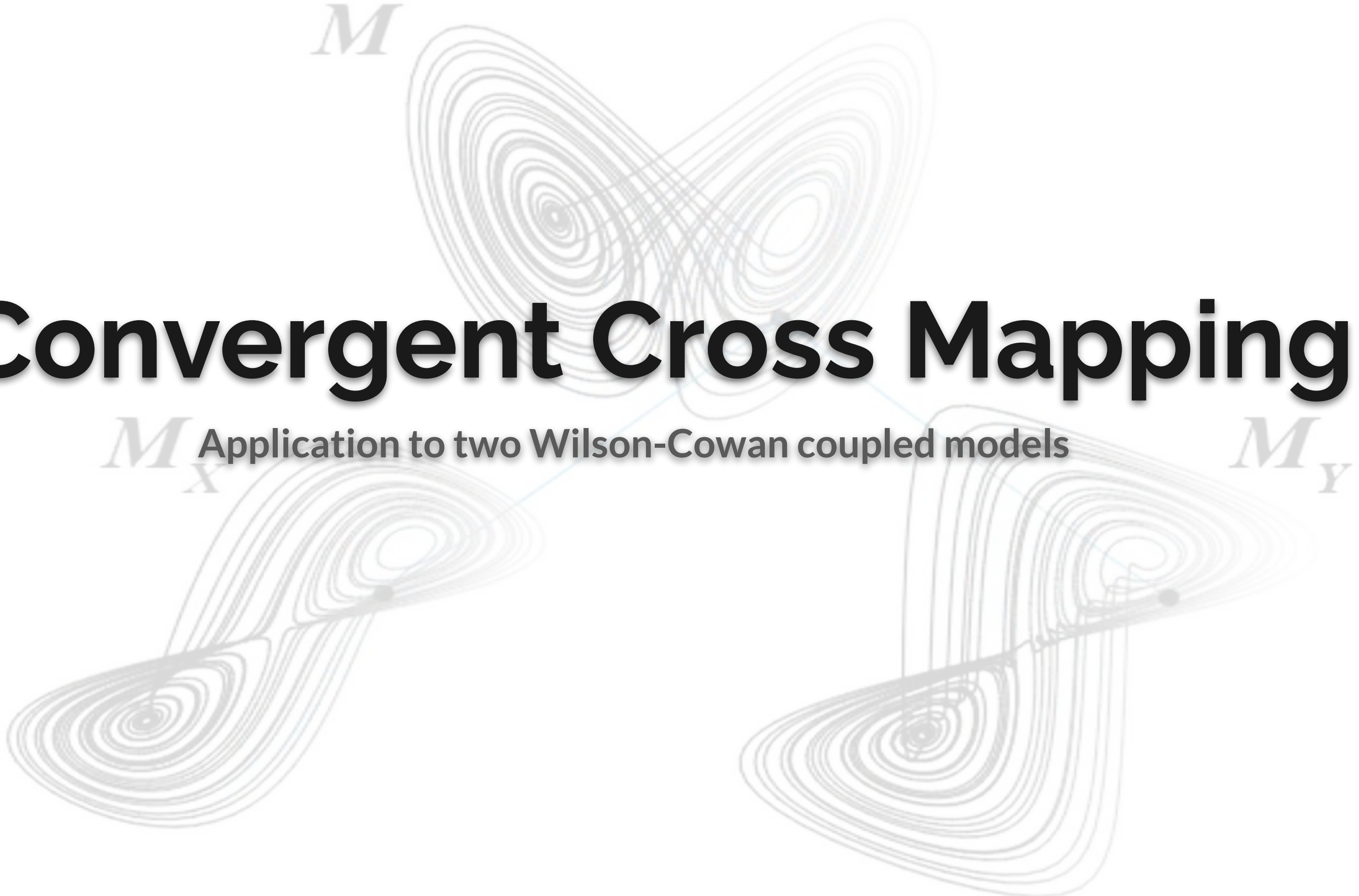
Convergent Cross Mapping

Application to two Wilson-Cowan coupled models



Convergent Cross Mapping

- Most of the presentation is based on the following articles and numerical simulations in Python (https://github.com/DanieleMDiNosse/CCM_WilsonCowan)

Summary

- Introduction
- Outline of the CCM algorithm
- Wilson-Cowan model
- Results from CCM applied to WC time series
- Conclusion

Detecting Causality in Complex Ecosystems

George Sugihara,[1*] Robert May,[2] Hao Ye,[1] Chih-hao Hsieh,[3*] Ethan Deyle,[1] Michael Fogarty,[4] Stephan Munch[5]

Identifying causal networks is important for effective policy and management recommendations on climate, epidemiology, financial regulation, and much else. We introduce a method, based on nonlinear state space reconstruction, that can distinguish causality from correlation. It extends to nonseparable weakly connected dynamic systems (cases not covered by the current Granger causality paradigm). The approach is illustrated both by simple models (where, in contrast to the real world, we know the underlying equations/relations and so can check the validity of our method) and by application to real ecological systems, including the controversial sardine-anchovy-temperature problem.

Biol Cybern (2014) 108:355–363
DOI 10.1007/s00422-014-0604-8

ORIGINAL PAPER

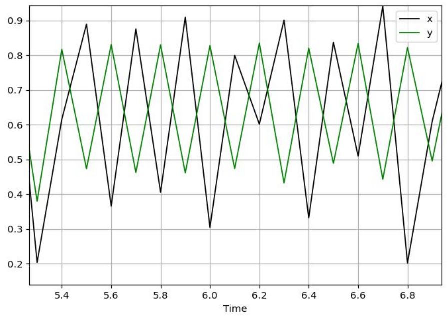
Biological Cybernetics

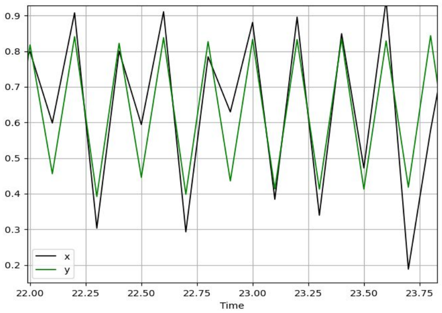
Analysis of chaotic oscillations induced in two coupled Wilson–Cowan models

