Convergent cross-mapping and pairwise asymmetric inference

James M. McCracken[*] and Robert S. Weigel[†]

School of Physics, Astronomy, and Computational Sciences, George Mason University, 4400 University Drive,
MS 3F3, Fairfax, Virginia 22030-4444, USA
(Received 8 July 2014; published 1 December 2014)

Convergent cross-mapping (CCM) is a technique for computing specific kinds of correlations between sets of times series. It was introduced by Sugihara *et al.* [Science **338**, 496 (2012).] and is reported to be "a necessary condition for causation" capable of distinguishing causality from standard correlation. We show that the relationships between CCM correlations proposed by Sugihara *et al.* do not, in general, agree with intuitive concepts of "driving" and as such should not be considered indicative of causality. It is shown that the fact that the CCM algorithm implies causality is a function of system parameters for simple linear and nonlinear systems. For example, in a circuit containing a single resistor and inductor, both voltage and current can be identified as the driver depending on the frequency of the source voltage. It is shown that the CCM algorithm, however, can be modified to identify relationships between pairs of time series that are consistent with intuition for the considered example systems for which CCM causality analysis provided nonintuitive driver identifications. This modification of the CCM algorithm is introduced as "pairwise asymmetric inference" (PAI) and examples of its use are presented.

I. INTRODUCTION

Modern time series analysis includes techniques meant to discern "driving" relationships between different data sets. These techniques have found application in a wide range of fields including neuroscience (e.g., Ref. [1]), economics (e.g., Refs. [2,3]), climatology (e.g., Ref. [4]), and others. General casual relationships in time series data are also being studied in an effort to understand causality itself (e.g., Ref. [5]).

To date, most techniques for "causal inference" in time series data fall into three broad categories: those related to transfer entropy, those related to Granger causality, and those related to lagged cross-correlation. Transfer entropy (introduced in Ref. [6]) and Granger causality (introduced in Ref. [7]) are known to be equivalent under certain conditions [8]. In this article, we investigate a casual inference technique, called convergent cross-mapping (CCM), that was recently introduced by Sugihara *et al.* [9]. (Currently, there is no evidence that CCM is related to either transfer entropy or Granger causality.)

Statements of causality beyond simple correlation that are derived from time series data alone fall under the broad term "time series causality." Most known time series causality measures have well-documented shortcomings. Granger causality (and its extensions) are model-based approaches that depend on, among other things, the validity of the model [1,10] and separability [9]. Transfer entropy, mutual information, and other information theoretic techniques rely on estimating entropies from empirical data, which may involve the computationally difficult task of estimating transition probability densities and generally require large amounts of data [11,12]. Lagged cross-correlation techniques can be unreliable in the presence of strongly autocorrelated data [13], and partial directed coherence (which is related to lagged cross-correlation techniques) can lead to spurious conclusions

given both simulated time series with known dynamics and empirical data [14]. The results documented here show that the recently introduced CCM technique also has shortcomings; in particular, the CCM technique may lead to conclusions that do not agree with intuitive notions of causality.

CCM is described as a technique that can be used to identify "causality" between time series and is intended to be useful in situations where Granger causality is known to be invalid (i.e., in dynamic systems that are "nonseparable" [9]). CCM is also intended to useful in determining causality in systems that experience "mirage" correlations [9] (i.e., correlations that are a function of time). Thus, CCM is introduced as a time series causality technique that specifically addresses some of the shortcomings discussed in the previous paragraph. The authors state that CCM is a "necessary condition for causation." It is well known that Granger causality is not causality as it is typically understood in physics [10,15,16]. Whether a similar conclusion can be drawn regarding CCM causality is currently an open question.

CCM has been used to draw conclusions regarding the "controversial sardine-anchovy-temperature" problem [9], confirm predictions of climate effects on sardines [17], compare the driving effects of precipitation, temperature, and solar radiation on the atmospheric CO_2 growth rate [18], and quantify cognitive control in developmental psychology [19]. The technique has also been presented as a useful tool in studying the causality of respiratory systems in insects [20]. The wide range of applications already appearing for the relatively new CCM technique is testament to the importance of time series causality studies. This work presents examples in which CCM does not provide consistent qualification of an intuitive notion of causality. However, the domain of applicability of CCM remains an open question; i.e., the method may work as expected for the authors cited above despite its apparent failure in the examples presented in this article.

We begin with a review of the work of Sugihara *et al.* [9], including an extended evaluation of the coupled logistic map example. We then introduce pairwise asymmetric inference (PAI) and use it to show that even though CCM causality may

---

[*]jmccrac1@masonlive.gmu.edu
[†]rweigel@gmu.edu

not be physical causality, it can still be a useful tool in the analysis of complex time series data.

II. CONVERGENT CROSS-MAPPING

CCM is closely related to simplex projection [21,22], which predicts a point in the times series X at a time $t + 1$, labeled X_{t+1}, by using the points with the most similar histories to X_t. Similarly, CCM uses points with the most similar histories to X_t to estimate Y_t. The CCM correlation is the squared Pearson correlation coefficient [23] between the original time series Y and an estimate of Y made using its convergent cross-mapping with X, which is labeled as $Y|\tilde{\mathbf{X}}$:

$$C_{YX} = [\rho(Y, Y|\tilde{\mathbf{X}})]^2 .$$

Any pair of times series, X and Y, will have two CCM correlations, C_{YX} and C_{XY}, which are compared to determine the CCM causality. For example, Sugihara *et al.* [9] define a difference of CCM correlations

$$\Delta = C_{YX} - C_{XY} \qquad (1)$$

and use the sign of Δ to determine the CCM causality between X and Y.

