Convergent Cross-Mapping and Causality Detection

McCracken,Weigel

June 4, 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 the calculation of CCM correlations can, however, be modified to identify asymmetric relationships between pairs of time series data. We introduce "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.

1 Introduction

Modern time series analysis includes a handful of techniques meant to discern "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 investigate 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 causality, 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 labeled as $Y|X$; i.e. the CCM correlation is given as

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

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

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

---

[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.
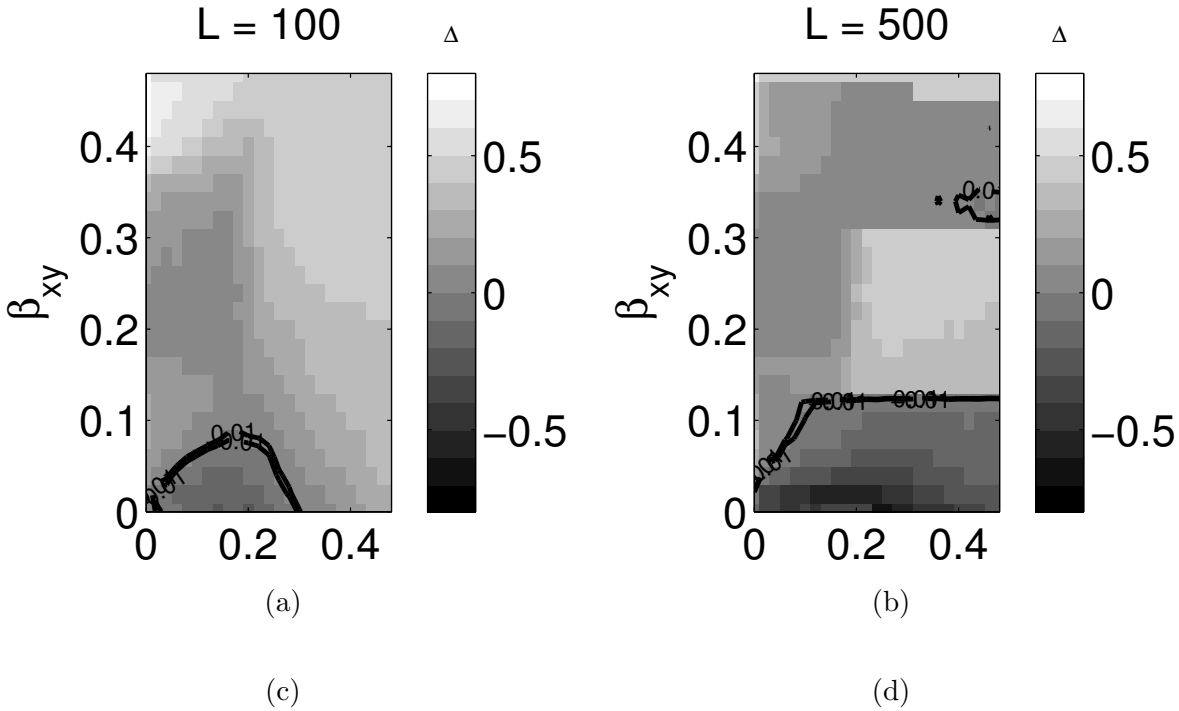

(a)   (b)

(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
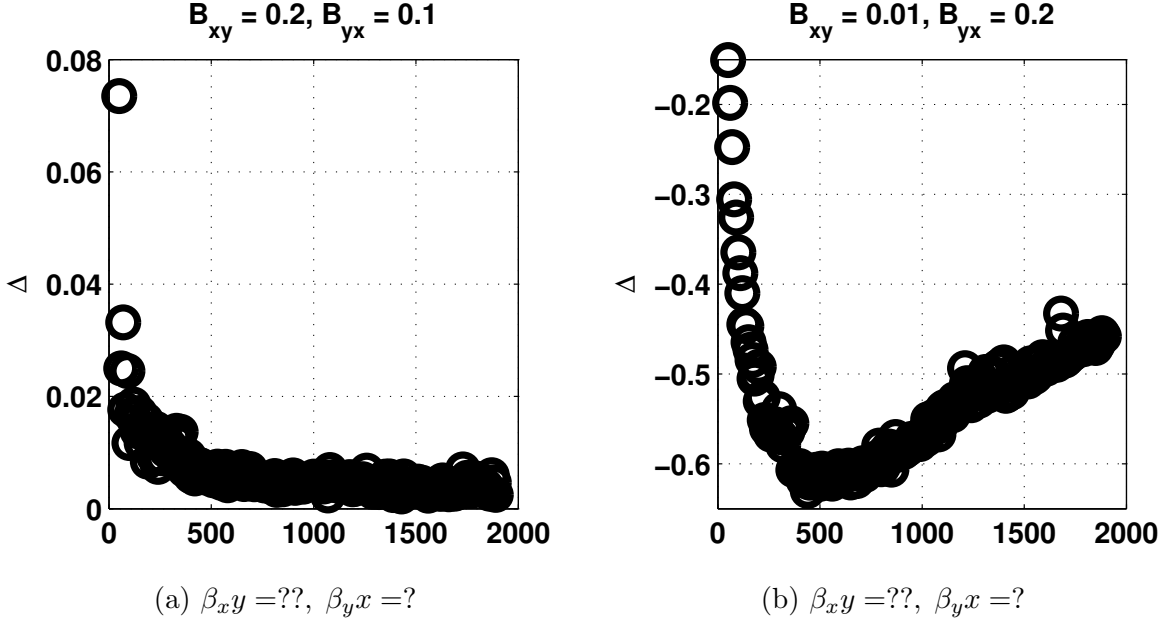
Figure 2: Reference figure.

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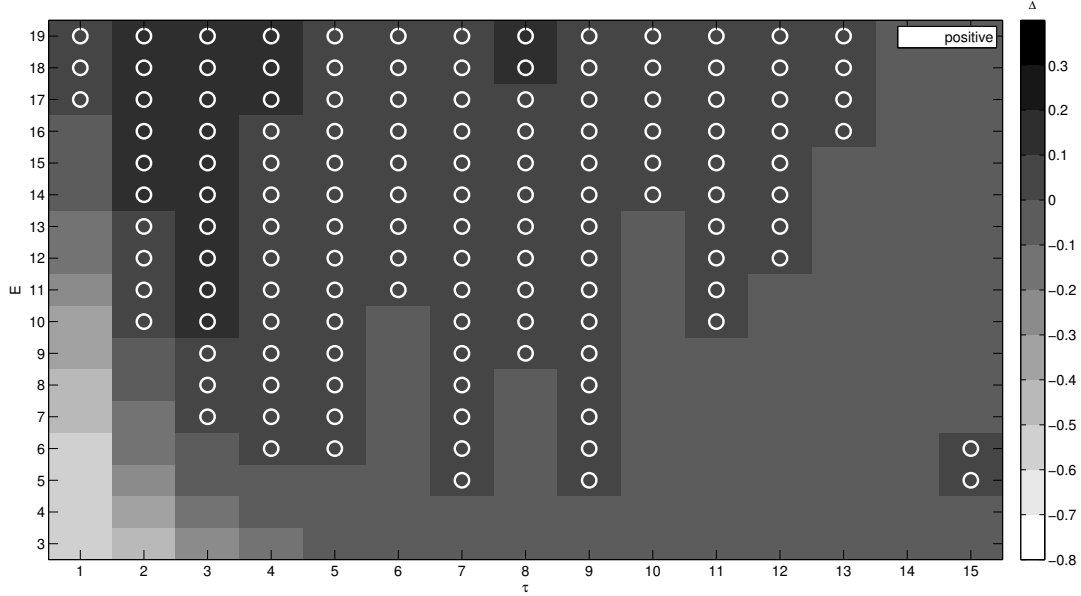

Figure 3: The determination of CCM causality in a system is dependent on the CCM parameters of embedding dimension E and lag time step τ. This plot show the dependence of Eqn. 1 on E and τ for the example system discussed in the text.

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). Other methods for determining E and τ for SSR algorithms can be found in the literature (e.g. cite ???).

4   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.

