Convergent Cross-Mapping and Causality Detection

McCracken,Weigel

April 24, 2014

Abstract

Convergent Cross-Mapping is a technique, introduced by Sugihara *et al.* [19], reported to be "a necessary condition for causation" capable of distinguishing causality from correlation in sets of time series data. 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 asymmetric relationships between pairs of time series data. We introduce a 2-vector called the "pairwise asymmetric inference" (PAI) and present examples of its use in inferring relationships within complex systems. The sensitivity of CCM correlations (and consequently, PAI) on embedding dimensions and lag times will be discussed and mitigation will be presented.

1 Introduction

Modern time series analysis includes a handful of techniques meant to discren "driving" or "cause and effect" relationships between different data sets. These techniques have found application in a wide range of fields including neuroscience (e.g. [8]), economics (e.g. [3, 2]), climatology (e.g. [13]), and others. General casual relationships in time series data are also being studied in an effort to understand causality itself (e.g. [4]).

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 [15]) and Granger causality (introduced in [5]) are known to be equivalent under certain conditions [1]. In this article, we investiagte a casual inference technique, called Convergent cross-mapping (CCM), that was recently introduced by Sugihara *et al.* in [19]. Currently, there is no evidence that CCM is related to either transfer entropy or Granger causality.

We will begin with a review of the work of Sugihara *et al.*, including working through the coupled logistic map example presented in [19]. We will then illustrate the dependence of CCM correlations on the embedding dimension and lag time parameters of the CCM algorithm. Finally, we will introduce "pairwise asymmetric inference" (PAI) and use it to show that, even though CCM causality may not be physical caulsality, it can still be a useful tool in the analysis of complex time series data.

2 Convergent Cross-Mapping

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" [19]). The authors state that CCM is a "necessary condition for causation". It is well known [6, 10, 14] that Granger causality is not 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 [18, 17], 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 labelled as $Y|X$; i.e. the CCM correlation is given as

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

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

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

---

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

and use the sign of Δ to determine the CCM causality between X and Y [19]. The CCM algorithm is explained in more detail in Appendix 8.

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

2.1 Simplified Two Population Dynamics

Consider the example system used by Sugihara *et al.* in [19]:

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

where $r_x, r_y, \beta_{xy}, \beta_{yx} \in \mathbf{R} \geq 0$. This pair of equations is a specific form of the two-dimensional coupled logistic map system, which is known to be chaotic [11].

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 necessary convergence. The dynamic parameters r_x and r_y are sampled from a normal distributions $N\left(\mu_{rx}, \sigma_{rx}\right)$ and $N\left(\mu_{ry}, \sigma_{ry}\right)$, respectively. The initial conditions X_0 and Y_0 are also sampled for normal distributions, specifically $N\left(\mu_{x0}, \sigma_{x0}\right)$ and $N\left(\mu_{y0}, \sigma_{y0}\right)$. 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 causality, identifying it as "a key property that distinguishes causation from simple correlation" [19]. Figure **??** 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.
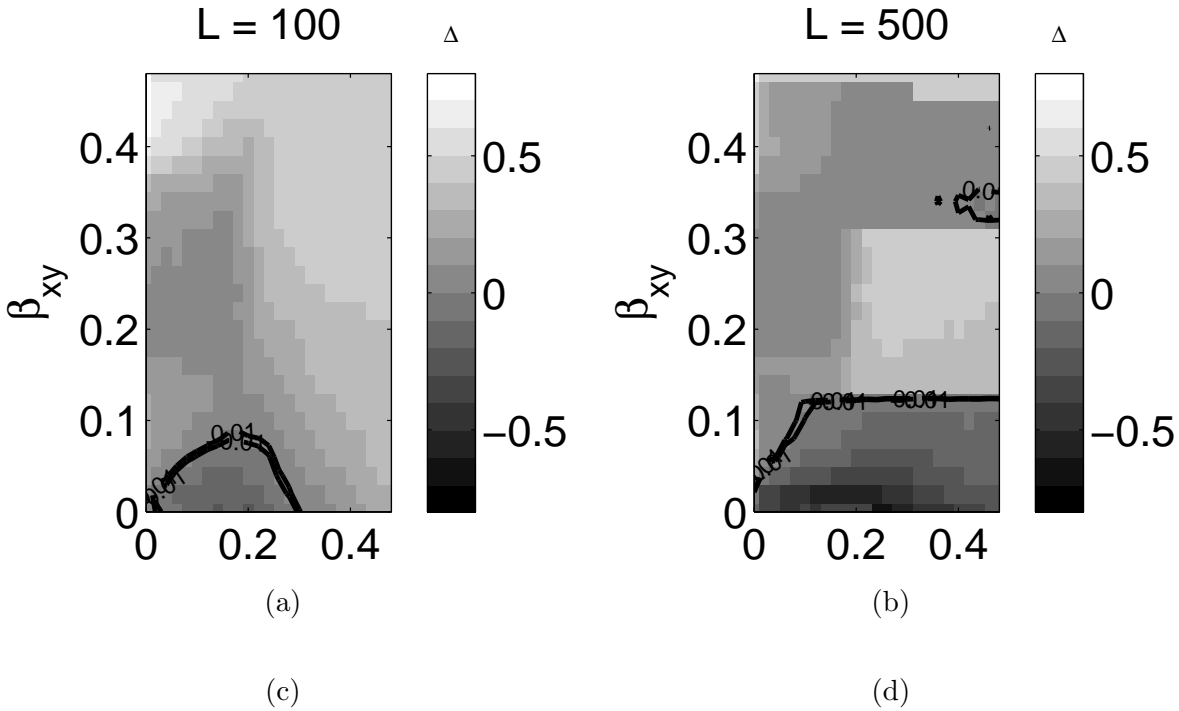

(a)

(b)

(c)

(d)

Figure 1: These plots show the dependence of Eqn. 1 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.

