Convergent Cross-Mapping and Causality Detection

James M. McCracken* and Robert S. Weigel†
School of Physics, Astronomy, and Computational Sciences
George Mason University
203 Planetary Hall, 4400 University Drive MS 3F3,
Fairfax,VA. 22030-4444
(Dated: June 30, 2014)

Convergent Cross-Mapping (CCM) is a technique for finding specific kinds of correlations between sets of times series data. It was introduced by Sugihara *et al.* [1] 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 in [1] do not, in general, agree with intuitive concepts of "driving", and as such, should not be considered indicative of causality. It is shown that CCM causality analysis implies causality is a function of system parameters for simple linear and nonlinear systems. For example, in a RL circuit both voltage and current can be identified as the driver depending on the frequency of the voltage. It is shown that CCM causality analysis can, however, be modified to identify asymmetric relationships between pairs of time series data that appear consistent with intuition for the systems for which CCM causality analysis provided non-intuitive driver identifications. To that end, we introduce "pairwise asymmetric inference" (PAI) and present examples of its use.

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. [2]), economics (e.g. [3, 4]), and climatology (e.g. [5]). General casual relationships in time series data are also being studied in an effort to understand causality itself (e.g. [6]).

To date, most techniques for "causal inference" in time series data fall into two broad categories, those related to transfer entropy and those related to Granger causality. Transfer entropy (introduced in [7]) and Granger causality (introduced in [8]) are known to be equivalent under certain conditions [9]. In this article, we investigate a casual inference technique, called Convergent Cross-Mapping (CCM), that was recently introduced by Sugihara *et al.* [1]. (Currently, there is no evidence that CCM is related to either transfer entropy or Granger causality.)

CCM is described as a technique that can be used to identify a necessary condition for 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 "nonseperable" [1]). Granger causality is not causality as it is typically understood in physics [10–12]. We show that a similar conclusion can be made regarding CCM causality.

CCM has been used to draw conclusions regarding the sardine-anchovy-temperature problem [1], confirm predictions of climate effects on sardines [13], compare the driving effects of precipitation, temperature, and solar radiation on the atmospheric CO_2 growth rate [14], and to quantify cognitive control in developmental psychology [15]. The technique has also been used to study the causality of respiratory systems in insects [16]. 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 have worked as expected in the above-cited papers despite its apparent failure in the examples presented in this article.

We begin with a review of the work of Sugihara *et al.* [1], including an extended evaluation of the coupled logistic map example. After showing examples where CCM analysis gives results that are inconsistent with intuitive notions of "driving", we introduce "pairwise asymmetric inference" (PAI) and use it to show that can be used to identify asymmetric relationships that are consistent with intuitive notions of "driving".

II. CONVERGENT CROSS-MAPPING

CCM is closely related to simplex projection [17, 18], 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 correlation coefficient [19] between the original time series Y and an estimate of Y made using the convergent cross-mapping with X, which is labeled as $Y|X$; ie. the CCM correlation is given as

$$C_{YX} = \left[\rho\left(Y, Y|X\right)\right]^2 \quad ,$$

where $\rho(A,B)$ is the Pearson correlation coefficient between A and B. Any pair of times series, X and Y, will

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

have two CCM correlations, C_{YX} and C_{XY}, which are compared to determine the CCM causality. For example, Sugihara *et al.* [1] define a difference of CCM correlations

$$\Delta = C_{YX} - C_{XY} \tag{1}$$

and use the sign of Δ to determine the CCM causality between X and Y [1]. The CCM algorithm is explained in more detail in Section II A.

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

A description of this algorithm is also available in [1] (supplementary materials). It is elucidating to partition the five (related) steps of the CCM algorithm:

1. the *shadow manifold* (defined in Section II A 1) 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. Use the nearest neighbors to create weights
4. Use the weights to estimate Y; that estimate is called $Y|\tilde{\mathbf{X}}$
5. Find the correlation between Y and $Y|\tilde{\mathbf{X}}$

The steps vary in complexity and are explained 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 to each point X_t that is constructed as $\tilde{\mathbf{X}}_t = \left(X_t, X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-(E-1)\tau}\right)$ (this vector is often called a *delay vector*). 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 (or *library length*).