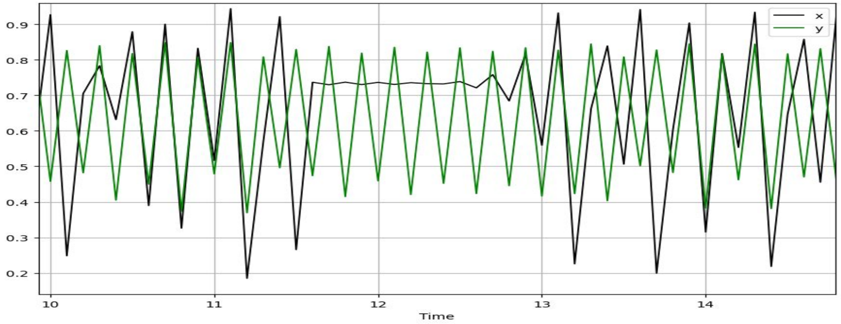
Yuya Maruyama · Yuta Kakimoto · Osamu Araki

Introduction

- “Correlation does not imply causation” (Berkeley, 1710, *A Treatise on Principles of Human Knowledge*).
- It is also true that lack of correlation does not imply lack of causation.
- Being able to identify causation in real world processes can lead to a much better control of the systems (ecology, climate changes, finance ecc.).
- “Mirage correlations” does not simplify our life:



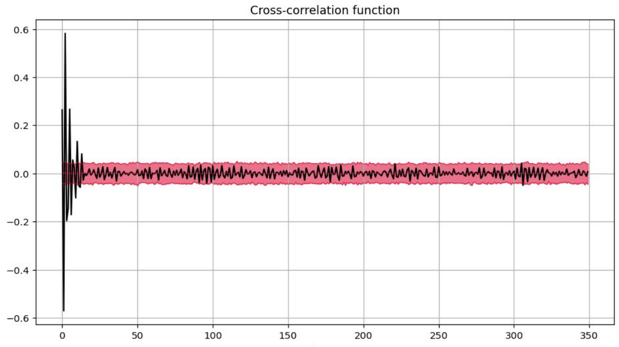




Logistic equations

$$X(t+1) = X(t)\,[3.8 - 3.8X(t) - 0.02Y(t)]$$
$$Y(t+1) = Y(t)\,[3.5 - 3.5Y(t) - 0.1X(t)]$$

- Sometimes we are just not looking to the whole picture



Introduction - Granger Analysis

- The standard causation identifier is the Granger Analysis.
- Let $x(t), y(t)$ two stationary time series. The central idea of G.A. is to compare predictions made by two different *linear* models: one that depends on past $x(t)(y(t))$ values alone and one that depends on both time series past values.

Univariate AR

$$x(t_i) = \sum_{k=1}^{p} a_k x(t_{i-k}) + \nu_1(t_i)$$

$$y(t_i) = \sum_{k=1}^{p} b_k y(t_{i-k}) + \omega_1(t_i)$$

Bivariate AR

$$x(t_i) = \sum_{k=1}^{p} c_k x(t_{i-k}) + \sum_{k=1}^{p} d_k y(t_{i-k}) + \nu_2(t_i)$$

$$y(t_i) = \sum_{k=1}^{p} e_k y(t_{i-k}) + \sum_{k=1}^{p} f_k x(t_{i-k}) + \omega_2(t_i)$$

- Let then $\overline{U}$ be the universe of all possible causative variables. If

$$\sigma^2\{(x|\bar{U})\} < \sigma^2\{(x|\bar{U} - y)\} \Longrightarrow$$ $y(t)$ is said to Granger causes $x(t)$

Introduction - Granger Analysis

- Key requirements are **stationarity** and **separability**.
- Stationary concerns the statistical properties of the process that must not change in time.
- Separability allows us to remove the effect of one variable from the other.
- In stochastic or linear dynamics G.A. works fine, but in nonlinear world (with weak to moderate coupling) separability can not be satisfied.

$$
\begin{aligned}
X(t+1) &= \alpha X(t)\left[1 - X(t) - \beta Y(t)\right] \\
Y(t+1) &= \omega Y(t)\left[1 - Y(t) - \delta X(t)\right]
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
X(t+1) &= f\left(X(t-1), X(t)\right) \\
Y(t+1) &= g\left(Y(t-1), Y(t)\right)
\end{aligned}
\quad \rightarrow
$$

G.A. should conclude, for example, that Y does not cause X.

- **Non separability arises from the redundant causative information already contained in the affected variables** (a consequence of Takens' Theorem).

We need something different...

Introduction - Granger Analysis

"It also follows from the definitions that a purely deterministic series, that is, a series which can be predicted exactly from its past terms such as a nonstochastic series, cannot be said to have any causal influences other than its own past. This may seem to be contrary to common sense in certain special cases but it is difficult to find a testable alternative definition which could include the deterministic situation." (Granger, 1969)

We need something different...

Convergent Cross Mapping - Takens' Theorem

- Of fundamental role is the following theorem by Floris Takens (1981)

Let $\boldsymbol{M}$ be a compact manifold of dimension d on which a smooth vector field ϕ and a smooth function X are defined. It is a generic property that