If X can be estimated using Y better than Y can be estimated using X (e.g., if $\Delta < 0$), then X is said to "CCM cause" Y.

A. CCM algorithm

The CCM algorithm [24] may be written in terms of five steps:

(1) create the *shadow manifold* for X, called $\tilde{\mathbf{X}}$;

(2) find the nearest neighbors to a point in the shadow manifold at time t, which is labeled $\tilde{\mathbf{X}}_t$;

(3) create weights using the nearest neighbors;

(4) estimate Y using the weights (this estimate is called $Y|\tilde{\mathbf{X}}$); and

(5) compute the correlation between Y and $Y|\tilde{\mathbf{X}}$.

The steps are described in more detail below.

1. Create shadow manifold $\tilde{\mathbf{X}}$

Given an embedding dimension E, the *shadow manifold* of X, called $\tilde{\mathbf{X}}$, is created by associating an E-dimensional vector (also called a *delay vector*) to each point X_t in X, i.e., $\tilde{\mathbf{X}}_t = (X_t, X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-(E-1)\tau})$. The first such vector is created at $t = 1 + (E - 1)\tau$ and the last is at $t = L$, where L is the number of points in the time series (also called the *library length*).

2. Find nearest neighbors

The minimum number of points required for a bounding simplex in an E-dimensional space is $E + 1$ [21,22]. Thus, the set of $E + 1$ nearest neighbors must be found for each shadow manifold $\tilde{\mathbf{X}}_t$. For each $\tilde{\mathbf{X}}_t$, the nearest neighbor search results in a set of distances that are ordered by closeness $\{d_1, d_2, \ldots, d_{E+1}\}$ and an associated set of times $\{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{E+1}\}$. The distances from $\tilde{\mathbf{X}}_t$ are

$$d_i = D(\tilde{\mathbf{X}}_t, \tilde{\mathbf{X}}_{\hat{t}_i}),$$

where $D(\mathbf{a}, \mathbf{b})$ is the Euclidean distance between vectors $\mathbf{a}$ and $\mathbf{b}$.

3. Create weights

Each of the $E + 1$ nearest neighbors are be used to compute an associated weight. The weights are defined as

$$w_i = \frac{u_i}{N},$$

where $u_i = e^{-d_i/d_1}$ and the normalization factor is $N = \sum_{j=1}^{E+1} u_j$.

4. Find $Y|\tilde{\mathbf{X}}$

A point Y_t in Y is estimated using the weights calculated above. This estimate is

$$Y_t|\tilde{\mathbf{X}} = \sum_{i=1}^{E+1} w_i Y_{\hat{t}_i}.$$

5. Compute the correlation

The CCM correlation is defined as

$$C_{YX} = [\rho(Y, Y|\tilde{\mathbf{X}})]^2,$$

where $\rho(A, B)$ is the standard Pearson's correlation coefficient between A and B.

The CCM algorithm depends on the embedding dimension E and the lag time step τ. A dependence on E and τ is a feature of most state space reconstruction (SSR) methods [25–27], so an E and τ dependence is not unexpected. Sugihara *et al.* mention that "optimal embedding dimensions" are found using univariate SSR [9] (Supplemental Material), and other methods for determining E and τ for SSR algorithms can be found in the literature (e.g., [25,27,28]).

B. CCM example

Consider the example system used by Sugihara *et al.* [9]:

$$X_t = X_{t-1}(r_x - r_x X_{t-1} - \beta_{xy} Y_{t-1}), \qquad (2)$$

$$Y_t = Y_{t-1}(r_y - r_y Y_{t-1} - \beta_{yx} X_{t-1}), \qquad (3)$$

where the parameters $r_x, r_y, \beta_{xy}, \beta_{yx} \in \mathbb{R} \geqslant 0$. This pair of equations is a specific form of the two-dimensional coupled logistic map system [29].

In this example, the CCM causality relationship between X and Y is determined using a sampling of both the initial conditions and the system parameters, calculating Δ, and demonstrating the necessary convergence. The term *convergence* is used here in the same sense as it was used in Ref. [9]; i.e., "convergence means that cross-mapped estimates improve in estimation skill with time-series length L" [30]. The dynamic parameters r_x and r_y are sampled from normal distributions $\mathcal{N}(\mu_{rx}, \sigma_{rx})$ and $\mathcal{N}(\mu_{ry}, \sigma_{ry})$, respectively. The initial conditions X_0 and Y_0 are also sampled from normal distributions, specifically $\mathcal{N}(\mu_{x0}, \sigma_{x0})$ and $\mathcal{N}(\mu_{y0}, \sigma_{y0})$. The coupling parameters β_{xy} and β_{yx} are then varied over the interval $[10^{-6}, 1]$ in steps of 0.02 to produce the plots seen in Fig. 1.