2. Find Nearest Neighbors

The minimum number of points required for a bounding simplex in an E-dimensional space is $E+1$ [17, 18]. Thus, the set of $E+1$ nearest neighbors must be found for each shadow manifold point $\tilde{\mathbf{X}}_t$. For each $\tilde{\mathbf{X}}_t$, the nearest neighbor search results is a set of (ordered) distances $\{d_1, d_2, \ldots, d_{E+1}\}$ and an associated set of (ordered) times $\{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{E+1}\}$ (where the subscript 1 denotes the closest neighbor, 2 denotes the next closest neighbor, and so on). The distances from $\tilde{\mathbf{X}}_t$ are defined as

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

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

3. Create Weights

Each nearest neighbor will be used to find an associated weight. The weights are defined as

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

where

$$u_i = e^{-\frac{d_i}{d_1}}$$

and the normalization factor is given as

$$N = \sum_{j}^{E+1} u_j \quad .$$

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

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

$$Y_t|\tilde{\mathbf{X}} = \sum_i w_i Y_{t_i} \quad .$$

5. Find the Correlation

The CCM correlation is defined as

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

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

6. Simplified Two-Population Dynamics

Consider the example system used by Sugihara *et al.* [1]:

$$X_t = X_{t-1}\left(r_x - r_x X_{t-1} - \beta_{xy} Y_{t-1}\right) \tag{2}$$
$$Y_t = Y_{t-1}\left(r_y - r_y Y_{t-1} - \beta_{yx} X_{t-1}\right) \tag{3}$$

where the parameters $r_x, r_y, \beta_{xy}, \beta_{yx} \in \mathbb{R} \geq 0$. This pair of equations is a specific form of the two-dimensional coupled logistic map system, which is known to be chaotic for certain choices of parameters [20].

In this example, the CCM causality of this system is determined by sampling both the initial conditions and the parameters, calculating Δ, and demonstrating the necessary convergence. The dynamic parameters r_x and r_y are sampled from a 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 Figure 1.

Sugihara *et al.* consider convergence to be critically important to determining CCM causality, identifying it as "a key property that distinguishes causation from simple correlation" [1]. Figure 1 shows plots created with several 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.

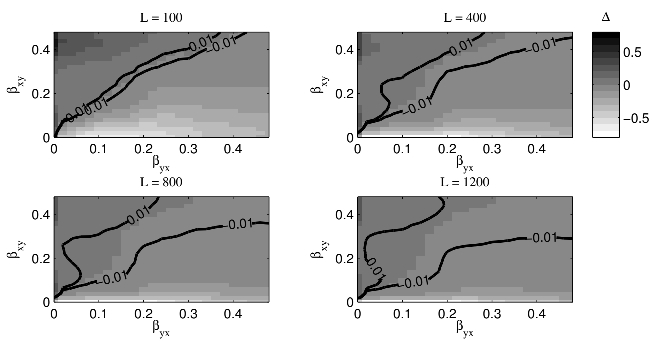

FIG. 1: The dependence of Eqn. 1 on β_{xy} and β_{yx}.

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.

The CCM algorithm depends on the embedding dimension E and the lag time step τ, which leads to Δ depending on E and τ. A dependence E and τ is a feature of most state space reconstruction (SSR) methods [21–23], so the E and τ dependence seen here is not unexpected. Sugihara *et al.* mention that "optimal embedding dimensions" are found using univariate SSR [1] (supplementary material), and other methods for determining E and τ for SSR algorithms can be found in the literature (e.g. [21, 23, 24]).

III. SIMPLE EXAMPLE SYSTEMS

The usefulness of the CCM algorithm in identifying causal or driving structure among sets of time series can be explored by using simple example systems. Each of the following examples intuitively supports the conclusion that X drives Y, and applying the CCM algorithm (with $E = 3$ and $\tau = 1$) leads to conclusions that do not always agree with this intuition. The CCM algorithm is used to calculate $\Delta = C_{YX} - C_{XY}$.

