Convergent Cross-Mapping and Causality Detection

McCracken,Weigel

February 6, 2014

Abstract

Convergent Cross-Mapping is a technique, introduced by Sugihara *et al.* [Detecting Causality in Complex Ecosystems, 2012], reported to be "a necessary condition for causation" capable of distinguishing causality from correlation in sets of time series data [2012]. We will show that CCM correlations do not in general agree with intuitive concepts of "driving" and "response", and as such, relationships among CCM correlations should not be considered indicative of causality. It is shown that CCM correlations can, however, be used to identify asymmetrical prediction capability between pairs of time series data. We introduce a 2-vector called the "directed correlation" and present several examples of its use for identifying asymmetrical prediction capability. The sensitivity of CCM correlations (and consequently, directed correlations) on embedding dimensions and lag times will be discussed and mitigation will be presented.

1 Introduction

2 Convergent Cross-Mapping

Convergent cross-mapping (CCM) is introduced in [**?**, **?**] by Sugihara *et. al.* . CCM is described as a technique 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 "nonseperable"). The authors state that CCM is a "necessary condition for causation". It is well known [**?**] that Granger causality is unrelated to causality as it is typically understood in physics. It will be shown that a similar conclusion can be drawn regarding CCM causality.

CCM is closely related to simplex projection, introduced by Sugihara and May in [**?**], which predicts a point in the times series X at a time $t+1$, i.e. 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[1] between the original time series Y and estimate of Y made using the convergent cross-mapping with X, which is labeled as $Y|X$; i.e. 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 have two CCM correlations, C_{YX} and C_{XY}, which are compared to determine the CCM causality. For example, Sugihara *et al.* define a difference of CCM correlations

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

and use the sign of Δ to determine the CCM casuality between X and Y [**?**]. The CCM algorithm is explained in more detail in Appendix **??**.

If X can be estimated from the shadow manifold of Y better than Y can be estimated from the shadow manifold of X (i.e. $\Delta < 0$), then X is said to "CCM cause" Y.

2.1 Simplified Two Population Dynamics

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

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

where $r_x, r_y, \beta_{xy}, \beta_{yx} \in \mathbf{R}$. This pair of equations is a specific form of the two-dimensional coupled logistic map system, which is known to be chaotic [**?**].

[1] This definition differs slightly from the original definition in [**?**], which just uses Pearon's 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 [**?**], as can be seen in our reproduction of key plots from that work.

In this example, the CCM causality of this system is determined by sampling both the initial conditions and the dynamics parameters, calculating Δ, and demonstrating the nessecary convergence. The dynamic parameters r_x and r_y are sampled from a normal distributions $N(\mu_{rx}, \sigma_{rx})$ and $N(\mu_{ry}, \sigma_{ry})$, respectively. The initial conditions X_0 and Y_0 are also sampled for normal distributions, specifically $N(\mu_{x0}, \sigma_{x0})$ and $N(\mu_{y0}, \sigma_{y0})$. The coupling parameters β_{xy} and β_{yx} are then varied over the interval $[10^{-6}, 1]$ (in steps of **????**) to produce the plots seen in Figure **??**.

Sugihara *et al.* consider convergence to be critically important to determining CCM casuality, identifying it as "a key property that distinguishes causation from simple correlation" [**?**]. Figure **??** shows plots created with several different library lengths to illustrate the convergence of Δ for this example. The idea is that $\beta_{xy} > \beta_{yx}$ intuitativly implies Y "drives" X more than X "drives" Y. Stated more formally, $\beta_{yx} > \beta_{xy} \Rightarrow \Delta > 0$, 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 nessecarily related to causality as it is typically understood in physics.

Figure 1: These plots show the dependence of Eqn. **??** on β_{xy} and β_{yx}. See the text for details on how these plots were created along with a discussion of the interpretation of these plots in terms of the CCM causality.

3 Embedding Dimension and Lag Time

The CCM algorithm depends on the embedding dimension E and the lag time step τ (see Appendix **??**). Consider the simplified two population system (Eqn. **??**) with $r_x = 3.8$, $r_y = 3.5$, $\beta_{xy} = 0.01$, $\beta_{yx} = 0.2$, $X_0 = 0.4$, and $Y_0 = 0.2$ with X and Y library lengths of $L = 1000$. Figure **??** shows the effect of varying E and τ on Δ for this system. E is varied over the interval $[2, 20]$ (in steps of 1), and τ is varied over the interval $[1, 50]$ (also in steps of 1). This figure makes it clear that a statement of CCM causality (and, subsequently, that statement's agreement with "intuition") depends strongly upon how the CCM technique is used.