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.

3 Embedding Dimension and Lag Time

The CCM algorithm depends on the embedding dimension E and the lag time step τ (see Appendix 8). Consider the simplified two population system (Eqn. 2) 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 $L = 2000$. 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

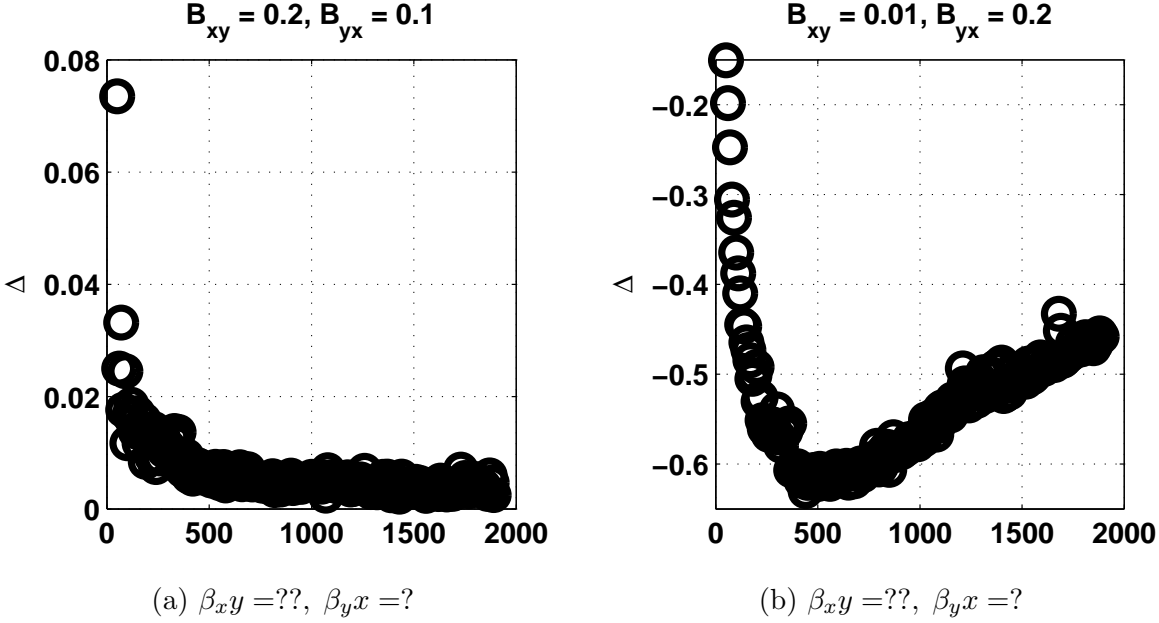(a) $\beta_x y = ??, \ \beta_y x = ?$ (b) $\beta_x y = ??, \ \beta_y x = ?$

Figure 2: Reference figure.

[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. A dependence E and τ is a feature of most state space reconstruction (SSR) methods [12, 20, 16]. CCM is related to state space reconstruction [19], 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 [19] (supplementary material).

Define the l-lagged autocorrelation A_l^X of X_t as $A_l^X = \rho\left(X_t, X_{t-l}\right)$ 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.

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

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 [16]. This method, however, is convenient, and optimality 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 [9]. The details of our method are outlined in Appendix 9. 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 PAI

The use of Eqn. 1 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}) is a point in the unit square. The "pairwise asymmetric inference" (PAI), $\vec{P}$, 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

$$P_m = \sqrt{C_{XY}^2 + C_{YX}^2} \quad , \tag{5}$$

and its angle with the base of the square is

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

As an example, the PAI for the simplified two population example discuss in Section 2.1 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 $\vec{P}$ lies in the lower triangular quadrant on the unit square, which indicates X CCM causes Y. This conclusion, as discussed in Section 2.1, 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 [7]:

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

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. 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 described by a simple sine wave, i.e.

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

and the resistance is given by

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

Consider $A = R_0 = 0.5$ with $\omega = 1$ and $\phi = 0$. These values lead to the PAIs $\vec{P}$ in Figure ??. These PAIs seem to imply CCM causality is at least representative of physical causality. Notice, however, that changing the phase, ω, of the resistance $R(t)$ can lead to situations that do not support the intuitive result of R CCM causes I. For example, consider $A = R_0 = \omega = 0.5$. These values lead the the PAIs in Figure ??.