$$\mathbf{\Phi}_{(\phi,X)}(\vec{m}) : \boldsymbol{M} \to \mathbb{R}^{2d+1}$$

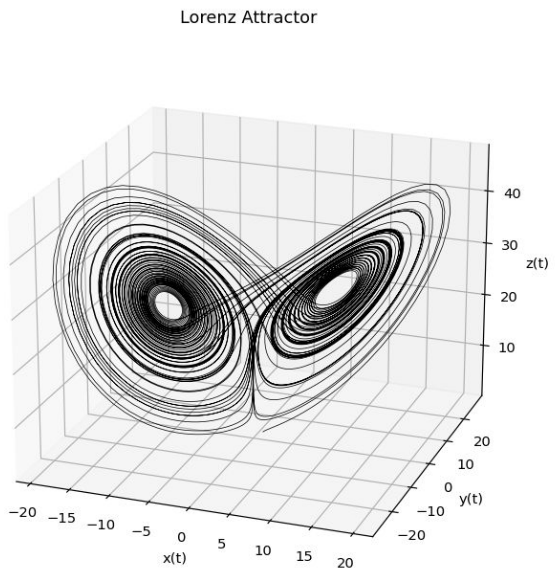
is an embedding, where

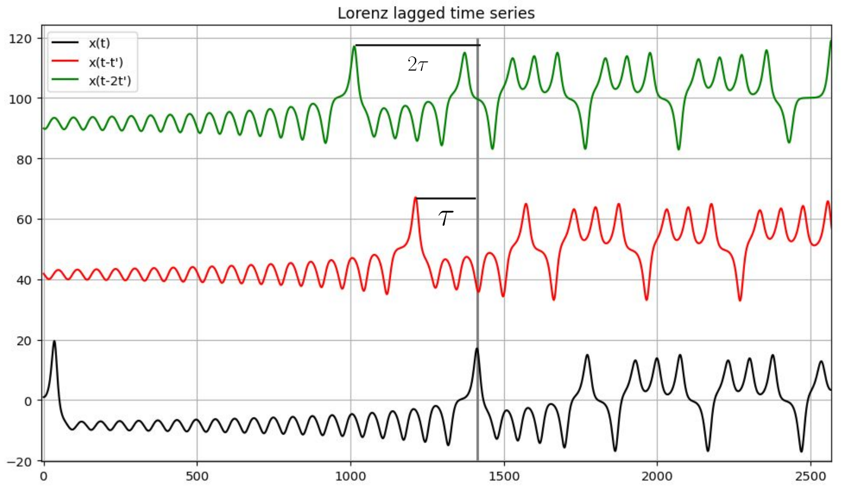
$$\mathbf{\Phi}_{(\phi,X)}(\vec{m}) = <X(\vec{m}), X(\phi(\vec{m})), X(\phi^2(\vec{m})), ..., X(\phi^{2d}(\vec{m}))>$$

- In our framework it allow us to say that if we have a certain manifold, we can reconstruct it by just looking at one of its projection on one axis.

Convergent Cross Mapping - Lagged time series

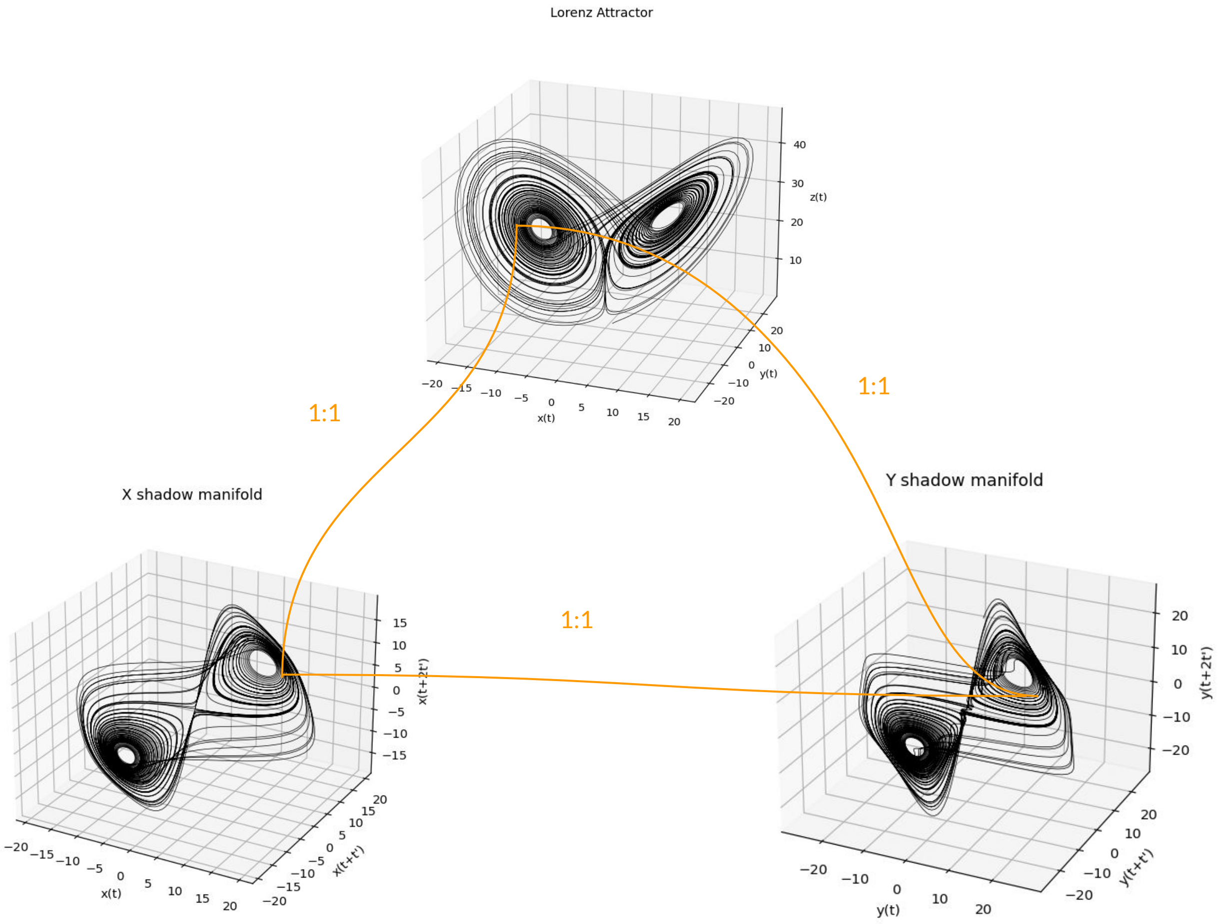
- In particular, we are going to consider lagged version of one time series obtained by projection of the original manifold on a specified axis.
- Let consider, for example, the Lorenz system:





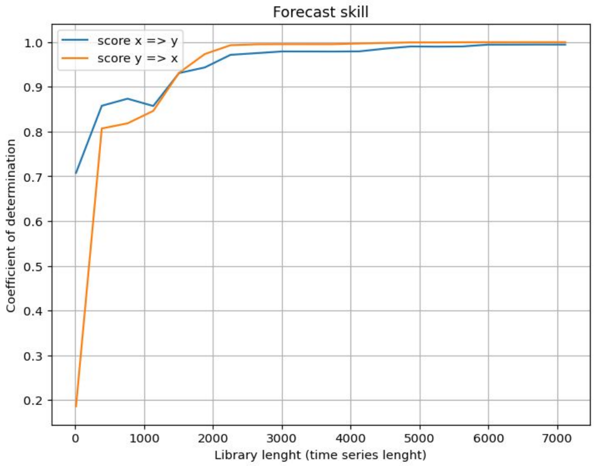
Convergent Cross Mapping - Shadow manifolds

- In the embedding space we can construct shadow manifolds of the original one.
- There is a 1:1 relation between them.
- The idea is to use states of the X shadow manifold to estimate states on the Y shadow manifold via a k-nearest-neighbors procedure.