Figure 2: The determination of CCM causality in a system is dependent on the CCM parameters of embedding dimension E and lag time step τ. This plot show the dependence of Eqn. **??** on E and τ for the example system discussed in the text.

A dependence E and τ is feature of most state space reconstruction (SSR) methods [**?**, **?**, **?**, **?**]. CCM is related to state space reconstruction [**?**], so the E and τ dependence seen here is not unexpected. Sugihara *et al.* do not discuss in depth how to determine E and τ, but they do mention that "optimal embedding dimensions" are found using univariate SSR [**?**].

Define the l-lagged autocorrelation A^X_l of X_t as $A^X_l = \rho(X_t, X_{t-l})$ given $t \in [l, L]$ and $l \in [0, L/2]$ where L is the library length. The lag time step τ will be set equal to the value of l that yields the maximum l-lagged autocorrelation; i.e.

$$\tau = l \mid \left|A^X_l\right| = \max_l \left|A^X_l\right| \quad . \tag{4}$$

In all of the examples discussed so far, this method leads to setting $\tau = 1$. Setting τ independently of E may lead to non-optimal choices for both due to their interdependence [**?**]. This method, however, is convinient, and optimatily is not required for the present study.

The embedding dimension E will be determined using a method very similar to the "false nearest neighbor" method commonly used in SSR [**?**]. The details of our method are outlined in Appendix **??**. The method requires setting a tolerance level δ and set of time steps to be checked T. All of the embedding dimensions used in this work were found using $\delta = 10^{-6}$ and $T = L/2$ where L is the library length of the time series being investigated.

4 Directed Correlation

The use of Eqn. **??** to study CCM causality hides information present in the individual CCM correlations that may be of some use. For example, $\Delta = 0.01$ could result from either $C_{XY} = 0.98$ and $C_{YX} = 0.99$ or $C_{XY} = 0.05$ and $C_{YX} = 0.06$. These two different sets of CCM correlations indicate very different relationships between the times series and their shadow manifold predicted counterparts.

Notice that the pair (C_{XY}, C_{YX}) as a point in the unit square. The "directed correlation", $\vec{D}$, is defined as a 2-vector on the unit square from the origin to the point (C_{XY}, C_{YX}). The magnitude of $\vec{D}$ is

$$D_m = \sqrt{C^2_{XY} + C^2_{YX}} \quad , \tag{5}$$

and its angle with the base of the square is

$$D_\theta = \arctan\left(\frac{C_{YX}}{C_{XY}}\right) \quad . \tag{6}$$

As an example, the directed correlation for the simplified two population example discuss in Section **??** is shown in Figure **??**. The same sampling procedures described in that section were used and the coupling constants were set to $\beta_{xy} = 0.02$ and $\beta_{yx} = 0.1$. Figure **??** shows $\vec{D}$ lies in the lower triangular quadrant

Figure 3: The directed correlation can be used to study the CCM causality in a system. See the text for an explanation of the dynamics that yield the $\vec{D}$ shown here.

on the unit square, which indicates X CCM causes Y. This conclusion, as discussed in Section **??**, seems to agree with intuition. Notice that a $\vec{D}$ aligned along the unit square bisector (i.e. the line $C_{YX} = C_{XY}$) would indicate no CCM causality in the system.

5 Driven System Example: RL Circuit

A circuit containing only a resistor and a inductor is described by the following differiential equation:

$$\frac{dI}{dt} = \frac{V(t)}{L} - \frac{R(t)}{L} I(t) \quad , \tag{7}$$

where $I(t)$ 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. The voltage and resistance are considered to be under the control of the experimenter. Consider, for example, that the resistance R is provided by a potentiometer controlled by the experimenter. The inductance is considered fixed. This scenario outline implies that the times series for V and R should be considered the driving time series and I is the response times series.

