

On Causal Discovery with Convergent Cross Mapping

Kurt Butler, *Student Member, IEEE*, Guanchao Feng, and Petar M. Djurić, *Life Fellow, IEEE*

Convergent cross mapping is a principled causal discovery technique for signals, but its efficacy depends on a number of assumptions about the systems that generated the signals. We present a self-contained introduction to the theory of causality in state-space models, Takens' theorem, and cross maps, and we propose conditions to check if a signal is appropriate for cross mapping. Further, we propose simple analyses based on Gaussian processes to test for these conditions in data. We show that our proposed techniques detect when convergent cross mapping may conclude erroneous results using several examples from the literature, and we comment on other considerations that are important when applying methods such as convergent cross mapping.

Index Terms—attractors, causality, convergent cross mapping, nonlinear systems, state space reconstruction

I. INTRODUCTION

The inference of causation from dynamic systems is an important task for engineers and scientists alike. A major challenge here is to discover causation using *observational* methods, where experimenters need not perturb or intervene upon the system under study [29]. This is especially relevant in fields such as ecology, finance, and neuroscience where controlled interventions could be costly, dangerous or infeasible. One of the first notions of causality for signals was proposed by Wiener [48], who argued that causation could be inferred if the history of the cause provides unique information about the future effect. Wiener's ideas were later made practical and popularized by Granger [14]. Granger causality, and related approaches [38], are especially effective in stochastic systems, but several authors have noted the challenges of applying Granger causality to systems with nonlinear or deterministic dynamics [7], [27], [41], and notably, systems with attractors [39]. Thus, Granger causality may be inadequate to analyze signals that arise from many physical systems, and this has motivated the exploration of alternative approaches.

Sugihara et al. proposed the convergent cross mapping (CCM) algorithm to study causality in systems with deterministic dynamics [41]. CCM uses a technique known as state-space reconstruction (SSR) to reconstruct the latent states that generated a signal, and then tests for causality by checking for the existence of a cross map, a mapping between reconstructed state-spaces that detects if two signals were generated by the same system. The CCM technique has been applied in several fields, including ecology [41], chemical engineering [26], informatics [25], cardiocartography [11], and neuroscience [17], [43].

Cross mapping, as a tool to detect synchronization and interdependence, has been known in the chaos theory community since the 1990s [35], where particular interest has been given to applications to neuroscience [2], [37], [32]. This work is comparatively unfamiliar, but relevant, to the signal processing community. While cross mapping methods like CCM can detect causation in cases where Granger causality fails, a number of concerns have been raised about the cross map

approach [3], [51], [21]. Particularly, cross mapping implicitly makes a number of assumptions which are potentially violated during blind use of the method. However, the CCM may remain a useful tool for observational causal inference when the user is aware of these issues and carefully checks their assumptions.

In this work, we make the following contributions:

- 1) We provide a tutorial on SSR, cross maps and CCM for the signal processing community.
- 2) We identify conditions to provide evidence that the observed signals were generated by a deterministic system.
- 3) We propose Gaussian process-based analyses to test for these conditions in data.
- 4) We discuss the caveats of CCM and subtleties of its application.
- 5) We validate our proposed methods on several examples where the ground truth is known.

We organize the rest of the paper as follows. In Section II, we introduce the mathematical formalism behind cross map theory. In Section III, we discuss CCM, and the caveats and assumptions of the approach. We then propose analyses to check these assumptions in Section IV. Finally, we simulate several systems and show that the failure to satisfy our conditions may anticipate certain failures of CCM.

II. BACKGROUND

This section is intended to be a self-contained introduction to causality, attractors, SSR and cross maps. We also introduce Gaussian process regression, which will be our tool-of-choice for learning smooth functions from data.

A. Causality in state-spaces

For us, a state space model of a system consists of a latent state $\mathbf{x} \in \mathbb{R}^D$, an observation $a \in \mathbb{R}^1$ that is a function of the latent state, and equations describing how these quantities evolve in time. In conventional signal processing and control, the observation signal can be multivariate, but one dimension will suffice for this discussion and simplify the notation later. Given an initial state $\mathbf{x}_0$, the system may evolve in discrete time by iterating a function

$$\mathbf{x}_{t+1} = F_x(\mathbf{x}_t), \quad (1)$$

The authors thank the support of NSF under Awards 2021002 and 2212506. Kurt Butler is supported by the U.S. Department of Education under the GAANN Fellowship.

The authors are with the Department of Electrical and Computer Engineering at Stony Brook University, Stony Brook, NY 11794-2350 USA.

where t is a discrete time index or in continuous time according to a differential equation

$$\dot{\mathbf{x}} \triangleq \frac{d\mathbf{x}}{dt} = F_x(\mathbf{x}), \quad (2)$$

where in either case we require F_x to be a smooth function. We can represent continuous time systems as discrete time systems by sampling the signals and approximating the evolution of the differential equation by a function as in (1), and generally this case will be our focus throughout the paper. The observation signal a is assumed to be a function of the system state, i.e.,

$$a_t = f(\mathbf{x}_t) + \varepsilon_t, \quad (3)$$

for some smooth function f . The term ε_t represents zero-mean additive noise. We will assume that our systems are **deterministic**, meaning that the latent state $\mathbf{x}_t$ is not a random quantity after specifying an initial condition $\mathbf{x}_0$.

We say that $\mathbf{x}$ **causes** $\mathbf{y}$, or $\mathbf{x}$ drives $\mathbf{y}$, if $\mathbf{y}_{t+1}$ depends on $\mathbf{x}_t$ in the dynamical equation,

$$\mathbf{y}_{t+1} = F_y(\mathbf{x}_t, \mathbf{y}_t), \quad (4)$$

where F_y is some function that is not constant in $\mathbf{x}$. If $\mathbf{y}_{t+1}$ only depends on $\mathbf{y}_t$, then $\mathbf{y}$ is called **autonomous**.

We introduce the notation $\mathbf{x} \Rightarrow \mathbf{y}$ to mean that $\mathbf{x}$ causes $\mathbf{y}$. If both $\mathbf{x} \Rightarrow \mathbf{y}$ and $\mathbf{y} \Rightarrow \mathbf{x}$, then we write $\mathbf{x} \Leftrightarrow \mathbf{y}$. If there is no causation, we will say that $\mathbf{x}$ and $\mathbf{y}$ are (causally) independent, denoted $\mathbf{x} \perp \mathbf{y}$.

Given two states $\mathbf{x}$ and $\mathbf{y}$, we may form a joint state by concatenating the vectors. We write $(\mathbf{x}, \mathbf{y})$ to mean the vector $[\mathbf{x}^\top \ \mathbf{y}^\top]^\top$. If the state $\mathbf{y}$ is driven by an autonomous system $\mathbf{x}$, then their joint state $(\mathbf{x}, \mathbf{y})$ forms another autonomous system described by

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{y}_{t+1} \end{bmatrix} = \begin{bmatrix} F_x(\mathbf{x}_t) \\ F_y(\mathbf{x}_t, \mathbf{y}_t) \end{bmatrix} = F_{xy}(\mathbf{x}_t, \mathbf{y}_t). \quad (5)$$

Note that by joining $\mathbf{y}$ with its drivers $\mathbf{x}$, the new system $(\mathbf{x}, \mathbf{y})$ is an autonomous system. This fact will lead us to the notion of a cross map in later sections.

B. Attractors

Attractors are subsets of the state space which both attract and trap trajectories [40]. Given an autonomous system with latent state $\mathbf{x}_t$, a closed subset A of the state space is called an **attractor** if three axioms are satisfied:

- 1) *Attraction*: There is an open set U containing A such that if $\mathbf{x}_t \in U$ then $\text{dist}(A, \mathbf{x}_t)$ tends to 0 as $t \rightarrow \infty$.
- 2) *Trapping*: If $\mathbf{x}_0 \in A$ then $\mathbf{x}_t \in A$ for all $t \geq 0$.
- 3) No proper subset of A also satisfies these conditions.

If $\mathbf{x}_t$ is inside the attractor A at any point, then the invariance axiom implies that we may model the future behavior of $\mathbf{x}_t$ entirely within the set A , without considering the rest of the ambient space $\mathbb{R}^D$. Since attractors tend to be lower-dimensional subsets of the state-space [10], the dynamics in the attractor region may sometimes be much simpler than what the state-space dimension would suggest. We show some commonly discussed attractors in Fig. 1.

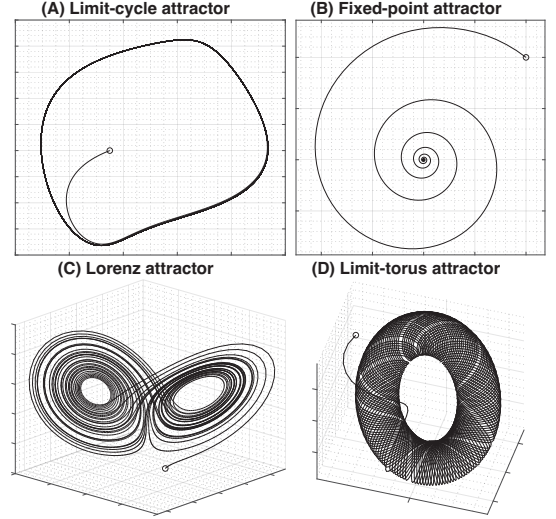


Fig. 1: Some well-known low-dimensional attractors that arise from systems of ordinary differential equations. In each plot, we show a trajectory in a state-space that starts at the dot. (A) A limit cycle is a one-dimensional attractor that corresponds to an asymptotically periodic behavior [40]. (B) An attracting fixed point is an example of a zero dimensional attractor. (C) The Lorenz attractor appears in the well-studied Lorenz system [24], and is known to have fractional dimension [40]. (D) Limit tori correspond to a superposition of asymptotically periodic signals [10], where the dimension of the torus is the number of fundamental frequencies in the superposition.