4.1 Linear Example

Consider the linear example dynamical system of

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

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 5 shows Δ for this example given a library length of $L = 2000$.
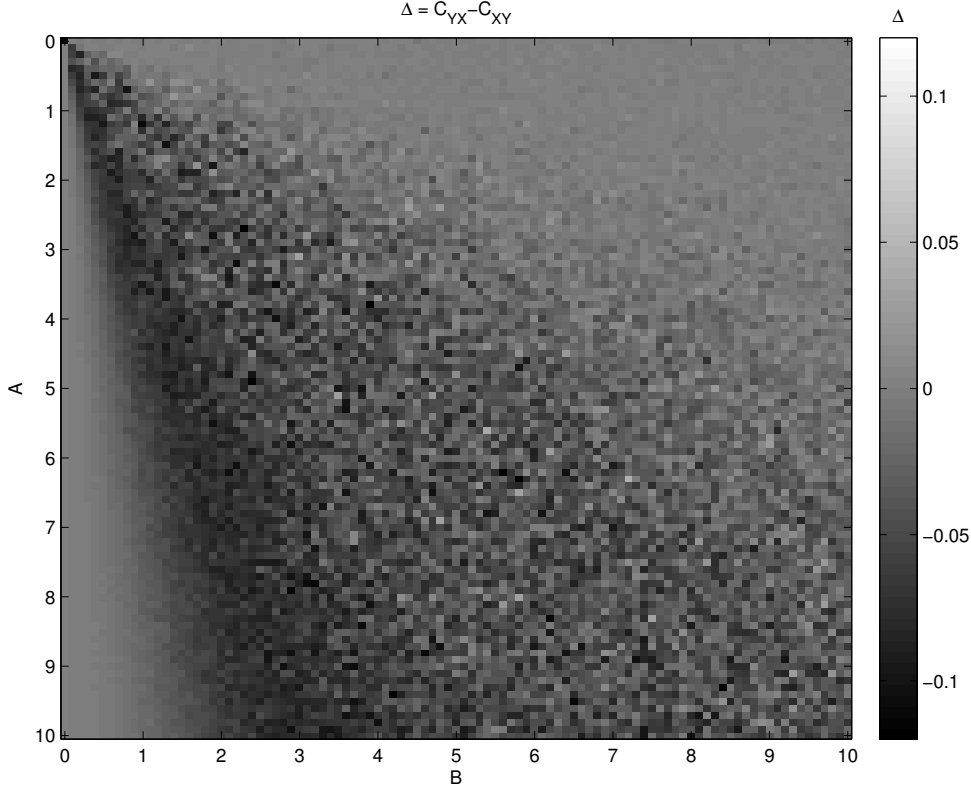
Figure 4: Changing A and B. Δ

The convergence of two specific points in Figure 5, $(A, B) = (2.6, 2.6)$ and $(A, B) = (3.0, 2.6)$, is shown in Figure 6.

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 6 shows (for the two specific points plotted) that Δ is more negative at shorter library length but appears to converge to a point near zero as the library length is increased. The convergence of CCM correlations is emphasized [19], so the seemingly counter intuitive behavior of Δ (and C_{XY} and C_{YX}) in Figure 6 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.
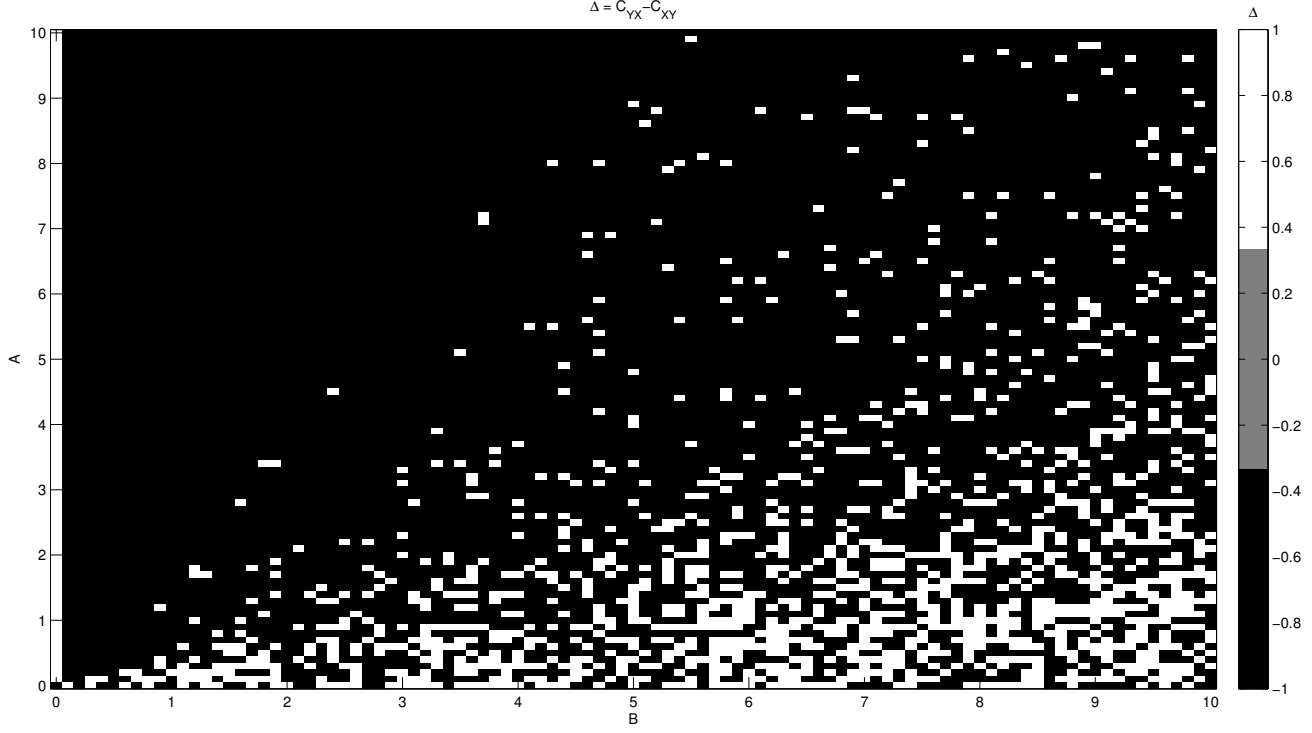
4

Figure 5: Changing A and B. Δ

4.2 Non-Linear Example

Consider the nonlinear example dynamical system of

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

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 7 shows Δ for specific values of C given a library length of $L =??$. 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 example; i.e. Δ does not appear to reliably reflect intuitive notions of driving.

5 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

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

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

$$
\begin{align}
I_{t+1} &= \frac{V_t}{L} + I_t\left(1 - \frac{R_t}{L}\right), \tag{10}\\
V_t &= L\left(I_{t+1} - I_t\left(1 - \frac{R_t}{L}\right)\right), \tag{11}
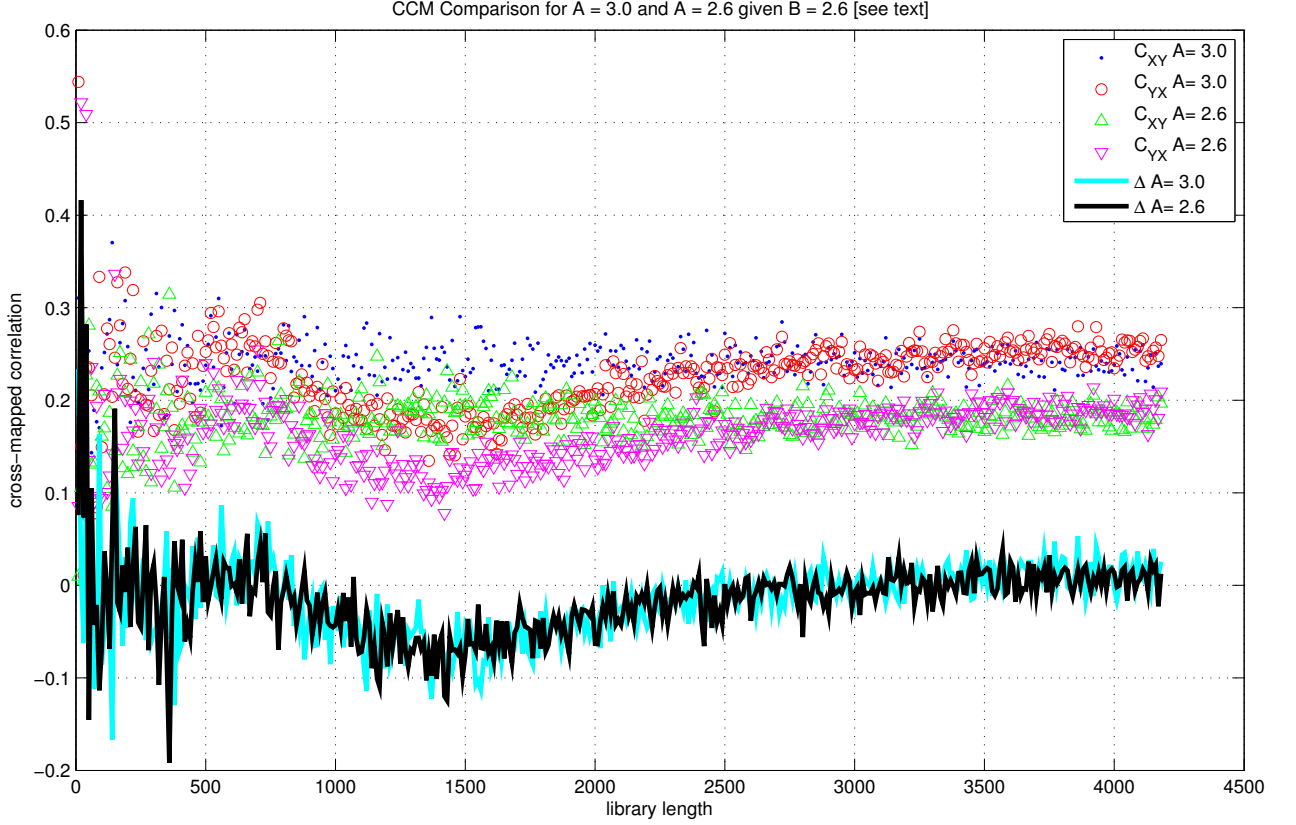\end{align}
$$

Figure 6

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.