Consider the situation where the voltage is driven with the sinusoidal signal shown in Eqn. 8 and the resistance is driven by the triangular pulse

$$??? \tag{10}$$

where .... Such a scenario leads to the PAIs shown in Figure ??. These results imply V CCM causes I as expected, but they also imply I CCM causes R and V CCM causes R which obeys transitivity but seems counter-intuitive. These results seem to imply that the voltage is the primary driver of the system and can, under certain scenarios, even be seen as a "driver" of the resistance through the current. Consider approximating Eqn. 7 as

$$\dot{I} = \frac{V(t)}{L} - \frac{R(t)}{L}I \Rightarrow I_{t+1} - I_t = \frac{V_t}{L} - \frac{R_t}{L}I_t \ . \tag{11}$$

Rearranging leads to

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

$$V_t = L\left(I_{t+1} - I_t\left(1 - \frac{R_t}{L}\right)\right) \ , \tag{13}$$

and

$$R_t = L\left(I_t - I_{t+1} + \frac{V_t}{L}\right) \ . \tag{14}$$

These equations imply that if the amplitude of the voltage is significantly larger than both the resistance and the current, then it can appear to drive both the current and the resistance, as seen above.

These examples help put a fine point on the problem of using language should as "driving" and "causality" to describe the relationship indicated by the PAI (and CCM causality in general). Suppose the voltage, resistance, and current were all completely out of the experimenter's control, and the only data available is the time series described above. In such a setting, it is reasonable that certain scenarios would allow the voltage to be used to predict both the resistance and the current because the voltage physically drives the current and shares a similar, repetitive mathematical structure with the resistance. The prediction ability of the voltage for the two quantities may come from different mechanisms (i.e. physical driving or mathematical structure), but the prediction ability itself is what is of interest to us here. It is not reasonable to state that the voltage physically drives the resistance in this particular case because by construction, the voltage and the resistance are independent.

6 PAI and Non-Linear Data Sets

The PAI does not need a physical causality interpretation to be useful. It is can helpful in inferring relationships between time series from complicated, non-linear dynamics.

For example, consider the extension of the simplified two population example presented in Section 2.1 to a "three population" system as follows:

$$X_{t+1} = X_t\left(r_x - r_x X_t - B_{xy}Y_t - B_{xz}Z_t\right) \tag{15}$$

$$Y_{t+1} = Y_t\left(r_y - r_y Y_t - B_{yx}X_t - B_{yz}Z_t\right) \tag{16}$$

$$Z_{t+1} = Z_t\left(r_z - r_z Z_t - B_{zx}X_t - B_{zy}Y_t\right) \tag{17}$$

where $r_x, r_y, r_z, \beta_{xy}, \beta_{yx}, \beta_{xz}, \beta_{zx}, \beta_{zy}, \beta_{yz} \in \mathbf{R} \geq 0$. There is no readily apparent physical intuition for this system unlike the two population example (which was, likewise, non-linear). For instance, if $\beta_{xz} = \beta$ **write this to specifically reflect the example you plan on plotting.**

7 Conclusion

Remember to state E and τ for all the plots!

A reasonable interpretation of the PAI and CCM causality is as the identification (and quantification) of asymmetrical prediction capability between pairs of time series data.

A set of three time series data sets X, Y, and Z, would have three directed correlations (which is pairwise in the data sets by definition). Each of the three directed correlations would determine the CCM causality as, for example, whether X CCM causes Y, Y CCM causes X, or "there is no CCM causality in this system". The relationship stated using the phrase "CCM causes" indicates which time series history in the pair of data sets is the better predictor of its partner. For example, the statement "X CCM causes Y" indicates the histories (i.e. delay vectors) of Y_t are better predictors of X_t than the histories of X_t are of Y_t. Such information can be useful, but should not be confused with physical causality.

8 Appendix A: CCM Algorithm

A description of this algorithm is also available in [19] (supplementary materials). It is elucidating to partition the CCM algorithm into five distinct (though related) steps:

1. Create the shadow manifold for X, called $\tilde{X}$

2. Find the nearest neighbours to $\tilde{X}_t$

3. Use the nearest neighbours to create weights

4. Use the weights to estimate Y, called $Y|X$

5. Find the correlation between Y and $Y|X$

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

8.1 Create $\tilde{X}$

Given an embedding dimension E, the shadow manifold of X, called $\tilde{X}$, 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 is 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, and so on). The distances from $\vec{X}_t$ are defined as

$$d_i = D\left(\vec{X}_t, \vec{X}_{t_i}\right) \ ,$$

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

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

The weights are defined as

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

where the normalization factor is given as

$$N = \sum_j u_j \ .$$

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

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

8.5 Find the Correlation

The CCM correlation is defined as

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

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 algorithm we use for determining the embedding dimension E can be outlined as follows:

1. Embed X in a shadow manifold $\tilde{X}$ with dimension E (see Appendix 8)
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 chosen to be "sufficiently small" while still being computationally 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 neighbour 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 algorithm will become much more computationally 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 algorithm begins. This algorithm can be modified to concurrently determine τ and E by repeating Steps $1 - 3$ for successive 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 computational cost of the algorithm.

# References

[1] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett.*, 103:238701, Dec 2009.

[2] Jean-Marie Dufour, Denis Pelletier, and Éric Renault. Short run and long run causality in time series: inference. *Journal of Econometrics*, 132(2):337–362, 2006.

[3] Jean-Marie Dufour and Eric Renault. Short run and long run causality in time series: theory. *Econometrica*, pages 1099–1125, 1998.

[4] Michael Eichler and Vanessa Didelez. Causal reasoning in graphical time series models. *arXiv preprint arXiv:1206.5246*, 2012.