A. Linear Example

Consider the linear example dynamical system of

$$X_t = \sin(t) \tag{4}$$
$$Y_t = AX_{t-1} + B\eta_t, \tag{5}$$

with $A, B \in \mathbb{R} \geq 0$ and $\eta_t \sim \mathcal{N}(0,1)$. Specifically, consider $A, B \in [0, 10]$ in increments of 0.1. Figure 2 shows Δ for this example given a library length of $L = 2000$. The convergence of two specific points in Figure 2, $(A, B) = (2.6, 2.6)$ and $(A, B) = (3.0, 2.6)$, are shown in Figure 2 (c).

The expected conclusion of X drives Y would correspond to X CCM causes Y, which implies $\Delta < 0$. But, it can be seen from the plots above that the sign of Δ changes as A and B change. Given that the intuitive conclusion of X drives Y in Eqn. 4 does not depend on A and B, it would seem that Δ does not reliably reflect the intuitive conclusion in this linear example system.

Figure 2 (c) shows (for the two specific points plotted) that Δ is more negative at shorter library lengths but appears to converge to a point near zero as the library length is increased. The convergence of CCM correlations is emphasized [1], so the seemingly counter intuitive behavior of Δ (and C_{XY} and C_{YX}) in Figure **??** again seems to imply that the CCM correlations may not be a reliable measure of "driving" (at least not the intuitive definition) for this simple linear example system.

B. Non-Linear Example

Consider the non-linear example dynamical system of

$$X_t = \sin(t) \tag{6}$$
$$Y_t = AX_{t-1}\left(1 - BX_{t-1}\right) + C\eta_t, \tag{7}$$

with $A, B, C \in \mathbb{R} \geq 0$ 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 given a library length of $L = 2000$. Just as in the previous linear example, the expectation for this example system is that $\Delta < 0$ 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 non-linear example leads to a similar conclusion to the previous linear

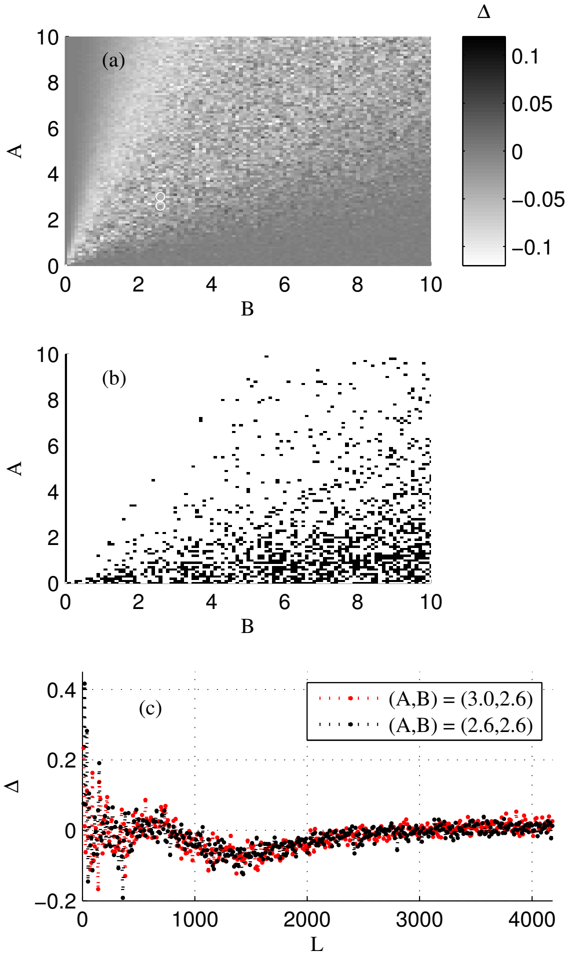

FIG. 2: The sign of Δ, and thus the CCM causality, depends on A and B. (a) Δ is calculated as described in the text; (b) Two color version of (a) where black indicates $\Delta > 0$, i.e. Y CCM causes X, and white indicates $\Delta < 0$, i.e. X CCM causes Y; (c) The convergence of points $(A, B) = (2.6, 2.6)$ and $(A, B) = (3.0, 2.6)$ marked in (a).

example; i.e. Δ does not appear to reliably reflect intuitive notions of driving.