5.1 Changing V(t)

Consider the situation where $R(t)$ is constant. Physical intuition is that V drives I, so we expect to find V CCM causes I (i.e. $C_{VI} > C_{IV}$). For this example, the voltage is described by

$$V(t) = A_v \sin (f_v t + \phi_v) + O_v, \tag{13}$$

where A_v is the amplitude, f_v is the frequency, ϕ_v is the phase, and O_v is the offset voltage.

5.2 Changing A_v

Consider evaluating the CCM correlations C_{VI} and C_{IV} for each $A_v \in [0.01, 2.0]$ in steps of 0.01. The CCM correlations are each plotted in Figure 8 along with the corresponding PAI elements P_θ and $|P|$.

5.3 Changing f_v

Consider evaluating the CCM correlations C_{VI} and C_{IV} for each $f_v \in [0.01, 2.0]$ in steps of 0.01. The CCM correlations are each plotted in Figure 9 along with the corresponding PAI elements P_θ and $|P|$.

5.4 Changing ϕ_v

Consider evaluating the CCM correlations C_{VI} and C_{IV} for each $\phi_v \in [0.01, 2.0]$ in steps of 0.01. The CCM correlations are each plotted in Figure 10 along with the corresponding PAI elements P_θ and $|P|$.
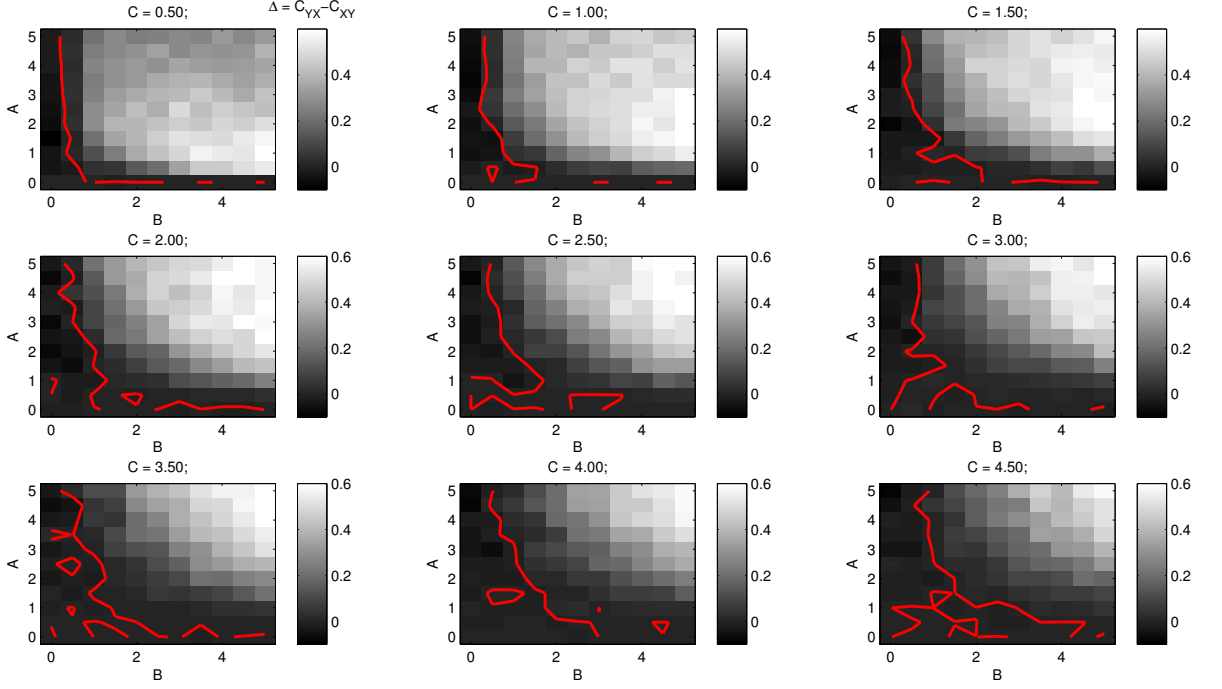
Figure 7

5.5   Changing O_v

Consider evaluating the CCM correlations C_{VI} and C_{IV} for each $O_v \in [0.01, 2.0]$ in steps of 0.01. The CCM correlations are each plotted in Figure 11 along with the corresponding PAI elements P_θ and $|P|$.

Figure 12 shows the effect of increasing the library length from 2×10^3 (i.e. `tspan = [0:0.5:1000];`) to 10^4 (i.e. `tspan = [0:0.5:5000];`), and Figure 13 extends the above plots to $O_v \in [0.01, 10.0]$ in steps of 0.05.

6   PAI

Consider the system (Sugihara Figure 3 C and D)

$$X_{t+1} = X_t \left(r_x - r_x X_t - \beta_{xy} Y_t \right) \tag{14}$$
$$Y_{t+1} = Y_t \left(r_y - r_y Y_t - \beta_{yx} X_t \right), \tag{15}$$

with $r_y = r_y = 3.7$, $X_0 = 0.2$, $Y_0 = 0.4$, $\beta_{xy} = 0$, and $\beta_{yx} = 0.32$. Plots of the correlation between X and $X|Y$, as well as, Y and $Y|X$ are shown below.

This leads to

C_{XY}	C_{YX}	$\Delta = C_{YX} - C_{XY}$
0.973921	0.110698	-0.8632

But, these are not the only correlations that can be tested:

This leads to

C_{XX}	$C_{X(XY)}$	C_{YY}	$C_{Y(YX)}$	$\Delta = C_{Y(YX)} - C_{X(XY)}$
0.999841	0.998989	0.999908	0.998693	-2.9548e-04

Consider a comparison of PAI and CCM given the linear example system from above.

7 Conclusion

Remember to state E and τ for all the plots!

   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 neighbors to $\tilde{X}_t$
3. Use the nearest neighbors 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 Neighbors

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 neighbor 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 neighbor, 2 denotes the next closest neighbor, and so on). 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 neighbor 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

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

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, $C_{YX} \neq C_{XY}$.