[5] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

[6] C.W.J. Granger. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2(0):329 – 352, 1980.

[7] D. Halliday, R. Resnick, and J. Walker. *Fundamentals of Physics Extended, 10th Edition*. Wiley Global Education, 2013.

[8] Maciej Kaminski, Mingzhou Ding, Wilson A. Truccolo, and Steven L. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics*, 85(2):145–157, 2001.

[9] Matthew B. Kennel, Reggie Brown, and Henry D. I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, 45:3403–3411, Mar 1992.

[10] Yan Liu and Mohammad Taha Bahadori. A survey on granger causality: A computational view. *University of Southern California, http://www-bcf. usc. edu/~ liu32/granger. pdf (12. 03.2013)*, 2012.

[11] Alun L. Lloyd. The coupled logistic map: a simple model for the effects of spatial heterogeneity on population dynamics. *Journal of Theoretical Biology*, 173(3):217 – 230, 1995.

[12] Hong-guang Ma and Chong-zhao Han. Selection of embedding dimension and delay time in phase space reconstruction. *Frontiers of Electrical and Electronic Engineering in China*, 1(1):111–114, 2006.

[13] Timothy J Mosedale, David B Stephenson, Matthew Collins, and Terence C Mills. Granger causality of coupled climate processes: Ocean feedback on the north atlantic oscillation. *Journal of climate*, 19(7), 2006.

[14] David L. Roberts and Stephen Nord. Causality tests and functional form sensitivity. *Applied Economics*, 17(1):135–141, 1985.

[15] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.

[16] Michael Small and C.K. Tse. Optimal embedding parameters: a modelling paradigm. *Physica D: Nonlinear Phenomena*, 194(3–4):283 – 296, 2004.

[17] G. Sugihara and R. M. May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734–741, April 1990.

[18] George Sugihara, Bryan Grenfell, Robert M. May, P. Chesson, H. M. Platt, and M. Williamson. Distinguishing error from chaos in ecological time series [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 330(1257):235–251, 1990.

[19] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, 2012.