Sugihara *et al.* consider convergence to be critically important for determining CCM causality and note that it is "a key property that distinguishes causation from simple correlation" [9]. Figure 1 shows plots created with several
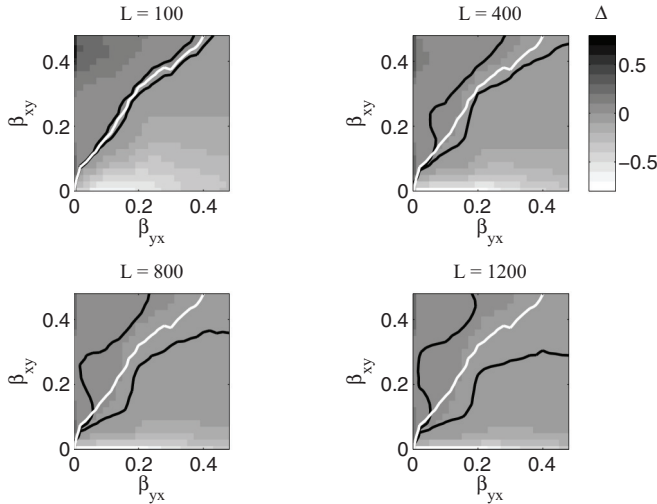
FIG. 1. The dependence of Eq. (1) on the dimensionless parameters β_{xy} and β_{yx}. The white line is a contour for $\Delta = 0$; black lines are contours for $\Delta = \pm 0.01$.

different library lengths to illustrate the convergence of Δ for this example. Typically, for convenience, the (approximately) converged CCM correlation values will be reported and proof of convergence will be implied, rather than shown.

The idea is that $\beta_{xy} > \beta_{yx}$ intuitively implies Y "drives" X more than X "drives" Y. Stated more formally, $\beta_{xy} > \beta_{yx} \Rightarrow \Delta > 0$, which is reported as "$Y$ CCM causes X." Likewise, $\beta_{xy} < \beta_{yx}$ implies X CCM causes Y and $\beta_{xy} = \beta_{yx}$ implies no CCM causality in the system. It will be shown below that CCM causality is not necessarily related to causality as it is typically understood in physics.

III. SIMPLE EXAMPLE SYSTEMS

The usefulness of the CCM algorithm in identifying drivers among sets of time series can be explored by using simple example systems. The benefit to this approach is that the driver is known beforehand. Thus, CCM causality can be validated by comparison against intuition in simple systems where intuition can be trusted. Intuitively identifying the driving signal in the coupled logistic map presented in Ref. [9] is difficult due to the complexity of the system. Commonly used chaotic dynamic systems such as the Lorentz system present similar difficulties. Thus, the examples here are intended to explore (and validate) the CCM method by comparing the results to expected conclusions.

Each of the following examples intuitively supports the conclusion that X drives Y, and CCM analysis (with $E = 3$ and $\tau = 1$) yields values of Δ that support conclusions that do not agree with intuition for all parameter choices.

A. Linear example

Consider the linear example dynamical system of

$$X_t = \sin(t), \tag{4}$$

$$Y_t = X_{t-1} + B\eta_t, \tag{5}$$

with $B \in \mathbb{R} \geqslant 0$ and $\eta_t \sim \mathcal{N}(0,1)$. Specifically, consider $B \in [0,2]$ in increments of 0.02. The noise term $B\eta_t$ is used to show how the conclusions drawn with the CCM algorithm depend on the parameters of the system. In this case, the response system Y is just a lagged version of the driving signal with varying levels of standard Gaussian noise applied at each time step. Figure 2 shows Δ for this example given a library length of $L = 2000$.

Figure 2(b) shows (for $B = 1.66$) that Δ is more negative at shorter library lengths but converges to a point near zero as the library length is increased. The convergence of CCM correlations is emphasized [9], so the seemingly counterintuitive behavior of Δ (and C_{XY} and C_{YX}) in Fig. 2 implies that the CCM correlations may not be a reliable measure of "driving" (at least not the intuitive definition) for this simple linear example system.

The expected conclusion that X drives Y, corresponding to X CCM causes Y, requires $\Delta < 0$. But it can be seen from Fig. 2(a) that the sign of Δ depends on B. Given that the intuitive conclusion of X drives Y in Eq. (4) does not depend on B, it would seem that Δ does not consistently reflect the intuitive conclusion in this linear example system.

These results do not imply that the primary failure of CCM causality is not being reliable at certain signal-to-noise ratios. The Eq. (4) system with no noise, i.e., $B = 0$, leads to $\Delta = 2.5 \times 10^{-3}$, which does not agree with intuition. However, Fig. 2 shows CCM causality for this example, which does agree with intuition for certain noise levels. The key idea is that the agreement of CCM causality with intuition depends on the system parameters. The value of Δ is strongly dependent on the system parameters even in noise-free systems (see Sec. III C).
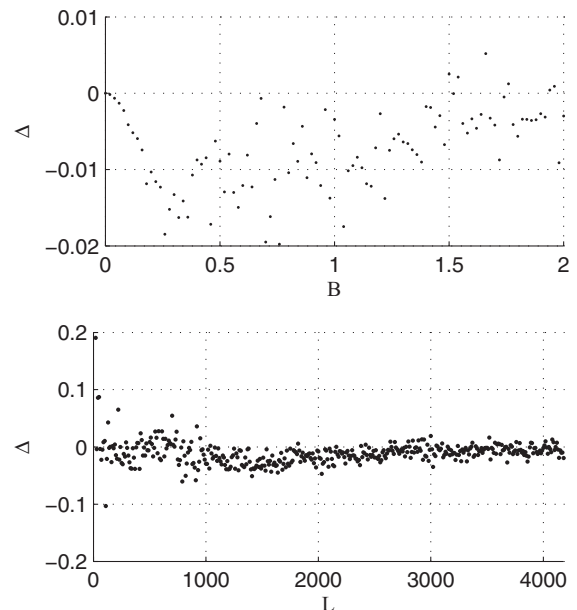