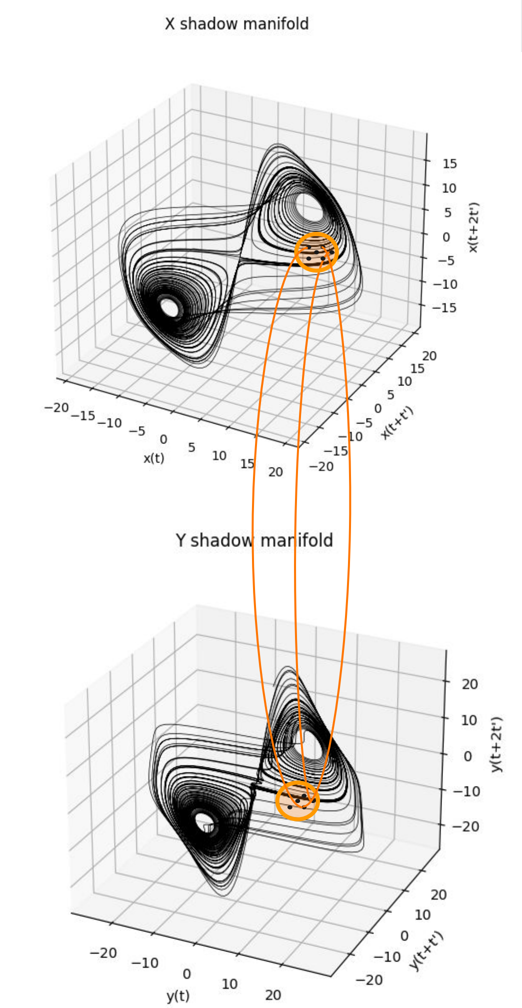


Convergent Cross Mapping - KNN

- Via a KNN we can determine nearby time indices of the X shadow manifold.
- We use these indices to estimate states on the Y shadow manifold.
- The goodness of the estimate can be calculated through the coefficient of determination.
- **The better the estimate the stronger the causal effect of one variable over the other is**.
- Additionally, **the longer the time series, the better the estimate is** (convergence).



➔ In the Lorenz system both x(t) and y(t) are coupled and so there is a bidirectional causal effect



Convergent Cross Mapping - Convergence (example)

- Intuitively, the denser the attractors are the lesser the estimation error would be, due to the narrower nearest neighbors.
- To see this more quantitatively, reconsider the two coupled logistic equations:

$$\begin{aligned} X(t+1) &= \alpha X(t)\left[1 - X(t) - \beta Y(t)\right] \\ Y(t+1) &= \omega Y(t)\left[1 - Y(t) - \delta X(t)\right] \end{aligned} \Longrightarrow \begin{aligned} X(t) &= \frac{\alpha}{\delta}\left[(1-\beta Y(t-1))\left(1 - Y(t-1) - \frac{Y(t)}{\omega Y(t-1)}\right) - \frac{1}{\delta}\left(1 - Y(t-1) - \frac{Y(t)}{\omega Y(t-1)}\right)^2\right] \\ Y(t) &= \frac{\omega}{\beta}\left[(1-\delta X(t-1))\left(1 - X(t-1) - \frac{X(t)}{\alpha X(t-1)}\right) - \frac{1}{\beta}\left(1 - X(t-1) - \frac{X(t)}{\alpha X(t-1)}\right)^2\right] \end{aligned}$$

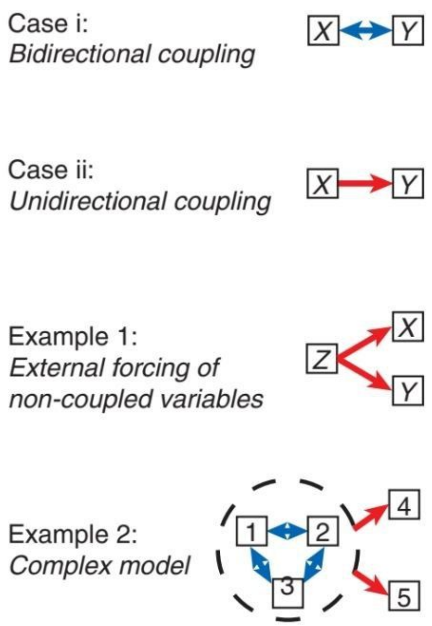
- Fixing $\alpha, \omega, \delta \implies \beta$ is the only parameter that can change and represents the strength of influence of Y over X.

- We have that (at the first order) $Var(Y) \propto \dfrac{Var(X)}{\beta^2}$

- The uncertainty of Y will be greater when the attractor of X is less dense. Furthermore, as β increases the variance of Y decreases . The rate of convergence provides an index of interaction strength.

CCM - Distinction between correlation and causation

- Causation is a transitive
  - If $X \Leftrightarrow Y \Leftrightarrow Z$ implies $X \Leftrightarrow Z$, as well as $X \Rightarrow Y$ and $Y \Rightarrow Z$ implies $X \Rightarrow Z$
- Transitivity can be used to distinguish variables that are coupled from those sharing a common driver.
- Complex networks with different couplings between their part can be unfolded in direct and indirect interaction.



Convergent Cross Mapping

Take home messages:

- ❖ Granger Analysis can not be used in non separable systems (like weakly coupled ones).
- ❖ Time series are causally related if they are coupled and are part of the same dynamical system.
- ❖ If X => Y, then information about X must be encoded in the shadow manifold of Y.

- There are two details not already explained:
  - How do we choose the value of the lag?
  - How do we choose, then, the dimension of the embedding space?
- We are going to tackle these problems in the framework of causal relations in two coupled Wilson-Cowan models.