[20] I Vlachos and D Kugiumtzis. State space reconstruction from multiple time series. In *Topics on Chaotic Systems: Selected Papers from Chaos 2008 International Conference*, page 378. World Scientific, 2009.
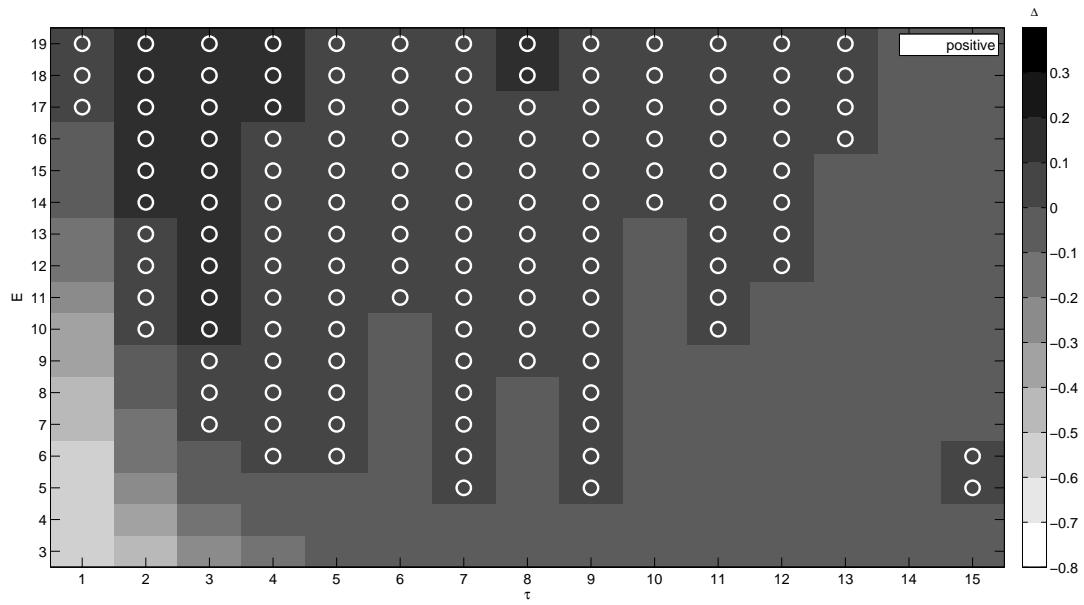
(a) Bug Fix