FIG. 2. Dependence of the sign of Δ from Eq. (1) on the dimensionless parameter B. (a) The dimensionless Δ is calculated as described in the text; (b) The dependence of Δ on library length L (dimensionless) is shown for the points $B = 1.66$.
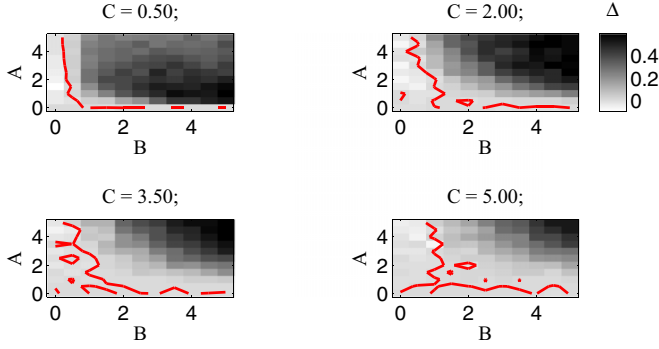
FIG. 3. (Color online) The sign of Δ, and thus the CCM causality, depends on the dimensionless parameters A, B, and C. Contour lines indicate where $\Delta = 0$.

B. Nonlinear example

Consider the nonlinear dynamical system of

$$X_t = \sin(t), \tag{6}$$

$$Y_t = AX_{t-1}(1 - BX_{t-1}) + C\eta_t, \tag{7}$$

with $A, B, C \in \mathbb{R} \geqslant 0$ and $\eta_t \sim \mathcal{N}(0,1)$. Specifically, consider $A, B, C \in [0,5]$ in increments of 0.5. Figure 3 shows Δ for specific values of C for a library length of $L = 2000$. As in the previous example, the expectation for this system is that Δ should be negative, independent of the parameters A, B, and C. However, it can be seen from the plots that the sign of Δ can depend on all three parameters. Thus, this simple nonlinear example leads to a conclusion similar to the previous linear example, i.e., Δ does not reliably reflect intuitive notions of driving.

C. Resistor-inductor circuit example

Both of the previous examples included a noise term, η_t. The failure of CCM analysis to give expected conclusions about the drivers in the previous examples is not due to a limitation of the algorithm with respect to noise. This can be investigated by considering a system without noise. Consider a series circuit containing a resistor, inductor, and time-varying voltage source related by

$$\frac{dI}{dt} = \frac{V(t)}{L} - \frac{R}{L}I, \tag{8}$$

where I is the current at time t, $V(t) = \sin(\Omega t)$ is the voltage at time t, R is the resistance, and L is the inductance. Equation (8) was solved using the *ode45* integration function in MATLAB. The time series $V(t)$ is created by defining values at fixed points and using linear interpolation to find the time steps required by the ODE solver.

Consider the situation where $L = 10$ H and $R = 5$ ohm are constant. Physical intuition is that V drives I, and so we expect to find that V CCM causes I (i.e., $C_{VI} > C_{IV}$ or $\Delta = C_{VI} - C_{IV} > 0$).

Consider evaluating the CCM correlations C_{VI} and C_{IV} for each $\Omega \in [0.01, 2.0]$ in steps of 0.01. The CCM correlations are found using $E = 2$ and $\tau = 1$ and are used to calculate $\Delta =$
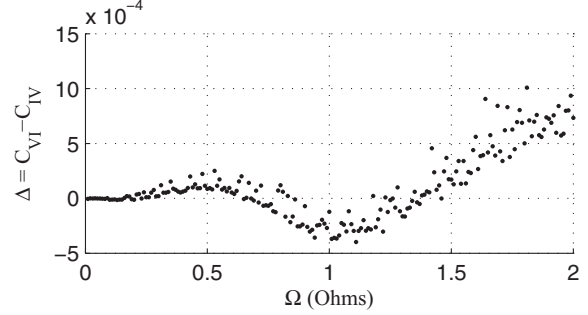

FIG. 4. Δ dependence on Ω for a series RL circuit with a sinusoidal voltage source. The sign of Δ implies V CCM causes I and I CCM causes V depending on the voltage frequency.

$C_{VI} - C_{IV}$, which is plotted in Fig. 4. Δ does not consistently agree with intuition in this example either. Changing the embedding dimension, E, used to calculate Δ leads to plots that are different than in Fig. 4, but in all of the cases tested (i.e., $E = 2, 3, 4$), the sign of Δ changes over the domain $\Omega \in [0.1, 1.5]$.

The resistance and inductance of the circuit are fixed and the voltage is varied from 1×10^{-2} to 2.0 V in discrete steps of 0.01 V as described by Eq. (8). Physically changing the voltage and witnessing a resulting change in the current is enough to convince most people that the voltage "drives" the current. Rigorous statistical hypothesis testing can be performed to strengthen the confidence in such a conclusion. Yet, from Fig. 4, the voltage does not consistently "CCM cause" the current as Ω is changed.

