models for stochastic climate prediction*, Proceedings of the National Academy of Sciences, 96 (1999), pp. 14687–14691.
[57] A. J. Majda, I. Timofeyev, and E. Vanden Eijnden, *A mathematical framework for stochastic climate models*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 54 (2001), pp. 891–974.
[58] N. J. Mantua and S. R. Hare, *The Pacific decadal oscillation*, Journal of oceanography, 58 (2002), pp. 35–44.
[59] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic processes and their applications, 101 (2002), pp. 185–232.
[60] J. K. Møller and H. Madsen, *From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform*, tech. report, Technical University of Denmark, DTU Informatics, Building 321, 2010.
[61] D. Mønster, R. Fusaroli, K. Tylén, A. Roepstorff, and J. F. Sherson, *Causal inference from noisy time-series data—testing the convergent cross-mapping algorithm in the presence of noise and external influence*, Future Generation Computer Systems, 73 (2017), pp. 52–62.
[62] S. L. Morgan and C. Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge University Press, Nov. 2014, https://doi.org/10.1017/cbo9781107587991, http://dx.doi.org/10.1017/CBO9781107587991.
[63] K. R. Parthasarathy, *Probability measures on metric spaces*, vol. 352, American Mathematical Society, 2005.
[64] J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Sept. 2009, https://doi.org/10.1017/cbo9780511803161, http://dx.doi.org/10.1017/CBO9780511803161.

[65] R. Penrose, *Techniques in Differential Topology in Relativity*, Society for Industrial and Applied Mathematics, Jan. 1972, https://doi.org/10.1137/1.9781611970609, http://dx.doi.org/10.1137/1.9781611970609.

[66] B. J. Pettis, *On integration in vector spaces*, Transactions of the American Mathematical Society, 44 (1938), p. 277–304, https://doi.org/10.1090/s0002-9947-1938-1501970-8, http://dx.doi.org/10.1090/S0002-9947-1938-1501970-8.

[67] Y. Polyanskiy and Y. Wu, *Information theory: From coding to learning*, Cambridge university press, 2024.

[68] P. J. Roebber, *Climate variability in a low-order coupled atmosphere-ocean model*, Tellus A, 47 (1995), pp. 473–494.

[69] K. J. Rothman and S. Greenland, *Causation and Causal Inference in Epidemiology*, American journal of public health, 95 (2005), pp. S144–S150.

[70] B. L. Rozovsky and S. V. Lototsky, *Stochastic Evolution Systems: Linear Theory and Applications to Non-Linear Filtering*, vol. 89, Springer Science & Business Media, 2018.

[71] D. B. Rubin, *Estimating causal effects of treatments in randomized and nonrandomized studies*, Journal of educational Psychology, 66 (1974), p. 688.

[72] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, *Detecting and quantifying causal associations in large nonlinear time series datasets*, Science advances, 5 (2019), p. eaau4996.

[73] A. Rupe, D. DeSantis, C. Bakker, P. Kooloth, and J. Lu, *Causal Discovery in Nonlinear Dynamical Systems using Koopman Operators*, arXiv preprint arXiv:2410.10103, (2024).

[74] R. Salmon, *Lectures on Geophysical Fluid Dynamics*, Oxford University Press, 1998.

[75] R. Sard, *Relativistic Mechanics: Special Relativity and Classical Particle Dynamics*, Lecture notes and supplements in physics, W.A. Benjamin, 1970, https://books.google.com/books?id=RFrvAAAAMAAJ.

[76] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, vol. 10, Cambridge University Press, 2019.

[77] S. Särkkä and L. Svensson, *Bayesian Filtering and Smoothing*, vol. 17, Cambridge university press, 2023.

[78] B. Scholkopf, J. Peters, and D. Janzing, *Elements of Causal Inference: Foundations and Learning Algorithms*, Adaptive Computation and Machine Learning series, MIT Press, London, England, Nov. 2017.

[79] T. Schreiber, *Measuring information transfer*, Physical review letters, 85 (2000), p. 461.

[80] L. Shams and U. Beierholm, *Bayesian causal inference: A unifying neuroscience theory*, Neuroscience & Biobehavioral Reviews, 137 (2022), p. 104619.

[81] K. Shanmugam, M. Kocaoglu, A. G. Dimakis, and S. Vishwanath, *Learning Causal Graphs with Small Interventions*, Advances in Neural Information Processing Systems, 28 (2015).

[82] J. Shi and B. Norgeot, *Learning Causal Effects From Observational Data in Healthcare: A Review and Summary*, Frontiers in Medicine, 9 (2022), p. 864882.

[83] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, *Detecting causality in complex ecosystems*, science, 338 (2012), pp. 496–500.

[84] J. Sun and E. M. Bollt, *Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings*, Physica D: Nonlinear Phenomena, 267 (2014), pp. 49–57.

[85] J. Sun, D. Taylor, and E. M. Bollt, *Causal network inference by optimal causation entropy*, SIAM Journal on Applied Dynamical Systems, 14 (2015), pp. 73–106.

[86] J. M. T. Thompson and J. Sieber, *Predicting climate tipping as a noisy bifurcation: a review*, International Journal of Bifurcation and Chaos, 21 (2011), pp. 399–423.

[87] S. Thual, A. J. Majda, and S. N. Stechmann, *A Stochastic Skeleton Model for the MJO*, Journal of the Atmospheric Sciences, 71 (2014), pp. 697–715.

[88] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics*, Cambridge University Press, 2017.

[89] O. Van Gaans, *Probability measures on metric spaces*, Lecture notes, (2003).

[90] L. van Veen, *Baroclinic flow and the Lorenz-84 model*, International journal of Bifurcation and Chaos, 13 (2003), pp. 2117–2139.

[91] S. Varambally, Y.-A. Ma, and R. Yu, *Discovering Mixtures of Structural Causal Models from Time Series Data*, arXiv preprint arXiv:2310.06312, (2023).

[92] S. Weichwald and J. Peters, *Causality in cognitive neuroscience: concepts, challenges, and distributional robustness*, Journal of Cognitive Neuroscience, 33 (2021), pp. 226–247.
[93] N. Wiener, *The theory of prediction*, in Modern Mathematics for Engineers, E. Beckenbach, ed., First Series, McGraw-Hill, New York, 1956, ch. 8.
[94] D. D. Zhang, H. F. Lee, C. Wang, B. Li, Q. Pei, J. Zhang, and Y. An, *The causality analysis of climate change and large-scale human crisis*, Proceedings of the National Academy of Sciences, 108 (2011), pp. 17296–17301.