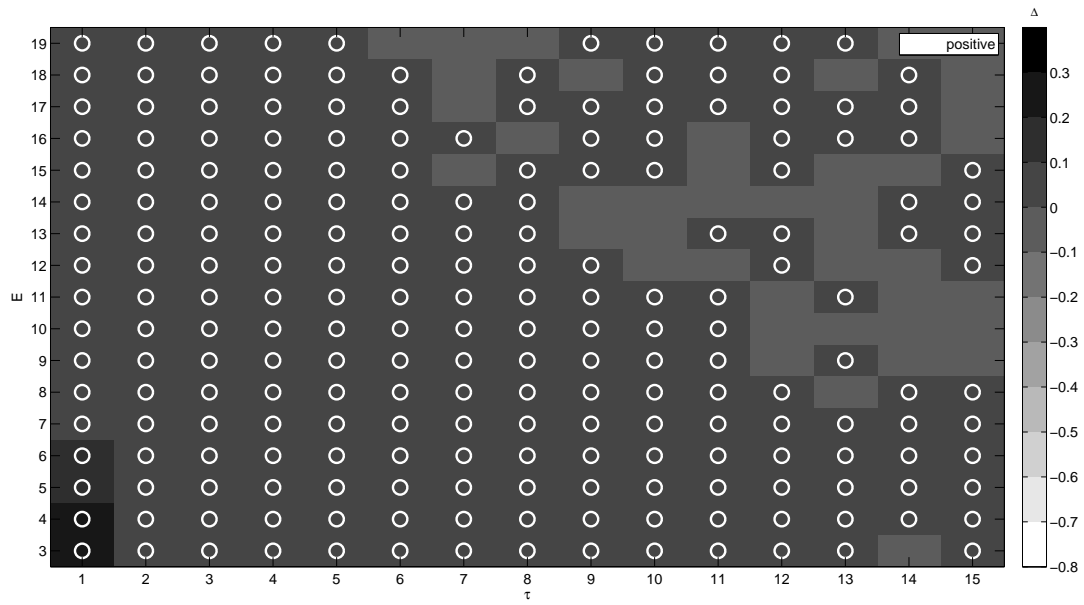
(b) Demonstration of alternate correlation calculation
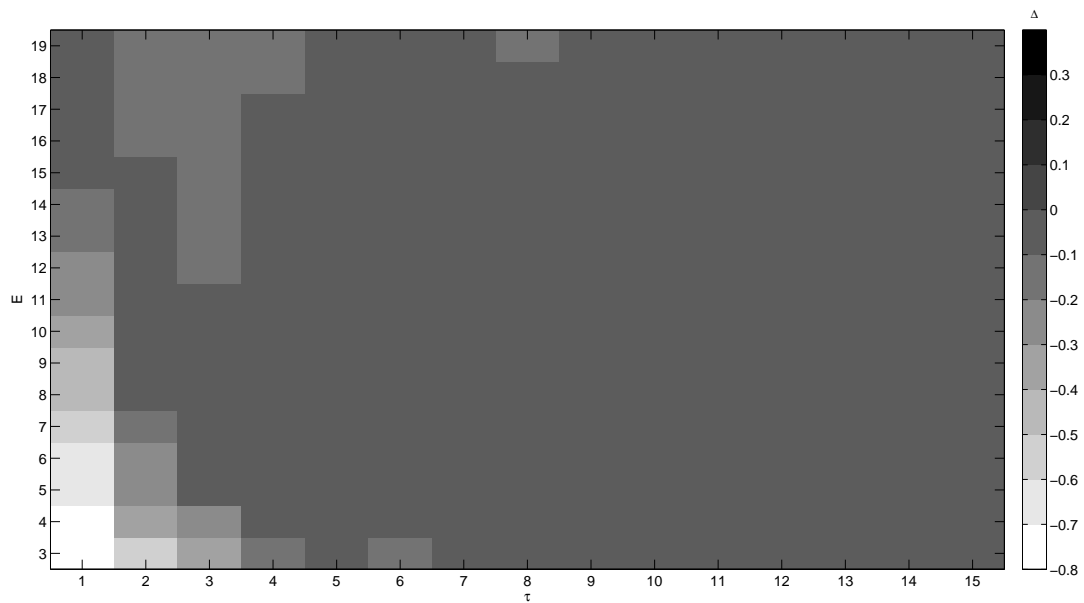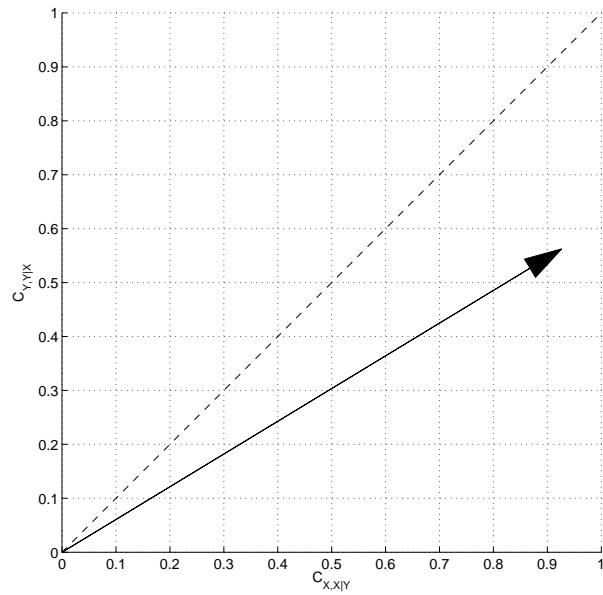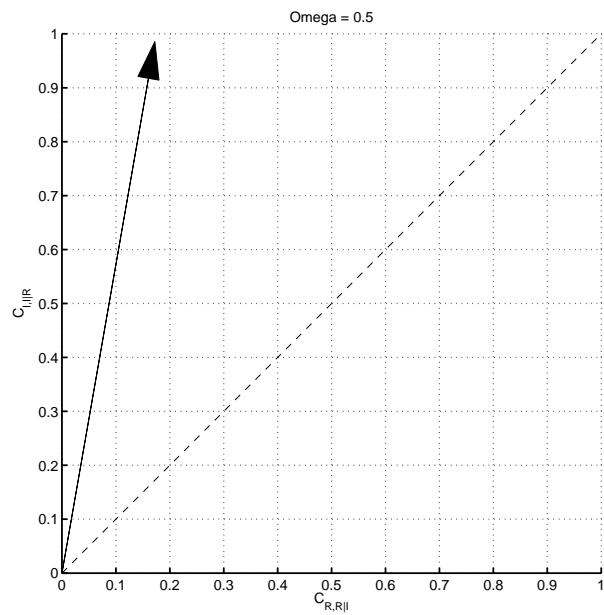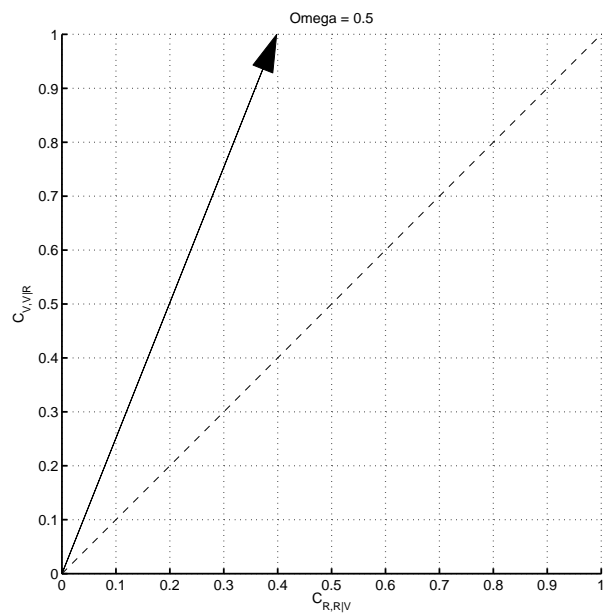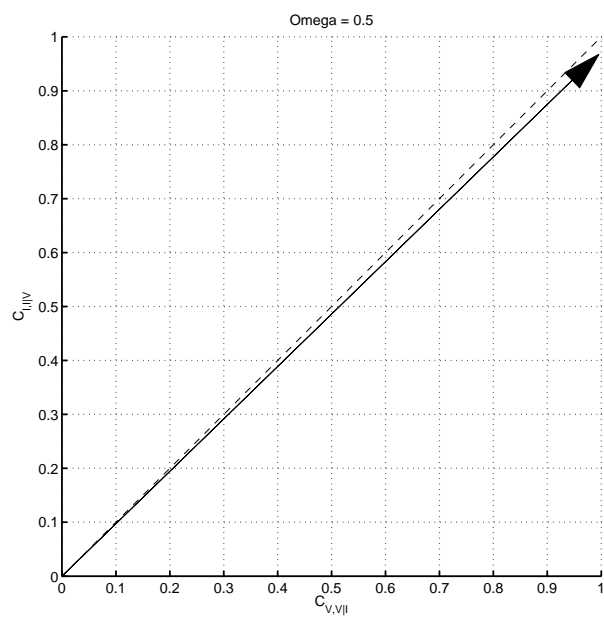