It may be argued that the relatively small values (as compared to the previous examples) of Δ plotted in Fig. 4 indicate that the correct conclusion should be either (1) there is no CCM causality in the system or (2) CCM causality is not applicable to this system. However, conclusion (1) conflicts with the intuitive notion of an RL circuit as a strongly driven system and conclusion (2) conflicts with identifying CCM causality as a general qualifier of "driving" in dynamical systems.

IV. PAIRWISE ASYMMETRIC INFERENCE

Consider the example system of Eq. (2) with $r_y = r_y = 3.7$, $X_0 = 0.2$, $Y_0 = 0.4$, $\beta_{xy} = 0$, and $\beta_{yx} = 0.32$. These parameters correspond to Figs. 3(c) and 3(d) of Ref. [9] (with $E = 2, \tau = 1$, and $L = 1000$). Plots of the correlation between X and $X|\tilde{Y}$ (i.e., X estimated using the weights found from the shadow manifold of Y) as well as Y and $Y|\tilde{X}$ are shown in Fig. 5. This leads to $\Delta = C_{YX} - C_{XY} \approx 0.11 - 0.97 = -0.86 < 0$, which implies X CCM causes Y. This result agrees with intuition because $\beta_{xy} = 0 < \beta_{yx} = 0.32$.

The correlations shown in Fig. 5 are not the only correlations that can be tested. Consider, for example, the correlation between X and the corresponding $X|\mathbf{X}$, which is estimated using weights found from the shadow manifold of X itself. The time series X may also be estimated using a multivariate shadow manifold consisting of points from both X and Y [17]. For example, an $E + 1$ dimensional point in the a multivariate shadow manifold constructed using both X and Y may be defined as $\tilde{\mathbf{X}}_t = (X_t, X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-(E-1)\tau}, Y_t)$.
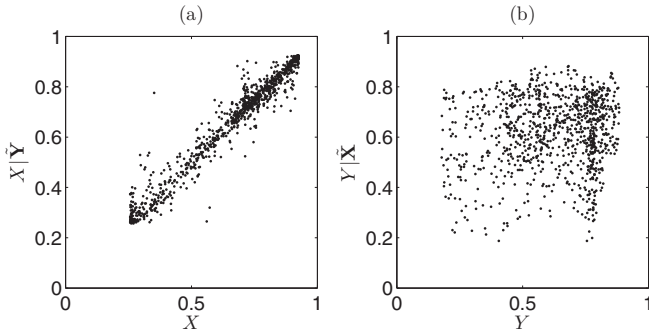
FIG. 5. Correlation plots between a given time series and its convergent cross-mapped estimate. (a) Reproduction of Fig. 3(c) from Ref. [9]. (b) Reproduction of Fig. 3(d) from Ref. [9].

An estimate of X using weights from a shadow manifold using this specific construction will be referred to as $X|(XY)$ and the correlation between this estimate and the original time series will be labeled $C_{X(XY)}$. See Fig. 6.

A difference in CCM correlations similar to Δ can be defined using the multivariate embedding. Consider $\Delta' = C_{Y(YX)} - C_{X(XY)}$. It might be argued, in close parallel to the arguments given in Ref. [9] for Δ, that an intuitive definition of "driving" might be captured by the sign of Δ'. For example, if $\Delta' < 0$, then the single time step of Y added to the delay vectors constructed from X create stronger estimators of X than the single time step of X added to the delay vectors constructed from Y do for Y. Thus, it might be argued that Y contains more "information" about X, which leads to the conclusion X drives Y. The example system and parameters (including E, τ, and L) described at the beginning of this section leads to $\Delta' \approx -3 \times 10^{-4}$, which agrees with the pre-

FIG. 7. (a) Reproduction of Fig. 2(a) using PAI rather than CCM. $\Delta' < 0 \ \forall B$, implying X "PAI drives" Y. (b) Reproduction of Fig. 2(c) using PAI rather than CCM. Δ' does not display the apparent erratic behavior seen in Δ in Fig. 2.

viously discussed conclusions of "X CCM causes Y" and "X drives Y." Using the multivariate embedding described above to explore "driving" relationships between pairs of time series will be referred to as *pairwise asymmetric inference* or PAI.