C. State-space reconstruction

A smooth function from a higher dimensional space to a lower dimensional space is never one-to-one, but a function from a low dimensional space to a higher dimensional one will usually be one-to-one if the difference in dimensions is big enough. Takens' theorem [44] exploits this idea by 'extending the dimension' of an observation signal by considering lagged copies of the signal as new observations. There are several versions of Takens' theorem, and here we state one version based on [36].

Theorem (Takens' theorem). *Let $\mathbf{x}_{t+1} = F_x(\mathbf{x}_t)$ be an autonomous deterministic system with a one-dimensional observation $a_t = f(\mathbf{x}_t)$ and suppose that there is an attractor $A \subset \mathbb{R}^D$ such that $\mathbf{x}_t \in A$. Let Q and τ be fixed positive integers where $Q > 2 \dim(A)$. Then the Q -dimensional delay embedding of a , defined by*

$$\mathbf{m}_t^a \triangleq [a_{t-(Q-1)\tau} \ \cdots \ a_{t-\tau} \ a_t]^\top \in \mathbb{R}^Q \quad (6)$$

is a smooth embedding $A \rightarrow \mathbb{R}^Q$ for almost-every F_x and f .

Several remarks are in order. Embedding means that the attractor and its image under the Q -dimensional delay embed-

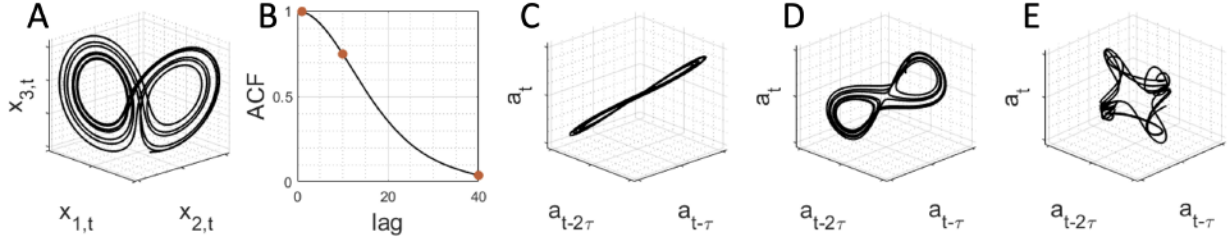


Fig. 2: A demonstration of SSR and the effect of the embedding delay parameter τ . (A) An attractor in the Lorenz system [24]. The complete system state is latent, but we observe a signal $a_t = x_{1,t}$. (B) The autocorrelation function (ACF) of the signal a_t . The values of the ACF at 1, 10 and 40 are marked by the orange dots. (C) Shadow manifold $\mathcal{M}_a$ produced by SSR when $Q = 3$ and $\tau = 1$. Since the ACF at this τ is close to 1, the coordinates of $\mathbf{m}_t^a$ are highly correlated and $\mathcal{M}_a$ is compressed to a line. (D) SSR result when $Q = 3$ and $\tau = 10$. We see the shadow manifold appears to reconstruct the shape of the latent attractor in (A). (E) SSR result when $Q = 3$ and $\tau = 40$. Although $\mathcal{M}_a$ embeds the latent attractor, the embedding becomes more complicated as τ increases. The ideal value of τ for reconstructing the latent attractor lies between the extremes of (C) and (E) [18]. However, the shadow manifolds in (C-E) are equivalent to the original attractor in (A) up to nonlinear transformation.

ding are related by a diffeomorphism.¹ Namely, since each $a_t = f(\mathbf{x}_t)$, the delay-embedding in (6) corresponds to a smooth function Φ ,

$$\Phi(\mathbf{x}_t) = \begin{bmatrix} f(F_x^{(-(Q-1)\tau)}(\mathbf{x}_t)) \\ \vdots \\ f(F_x^{(-\tau)}(\mathbf{x}_t)) \\ f(\mathbf{x}_t) \end{bmatrix} = \begin{bmatrix} f(\mathbf{x}_{t-(Q-1)\tau}) \\ \vdots \\ f(\mathbf{x}_{t-\tau}) \\ f(\mathbf{x}_t) \end{bmatrix} = \mathbf{m}_t^a \quad (7)$$

which Takens' theorem states to be differentiable and invertible (here $F^{(-k)}$ denotes F^{-1} composed with itself k times). The image of the attractor under the diffeomorphism is called the *shadow manifold*² and is denoted $\mathcal{M}_a$ [41]. The reconstruction of latent states $\mathbf{x}_t$ from the observations y_t is called *state-space reconstruction* (SSR) [8].

The parameter Q is called the *embedding dimension* and the idea is that picking a large enough Q implies that we have embedded the latent attractor. The parameter τ is called the *embedding lag*, and it is used to improve the reconstruction by spacing out correlated samples in time. We discuss our preferences for parameter selection in Section V, and more discussion on the matter can be found in [18], [19]. In Fig. 2, we demonstrate how SSR is affected by the parameter τ .

The phrase “almost-every” in the statement of Takens' theorem is also important. One may fabricate systems that do not satisfy Takens' theorem when the dynamics are explicitly written down. A noteworthy example occurs in the Lorenz system, (26), when the observation function is taken to be the x_3 -coordinate projection (see the supplemental material of [41]); however, this type of counterexample is non-generic and may be rare in natural systems [36], [51].

¹A nonlinear transformation Φ is a *diffeomorphism* if Φ is invertible, and both Φ and Φ^{-1} are smooth functions. A smooth embedding $\Phi : \mathbb{R}^D \rightarrow \mathbb{R}^Q$ becomes a diffeomorphism when we restrict the domain of Φ^{-1} to the image of Φ , since $\Phi(\mathbb{R}^D)$ may be a lower dimensional subset of $\mathbb{R}^Q$. We say that two sets A and B are *diffeomorphic* if there is a diffeomorphism Φ such that $\Phi(A) = B$.

²Despite the terminology, shadow manifolds are not usually topological manifolds.

D. Cross maps

A *cross map* is a special function that maps between shadow manifolds. The existence of cross maps can be interpreted as a sign of causality, which is the key insight of the CCM method. In short, if $\mathbf{x} \Rightarrow \mathbf{y}$, then there is a smooth map $\mathcal{M}_a \leftarrow \mathcal{M}_b$ for almost-every pair of signals a, b that are functions of $\mathbf{x}$ and $\mathbf{y}$, respectively [7]. The reason that the cross map arrow goes in the opposite direction of the causality is explained below.

Suppose we observe a_t from an autonomous latent system $\mathbf{x}_t$ and b_t from a latent system $\mathbf{y}_t$ that is driven by $\mathbf{x}_t$. As noted previously, (5) implies that $(\mathbf{x}, \mathbf{y})$ is an autonomous system. If the $(\mathbf{x}, \mathbf{y})$ system has an attractor $A_{\mathbf{xy}}$, then the $\mathbf{x}$ system also has an attractor $A_{\mathbf{x}}$ obtained by projecting $A_{\mathbf{xy}}$ onto the $\mathbf{x}$ coordinates. The reasoning is that the dynamics of $\mathbf{x}$ depend only on $\mathbf{x}$ itself, so attraction/invariance behavior in the joint system $(\mathbf{x}, \mathbf{y})$ imply that the $\mathbf{x}$ system behaves the same. If we denote the $\mathbf{x}$ -coordinate projection by π , then the function π is a smooth function from $A_{\mathbf{xy}}$ to $A_{\mathbf{x}}$, shown in Fig. 3.

The existence of the smooth function π is uninteresting on its own, but it becomes more powerful with Takens' theorem. Since a observes the state $\mathbf{x}$, Takens' theorem says that there is a diffeomorphism $\Phi_a : A_{\mathbf{x}} \rightarrow \mathcal{M}_a$. Since $\mathbf{x} \Rightarrow \mathbf{y}$, b is an observation of the joint system $(\mathbf{x}, \mathbf{y})$, and so there is a diffeomorphism $\Phi_b : A_{\mathbf{xy}} \rightarrow \mathcal{M}_b$. Given these functions, we define the *cross map* σ to be the function

$$\sigma : \mathcal{M}_b \rightarrow \mathcal{M}_a, \\ \sigma(\mathbf{m}) \triangleq \Phi_a(\pi(\Phi_b^{-1}(\mathbf{m}))).$$

By construction, σ satisfies the *time-invariance* property,

$$\sigma(\mathbf{m}_t^b) = \mathbf{m}_t^a, \quad (8)$$

for all t . The time-invariance property encodes the notion that the two shadow manifolds $\mathcal{M}_a$ and $\mathcal{M}_b$ are dynamically synchronized. In the case of no causality, $\mathbf{x} \perp \mathbf{y}$, the shadow manifold $\mathcal{M}_b$ reconstructs the $\mathbf{y}$ state space, but contains no information about $\mathbf{x}$. Thus, if both systems have attractors $A_{\mathbf{x}}$ and $A_{\mathbf{y}}$, it is unlikely for a cross map between $A_{\mathbf{x}}$ and $A_{\mathbf{y}}$ to

exist. Hence, the existence of a cross map suggests that $\mathbf{x}$ and $\mathbf{y}$ are dynamically coupled by some mechanism.³

E. Gaussian process regression

The previous sections on state-space reconstruction and cross mapping rely heavily on the notion of a smooth function, but we haven't addressed the problem of how one may discover smooth functions from data. Gaussian processes provide an interpretable, non-parametric, and Bayesian approach to learning functions from data [49]. The notation in this section will differ from the rest of the paper, since we will borrow notation from [49].