(c) Difference between (a) and (b)

Figure 4: The PAI 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.
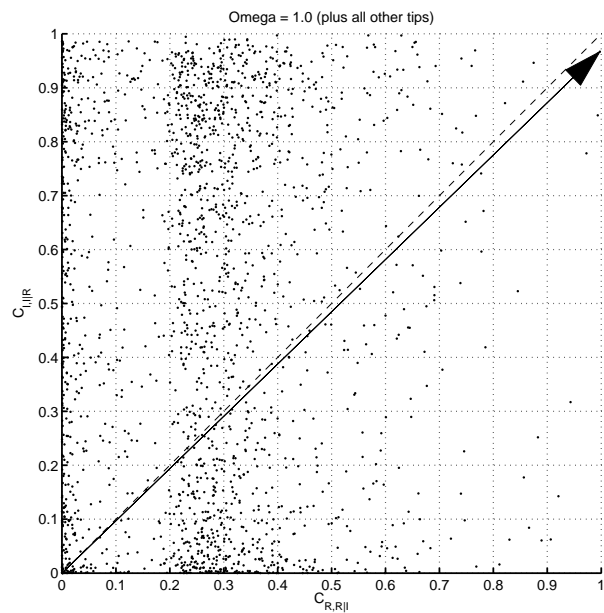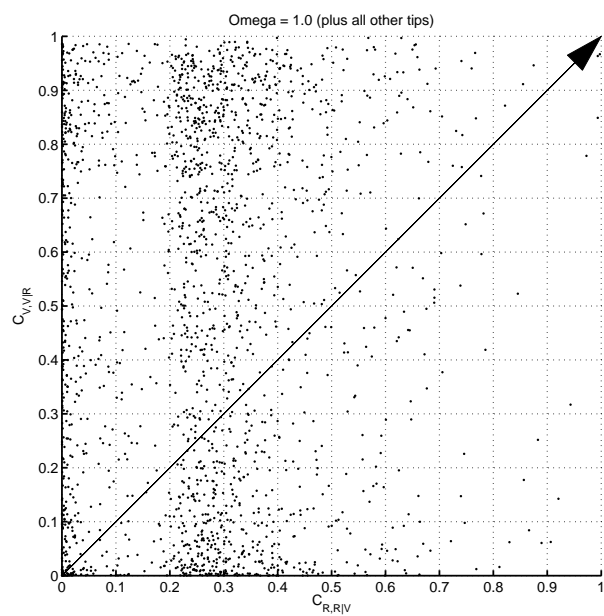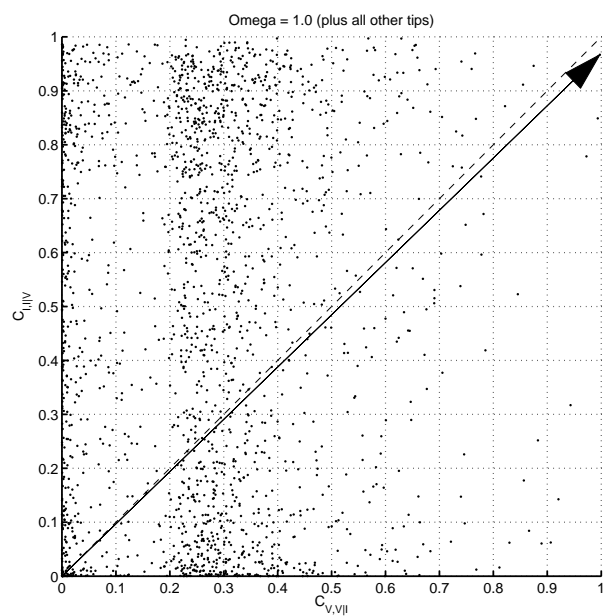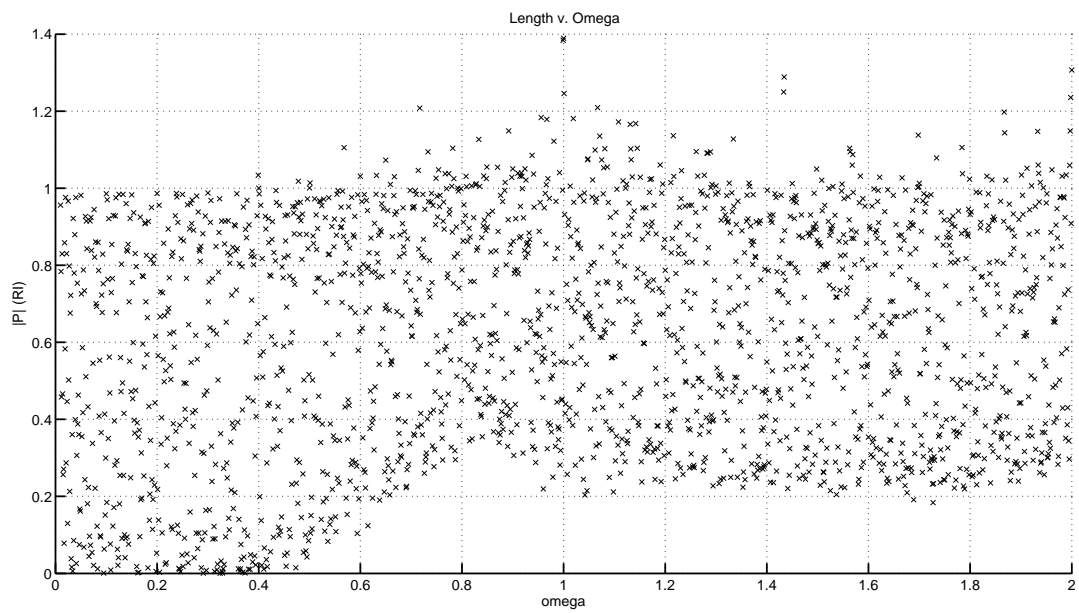
(a) RI

(b) RV

(c) VI

10

Omega = 1.0 (plus all other tips)

(A) RI

Omega = 1.0 (plus all other tips)