Consider a comparison of PAI and CCM given the linear example system from above, i.e., Eq. (4). Figure 7 shows Δ' as a function of B using of the same E, τ, L, and step sizes as was used to produce Fig. 2. $\Delta' < 0 \ \forall B$ in the domains shown in the figure. Thus, the sign of Δ' is in agreement with an intuitive notion of driving more consistently than Δ. Δ' is significantly smaller than Δ, which is expected since the correlation of X and Y with their "self-estimation" counterparts of $X|\tilde{\mathbf{X}}$ and $Y|\tilde{\mathbf{Y}}$ are initially very high, even without the multivariate additions. But if the concept of driving is determined solely on the sign of Δ', then, at least for the simple linear example presented here, PAI is a consistent qualifier of "driving."
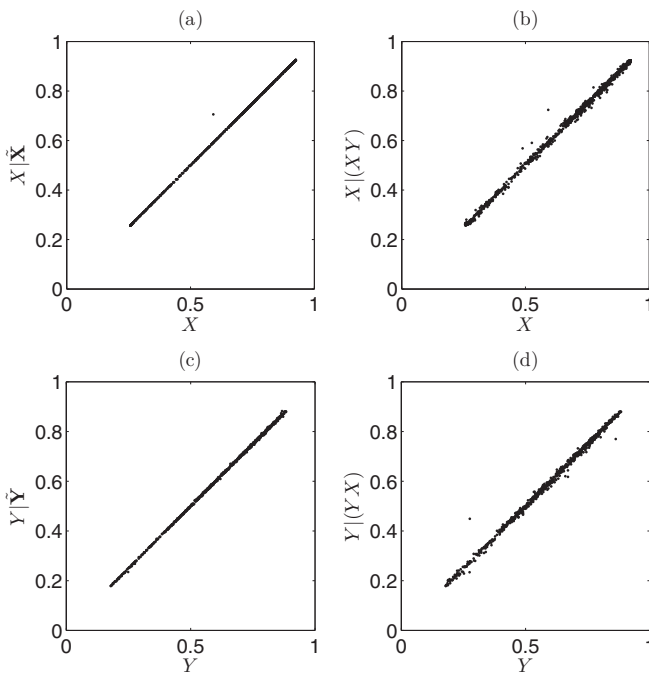

FIG. 6. Stronger correlations, as compared to Fig. 5, can be seen between a time series and its estimate when the shadow manifold includes points from the time series it is estimating.
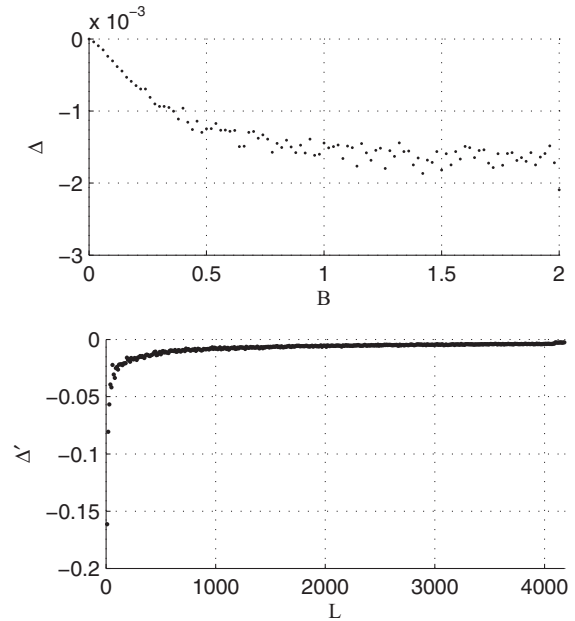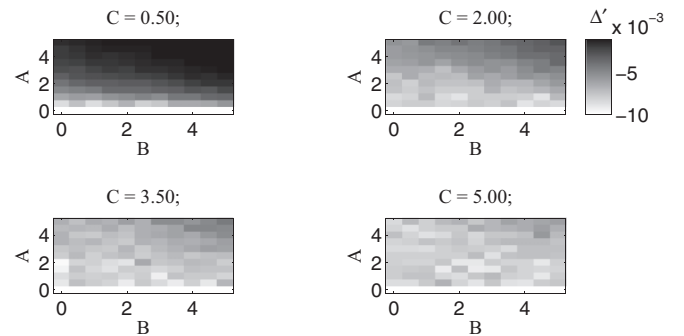

FIG. 8. Reproduction of Fig. 3 using PAI rather than CCM. $\Delta' < 0 \ \forall A, B, C$, implying X "PAI drives" Y consistently in the plotted parameter domains.
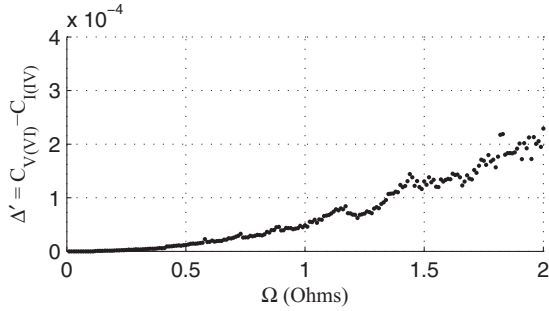
FIG. 9. Reproduction of Fig. 4 using PAI rather than CCM. $\Delta' < 0 \; \forall \Omega$, implying V "PAI drives" I consistently across the plotted domain for Ω.

Reproducing Fig. 2(c) using PAI shows an apparent reduction in some of the erratic behavior seen in CCM. See Fig. 7.

The conclusions that PAI agrees with intuition more consistently than CCM is also supported by the nonlinear example system, Eq. (6). Figure 8 plots Δ' as a function of A, B, and C using of the same E, τ, L, and step sizes that were used to produce Fig. 8. Again in contrast to the CCM figure, PAI agrees with intuition for all the plotted values of A, B, and C (i.e., $\Delta' < 0 \; \forall A, B, C$ in the domains shown).

Finally, a comparison of PAI and CCM for the RL circuit example leads to similar conclusions. The expectation is the V drives I; thus, it is expected that V PAI drives I, which implies $\Delta' = C_{V(VI)} - C_{I(IV)} > 0$ (which is what is observed). See Fig. 9.