Given a set of input points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ and their corresponding values $y_1, y_2, \dots, y_N$, we require a technique to find a smooth nonlinear function f that will describe the data, i.e.,

$$y_n \approx f(\mathbf{x}_n) \quad n = 1, \dots, N. \quad (9)$$

Suppose that we want to predict some datum $y_* = f(\mathbf{x}_*)$ for a given point $\mathbf{x}_*$. In Gaussian process regression (GPR), we model the previously observed data and the interpolant jointly with a zero-mean normal distribution,

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{k}_* \\ \mathbf{k}_*^\top & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right), \quad (10)$$

where $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{k}_{*,i} = k(\mathbf{x}_i, \mathbf{x}_*)$ and $k(\cdot, \cdot)$ is a covariance function [49]. Since $y_1, \dots, y_N$ have already been observed, we generate our predictive density for y_* by conditioning on $y_1, \dots, y_N$. For a single one-dimensional prediction, [49] provides the mean and variance of the predictive density of an interpolated datum y_* , where

$$\mu_* = \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} = \mathbb{E}(y_* | y_1, \dots, y_N), \quad (11)$$

$$\sigma_*^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_* = \text{Var}(y_* | y_1, \dots, y_N). \quad (12)$$

To describe the GPR predictive density for y_* as a function of $\mathbf{x}_*$, we write

$$y_* | \mathbf{x}_* \sim \mathcal{N}(\mu_*, \sigma_*^2). \quad (13)$$

Implicitly, we also conditioned on the training data $\{(\mathbf{x}_n, y_n)\}$.

In practice, the covariance function $k(\cdot, \cdot)$ will depend on some hyperparameters θ . Training the GPR regression typically refers to optimizing the value of θ to fit the distribution in (10), often by maximum likelihood. The mean of the distribution in (10) may be replaced with a parametric, e.g. polynomial, function of $\mathbf{x}$, in which case θ and the parameters of the mean function are optimized jointly.

In large data sets, the basic form of GPR presented here may suffer from a high computational complexity. Inducing-point GPs [31] and sparse spectrum GPs [23] are popular approximations to reduce the computational complexity of the GPR approach.

³This principle might be considered as a dynamical analogue to Reichenbach's common cause principle [33]. It can theoretically be violated, e.g., by designing two systems to produce periodic signals with the same frequency, but such situations are not generic [4].

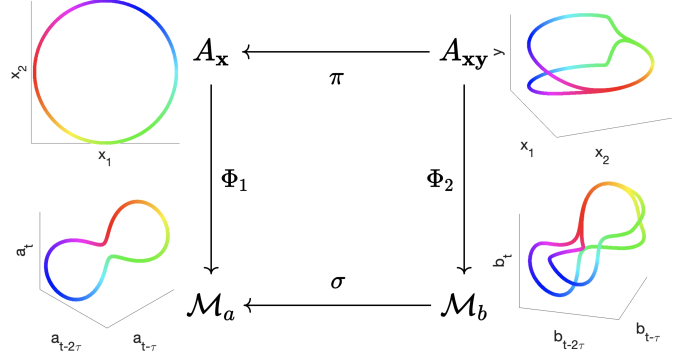


Fig. 3: Visual explanation of the cross map. We consider a two dimensional state $\mathbf{x} \in \mathbb{R}^2$ with an attractor A_x . We then suppose that $\mathbf{x}$ causes $\mathbf{y} \in \mathbb{R}^1$, and the joint state $(\mathbf{x}, \mathbf{y})$ has an attractor A_{xy} . The signals a_t and b_t are functions of $\mathbf{x}_t$ and $\mathbf{y}_t$, respectively. The projection $\pi(\mathbf{x}, \mathbf{y}) = \mathbf{x}$ maps A_{xy} onto A_x . The mappings Φ_1, Φ_2 are diffeomorphisms which exist by Takens' theorem. The colors of points on each attractor and shadow manifold correspond uniquely to points in A_x . Using the colors to map points in M_b to points in M_a describes the cross map σ between the two shadow manifolds. Since Φ_2 is invertible, we see that the cross map is defined by $\sigma(\mathbf{m}) = \Phi_1(\pi(\Phi_2^{-1}(\mathbf{m})))$.

III. PROBLEM FORMULATION

In this section, we introduce CCM as a causality test, and we discuss its caveats and our approach to work around them.

A. Convergent cross mapping

Since a cross map typically only exists if the underlying latent systems are coupled, several authors have proposed to use cross maps to infer causation [2], [5], [32], [35], [37]. The CCM method proposed by Sugihara et al [41] tests for causality by detecting if a cross map between shadow manifolds exists. Namely, if $\mathbf{x} \Rightarrow \mathbf{y}$, then a cross map $M_a \leftarrow M_b$ exists in the opposite direction.

Proving that a cross map exists, using a finite set of data points, is a delicate procedure. The CCM method aims to do this by learning the cross map in an online manner. Sugihara et al. proposed to learn the cross map using the simplex projection method [41] (which is essentially k -nearest-neighbor regression), although other regression techniques may be useful as well [12]. Briefly, simplex projection will interpolate $f(\mathbf{x}_t)$ as a weighted sum of data $f(\mathbf{x}_q)$, where $\mathbf{x}_q$ are the $Q + 1$ nearest neighbors to $\mathbf{x}_t$ and the weights w_q are obtained by normalizing $\tilde{w}_q = \exp(-\|\mathbf{x}_{tq} - \mathbf{x}_t\|^2)$. Given two signals a_t and b_t , the basic CCM test has three steps:

- 1) Perform SSR by delay embedding the signal b_t to produce SSR vectors $\mathbf{m}_t^b = (b_t, \dots, b_{t-(Q-1)\tau})$.
- 2) Estimate a_t as $\hat{a}_t = g(\mathbf{m}_t^b)$, where the function g is learned using simplex projection.
- 3) Repeat steps 1 and 2 sequentially as new data are received. If the "skill" of the cross map, measured by the

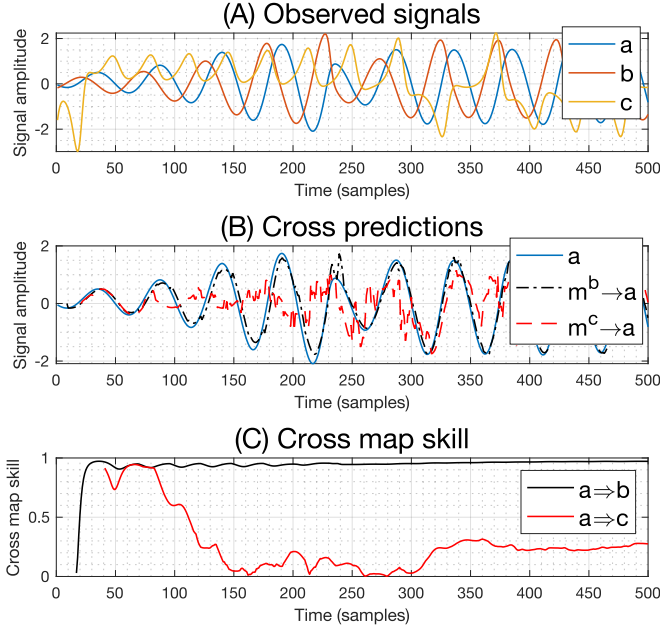


Fig. 4: Given a triplet of signals a_t, b_t, c_t , we use CCM to test $a \Rightarrow b$ and $a \Rightarrow c$. **(A)** Observed signals. a_t, b_t, c_t are the x_1, x_2 and y_2 coordinates respectively of the Rössler-Lorenz system in (30), with zero coupling ($C = 0$). **(B)** The signal a_t is estimated using $\mathbf{m}_t^b$ to test $a \Rightarrow b$. Similarly, we estimate a_t from $\mathbf{m}_t^c$ to test $a \Rightarrow c$. **(C)** For $a \Rightarrow b$, the cross map skill, $\rho_t = \text{corr}(a_{1:t}, \hat{a}_{1:t})$, improves over time. Since the cross map skill converges to something near 1, we conclude that a_t causes b_t . In contrast, the asymptotic cross map skill for $a \Rightarrow c$ is poor, so we conclude that a_t doesn't drive c_t . As $t \rightarrow \infty$, the cross map skill should tend to 0, but it may take some time for the cross map skill to decay sufficiently.

correlation coefficient between $\hat{a}_t$ and a_t , converges to a nonzero number, then we conclude that a causes b .⁴

CCM approximates the cross map by estimating one of its coordinates, since a_t is the first coordinate of $\mathbf{m}_t^a$ by (6). The estimator $\hat{a}$ is expected to improve if and only if the cross map σ between the shadow manifolds exists. In general, the estimator can improve or degrade in performance with additional data depending on the ground truth. However, as we increase the number of observations we also increase the number of points that are sampled along the shadow manifold. If new points on the manifold agree with old points in how they map between the shadow manifolds, then it indicates that the cross map is real and not just a statistical anomaly. If there are nearby points which do not make nearby predictions, then the cross map does not exist,⁵ and we will see a degradation in the cross map performance. We show a case in which CCM detects the correct causality in Fig. 4.

⁴To test the reverse causality, $a \Leftarrow b$, we apply CCM again with the roles of a and b reversed.