Wilson-Cowan model

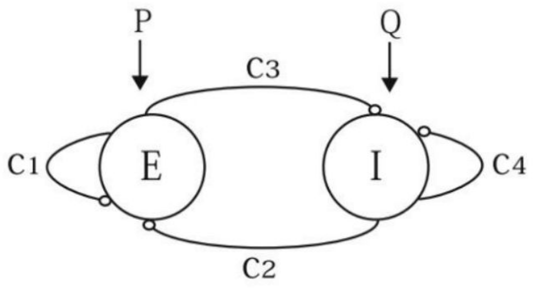
- Wilson-Cowan model (1972, 1973) derived an effective model composed by two differential equations that describes the macroscopic activity of a large population of neurons (neural mass models), where

$$A(t) = lim_{\Delta t \to \infty} \frac{1}{\Delta t} \frac{n_{act}(t; t + \Delta t)}{N} = \frac{1}{N} \sum_{j=1}^{N} \sum_{f} \delta(t - t_j^f)$$

- Coupling between excitatory and inhibitory subpopulations is crucial.

$$\tau_e \frac{dE(t)}{dt} = -E(t) + (k_e - E(t))S_e(c_1 E(t) - c_2 I(t) + P)$$

$$\tau_i \frac{dI(t)}{dt} = -I(t) + (k_i - I(t))S_i(c_3 E(t) - c_4 I(t) + Q)$$

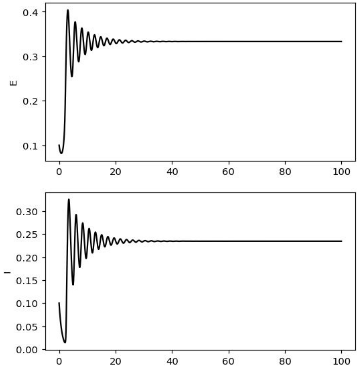


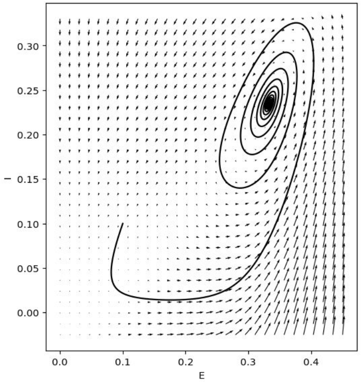
- Excitatory (inhibitory) neurons make their neighbors more (less) likely to become active and inhibit the other population through a sigmoid activation function $S_e(S_i)$ of the present state and external input $P(Q)$.

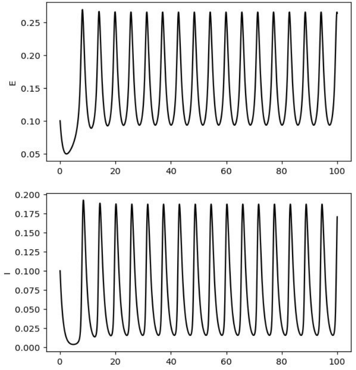
- WC models has been used for various areas such as the visual cortex, thalamus, hippocampus.

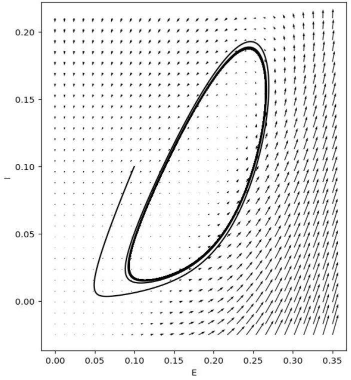
Wilson-Cowan model

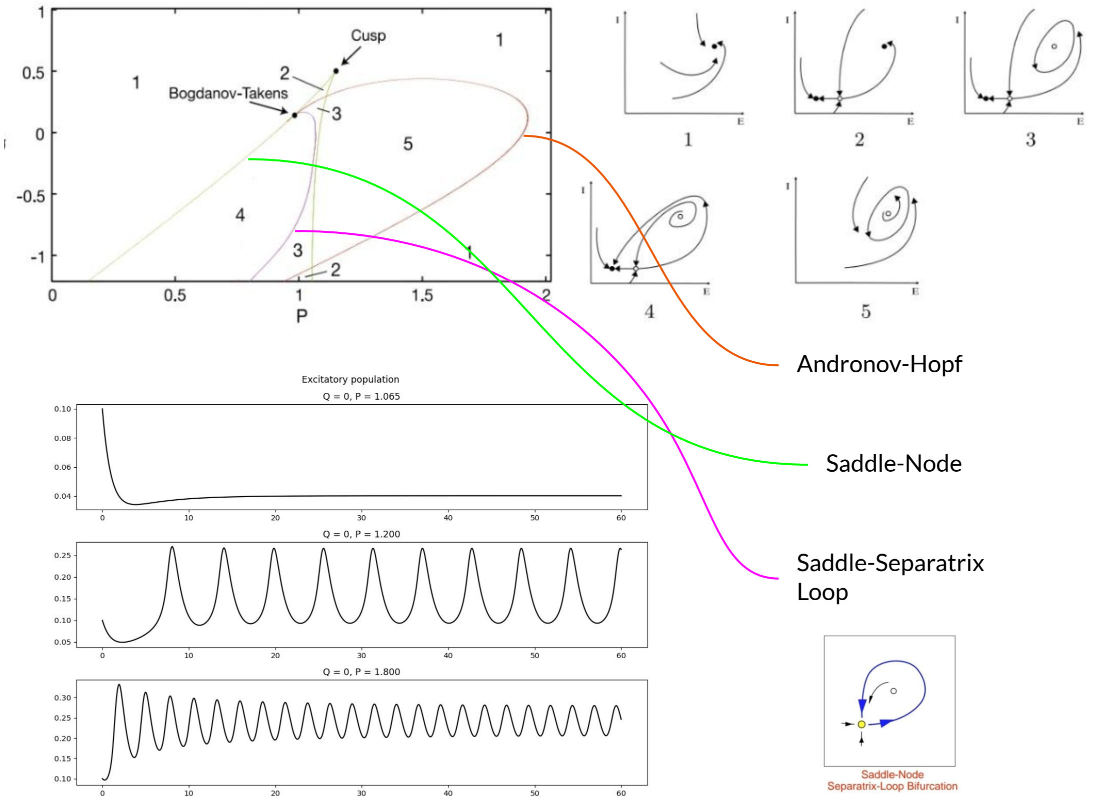
- The model experiences different dynamical behaviours with respect to variations of P and Q