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.
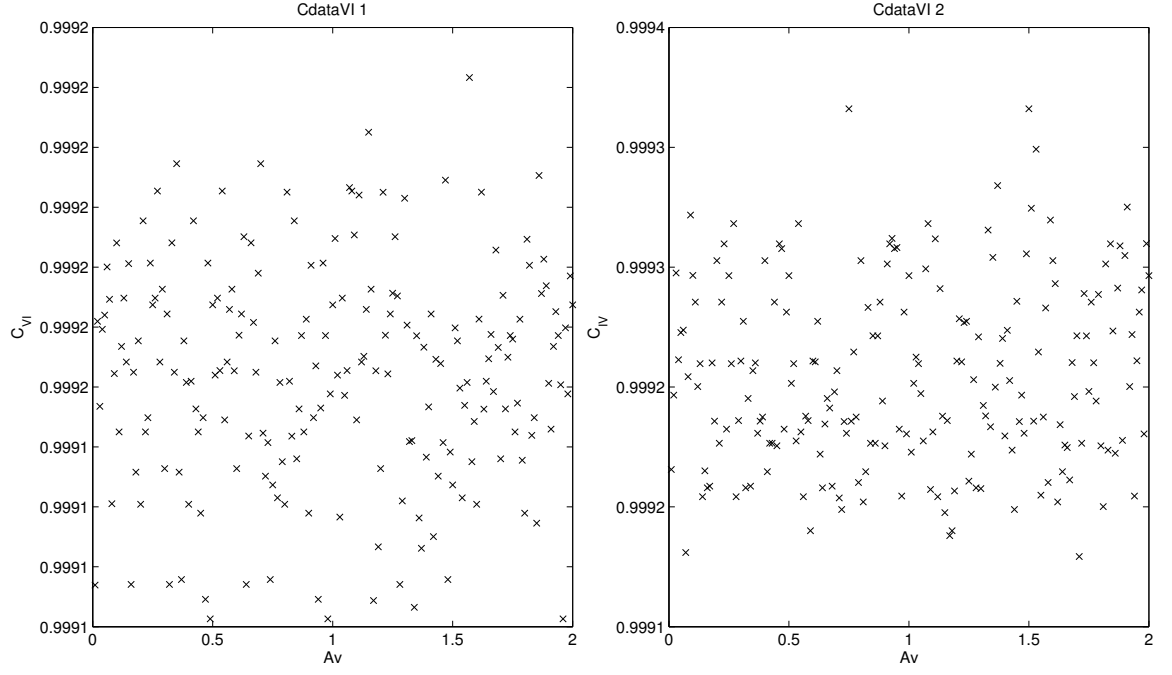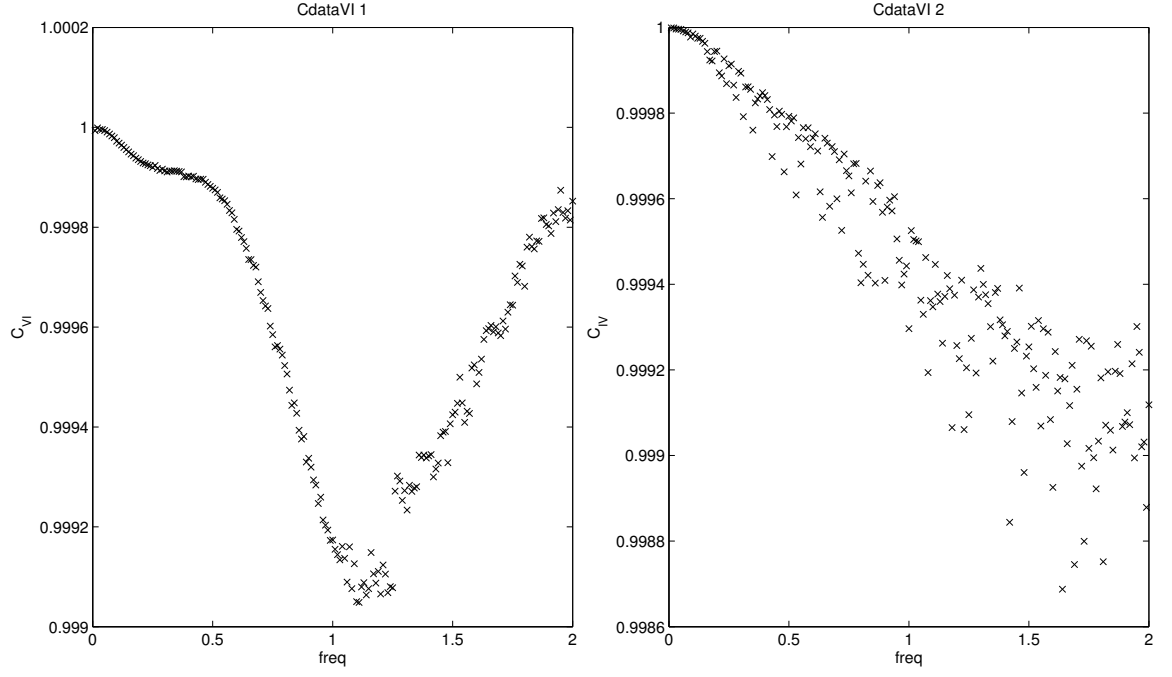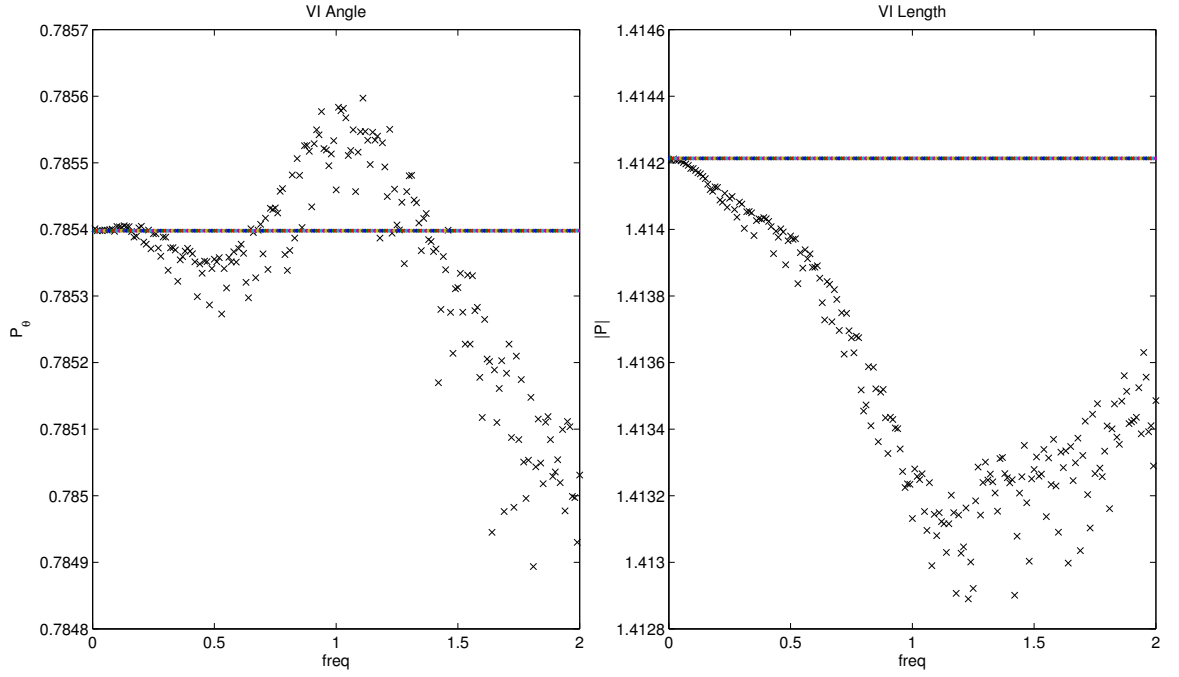
(a) C_{VI} and C_{IV}
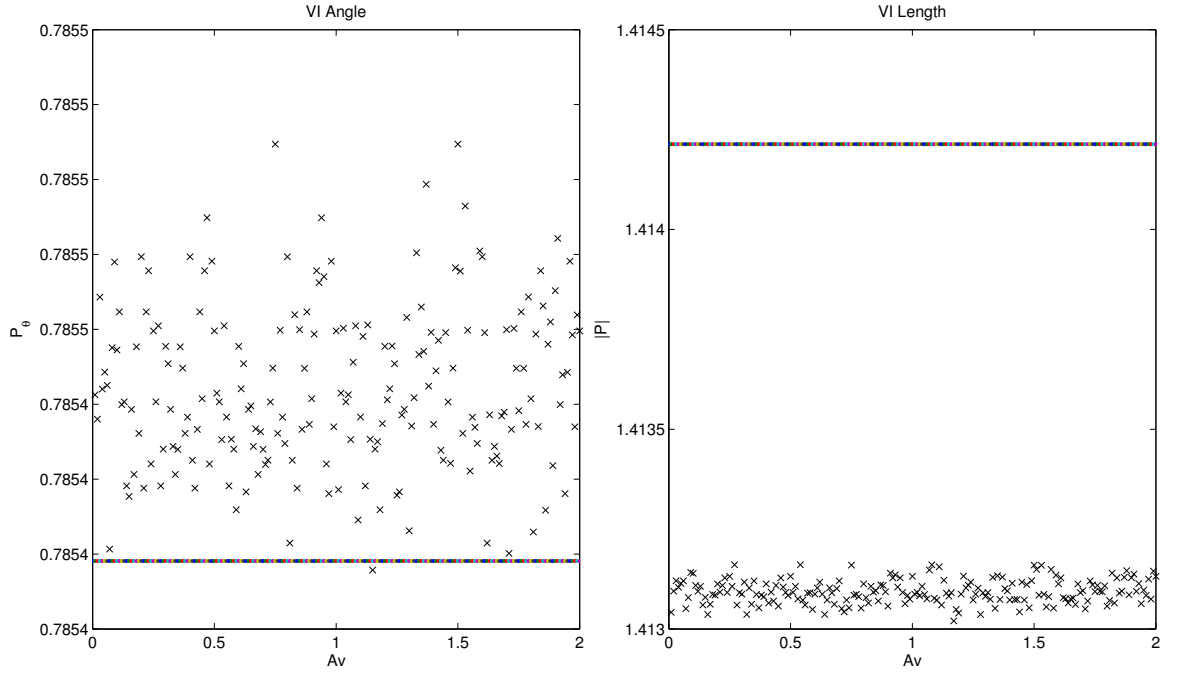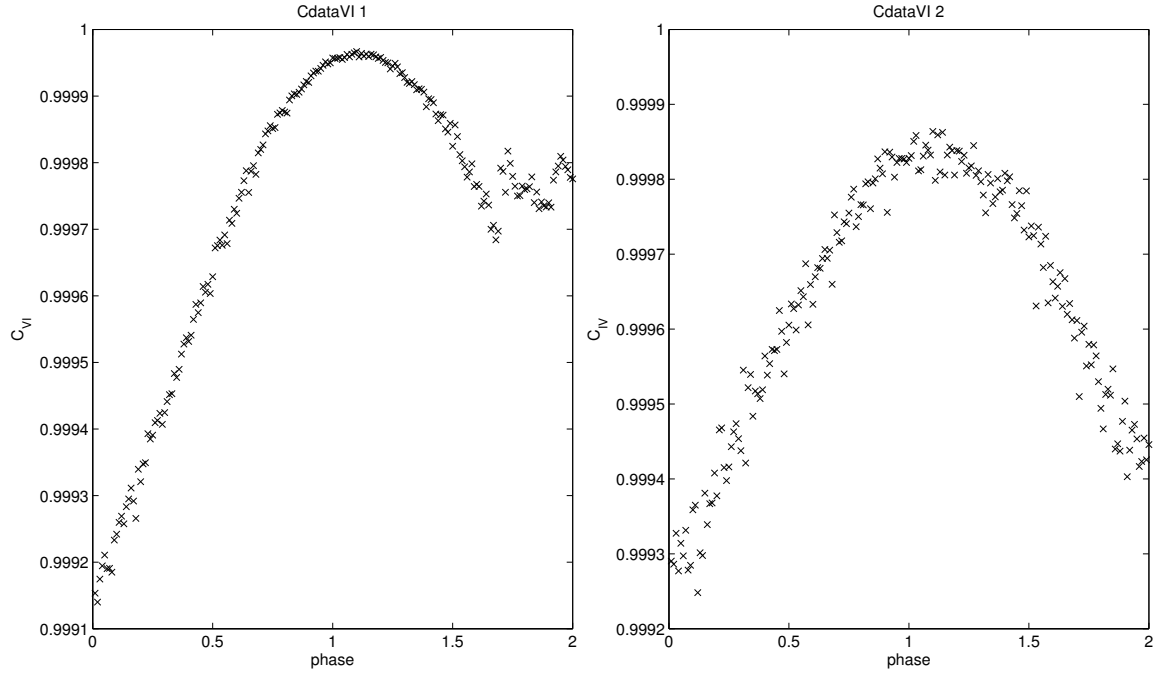

(b) P_θ and $|P|$

Figure 8: Changing A_v.