⁵This technically means that the cross map does not exist for a particular pair of shadow manifolds $\mathcal{M}_a, \mathcal{M}_b$.

B. Subtleties of CCM

The applicability of cross-map-based tests for causality rests primarily on whether or not the SSR result contains a shadow manifold, that is, a reconstructed attractor or invariant set. If there is not really a shadow manifold, CCM and related methods cannot be guaranteed to reproduce the ground truth. Even when the shadow manifold exists, it can be distorted, by noise, trends and other effects. Furthermore, there are a few cases where even with long, clean time series, cross mapping may not be an effective solution. We now discuss some of the reasons that cross map analysis may struggle in practice.

1) Small observation windows

Convergence is an important consideration in CCM. While some pairs of signals may converge faster than others, even randomly generated signals may feign convergence when observed over a short time window. In Fig. 4, we notice that the $a \Rightarrow c$ cross map skill is good initially, but this performance deteriorates as more data is observed. An opposite issue occurs when the shadow manifold does exist, but the current data set was insufficient to explore the whole manifold. In the best case, we should have enough data to resolve the entire shadow manifold and determine if it is structured enough for analysis. The number of samples required make a statistically meaningful conclusion depends upon the sampling rate and the complexity of the signals.

2) Distortions

Sometimes, the signal of interest might be superimposed with noise or external influences to the measurement. These effects could cause trends, seasonalities or noisy residuals which obscure the observation signal and distort the results of cross map analysis [51].

3) SSR failure

As noted in Sec. II-D, Takens' theorem holds for 'almost-every' attractor system. When producing synthetic data, one should be careful not to pick an observation function with any special symmetries or properties [4], [36], [51]. SSR failure could in principle occur due to improper selection of Q or τ .

4) Generalized synchrony

A major issue with all cross map methods is the problem of generalized synchrony [41]. Synchrony occurs when the strength of a unidirectional relationship $\mathbf{x} \Rightarrow \mathbf{y}$ is so strong that $\mathbf{y}$ does not show independent behavior from $\mathbf{x}$. As a result, the cross map $\sigma : \mathcal{M}_b \rightarrow \mathcal{M}_a$ is actually reversible, and the inverse function $\sigma^{-1} : \mathcal{M}_a \rightarrow \mathcal{M}_b$ is a cross-map that goes in the anti-causal direction, and one will mistake unidirectional causation $\mathbf{x} \Rightarrow \mathbf{y}$ as bidirectional $\mathbf{x} \Leftrightarrow \mathbf{y}$.

When synchrony occurs, the underlying causal structure of the latent system becomes non-identifiable. Although detecting a cross map in this case can be useful as a way to rule out independence, any further deductions would require prior knowledge about the specific system under investigation. The detection of synchrony best addressed by other work [16], [34]. Ye et. al. proposed a partial solution to the problem of synchrony in the CCM framework in [50] by studying the time-delay of the causal interactions.

C. When can we use CCM?

Suppose that we have a pair of signals which we suspect to arise from nonlinear deterministic systems and the known distortions have been removed. We want to know if cross map analysis is applicable. Since cross mapping attempts to detect if two shadow manifolds are synchronized, a natural requirement is to check that the SSR actually produces a shadow manifold, i.e. a reconstructed autonomous system. If a signal observes an autonomous system, then the shadow manifold produced by SSR should be useful to predict future observations [20], which will be our first condition. This kind of predictability should be possible, since

$$\mathbf{m}_{t+1}^a = \Phi(F_x(\Phi^{-1}(\mathbf{m}_t^a))) = \tilde{F}(\mathbf{m}_t^a), \quad (14)$$

$$a_{t+1} = [0 \quad \cdots \quad 0 \quad 1] \mathbf{m}_{t+1}^a \quad (15)$$

where Φ and F_x are defined as in Takens' theorem (Section II-C). If a function $\tilde{F} = \Phi \circ F_x \circ \Phi^{-1}$ can be well-approximated, at least on shorter time frames, then we will say that the signal a_t is **auto-predictable** (since it is being predicted from its own history). Exploiting (14) to use $\mathbf{m}_t^a$ as a proxy for the latent state $\mathbf{x}_t$, we recognize that an auto-predictable signal represents an approximately deterministic system by this construction. This provides evidence that the observed shadow manifold isn't just a statistical anomaly, but a true reconstruction of the states that produced the signal [18].

Our second condition is intended to mitigate the small observation window problem. The main issue with small windows is that we cannot be certain that we have explored the attractor enough to conclude that there is a cross map. As we receive new data, and hence new points along the shadow manifold which we use to make cross predictions, we can either learn that the current model is predictive and it continues to be useful, or we may find that the additional data proves that our model is wrong. In the latter case, CCM will eventually show non-convergence, but it takes time to do so, as seen in Fig. 4.

One reason for this is that when we have few data, our cross map model is over-fitted to the sparsely sampled manifold, but as we receive more data and more dense samples along the shadow manifold, our cross predictions are forced to reconcile with more and more data. Similarly, if we attempt to measure auto-predictability and make self-predictions about how the state changes on the shadow manifold, we are forced to reconcile how multiple trajectories had evolved when they were at a similar point in space. Unless there are other trajectories nearby, when making a cross prediction, we do not have any confidence that our cross-prediction is not overfitted to only the local data. Thus, an important condition is that after enough time has passed, we return to previously sampled locations on the shadow manifold, so that we observe more than one trajectory at each location. This property, called the **recurrence** property, provides a condition on the data set that is important for cross mapping to have statistical significance.

We summarize our two conditions for asserting that a signal, or specifically the shadow manifold produced by a finite set of samples from the signal, is appropriate for cross mapping analysis:

- 1) **Auto-predictability.** If a_t reconstructs a shadow manifold, then the signal a_t should be auto-predictable, that is, it can be well-predicted by its own history.
- 2) **Recurrence.** At every point in the shadow manifold, there should be nearby points from another trajectory, i.e. that are not nearby in time.

While we are primarily concerned about studying the SSR result, it has also been suggested that additional analysis should be done to validate the statistical significance of the CCM result [6], [47]. Methods, such as surrogate analysis [6] or the comparison of cross and self-predictions [20], have been proposed to measure the confidence of a cross map result.

IV. PROPOSED SOLUTION

In the previous section we proposed the auto-predictability and recurrence conditions to provide evidence that the observed signals are appropriate for cross map analysis. We propose the AF and RF statistics, respectively, to test for these conditions. We use a procedure based on Gaussian processes (Section II-E) to estimate these statistics from data.

A. Preprocessing

The first step in any causal analysis should be to isolate the signal of interest, by denoising, detrending and otherwise removing any distortions that known or expected. Since attractors represent stationary behavior, the cleaned signal should look somewhat stationary for cross map analysis. Furthermore, the nonlinearity of the resultant signals should also be checked. If the signals are linear, in the sense that they are well-described by a linear autoregressive model, then Granger causality [14], [15] is readily applicable.

B. Testing for auto-predictability

Generally a random signal is more challenging to predict than one arising from a deterministic system, because there is no latent state that can be reconstructed from the history of the signal. Thus, if we use a regression model to predict the future of a random signal, we would expect a larger prediction error than for a non-random one. The prediction error is defined to be the out-of-sample mean-square-error (MSE) of the estimate of the function $\tilde{F}$, in (14), which we estimate using GPR.⁶

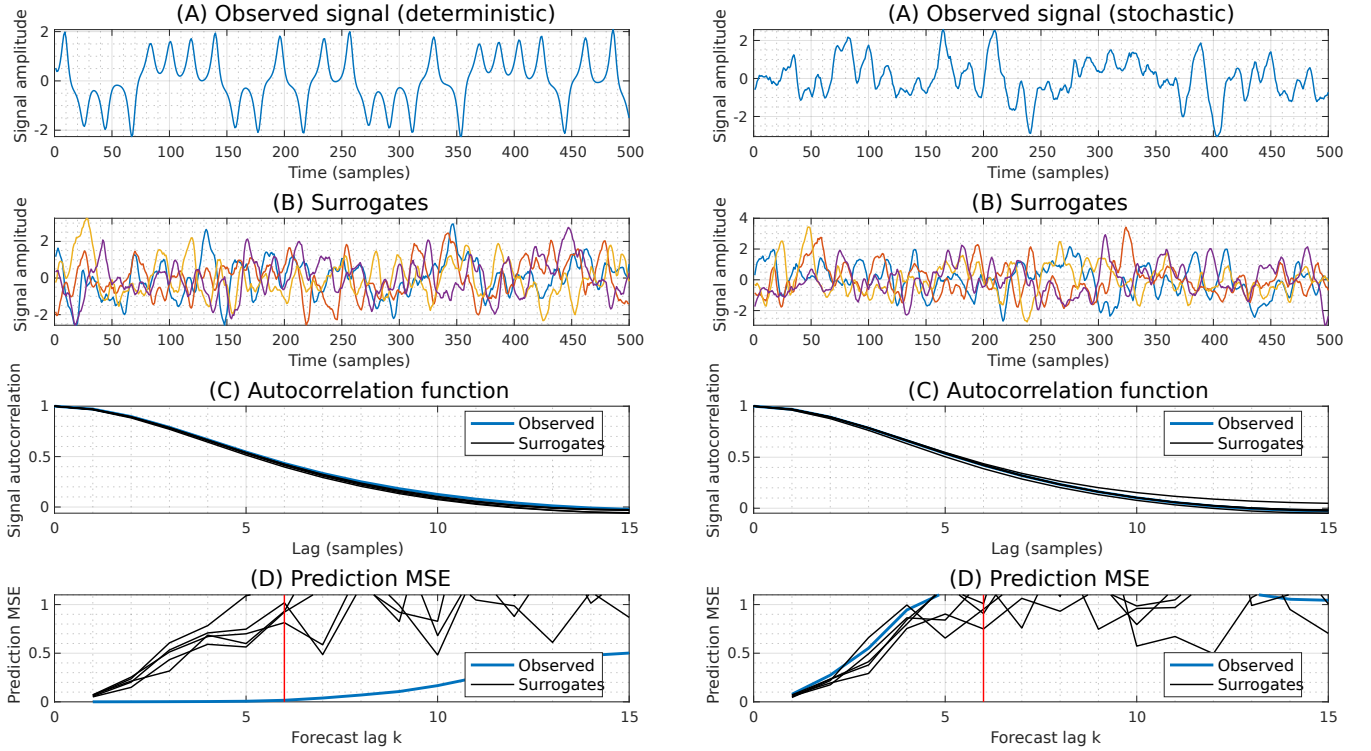
Suppose that we have constructed a number of SSR vectors $\mathbf{m}_t^a$ for times $t = 1, \dots, T$. Let us partition the set of all such points into a training set $\mathcal{T}$ and a validation set $\mathcal{V}$, e.g. $\mathcal{T} = \{1, \dots, t_*\}$ and $\mathcal{V} = \{t_* + 1, \dots, T\}$. Using GPR, we may produce a non-parametric estimate of the function $\tilde{F}$ by training on the points in $\mathcal{T}$. As discussed in Section II-E, Gaussian processes give us an explicit predictive probability distribution for each $a_{t+\tau}$, $t \in \mathcal{T}$:

$$a_{t+\tau} | \mathbf{m}_t^a \sim \mathcal{N}(\mu_*(\mathbf{m}_t^a), \sigma_*^2(\mathbf{m}_t^a)) \quad (16)$$

where μ_* and σ_*^2 are functions of $\mathbf{m}_t^a$ given by (11) and (12). Predictions are then given by the mean,

$$\hat{a}_{t+\tau} = \mathbb{E}(a_{t+\tau} | \mathbf{m}_t^a) = \mu_*(\mathbf{m}_t^a),$$

⁶Technically, we will estimate the τ -th iterate of the function, $\tilde{F}^{(\tau)}$.



(a) The surrogate test with a Lorenz signal. The prediction MSE of the observed signal is significantly less than any of the surrogates produced from it, indicating that the observed signal is much more predictable than a randomly generated signal.

(b) The surrogate test with a random signal. The prediction MSE of the observed signal accumulates quickly as the forecast lag is increased, and it does so as fast as the surrogate signals, so it is unlikely that the observed signal arose from a deterministic system.

Fig. 5: Two demonstrations of the surrogate predictability test. (A) Observed signal. (B) Surrogate signals produced by randomly phase shifting the power spectrum of the observed signal. (C) Empirical autocorrelation functions of the observed signal and its surrogates. (D) Prediction MSE of $\hat{a}_{t+k}$ as a function of the prediction length k . The red vertical line marks $k = \tau$. In both cases, the surrogate time series demonstrate high predictability for small k ; indeed, an autocorrelated random signal will be predictable for a short period into the future. However, we observe that the deterministic signal exhibits a noticeably slower accumulation of error as k is increased.

and the prediction MSE (PMSE) is estimated by

$$\text{PMSE} = \frac{1}{|\mathcal{V}|} \sum_{t \in \mathcal{V}} (a_{t+\tau} - \hat{a}_{t+\tau})^2. \quad (17)$$

The PMSE measures how predictable the signal a_t is in an interpretable way [18]. When the PMSE is very small compared to the energy of the signal, then auto-predictability can be concluded. However, when the PMSE is not close to zero, the statistical significance of the figure is not obvious, and so we need a way to understand how probable it is. To this end, we introduce surrogate data analysis.

A *surrogate* is a random signal which is desired to share common traits with the original signal, such as the ACF or power spectral density (PSD), but is otherwise random in nature [45]. The ACF, or equivalently the PSD, does not identify the dynamical system that produced it [30]. As a result, two signals may have the same PSD but only one of them produces a shadow manifold with dynamical structure. Surrogate analysis has been used extensively in nonlinear systems and chaos theory to bootstrap the significance of our test statistics [1]. To sample surrogate signals, there are a number of candidate methods. A popular technique is to

convert the observed signal to the frequency domain, and to apply random phase shifts [1]. This approach preserves the ACF and spectrum of the signal, but the surrogates are usually not dynamical in the sense of the AF. The random phases approach is particularly effective when the signal spectrum is broadband, as is the case in many chaotic systems [10]; for periodic or quasiperiodic signals, other surrogate methods may be desirable.

To use surrogates to improve our understanding of the PMSE measure, we can sample K surrogate signals, repeat the GPR analysis above, and compare the histogram of the surrogate PMSE with the observed signal's PMSE. If the observed signal is deterministic, then the surrogates should be less predictable because they are random, and so the probability that $\text{PMSE}(\text{surrogate}) > \text{PMSE}(\text{observed})$ is high. We define the **auto-predictability fraction (AF)** of the signal a_t to be

$$\text{AF} = \frac{1}{K} \sum_{k=1}^K \vartheta_k \quad (18)$$

where

$$\vartheta_k = \begin{cases} 1, & \text{if } \text{PMSE}(s_t^{(k)}) \geq \text{PMSE}(a_t) \\ 0, & \text{otherwise} \end{cases}$$

and $s_t^{(k)}$ is the k -th surrogate signal. AF measures the probability that a surrogate is less predictable than the observed signal. If the signal is auto-predictable, then $\text{AF} \approx 1$. In Fig. 5a and 5b, we demonstrate the surrogate predictability test with Lorenz and random signals, respectively. The accuracy of the AF measure depends on K . We note that parallel computing can be used to compute $\text{PMSE}(s_t^{(k)})$ more efficiently, since each surrogate can be independently processed.

C. Detecting recurrence

The recurrence condition says that we observe points in the shadow manifold that are nearby in space but not in time. In order to detect pairs of nearby points in the shadow manifold, we automate the interpretation of recurrence plots [9], a widely known tool in behavioral sciences [42] and dynamics [28].

For every pair of points $\mathbf{m}_{t_i}^a, \mathbf{m}_{t_j}^a$ on the shadow manifold $\mathcal{M}_a$, we define

$$\mathbf{D}_{ij} = \|\mathbf{m}_{t_i}^a - \mathbf{m}_{t_j}^a\| \quad (19)$$

to be the pairwise distances between points in $\mathcal{M}_a$. Given a parameter r which we call the *masking radius*, we say that $\mathbf{m}_{t_i}^a$ and $\mathbf{m}_{t_j}^a$ are neighbors if $\mathbf{D}_{ij} \leq r$. If one thinks of $\mathbf{D}_{ij}$ as the pixels of an image, as in Fig. 6, then the recurrence matrix is a mask of the image. Points that are nearby in time will be neighbors, but these are not useful for checking the recurrence condition. Thus, we define the recurrence matrix $\mathbf{R}_{ij}$ to record when neighbors are more than τ samples apart in time, i.e.,

$$\mathbf{R}_{ij} = \begin{cases} 1, & \mathbf{D}_{ij} \leq r \text{ and } |t_i - t_j| > \tau \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

We define the **recurrence fraction** (RF) to be the fraction of points in the shadow manifold that have at least one neighbor of this form,

$$\text{RF} = \frac{1}{T} \sum_{i=1}^T \max_j \mathbf{R}_{ij}. \quad (21)$$

Ideally $\text{RF} = 1$, but it isn't strictly required for most data sets. In Fig. 7, we show a simple system where CCM's accuracy depends noticeably on RF. In Alg. 1, we summarize the algorithm.

Algorithm 1 Computation of RF

- 1: **input:** Matrix M whose rows are $(\mathbf{m}_t^a)^\top$, parameters r, τ
 - 2: T = number of columns in M
 - 3: $D_{ij} = \|\mathbf{m}_{t_i}^a - \mathbf{m}_{t_j}^a\|$
 - 4: $R_{ij} = 1$ if $D_{ij} \leq r$ and $|t_i - t_j| > \tau$
 - 5: $\text{RF} = \frac{1}{T} \sum_{i=1}^T \max_j R_{ij}$
 - 6: **return** RF
-

The efficacy of the RF measure depends on the choice of masking radius r . If r exceeds the diameter of $\mathcal{M}_a$, i.e. $r > \max_{s,t} \|\mathbf{m}_s^a - \mathbf{m}_t^a\|$, then $\text{RF} = 1$. Similarly, $\text{RF} = 0$ whenever

$r < \min_{s,t} \|\mathbf{m}_s^a - \mathbf{m}_t^a\|$. There are several ways one may select a value for r , but we now introduce a method based on the GP model learned when we computed AF.