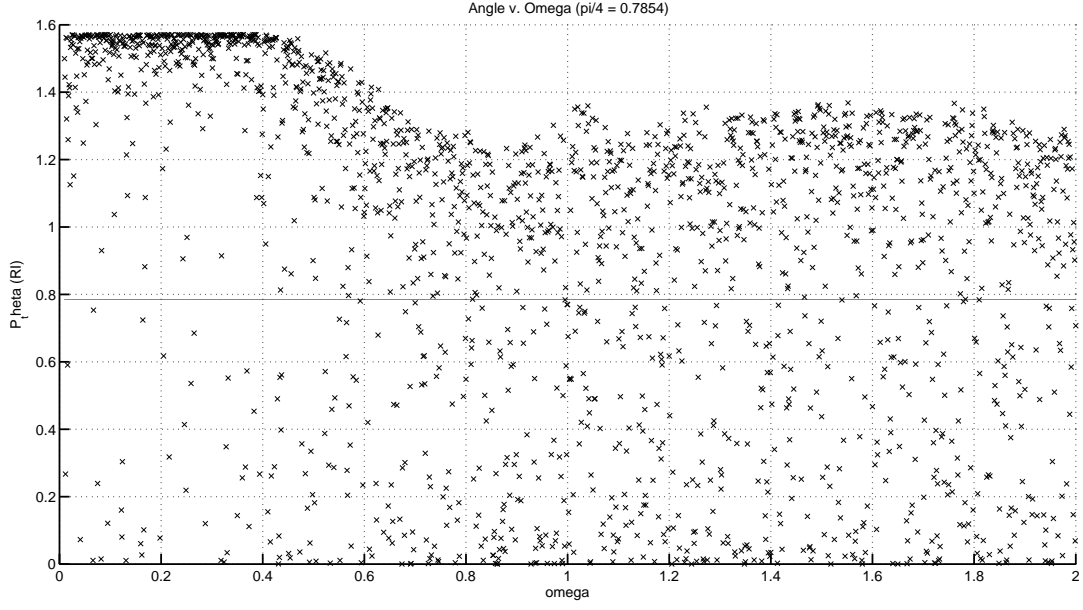(B) RV

Omega = 1.0 (plus all other tips)
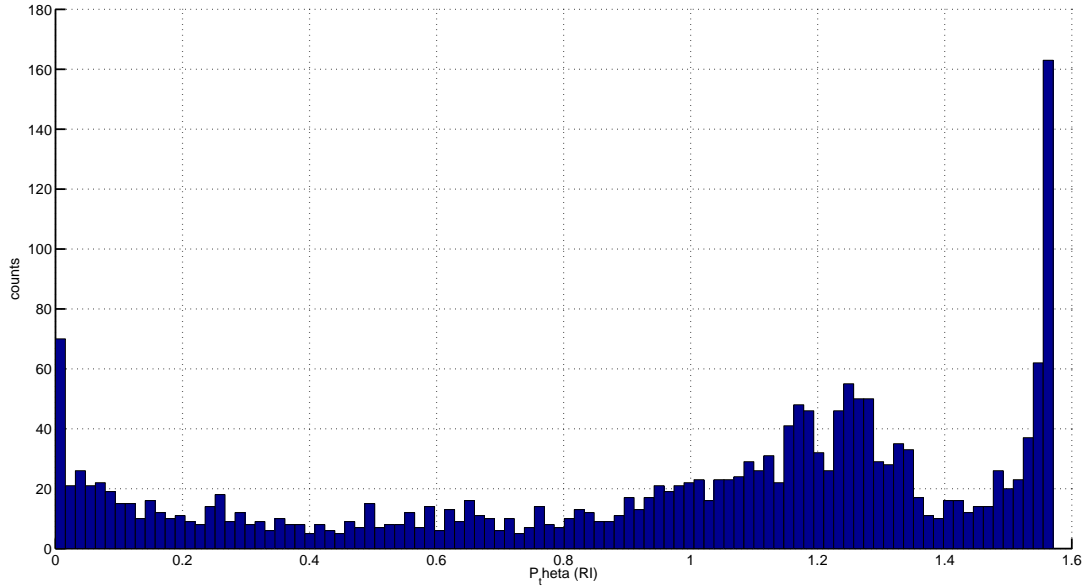
(C) VI

11

(a)
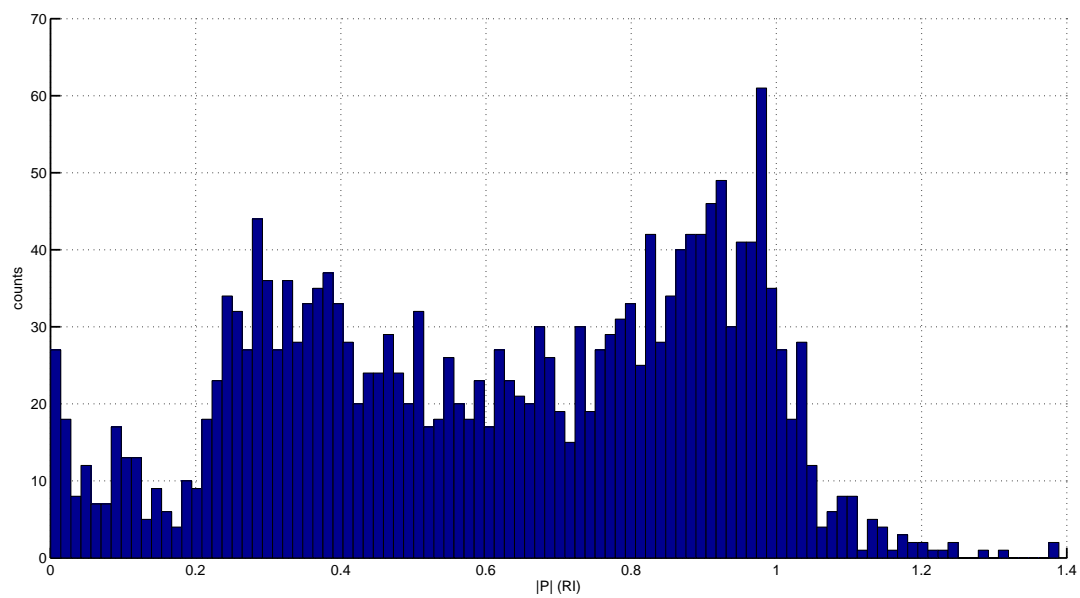
(b)

(c)

(d)

Figure 6: The PAIs of Figure **??** change dramatically when the phase between $R(t)$ and $V(t)$ is changed. See the text for the parameters used to produce $\vec{P}$.

Figure 7: The PAIs of the RL circuit system with $V(t)$ given by Eqn. 8 and $R(t)$ given by Eqn. 10.

13