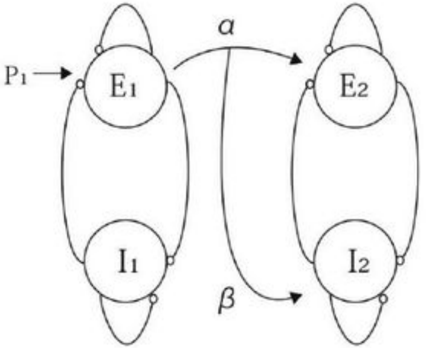




Wilson-Cowan model

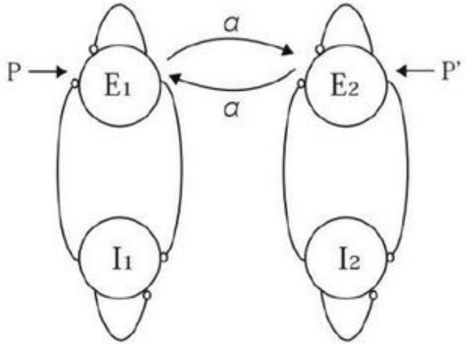
- In order to have interesting (chaotic) time series on which we can test the CCM algorithm, consider two Wilson-Cowan oscillators coupled in two different way:

<u>Unidirectional</u>



$$\tau_e \frac{dE_1(t)}{dt} = -E_1(t) + (k_e - E_1(t))S_e(c_1E_1(t)) - c_2I_1(t) + P_1)$$
$$\tau_i \frac{dI_1(t)}{dt} = -I_1(t) + (k_i - I_1(t))S_i(c_3E_1(t)) - c_4I_1(t))$$
$$\tau_e \frac{dE_2(t)}{dt} = -E_2(t) + (k_e - E_2(t))S_e(c_1E_2(t)) - c_2I_2(t) + \alpha E_1)$$
$$\tau_i \frac{dI_2(t)}{dt} = -I_2(t) + (k_i - I_2(t))S_i(c_3E_2(t)) - c_4I_2(t) + \beta E_1)$$

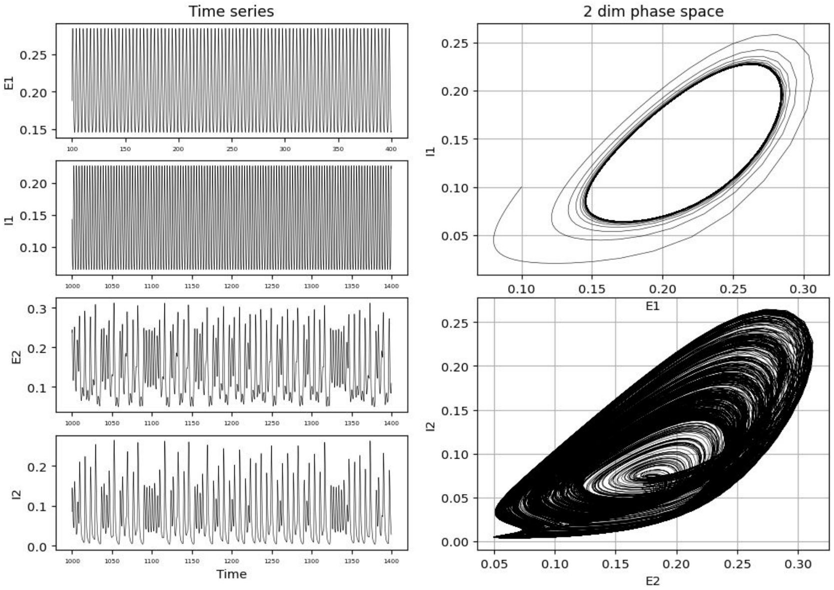
<u>Bidirectional</u>



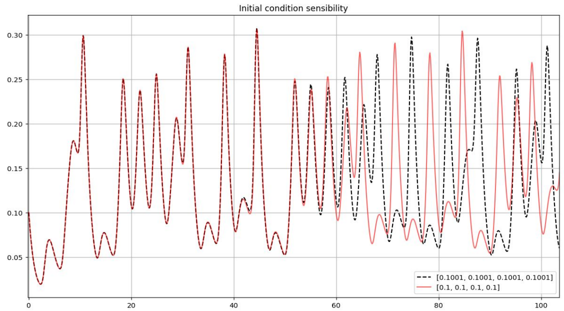
$$\tau_e \frac{dE_1(t)}{dt} = -E_1(t) + (k_e - E_1(t))S_e(c_1E_1(t)) - c_2I_1(t) + P + \alpha E_2)$$
$$\tau_i \frac{dI_1(t)}{dt} = -I_1(t) + (k_i - I_1(t))S_i(c_3E_1(t)) - c_4I_1(t))$$
$$\tau_e \frac{dE_2(t)}{dt} = -E_2(t) + (k_e - E_2(t))S_e(c_1E_2(t)) - c_2I_2(t) + P' + \alpha E_1)$$
$$\tau_i \frac{dI_2(t)}{dt} = -I_2(t) + (k_i - I_2(t))S_i(c_3E_2(t)) - c_4I_2(t))$$

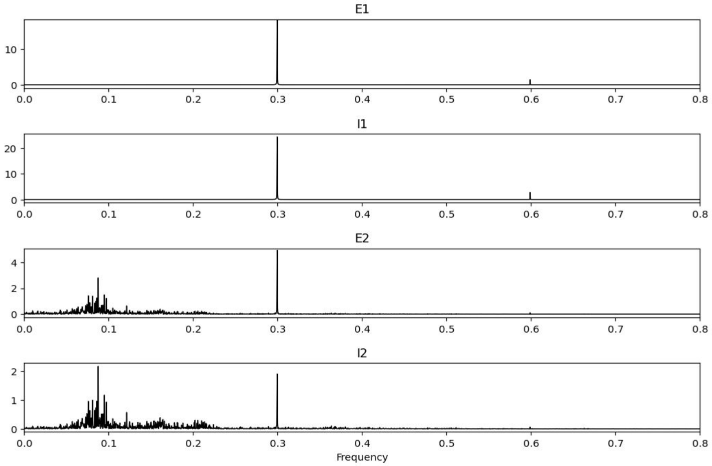
Wilson-Cowan model - Unidirectional case

- Chaotic regime is “verified” looking at power spectra and at strong initial condition sensibility.



➔ Oscillator 1 is the periodic external force for oscillator 2.