To learn a masking radius r in a more interpretable manner, we can exploit the GP kernel k . Since $\mathbf{K}_{ij} = k(\mathbf{m}_{t_i}^a, \mathbf{m}_{t_j}^a)$ encodes the covariance between $a_{t_i+\tau}$ and $a_{t_j+\tau}$, the entries $\mathbf{K}_{ij}$ encode a notion of recurrence from the perspective of the GPR model used to detect auto-predictability. Suppose that we want to conclude that two points are neighbors when their correlation exceeds a threshold ρ , i.e.

$$\tilde{\mathbf{K}}_{ij} = \frac{\mathbf{K}_{ij}}{\sqrt{\mathbf{K}_{ii}\mathbf{K}_{jj}}} \geq \rho. \quad (22)$$

The resulting parameter selection depends on the kernel in use. The default kernel choice is often the squared exponential kernel [49],

$$k(\mathbf{m}, \mathbf{m}') = \sigma_f^2 \exp\left(\frac{-(\mathbf{m} - \mathbf{m}')^\top (\mathbf{m} - \mathbf{m}')}{2\ell^2}\right), \quad (23)$$

where σ_f, ℓ are parameters. When using this kernel, (22) may be expressed as

$$\tilde{\mathbf{K}}_{ij} = \exp\left(\frac{-\mathbf{D}_{ij}^2}{2\ell^2}\right) \geq \rho. \quad (24)$$

and this is equivalent to

$$\mathbf{D}_{ij} \leq r_* = \sqrt{2\ell^2 \ln(1/\rho)}. \quad (25)$$

The range of values $\rho \in (0, 1)$ can define any $r_* > 0$, and so the distance or covariance-based recurrence measures are equally expressive. However, the kernel-based approach allows us to use a more intuitive notion of threshold to define our measure of recurrence, and relates directly back to our predictability analysis. Additionally, kernel-based analysis is not costly since the covariance matrix $\mathbf{K}$ is already computed when we fit the GPR model.

V. RESULTS

A. Simulated systems

To verify our approach, we implement several systems and consider the performance of CCM in comparison to the AF and RF measures. We briefly outline the generative process and true causality for each pair of signals, and we show the resulting CCM, AF and RF results in Table 1.

In each example, we use standard approaches to infer the τ and Q parameters for SSR. To get the embedding lag τ , we define τ to be the first value of the autocorrelation function (ACF) to drop below a threshold [22], which we will take to be 0.5 for the time being.⁷ Given τ , we select Q using the method of false-nearest neighbors [19]. Both approaches are standard, and we refer to [18] for more discussion. MATLAB code to reproduce these examples is made available in the repository for this paper.⁸

⁷Some authors prefer using 0 as the threshold [22], which is potentially more principled but also inflates the value of τ . Since the number of delay vectors that we can produce from a fixed size data set decreases as τ increases, we might prefer the smaller τ produced by our heuristic.

⁸See <https://github.com/KurtButler/2023-CCM-paper>.

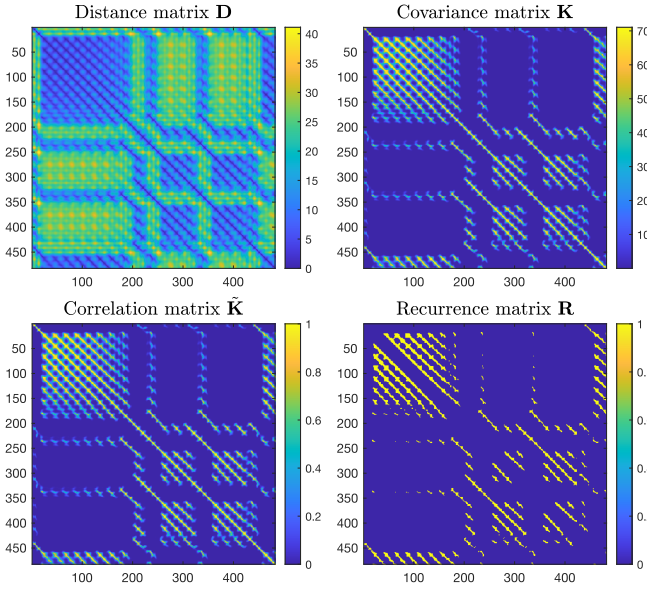


Fig. 6: Visualization of the matrices $\mathbf{D}$, $\mathbf{K}$, $\tilde{\mathbf{K}}$ and $\mathbf{R}$ (c.f. equations (19), (22) and (20)) for a signal obtained from a Lorenz system. To obtain covariances, we trained a GPR model to predict $\mathbf{m}_{t+\tau}^a$ from $\mathbf{m}_t^a$, using SE kernel. To obtain $\mathbf{R}$, we used the kernel-based threshold with $\rho = 0.5$, as in (25).

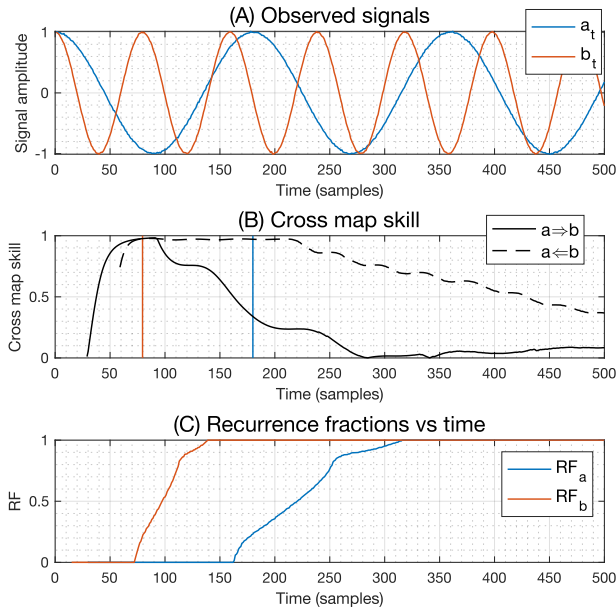


Fig. 7: Demonstration of the recurrence principle with a pair of sinusoids. (A) Two sine waves a_t and b_t signals are observed with frequencies f_a and f_b . The true causality is $a \perp b$. (B) The CCM convergence coefficient is observed against time. CCM asymptotically detects the true causality, but for a shorter time series (e.g. the first 200 samples), we would erroneously infer that $a \leftarrow b$. The vertical lines mark one period of the sinusoids in (A). (C) The recurrence fraction RF as new data is received, for a particular choice of masking radii r_a and r_b . We observe that the convergence coefficients begin to decline shortly after when RF begins to increase.

1) Lorenz system

The Lorenz system [24] consists of a latent state $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ and a system of differential equations,

$$\begin{aligned} \dot{x}_{1,t} &= a(x_{2,t} - x_{1,t}), \\ \dot{x}_{2,t} &= x_{1,t}(c - x_{3,t}) - x_{1,t}, \\ \dot{x}_{3,t} &= x_{1,t}x_{2,t} - bx_{3,t}, \end{aligned} \quad (26)$$

where we use the initial condition $\mathbf{x}_0 = (10, 0, 5)$.

In scenario L_1 , we observe the $x_{1,t}$ and $x_{2,t}$ coordinates of the Lorenz system, i.e., $a_t = x_{1,t}$ and $b_t = x_{2,t}$, for 500 time steps. Both a_t and b_t were normalized prior to processing. Since all variables in the Lorenz system interact, the true causality is $a \Leftrightarrow b$. The sampling period was 0.02. Simulation of each trajectory was completed using the MATLAB `ode45` solver. The SSR parameters were selected to be $Q = 5$ and $\tau = 5$ for a_t and $Q = 5$ and $\tau = 6$ for b_t .

In scenario L_2 , we observe $b_t = x_{2,t}$, but a_t is now a surrogate signal sampled from the PSD of $x_{1,t}$. The surrogate has no causal relationship to the original system, so the true causality is $a \perp b$. The SSR parameters were selected to be $Q = 5$ and $\tau = 6$ for both a_t and b_t .

CCM detected the correct causality in scenario L_1 but not L_2 . In scenario L_2 , we note that $RF_a = 0.57$ suggests that almost half of the reconstructed points do not have neighbors, indicating that a spurious cross map due to non-recurrence is possible. Since both the Lorenz signal and its surrogates are generally continuous (see Fig. 5a), their shadow manifolds are continuous and so the existence of a spurious cross map is possible. Additionally, AF_a indicates that the observed predictability is within-distribution, and so our signal is likely a random process with no latent attractor structure.

2) Drift processes

Based on [3], we consider two systems that combine a linear drift with bounded signals. In scenario B_1 , we observe signals a_t and b_t for $t = 501, \dots, 1000$ from the following system:

$$\begin{aligned} x_{t+1} &= 3.82x_t(1 - x_t) \\ y_{t+1} &= 3.74y_t(1 - y_t) \\ a_t &= 0.015x_t + 0.0003t \\ b_t &= y_t \end{aligned} \quad (27)$$

where $x_t, y_t \in [0, 1]$. The initial conditions for x_t, y_t were sampled from a uniform distribution $\mathcal{U}(0, 1)$. The SSR parameters were selected to be $Q = 3$ and $\tau = 33$ for a_t (due to the large linear trend) and $\tau = 1$ for b_t .

The signal a_t is largely driven by the linear trend in time, so its shadow manifold is expected to not be recurrent. The signal b_t settles into a periodic cycle with period 2, and so the shadow manifold $\mathcal{M}_b$ is asymptotically a two point set. However, it sometimes takes time for the states y_t to converge onto their attractors, so we burn the initial 500 samples to ensure that our trajectories reside on their attractors. The true causality in scenario B_1 is $a \perp b$.

Scenario B_2 is similar to B_1 , but now the logistic process x_t is substituted for a noise process. In scenario B_2 , we have

$$\begin{aligned} x_t &\stackrel{i.i.d.}{\sim} \mathcal{U}(0, 1), \\ y_{t+1} &= 3.4y_t(1 - y_t), \\ a_t &= 2 + \frac{0.3t}{500 + t} + 0.1x_t, \\ b_t &= y_t. \end{aligned} \quad (28)$$

The SSR parameters were selected to be $Q = 4$ and $\tau = 1$ for both signals. The true causality in scenario B_2 is again $a \perp b$.