IV. RL CIRCUIT EXAMPLE

Both of the previous examples included a noise term, η_t (which was not averaged over in any way). The failure of CCM correlations to meet expectations in the previous examples may be considered a failure of the algorithm's ability to deal with noise. This idea can be investigated by considered a system without stochastic noise terms. To that end, consider the familiar physical system of a electrical circuit containing a resistor and an inductor.

The continuous system is

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

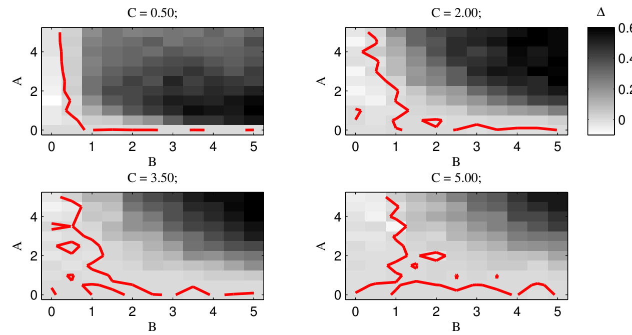

FIG. 3: The sign of Δ, and thus the CCM causality, depends on A, B, and C. The contour line indicates where $\Delta = 0$ and helps illustrate how the A and B dependence of Δ also depends on C.

where I is the current at time t, $V(t)$ is the voltage at time t, $R(t)$ is the resistance at time t, and L is the inductance (which is also constant in these examples), and it can be approximated as

$$\dot{I} = \frac{V(t)}{L} - \frac{R(t)}{L} I \Rightarrow I_{t+1} - I_t = \frac{V_t}{L} - \frac{R_t}{L} I_t. \tag{9}$$

Rearranging leads to

$$I_{t+1} = \frac{V_t}{L} + I_t \left(1 - \frac{R_t}{L}\right), \tag{10}$$

$$V_t = L\left(I_{t+1} - I_t\left(1 - \frac{R_t}{L}\right)\right), \tag{11}$$

and

$$R_t = L\left(I_t - I_{t+1} + \frac{V_t}{L}\right). \tag{12}$$

All of the plots of I seen below are produced by using MATLAB's *ode45* to solve Eqn. 8 (i.e. not using the discrete approximation shown). The time series $V(t)$ and $R(t)$ are created by defining values at fixed points and using linear interpolation (i.e. MATLAB's *interp1*) to find the time steps required by the ODE solver.

A. Changing V(t)

Consider the situation where $L = 10H$ and $R(t) = R_0 = 5\Omega$ are constant. Physical intuition is that V drives I, so we expect to find V CCM causes I (i.e. $C_{VI} > C_{IV}$ or $\Delta = C_{VI} - C_{IV} > 0$). For this example, the voltage is described by

$$V(t) = \sin(\Omega_v t), \tag{13}$$

where Ω_v is the frequency.

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

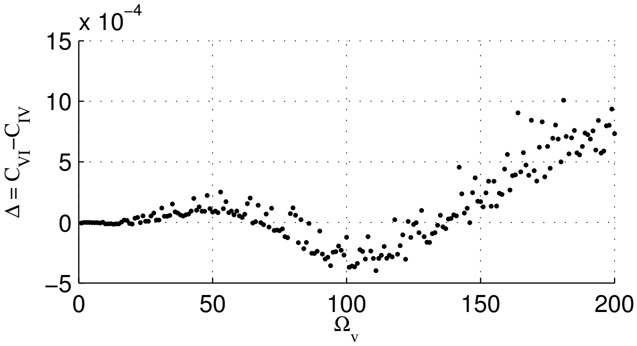

FIG. 4: Δ depends on Ω_v. Notice that the sign of Δ implies both V CCM causes I and I CCM causes V for different values $\Omega_v \in [0.01, 2.0]$.