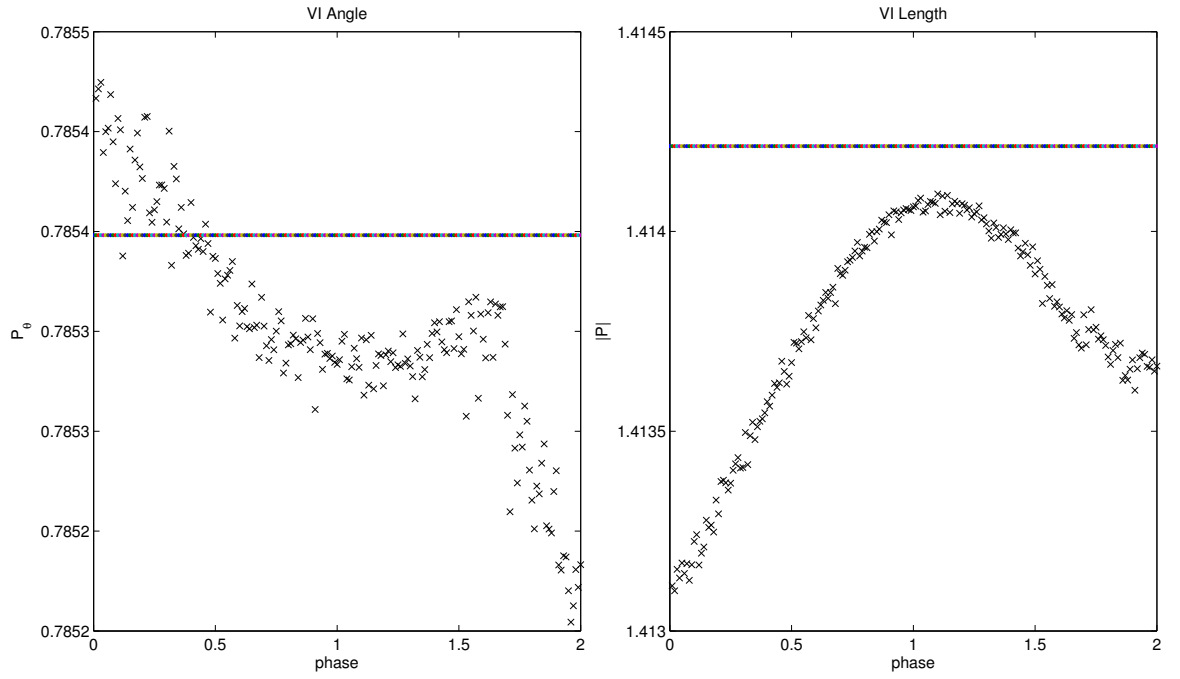(a) C_{VI} and C_{IV}

(b) P_θ and $|P|$

Figure 9: Changing f_v.

(a) C_{VI} and C_{IV}

(b) P_θ and $|P|$

Figure 10: Changing ϕ_v.
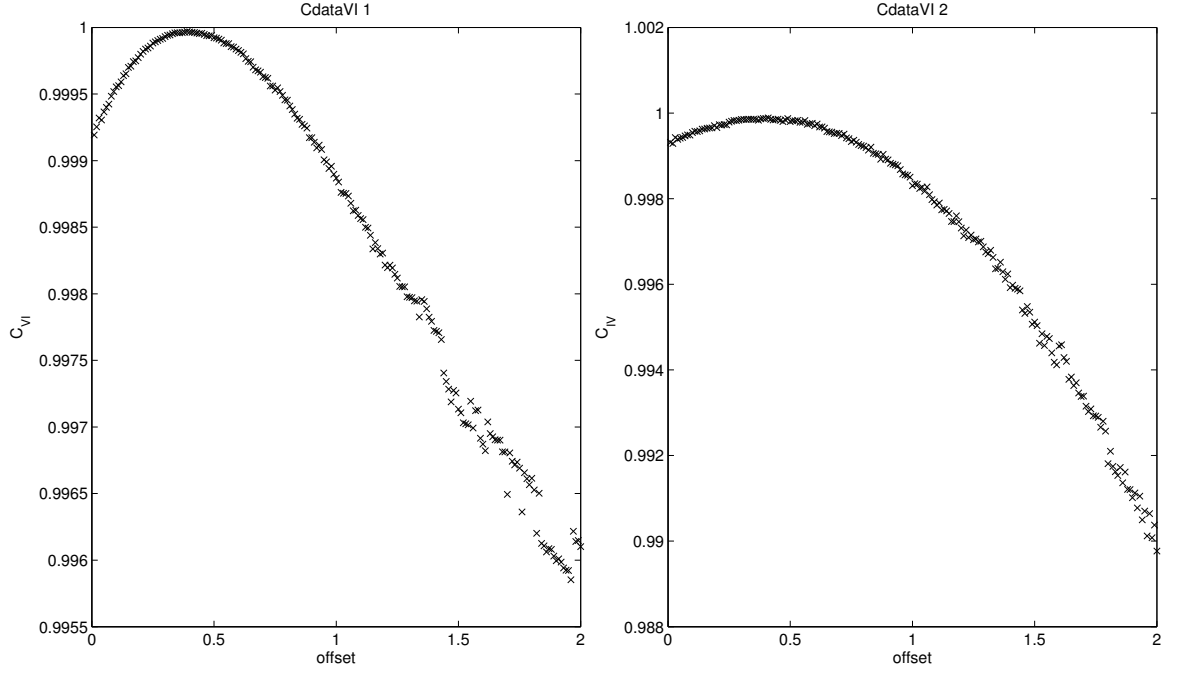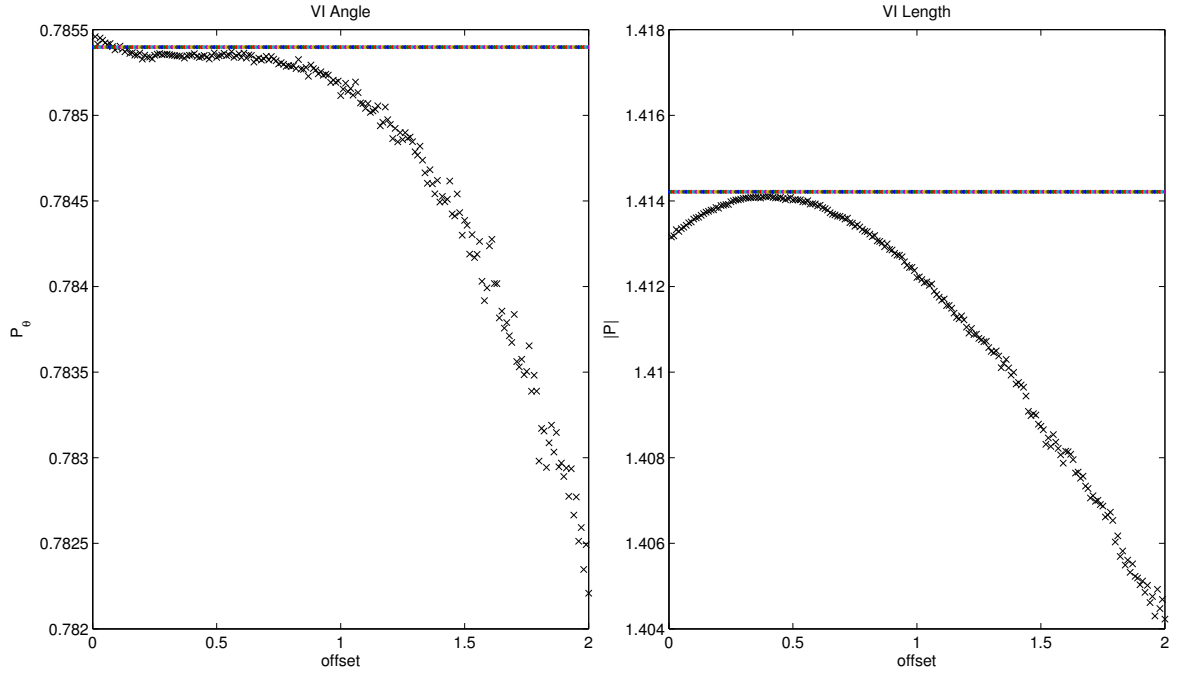
(a) C_{VI} and C_{IV}
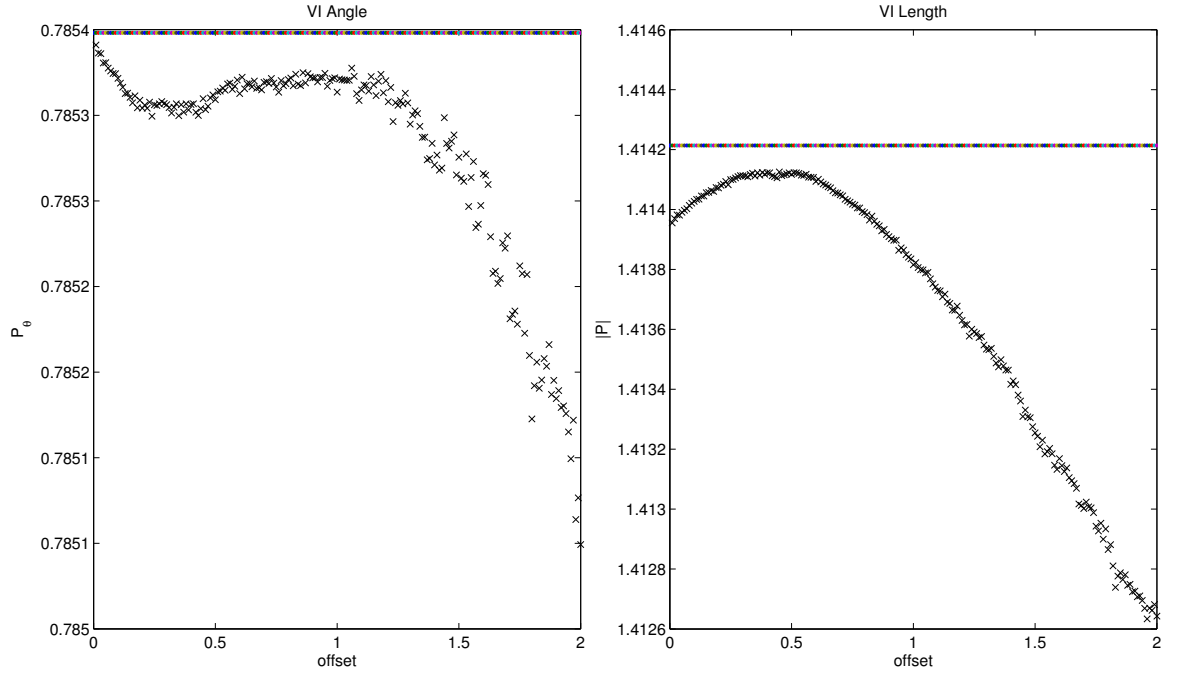