While CCM detected the correct causality for both scenarios, we note that a_t in scenario B_2 is not a signal arising from a deterministic system, and so the use of CCM would be unjustified if the ground truth was known. In practice, this would be evidenced by the low auto-predictability ($AF_a = 0.63$). In scenario B_1 the dynamics are deterministic, but the latent dynamics are uncoupled, and no cross map was detected.

3) Coupled autoregressive processes

In scenarios K_1 and K_2 , we consider the following system for different values of the coupling parameter C based on [21].

$$\begin{aligned} u_t, v_t &\sim \mathcal{N}(0, 1), \\ x_{t+1} &= 0.5x_t + 0.2y_t + \sqrt{0.1}u_t, \\ y_{t+1} &= Cx_t + 0.7y_t + \sqrt{0.1}v_t, \\ a_t &= x_t, \\ b_t &= y_t, \end{aligned} \quad (29)$$

for $t = 1, \dots, 500$. We randomly sample the initial conditions as $x_1, y_1 \sim \mathcal{N}(0, 1)$. In scenario K_1 , the coupling is unidirectional ($C = 0$) and so the true causality is $a \Leftarrow b$. In scenario K_2 , the coupling parameter is $C = 0.6$, and so the true causality is $a \Leftrightarrow b$. The SSR parameters were selected to be $Q = 4$ in all scenarios, and $\tau = 1$. In practice, exploratory data analysis should reveal that these processes are linear and stochastic, and hence appropriate for Granger causality. We analyze this case anyways to better understand what happens under a model mismatch.

In scenario K_1 , CCM detected no causality between the coupled autoregressive (AR) processes. Both AF_a and AF_b are suspect in this case, which indicates that the CCM result may be highly unreliable. Since AR processes are stochastic, they do not have attractors, and rightfully the GPDM struggles to find a dynamical rule on the shadow manifolds. In scenario K_2 , we also found small AF values, but CCM still detected that the signals were coupled because the strong coupling between the latent systems induced a large correlation between the two signals. Thus, scenario K_2 suggests that strong forcing between systems could permit CCM to detect bidirectional coupling even in stochastic systems. Although CCM obtained the correct result in the K_2 case, the low predictability indicates that we cannot trust the CCM result, because it may have only been obtained by chance.

4) Rössler-Lorenz

For scenarios K_3 and K_4 , we consider a composite system where a Rössler system $\mathbf{x}$ influences a Lorenz system $\mathbf{y}$. The

system from [21] is defined by

$$\begin{aligned} \dot{x}_{1,t} &= -6(x_{2,t} + x_{3,t}), \\ \dot{x}_{2,t} &= 6(x_{1,t}) + 0.2x_{2,t}, \\ \dot{x}_{3,t} &= 6(0.2 + x_{3,t}(x_{1,t} - 5.7)), \\ \dot{y}_{1,t} &= 10(y_{2,t} - y_{1,t}), \\ \dot{y}_{2,t} &= 28y_{1,t} - y_{2,t} - y_{1,t}y_{3,t} + Cx_{2,t}^2, \\ \dot{y}_{3,t} &= y_{1,t}y_{2,t} - 8y_{3,t}/3, \\ a_t &= x_{2,t}, \\ b_t &= y_{2,t}, \end{aligned} \quad (30)$$

where we observe the system for $t = 1, \dots, 1000$. The sampling period was 0.025, and the initial conditions were given by $(x_{1,0}, \dots, y_{3,0}) = (0, 0, 0.4, 0.3, 0.3, 0.3)$. Simulation of each trajectory was completed using the MATLAB `ode45` solver. The SSR parameters were selected to be $Q = 4$ and $\tau = 5$ for a and $Q = 8$ and $\tau = 3$ for b . The coupling parameter C controls the strength of the causal interaction. In scenario K_3 , we set $C = 0$ so the true causality is $a \perp b$. In scenario K_4 we set $C = 3$, so the causality is unidirectional. However, it is noted in [21] that general synchrony for this system occurs when $C = 3$, and so the cross map is invertible. Thus, the detected causation should be $a \Leftrightarrow b$, even though this does not reflect the true causality.

In both scenarios K_3 and K_4 , the AF and RF metrics detected enough structure for the application of cross mapping. Upon applying CCM, we detected the expected cross maps in both scenarios.

B. Electricity data

To test our approach on a real data set, we used the Electricity Load Diagrams 2011-2014 data set from the UCI Machine Learning Repository [46], which records the electricity consumption of 370 households in a Portuguese city. The data show a daily periodic behavior which we expect to be consistent across the city. To study these data, we partition all households into two groups, and we average the electricity trends in each group for each time. The average daily electricity usage for each group defines two signals a_t and b_t . Since both signals observe the same latent process (the city's collective electricity usage), we expect that the signals are synchronized and $a_t \Leftrightarrow b_t$.

To study the synchronization, we observe a_t, b_t for $t = 500, \dots, 1500$ (roughly 100 days), which we call scenario E_1 in the table. Prior to processing, each signal was smoothed by a Savitzky-Golay filter to remove daily noise. The SSR parameters were selected to be $Q = 5$ and $\tau = 6$ for both signals. Our analysis finds that the smoothed signals are both recurrent and predictable, which agrees with other work that has found limit cycle behavior in this data set [13]. Performing CCM on these signals, we deduce that $a_t \Leftrightarrow b_t$, indicating that their behavior is indeed synchronized.

C. Discussion of the Rössler-Lorenz system

In the previous analyses, we considered the Rössler-Lorenz system in (30) for somewhat extreme values of C . However,

Scenario	Correct result	CCM result	AF _a	RF _a	AF _b	RF _b	Appropriate for Cross-Mapping?
L ₁	$a \Leftrightarrow b$	$a \Leftrightarrow b$	1.00	1.00	1.00	1.00	Yes
L ₂	$a \perp b$	$a \Leftrightarrow b$	0.76	0.57	1.00	1.00	No
B ₁	$a \perp b$	$a \perp b$	1.00	1.00	1.00	1.00	Yes
B ₂	$a \perp b$	$a \perp b$	0.63	1.00	1.00	1.00	No
K ₁	$a \Leftarrow b$	$a \perp b$	0.39	1.00	0.72	1.00	No
K ₂	$a \Leftrightarrow b$	$a \Leftrightarrow b$	0.28	1.00	0.77	1.00	No
K ₃	$a \perp b$	$a \perp b$	1.00	1.00	1.00	0.99	Yes
K ₄	$a \Leftrightarrow b^*$	$a \Leftrightarrow b$	1.00	1.00	1.00	1.00	Yes
E ₁	$a \Leftrightarrow b$	$a \Leftrightarrow b$	1.00	0.81	1.00	0.87	Yes

TABLE I: A comparison of CCM results and the proposed metrics for a few systems. Incorrect CCM results are highlighted. The * denotes that due to generalized synchrony, only bidirectional causation can be detected.

any value of C between 0 and 3 also produces a valid dynamical system with attractors, and as long as $C > 0$ the causality $\mathbf{x} \Rightarrow \mathbf{y}$ is present. Thus, an important question is what happens for intermediate C , where the coupling is weak but still non-negligible. The transition between the $C = 0$ regime and the $C = 3$ regime requires a bifurcation of the system's attractor, but since this transformation occurs continuously, the distribution of test statistics for the cross map (e.g., the CCM convergence coefficient) will also vary continuously.

In Fig. 8, we compare the situations when $C = 0, 1$ and 3. For each plot, we consider 1,000 possible trajectories of 2,000 samples, where we discarded the initial transient (the first 1,000 samples) of each realization, so we have only the remaining 1,000 samples. The sampling period of the observation signal was 0.025, and the initial conditions were distributed according to a Gaussian: $(\mathbf{x}_0, \mathbf{y}_0) + 0.01\mathbf{n}$, where $(\mathbf{x}_0, \mathbf{y}_0) = (0, 0, 0.4, 0.3, 0.3, 0.3)$ and $(\mathbf{n})_i \sim \mathcal{N}(0, 1)$ i.i.d. Simulation of each trajectory was completed using the MATLAB `ode45` solver. SSR parameters were estimated for each realization, and their values varied across realizations, but the typical values for τ were between 4 and 6. We used $Q = 4$ and $Q = 8$ for the signals a and b , respectively.

We observe that for $C = 0$ and $C = 3$, the distributions agree with what we would expect a cross map causality test to produce in this situation. However, when $C = 1$ we see that the distribution corresponding to $a \leftarrow b$ is centered at 0.5. If one uses a threshold of 0.5 to decide if there is or is not causation, then the probability of making an error (given a random trajectory) is 50 percent, and thus CCM is completely unreliable in this situation. Even for a different threshold value, one may continuously vary C to find a value where the distribution average is on top of the threshold. In general, any system defined using a coupling parameter will likely undergo a bifurcation as the coupling parameter is varied, and so this behavior could be generally expected for systems with weaker coupling (i.e., near the bifurcation point).