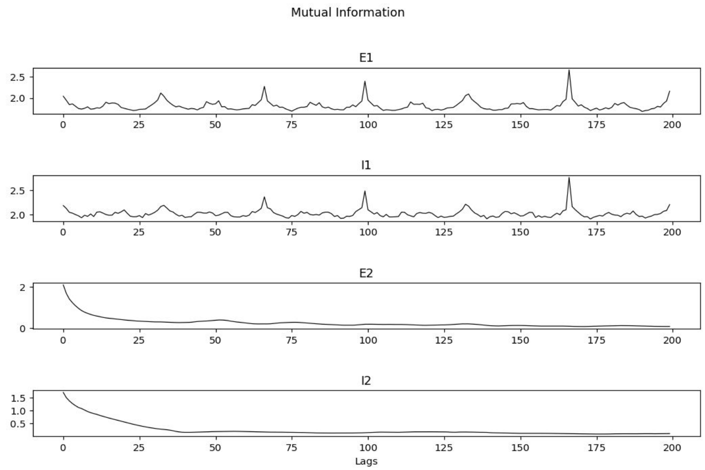


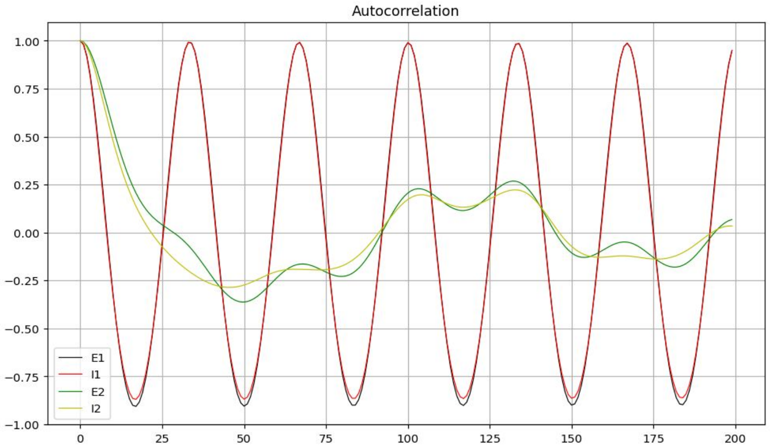
Wilson-Cowan model - Unidirectional case - Lag

- So, **how do we choose the optimal lag value** to be used for the embedding?
- The idea is to choose the one that leads to the least possible redundant information between lagged values in each time series.

$$MI(x(t), x(t+\tau)) = \sum_{x(t), x(t+\tau)} P(x(t); x(t+\tau)) \log_2 \frac{P(x(t); x(t+\tau))}{P(x(t))P(x(t+\tau))}$$

$$AC(x(t), \tau) = \frac{\sum_t [(x(t)-\bar{x})(x(t-\tau)-\bar{x})]}{\sqrt{\sum_t (x(t)-\bar{x})^2 (x(t)-\bar{x})^2}}$$





- MI is more suitable for non linear case, but AC gave similar results (lag ~ 38).

Wilson-Cowan model - Unidirectional case - Embedding dim

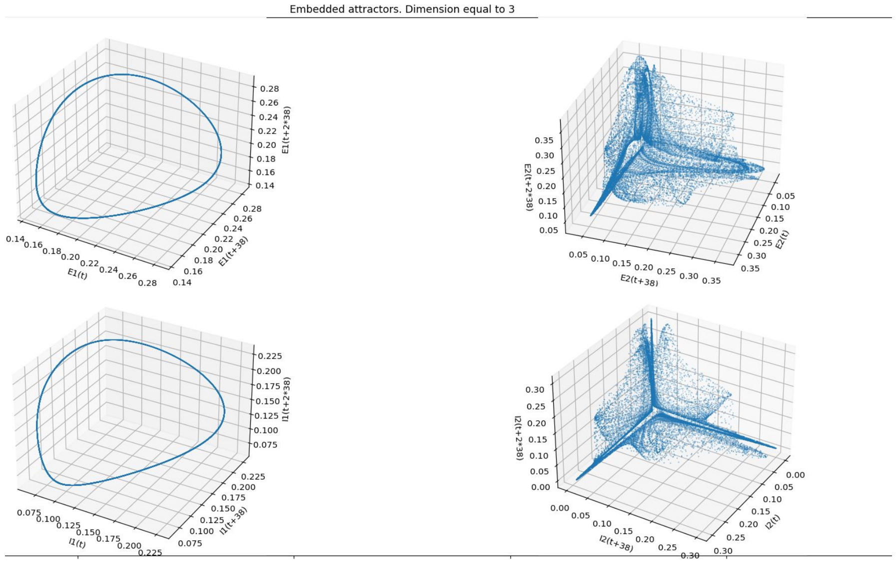
- Then, **how do we choose the optimal embedding dimension**?
- In principle it should be the one that best “unfold” the original manifold.
- For example, if the manifold is a ball of threads (d = 1), they are best recognizable in a 3D space (2d+1)(related to Whitney’s embedding theorem).
- In CCM we check a range of embedding dimension and choose the one that gives the best results.

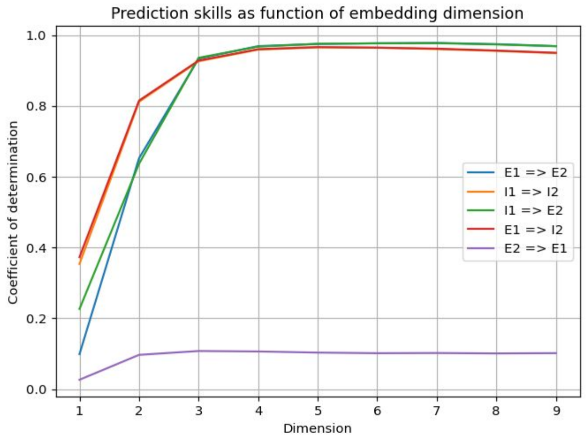
Procedure

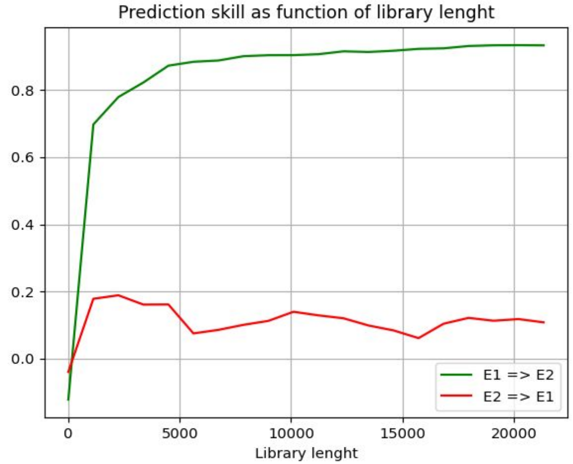
1. Choose an embedding dimension (d)
2. Choose a library length.
3. Divide the two time serie in train sets (75%) and test sets (25%):
   X_train, X_test, Y_train, Y_test
4. Calculate distances from every point in X_test to every point in X_train.
5. Sort these distances and choose the k smaller ones (k = 2d+1). These are time indices.
6. Use these time indices on Y_train to estimate states via a weighted average and compare with Y_test
7. Repeat for all library length.
8. Repeat for all embedding dimension you chose.