(b) P_θ and $|P|$

Figure 11: Changing O_v.
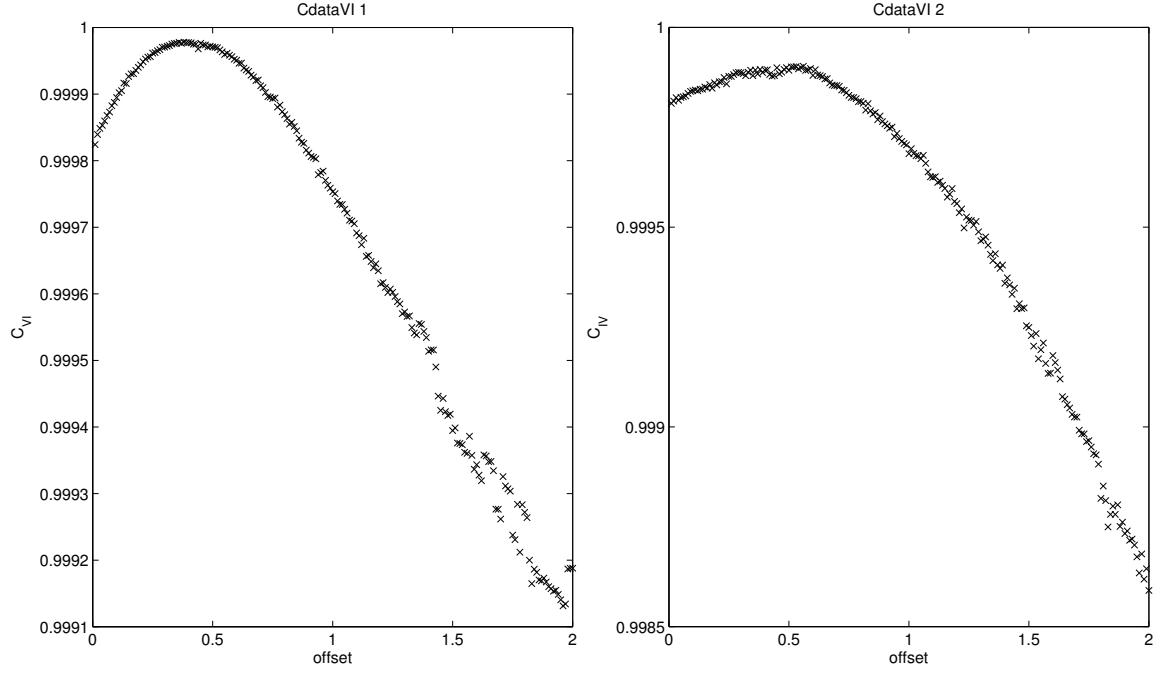
(a) C_{VI} and C_{IV}

(b) P_θ and $|P|$
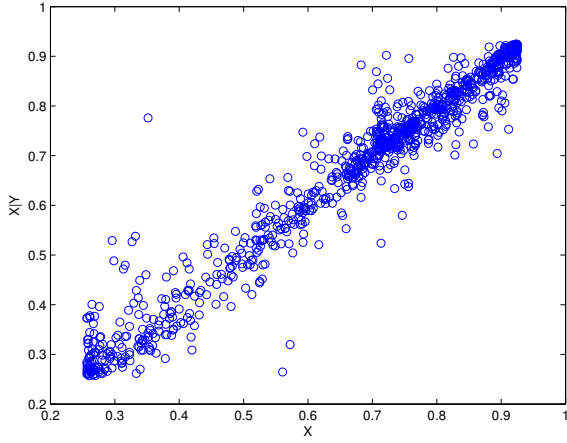
Figure 12: Changing O_v (longer library length).
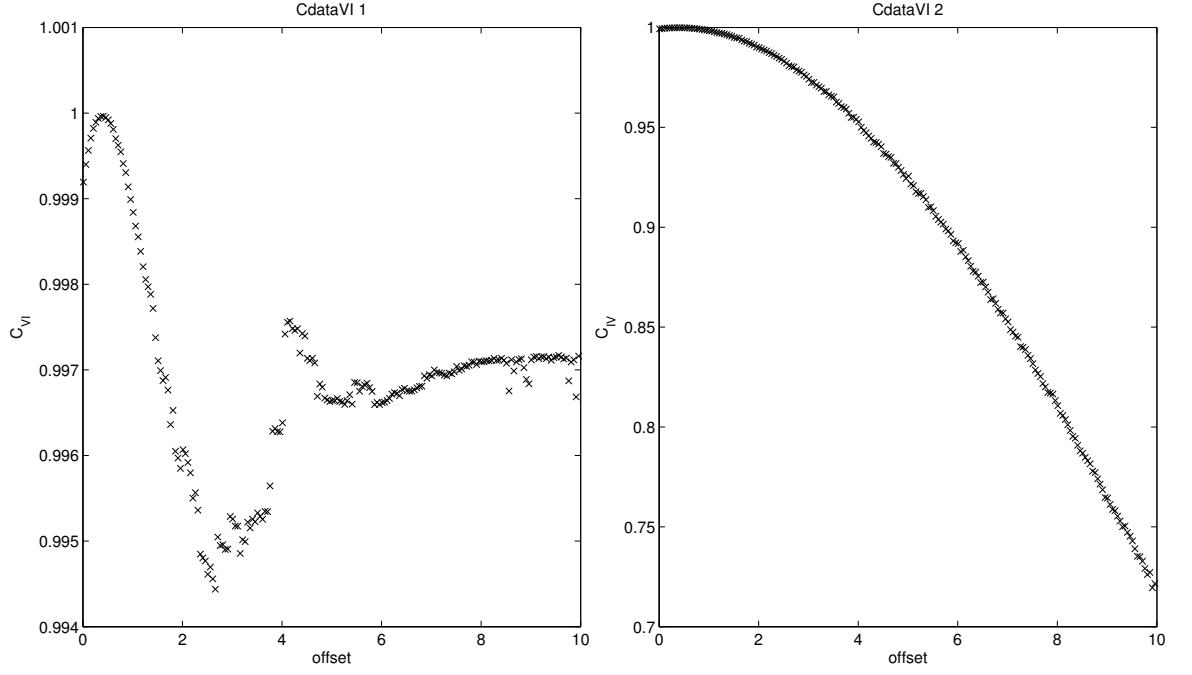
(a) C_{VI} and C_{IV}
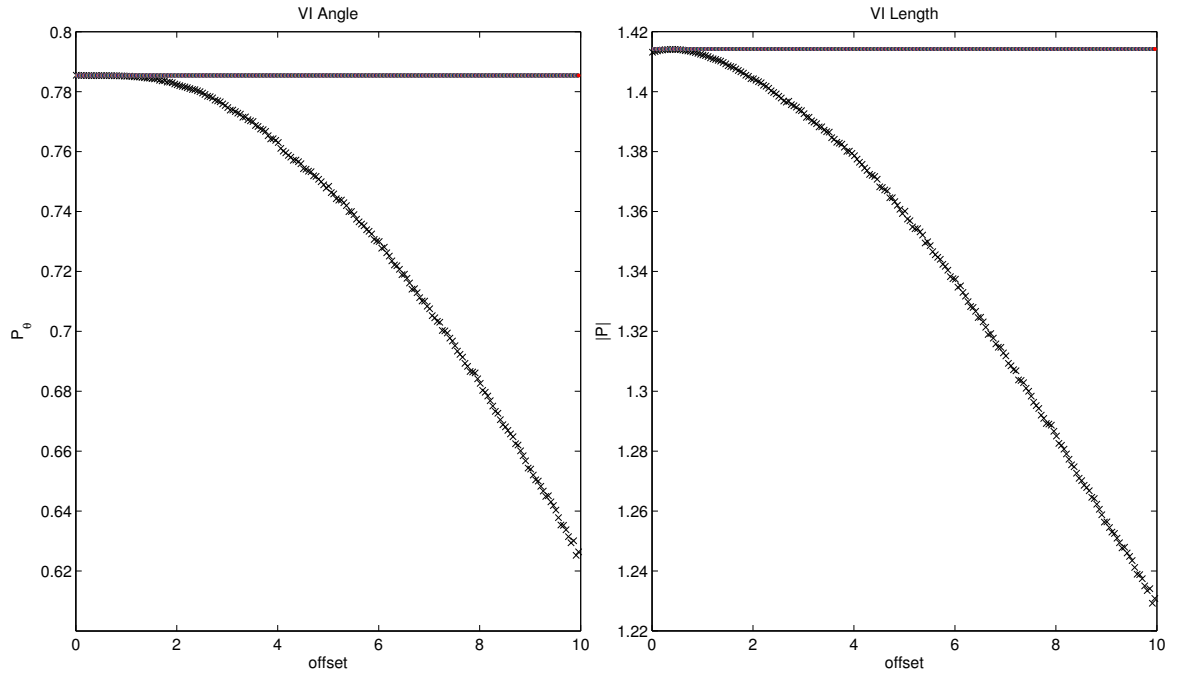
(c) P_θ and $|P|$