VI. CONCLUSION

The theory of cross mapping proposes a principled and appealing approach to causal inference in deterministic systems with attractors, which is especially attractive given evidence that Granger-style methods struggle in these settings. The assumption that we can reconstruct deterministic latent states from the observation signals is a necessary but often nontrivial one, and so we proposed to provide evidence for this by

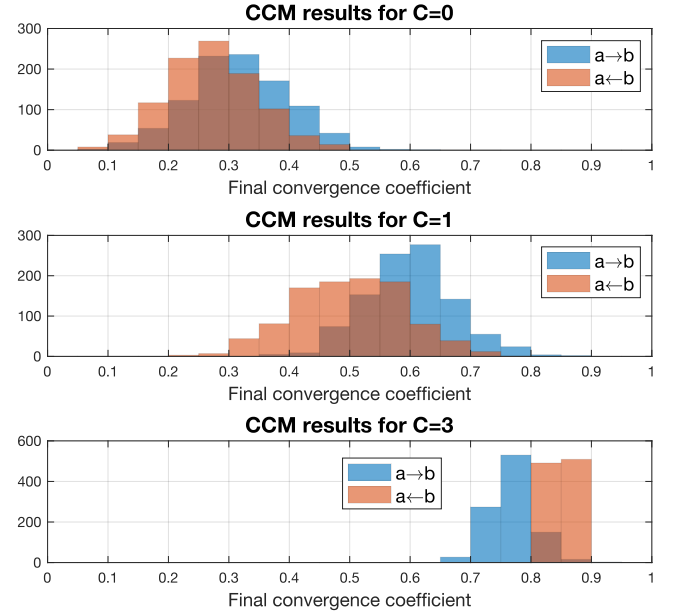


Fig. 8: Histogram of the CCM convergence coefficient across 1,000 trajectories with randomized initial conditions, visualized for different values of C in the Rössler-Lorenz system.

leveraging Gaussian processes to test for auto-predictability and recurrence. It was found that by performing these applicability tests, some naïve failure modes of cross mapping could be detected and avoided. This is especially relevant when attempting to use cross mapping in settings with an unclear ground truth.

Cross mapping still has several major challenges to overcome (e.g., generalized synchrony, the bifurcation problem), even after verifying that the signals are appropriate. The amelioration of these situations may require subtle dynamical and topological considerations; the details of which vary from case to case. Cross mapping may be a powerful tool when combined with other evidence and careful analysis, but overall our results suggest that it should never be applied blindly.

ACKNOWLEDGMENT

The authors thank the support of NSF under Awards 2021002 and 2212506, and the support of the U.S. Department of Education through the GAANN Fellowship. We thank Josue Nassar and Marija Iloska for providing feedback on

early drafts of this work. We also express our thanks to the reviewers, whose comments have greatly improved the style, rigor, and content of this paper.

REFERENCES

- [1] R. G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz. Bivariate surrogate techniques: necessity, strengths, and caveats. *Physical Review E*, 68(6):066202, 2003.
- [2] J. Arnold, P. Grassberger, K. Lehnertz, and C. E. Elger. A robust method for detecting interdependences: application to intracranially recorded EEG. *Physica D: Nonlinear Phenomena*, 134(4):419–430, 1999.
- [3] S. Bartsev, M. Saltykov, P. Belolipetsky, and A. Panykh. Imperfection of the convergent cross-mapping method. In *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2021.
- [4] M. Besserve, N. Shajarisales, B. Schölkopf, and D. Janzing. Group invariance principles for causal generative models. In *International Conference on Artificial Intelligence and Statistics*, pages 557–565. PMLR, 2018.
- [5] D. Chicharro and R. G. Andrzejak. Reliable detection of directional couplings using rank statistics. *Physical Review E*, 80(2):026217, 2009.
- [6] S. Cobey and E. B. Baskerville. Limits to causal inference with state-space reconstruction for infectious disease. *PloS one*, 11(12):e0169050, 2016.
- [7] B. Cummins, T. Gedeon, and K. Spendlove. On the efficacy of state space reconstruction methods in determining causality. *SIAM Journal on Applied Dynamical Systems*, 14(1):335–381, 2015.
- [8] E. R. Deyle and G. Sugihara. Generalized Theorems for Nonlinear State Space Reconstruction. *PLOS One*, 6(3):e18295, 2011.
- [9] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle. Recurrence Plots of Dynamical Systems. *Europhysics Letters*, 4:973–977, 1987.
- [10] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *The Theory of Chaotic Attractors*, pages 273–312, 1985.
- [11] G. Feng, J. G. Quirk, and P. M. Djurić. Discovering causalities from cardiocography signals using improved convergent cross mapping with Gaussian processes. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1309–1313. IEEE, 2020.
- [12] G. Feng, K. Yu, Y. Wang, Y. Yuan, and P. M. Djurić. Improving Convergent Cross Mapping for Causal Discovery with Gaussian Processes. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3692–3696. IEEE, 2020.
- [13] W. Gilpin. Deep reconstruction of strange attractors from time series. *Advances in neural information processing systems*, 33:204–216, 2020.
- [14] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
- [15] C. W. Granger. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352, 1980.
- [16] D. He, Z. Zheng, and L. Stone. Detecting generalized synchrony: An improved approach. *Physical Review E*, 67(2):026223, 2003.
- [17] L. Heskamp, A. Meel-van den Abeelen, J. Lagro, and J. Claassen. Convergent cross mapping: a promising technique for cerebral autoregulation estimation. *International Journal Clinical Neurosciences and Mental Health*, 20, 2014.
- [18] H. Kantz and T. Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge University Press, 2004.
- [19] M. B. Kennel, R. Brown, and H. D. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45(6):3403, 1992.
- [20] A. Krakovská and J. Jakubík. Implementation of two causal methods based on predictions in reconstructed state spaces. *Physical Review E*, 102(2):022203, 2020.
- [21] A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, and M. Paluš. Comparison of six methods for the detection of causality in a bivariate time series. *Physical Review E*, 97(4):042207, 2018.
- [22] D. Kugiumtzis. State space reconstruction parameters in the analysis of chaotic time series—the role of the time window length. *Physica D: Nonlinear Phenomena*, 95(1):13–28, 1996.
- [23] M. Lázaro-Gredilla, J. Quinero-Candela, C. E. Rasmussen, and A. R. Figueiras-Vidal. Sparse Spectrum Gaussian Process Regression. *The Journal of Machine Learning Research*, 11:1865–1881, 2010.
- [24] E. N. Lorenz. The statistical prediction of solutions of dynamical equations. In *Proceedings of the International Symposium on Numerical Weather Prediction, 1962*. Meteor. Soc. Japan, 1962.
- [25] C. Luo, X. Zheng, and D. Zeng. Causal inference in social media using convergent cross mapping. In *2014 IEEE Joint Intelligence and Security Informatics Conference*, pages 260–263. IEEE, 2014.
- [26] L. Luo, F. Cheng, T. Qiu, and J. Zhao. Refined convergent cross-mapping for disturbance propagation analysis of chemical processes. *Computers & Chemical Engineering*, 106:1–16, 2017.
- [27] B. Lusch, P. D. Maia, and J. N. Kutz. Inferring connectivity in networked dynamical systems: Challenges using Granger causality. *Physical Review E*, 94:032220, 2016.
- [28] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329, 2007.
- [29] J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017.
- [30] R. J. Povinelli, M. T. Johnson, A. C. Lindgren, F. M. Roberts, and J. Ye. Statistical models of reconstructed phase spaces for signal classification. *IEEE Transactions on Signal Processing*, 54(6):2178–2186, 2006.
- [31] J. Quinero-Candela and C. E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959, 2005.
- [32] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger. Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Physical Review E*, 65(4):041903, 2002.
- [33] H. Reichenbach. *The Direction of Time*, volume 65. University of California Press, 1956.
- [34] N. F. Rulkov and V. S. Afraimovich. Detectability of nondifferentiable generalized synchrony. *Physical Review E*, 67(6):066218, 2003.
- [35] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. Abarbanel. Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E*, 51(2):980, 1995.
- [36] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics*, 65(3):579–616, 1991.
- [37] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Physical Review E*, 54(6):6708, 1996.
- [38] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461, 2000.
- [39] G. Stepaniants, B. W. Brunton, and J. N. Kutz. Inferring causal networks of dynamical systems through transient dynamics and perturbation. *Physical Review E*, 102, 2020.
- [40] S. H. Strogatz. *Nonlinear Dynamics and Chaos: With applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, 2018.
- [41] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch. Detecting Causality in Complex Ecosystems. *Science*, 338(6106):496–500, 2012.
- [42] Y. Taghizadegan, N. J. Dabanloo, K. Maghooli, and A. Sheikhan. Obstructive sleep apnea event prediction using recurrence plots and convolutional neural networks (rp-cnns) from polysomnographic signals. *Biomedical Signal Processing and Control*, 69:102928, 2021.
- [43] S. Tajima, T. Yanagawa, N. Fujii, and T. Toyozumi. Untangling brain-wide dynamics in consciousness by cross-embedding. *PLoS Computational Biology*, 11(11):e1004537, 2015.
- [44] F. Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, pages 366–381. Springer, 1981.
- [45] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94, 1992.
- [46] A. Trindade. Electricity Load Diagrams 2011–2014.
- [47] A. A. Tsonis, E. R. Deyle, R. M. May, G. Sugihara, K. Swanson, J. D. Verbeten, and G. Wang. Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. *Proceedings of the National Academy of Sciences*, 112(11):3253–3256, 2015.
- [48] N. Wiener. *The Theory of Prediction*. In E. Beckenbach, editor, *Modern Mathematics for the Engineer*. Dover, 1956.
- [49] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*. MIT Press, 2006.
- [50] H. Ye, E. R. Deyle, L. J. Gilarranz, and G. Sugihara. Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports*, 5(1):1–9, 2015.
- [51] A. E. Yuan and W. Shou. Data-driven causal analysis of observational time series in ecology. *bioRxiv*, 2021.