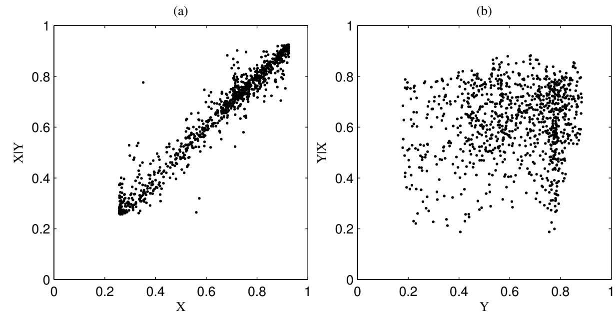

FIG. 5: Correlation plots between a given time series and its convergent cross-mapped estimate. (a) Reproduction of Figures 3C from [1]. (b) Reproduction of Figures 3D from [1].

used to calculate $\Delta = C_{VI} - C_{IV}$, which is plotted in Figure 4. It appears that Δ does not consistently agree with intuition in this example either. Notice that, unlike the previous examples, there are no noise terms in this system.

The resistance and inductance of the circuit are fixed and the voltage is varied from 1×10^{-2} to 2.0 volts in discrete steps of 0.01 volts as described by Eqn. 13 (with a fixed Ω_v). This traditional experiment is straightforward to imagine. 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 Figure 4, it appears that the voltage does not consistently "CCM cause" the current as Ω_v is changed. Thus, it seems as though CCM causality does not agree with physical causality (at least for the specific RL circuit experiment described here).

It may be argued that the relatively small values (as compared to the previous examples) of Δ plotted in Figure 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. 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.

V. PAIRWISE ASYMMETRIC INFERENCE

Consider the example system of Eqn. 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 Figure 3C and D of [1] (with $E = 2$, $\tau = 1$, and $L = 1000$). Plots of the correlation between X and $X|\tilde{\mathbf{Y}}$ (i.e. X estimated using the weights found from the shadow manifold of Y), as well as, Y and $Y|\tilde{\mathbf{X}}$ are shown below. This leads to $\Delta = C_{YX} - C_{XY} \approx 0.11 - 0.97 = -0.86$. Notice $\Delta < 0$ implies X CCM causes Y, which agrees with intuition because $\beta_{xy} = 0 < \beta_{yx} = 0.32$.

But, the correlations shown in Figure 5 are not the only correlations that can be tested. Consider, for example, the correlation between X and the corresponding $X|X$, 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 [13]. For example, an $E+1$ dimensional point in the a multivariate shadow manifold constructed using both X and Y may be defined as $\tilde{\mathbf{X}}_t = (X_t, X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-(E-1)\tau}, Y_t)$. 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 Figure 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 [1] 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 previously 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. Eqn. 4. Figure 7 plots Δ' as a function of A and B using of the same E, τ, L, and step sizes as was used to produce Figure 2. $\Delta' < 0\ \forall A, B$ in the domains shown in the figure. Thus,

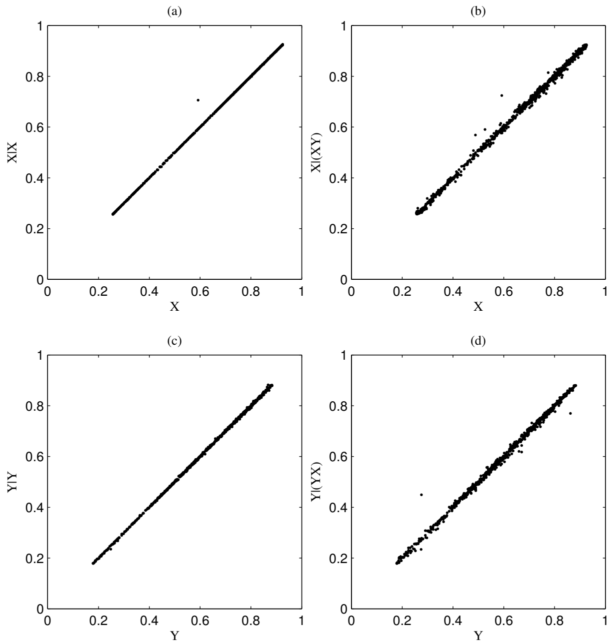

FIG. 6: Stronger correlations, as compared to Figure 5, can be seen between a time series and its estimate when the shadow manifold includes points from the time series it is estimating.