Figure 13: Changing O_v (larger domain for O_v).

(a) C_{XY}
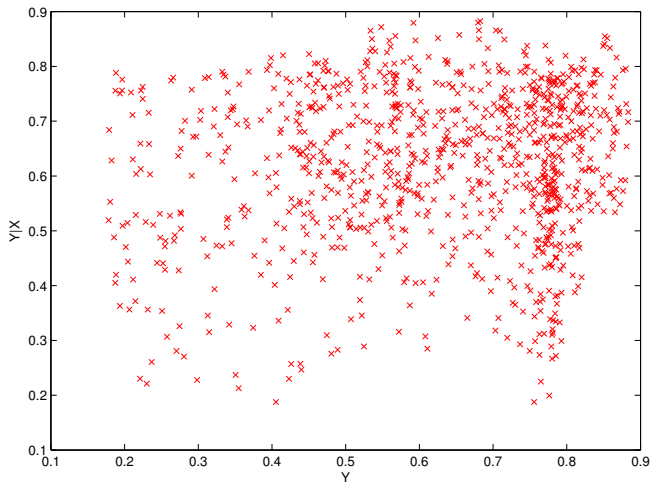
(b) C_{YX}

Figure 14: Changing A and B. C_{XY} and C_{YX}
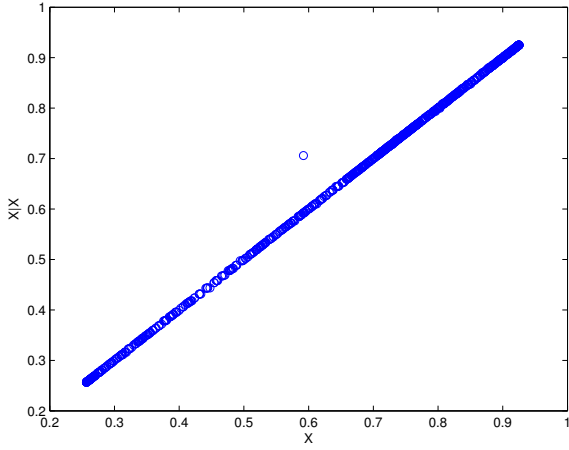
(a) C_{XX}
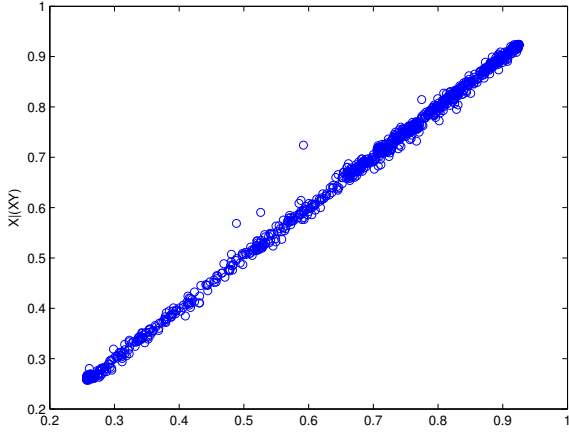
(b) $C_{X(XY)}$

Figure 15: Changing A and B. C_{XY} and C_{YX}
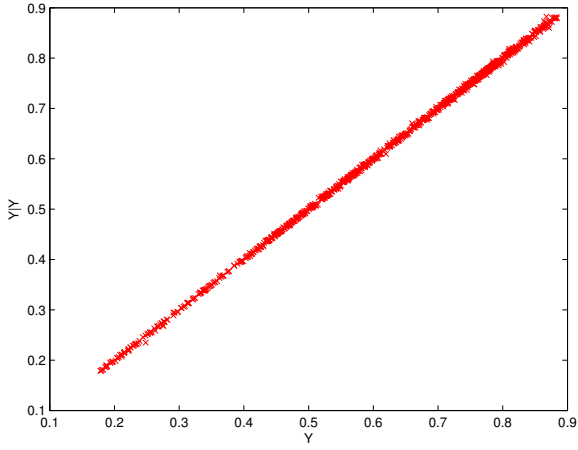
(c) C_{XX}
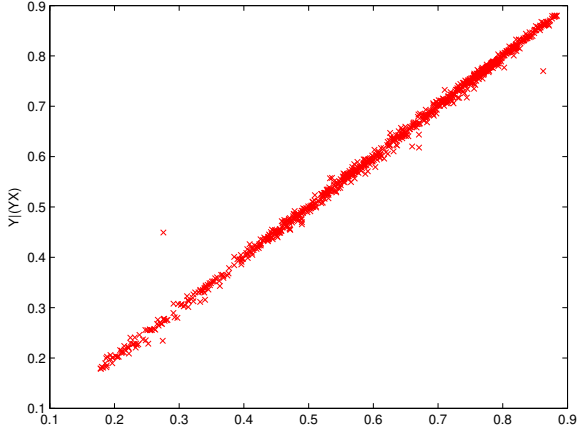
(d) $C_{X(XY)}$