V. EMPIRICAL DATA

The CCM and PAI methods can also be tested using empirical datasets with known causal relationships. Consider the measurements of altitude (in meters) paired with the mean temperature (from 1961 to 1990) at that altitude (in Celsius) [31]. Intuition is that the altitude drives the temperature in this scenario. Let A indicate the altitude and T indicate the mean temperature. For the CCM algorithm, the intuitive result is that $\Delta = C_{TA} - C_{AT} < 0$ and for PAI that $\Delta' = C_{T(TA)} - C_{A(AT)} < 0$. Figure 10 shows Δ and Δ' calculated for various library lengths, L, using $E = 4$ and $\tau = 1$. The CCM algorithm converges to $\Delta = 7.7 \times 10^{-3} > 0$ as the library length is increased, and the PAI algorithm converges to $\Delta' = -8.4 \times 10^{-3} < 0$ as the library length is increased. Thus, for this empirical data set, PAI agrees with intuition and CCM does not, as was the case for the synthetic data sets presented in the previous subsections.

VI. CONCLUSION

In this work we have shown that the recently introduced and frequently used convergent cross-mapping (CCM) method can lead to conclusions about a driver in a system that do not agree with intuition and the identified driver can depend on system parameters. For the examples presented in this article, PAI better indicates "driving" relationships that both agree with intuition and are consistent in the sense that the driver identification does not depend on system parameters.
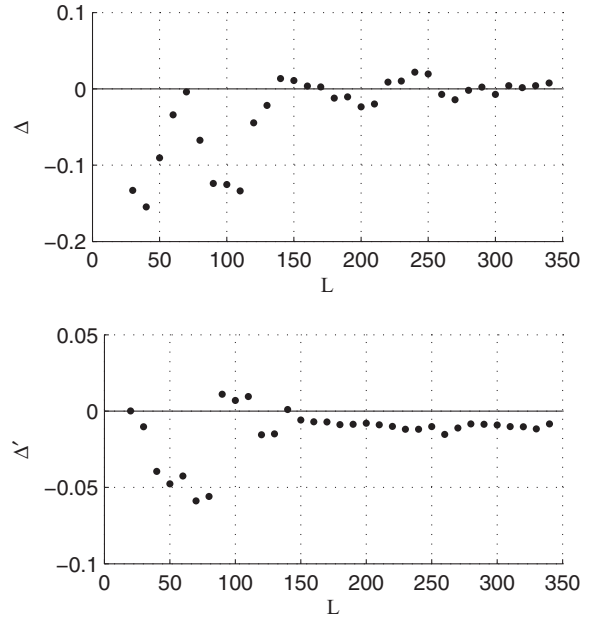

FIG. 10. The sign of Δ (Δ') indicates the CCM (PAI) driving between the altitude and mean temperature. The CCM algorithm converges to $\Delta > 0$ as the library length is increased, and the PAI algorithm converges to $\Delta' < 0$ as the library length is increased.

The introduced pairwise asymmetric inference method (PAI) attempts to keep the model-independent benefits of CCM while making it more robust. (SSR methods such as CCM and PAI are model independent, which may be seen as a benefit over Granger causality methods.) PAI may be useful exploratory data analysis. For example, PAI may help guide the development of physical causality models (e.g., by suggesting future experiments) in scenarios involving a large collection of simultaneous time series measurements of different variables in a system for which no *a priori* notions of causality in the system exist.

The given definition of Δ' in PAI, the sign of which is used to identify a dominant driver, is not without its own difficulties, however. For example, Δ' does not account for the differences between correlations between X and $X|\tilde{\mathbf{X}}$ and between Y and $Y|\tilde{\mathbf{Y}}$. Such differences may bias conclusions drawn from using Δ' without proper care. As a concrete example, consider the example system and parameters (including E, τ, and L) described at the beginning of Sec. IV. The value $\Delta' \approx -3 \times 10^{-4}$ was already discussed, but notice $C_{YY} - C_{XX} \approx 1.5 \times 10^{-3}$, indicating that Y is a better "self-estimator" than X (though both $C_{YY}, C_{XX} > 0.99$). How does this fact affect interpretations of the $\Delta' < 0$ result, which was that X PAI drives Y? Such questions are still open. It may be argued that a different measure may be more suitable, such as $\Delta'' = |C_{Y(YX)} - C_{YY}| - |C_{X(XY)} - C_{XX}|$. For this example, $\Delta'' \approx 3.9 \times 10^{-4}$, which does not agree with intuition, despite the agreement of both Δ and Δ'. There are still many open questions in the study of driving relationships among time series sets using state space methods.

Finally, care should be taken in any discussion of causality and especially in discussions of time series causality. We have made many statements about failure to agree with "intuition" in this paper. While some authors argue that any discussion of

causality will necessarily involve appeals to intuition [32], the possibility of intuition failing cannot be ignored completely.

Consider the RL circuit example of Sec. III C. The intuitive definition of causality was motivated by an example of the experimenter physically manipulating a voltage source to create the V and I times series. Suppose instead that two such experiments where conducted in isolation: one with an experimenter, Alice, physically manipulating a voltage source and measuring the current to create the V and I time series (call this set **VI**) and another, different experiment with an experimenter, Bob, physically manipulating a current source and measuring the voltage to create the V and I time series (call this set **IV**). Both **VI** and **IV** are handed to a third party, Charlie, who has no *a priori* knowledge of how the time series are created.