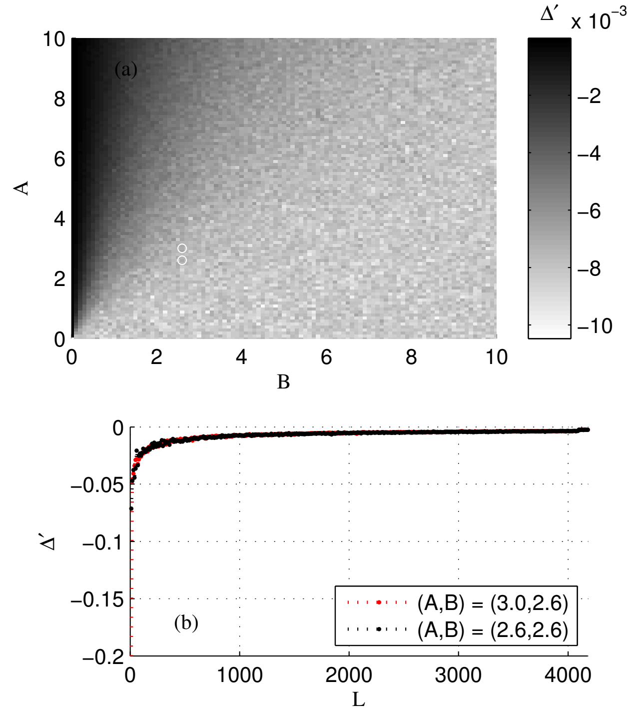

FIG. 7: (a) Reproducing Figure 2 (a) using PAI rather than CCM. Notice that $\Delta' < 0\ \forall A, B$ implying X "PAI drives" Y. (b) Reproducing Figure 2 (c) using PAI rather than CCM. Notice that Δ' does not display the apparent erratic behavior seen in Δ in Figure 2.

it appears Δ' is in agreement with an intuitive definition of driving more consistently than Δ. Notice, Δ' is significantly smaller than Δ, which is expected since the correlation of X and Y with their "self estimation" counterparts of $X|X$ and $Y|Y$ 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 appears to be a consistent qualifier of "driving".

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

The conclusions that PAI agrees with intuition more consistently than CCM is also supported by the non-linear example system, Eqn. 6. Figure **??** plots Δ' as a function of A, B and C using of the same E, τ, L, and step sizes as was used to produce Figure 8. Again in contrast to the CCM figure, PAI seems to agree 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 Figure 9.

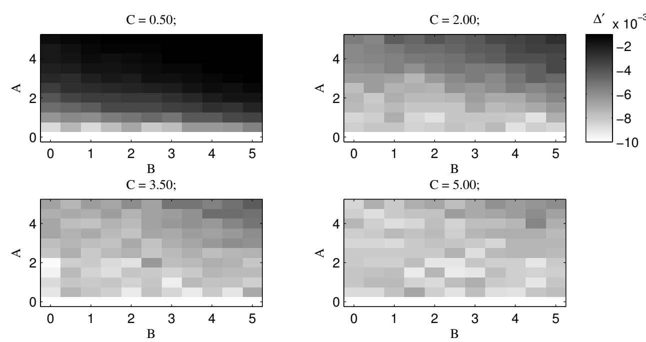

FIG. 8: Reproducing Figure 3 using PAI rather than CCM. Notice that $\Delta' < 0\ \forall A, B, C$ implying X "PAI drives" Y consistently in the plotted parameter domains.

VI. CONCLUSION

The examples presented in this article have shown that PAI can be used to indicate "driving" relationships that agree with intuition. Such information can be useful, but should not be confused with physical causality. However, even without notions of causality, PAI may be useful exploratory data analysis. For example, PAI may help guide the development of physical causality models (e.g.