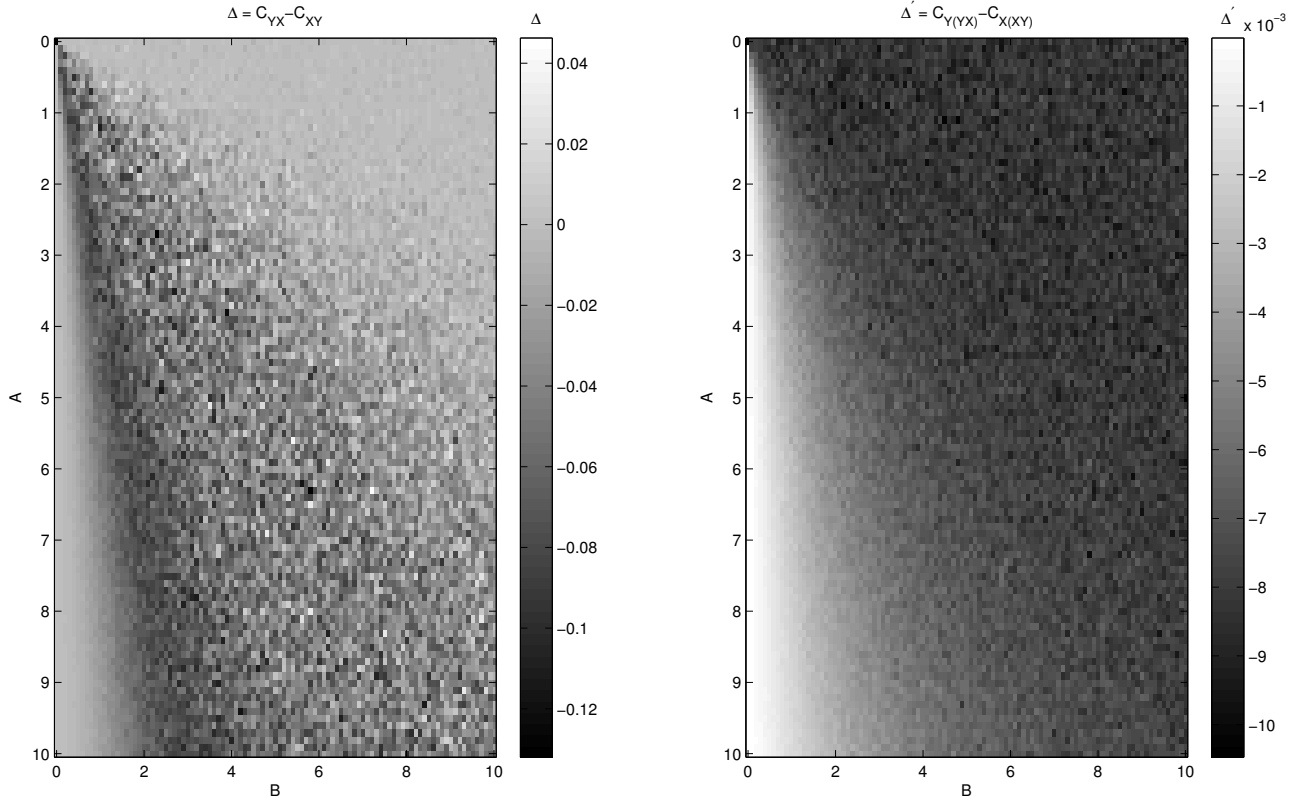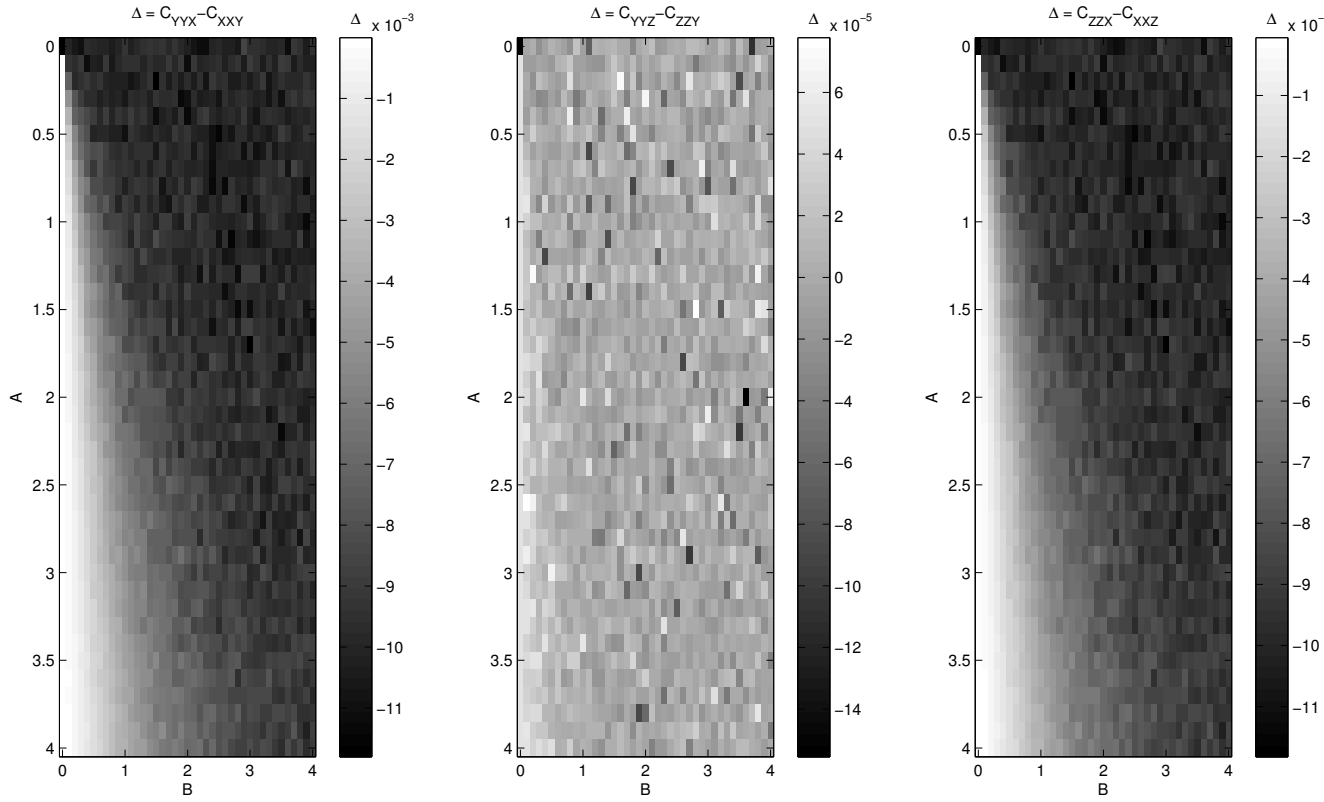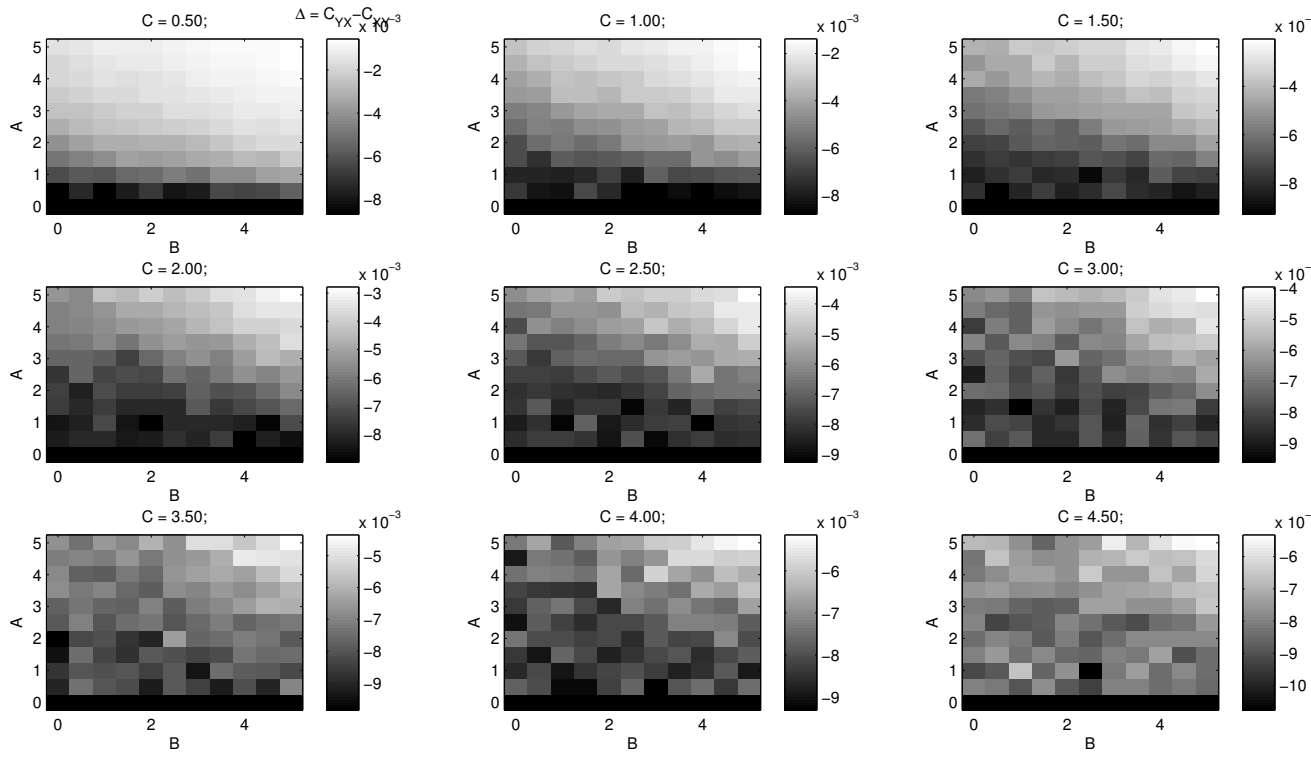Figure 16: Changing A and B. C_{XY} and C_{YX}

Figure 17: Linear Example 1 PAI

Figure 18: Linear Example 2 PAI

Figure 19: Non Linear Example PAI