Intuition for Alice is V causes I and she believes **VI** supports that conclusion. Likewise, Bob believes **IV** supports his intuition that I causes V. Charlie, however, must rely on time series analysis alone to determine the causality in the system. The argument we present here is not that CCM causality is insufficient because it does not provide Charlie with a definitive answer (which it does not). Such a task is difficult and may not even be possible with time series analysis alone [32]. The main problem is that the CCM method, as it has been explored in this work, is inconsistent. Any method Charlie uses must be consistent if it is to be useful. Neither Alice nor Bob would change their causality conclusions if they changed their respective input frequencies [i.e., Ω in Eq. (8)]. However, if Charlie used the CCM method, his causality conclusions would depend on the frequency of the signal controlled by Alice (as seen in Fig. 4). Thus, CCM causality would not be a *consistent* tool for Charlie. PAI was shown to give consistent results for the considered examples but does not address the ambiguity identified in this example (i.e., PAI may not help Charlie determine if it was Alice or Bob that created his time series).

---

[1] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, Biol. Cybern. **85**, 145 (2001).

[2] J.-M. Dufour and E. Renault, Econometrica **66**, 1099 (1998).

[3] J.-M. Dufour, D. Pelletier, and É. Renault, J. Econometrics **132**, 337 (2006).

[4] T. J. Mosedale, D. B. Stephenson, M. Collins, and T. C. Mills, J. Clim. **19**, 1182 (2006).

[5] M. Eichler and V. Didelez, arXiv:1206.5246.

[6] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).

[7] C. W. J. Granger, Econometrica **37**, 424 (1969).

[8] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).

[9] G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Science **338**, 496 (2012).

[10] C. Granger, J. Econ. Dynamics Control **2**, 329 (1980).

[11] A. Kaiser and T. Schreiber, Phys. D (Amsterdam, Neth.) **166**, 43 (2002).

[12] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).

[13] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*, Wiley Series in Probability and Statistics (Wiley, New York, 2013).

[14] R. D. Pascual-Marqui, R. J. Biscay, J. Bosch-Bayard, D. Lehmann, K. Kochi, T. Kinoshita, N. Yamada, and N. Sadato, Frontiers Human Neurosci. **8**, 448 (2014).

[15] J. Ghosh, Z. Obradovic, J. Dy, Z.-H. Zhou, C. Kamath, and S. Parthasarathy, *Proceedings of the 2013 SIAM International Conference on Data Mining* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013).

[16] D. L. Roberts and S. Nord, Appl. Econ. **17**, 135 (1985).

[17] E. R. Deyle, M. Fogarty, C.-H. Hsieh, L. Kaufman, A. D. MacCall, S. B. Munch, C. T. Perretti, H. Ye, and G. Sugihara, Proceed. Nat. Acad. Sci. (USA) **110**, 6430 (2013).

[18] X. Wang, S. Piao, P. Ciais, P. Friedlingstein, R. B. Myneni, P. Cox, M. Heimann, J. Miller, S. Peng, T. Wang, H. Yang, and A. Chen, Nature **506**, 212 (2014).

[19] J. R. Anastas, Ph.D. thesis, University of Connecticut, 2013.

[20] A. Bozorgmagham and S. Ross, in SIAM (Society for Industrial and Applied Mathematics), Student Chapter at Virginia Tech, Oct 2013 (2013), http://www.siamsc.org.vt.edu/archive.php.

[21] G. Sugihara, B. Grenfell, R. M. May, P. Chesson, H. M. Platt, and M. Williamson, Philos. Trans. R. Soc. London B **330**, 235 (1990).

[22] G. Sugihara and R. M. May, Nature (London) **344**, 734 (1990).

[23] This definition differs slightly from the definition in Ref. [9], which uses the unsquared Pearsons correlation coefficient. We use the square of this value to avoid dealing with negative correlation values. This subtle change in the definition does not affect the conclusions drawn in Ref. [9], as can be seen in our reproduction of key plots from that work: Figs. 1 and 5.

[24] The software used to implement both the CCM and the PAI algorithms is available at https://github.com/jmmccracken/PAI.

[25] H.-G. Ma and C.-Z. Han, Frontiers Elect. Electron. Engin. China **1**, 111 (2006).

[26] I. Vlachos and D. Kugiumtzis, in *Topics on Chaotic Systems: Selected Papers from Chaos 2008 International Conference* (World Scientific, Singapore, 2009), p. 378.

[27] M. Small and C. Tse, Phys. D (Amsterdam, Neth.) **194**, 283 (2004).

[28] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).

[29] A. L. Lloyd, J. Theor. Biol. **173**, 217 (1995).

[30] In terms of Δ, convergence to a value κ means $\lim_{L \to \infty} |\Delta - \kappa| = 0$. As the "estimation skill" of the CCM algorithm increases, each of the CCM correlations, C_{XY} and C_{YX}, converge to some fixed value. Thus, Δ converges to some fixed value κ. Δ depends on the library length L because the CCM algorithm depends on L. However, the complicated algorithmic dependence makes formally solving this limit difficult. As such, convergence is determined from plots, following the method used in Ref. [9].

[31] This data is available online ([33], cf. http://webdav.tuebingen.mpg.de/cause-effect/[pair0001]) as measurements from 349 different locations across Germany (http://www.dwd.de/).

[32] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, Cambridge, 2000).

[33] K. Bache and M. Lichman, "UCI machine learning repository", (2013), http://archive.ics.uci.edu/ml/citation_policy.html.