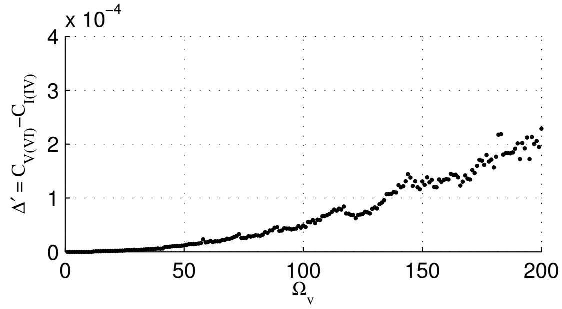

FIG. 9: Reproducing Figure 4 using PAI rather than CCM. Notice that $\Delta' < 0 \; \forall \Omega_v$ implying V "PAI drives" I consistently across the plotted domain for Ω_v.

by suggesting future experiments) in scenarios involving a large collection of simultaneous time series measurements of different variables in a system but no *a priori* notions of causality in the system.

It should also noted that the basic idea behind CCM may be useful in studying time series driving independently of the specific implementation using Δ (or the sign of Δ). SSR methods are model-independent, which may be seen as a benefit over the popular Granger causality measures.

PAI attempts to keep the benefits of CCM while making it slightly more robust. But, the given definition of Δ' is not without its own difficulties. For example, Δ' does not account for the differences between correlations between X and $X|X$ and Y and $Y|Y$. 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 Section V. 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'. Studying driving relationships among time series sets using state space methods is still full of open questions.

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 [25], the possibility of intuition failing cannot be ignored completely.

Consider the RL circuit example of Section IV. 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 the current 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 and no communication channels with Alice or Bob.

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 doesn't). Such a task is difficult and may not even be possible with time series analysis alone [25]. 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_v in Eqn. 13). 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 attempts to remedy this inconsistency, but it remains an open question to determine if PAI is consistent outside of the examples shown here.

[1] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Science **338**, 496 (2012).
[2] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, Biological Cybernetics **85**, 145 (2001).
[3] J.-M. Dufour and E. Renault, Econometrica , 1099 (1998).
[4] J.-M. Dufour, D. Pelletier, and É. Renault, Journal of Econometrics **132**, 337 (2006).
[5] T. J. Mosedale, D. B. Stephenson, M. Collins, and T. C. Mills, Journal of climate **19** (2006).
[6] M. Eichler and V. Didelez, arXiv preprint arXiv:1206.5246 (2012).
[7] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[8] C. W. Granger, Econometrica: Journal of the Econometric Society , 424 (1969).
[9] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).
[10] C. Granger, Journal of Economic Dynamics and Control

2, 329 (1980).

[11] Y. Liu and M. T. Bahadori, University of Southern California, http://www-bcf. usc. edu/˜ liu32/granger. pdf (12. 03.2013) (2012).

[12] D. L. Roberts and S. Nord, Applied Economics **17**, 135 (1985).

[13] E. R. Deyle, M. Fogarty, C.-h. Hsieh, L. Kaufman, A. D. MacCall, S. B. Munch, C. T. Perretti, H. Ye, and G. Sugihara, Proceedings of the National Academy of Sciences **110**, 6430 (2013).

[14] X. Wang, S. Piao, P. Ciais, P. Friedlingstein, R. B. Myneni, P. Cox, M. Heimann, J. Miller, S. Peng, T. Wang, *et al.*, Nature (2014).

[15] J. R. Anastas, *Individual Differences on Multi-Scale Measures of Executive Function*, Ph.D. thesis, University of Connecticut (2013).

[16] A. Bozorgmagham and S. Ross, in *SIAM (Society for Industrial and Applied Mathematics), Student Chapter at Virginia Tech, Oct 2013* (2013).

[17] G. Sugihara, B. Grenfell, R. M. May, P. Chesson, H. M. Platt, and M. Williamson, Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **330**, 235 (1990).

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

[19] This definition differs slightly from the original definition in [1], which just uses 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 [1], as can be seen in our reproduction of key plots from that work: Figure 1 and Figure 5.

[20] A. L. Lloyd, Journal of Theoretical Biology **173**, 217 (1995).

[21] H.-g. Ma and C.-z. Han, Frontiers of Electrical and Electronic Engineering in China **1**, 111 (2006).

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

[23] M. Small and C. Tse, Physica D: Nonlinear Phenomena **194**, 283 (2004).

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

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