Consider the scenario where the voltage is decribed by a simple sine wave, i.e.

$$V(t) = \sin(t) \tag{8}$$

and the resistance is given by

$$R(t) = A \sin(\omega t + \phi) + R_0 \quad . \tag{9}$$

6 Causality ...?

7 Conclusion

8 Appendix A: CCM Algorithm

It is elucidating tp partition the CCM algorithm into five distinct (though related) steps:

1. Create the shadow manifold for X, called M_X
2. Find the nearest neighbours to $X(t)$ in M_X
3. Use the nearest neighbours to create weights
4. Use the weights to estimate $Y(t)$, called $Y(t)|M_X$
5. Find the correlation between $Y(t)$ and $Y(t)|M_X$

The steps vary in complexity and are explained in more detail below.

8.1 Create M_X

Given an embedding dimension E, the shadow manifold of X, called M_X, is created by associating an E-dimensional vector to each point $X(t)$ that is constructed as $\vec{X}(t) = (X(t), X(t-\tau), X(t-2\tau), \ldots, X(t-(E-1)\tau)$ (this vector is often called a "delay vector"). The first such vector is created at $t = 1+(E-1)\tau \equiv t_s$ and the last is at $t = L \equiv t_l$ where L is the time series length (or "library length").

8.2 Find Nearest Neighbours

The minimum number of points required for a bounding simplex in an E-dimensional space is $E+1$ (find a non-Sugihara reference for this statement). Thus, the nearest neighbour search results in a set of distances $\{d_1, d_2, \ldots, d_{E+1}\}$ and an associated set of times $\{t_1, t_2, \ldots, t_{E+1}\}$ (where the subscript 1 denotes the closest neighbour, 2 denotes the next closest neighbour, etc.). The distances from $\vec{X}(t)$ are defined as

$$d_i = D\left(\vec{X}(t), \vec{X}(t_i)\right) \quad ,$$

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

8.3 Create Weights

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

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

The weights are defined as

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

where the normalization factor is given as

$$N = \sum_j u_j \quad .$$

8.4 Find $Y|X$

A point $Y(t)$ in Y can be estimated using the (normalized) distances to the points in X using the weights calculated above. This estimate is calculated as

$$Y(t)|M_X = \sum_i w_i Y(t_i) \quad .$$

8.5 Find the Correlation

The CCM correlation is defined as

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

where $\rho_{A,B}$ is the standard Pearson's correlation coefficient between A and B. It can be seen from the above algorithm that $X = Y \Rightarrow C_{YX} = C_{XY}$, but in general, $C_{YX} \neq C_{XY}$.

9 Appendix B: Determining E

The algrotihm we use for determining the mebedding dimension E can be outlined as follows:

1. Embed X in a manifold $\tilde{X}$ with dimension E
2. Record the weights $\{w_i^E\}$ of the $E+1$ nearest neighbors to $\tilde{X}_t$
3. Repeat steps 1 and 2 for $E = E+1$
4. Stop when $\exists w_i^E < \delta$
5. Record E (i.e. the embedding dimension for which the above iteration is halted) as E_t^H
6. Repeat steps 1-4 for all $t \leq T$ where T is dependent on the library length L of X
7. Find the mode of $\{E_t^H\}$

This algorithm requires the selection of some δ and T. Both parameters are usually choosen to be "sufficently small" while still being computionationally tractable.

Notice that all subjectivity can be removed from this algorithm by recording the locations $\tilde{t}_i^E$ associated to each w_i^E and setting the halt criterion in Step 4 to "Stop when $\{\tilde{t}_i^E\} \not\subset \{\tilde{t}_i^{E+1}\}$"; i.e. stop when, for example, a nearest neighbor of $\tilde{X}_t$ given $E = 3$ is no longer a nearest neighbor when $E = 4$. The set of time steps T can be set to $\tilde{L}$, the library length of $\tilde{X}$. These modifications will remove the need to set any (possibly arbitrary) parameters such as δ and T, but the algrotihm will become much more comutationally expensive.

It is assumed that the lag time step τ is determined (e.g. with the autocorrelation method discussed in the main text) before this algrotihm begins. This algorithm can be modified to concurrently determine τ and E by repeating Steps $1-3$ for succesive iterations of τ and recording the halting values of both E and τ in Step 5. But, again, such a modification will significantly increase the computional cost of the algrotihm.