Wilson-Cowan model - Unidirectional case - Embedding dim

- The embedding dimension was chosen looking at the coefficient of determination obtained with the whole time series.
- A 3D embedding space is a good compromise between good estimates and visualization purpose.

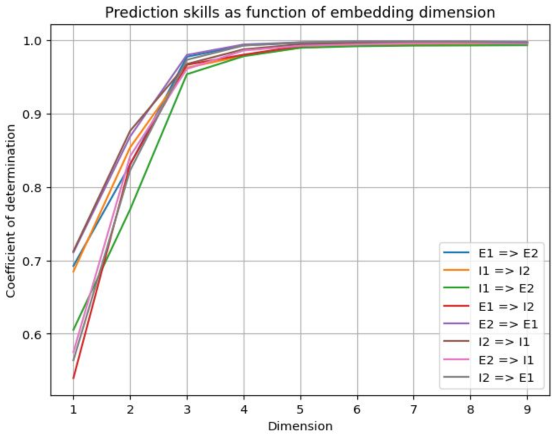


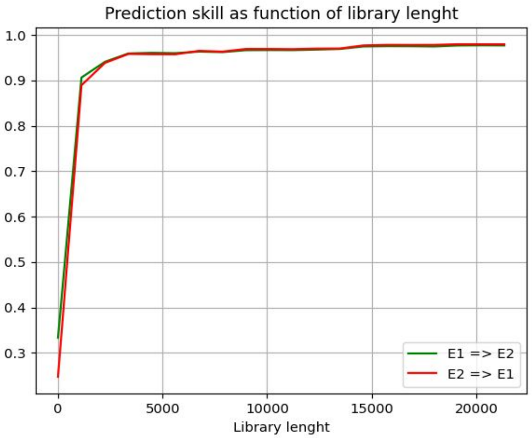




Wilson-Cowan model - Bidirectional case

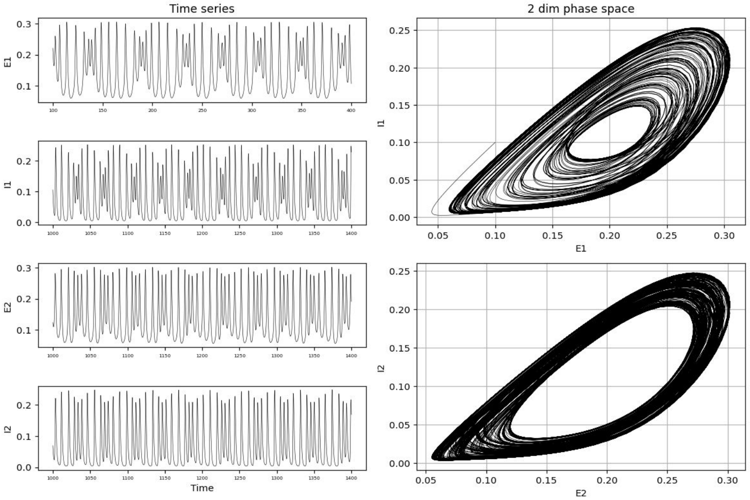
- The bidirectional case gave similar results and so was able to detect the causative relation in both directions.

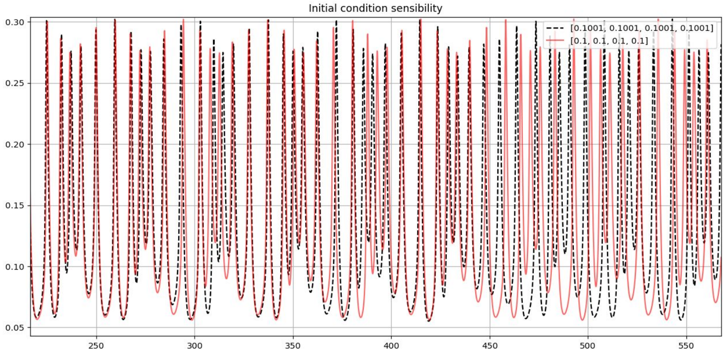


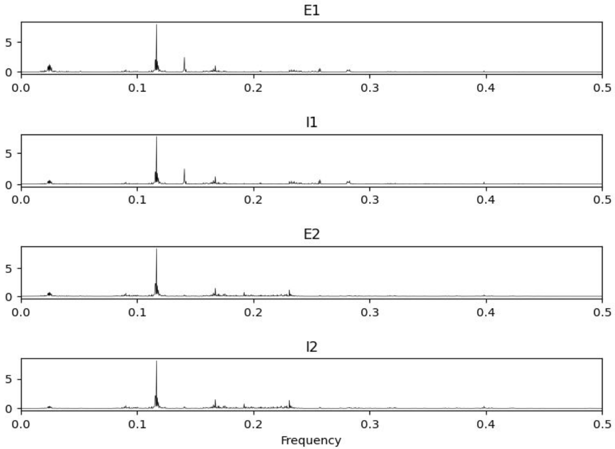


Wilson-Cowan model - Bidirectional case

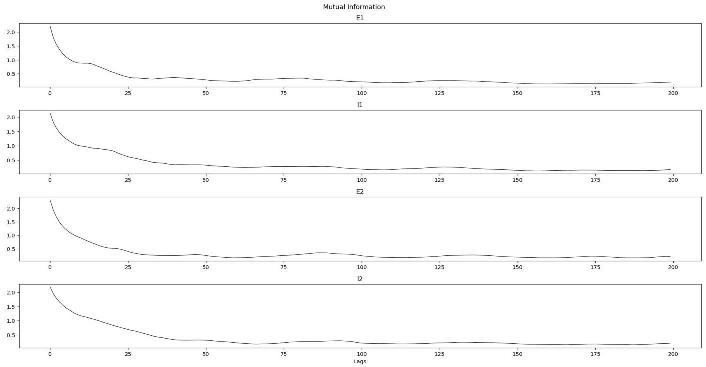
- The bidirectional case gave similar results and so was able to detect the causative relation in both directions.

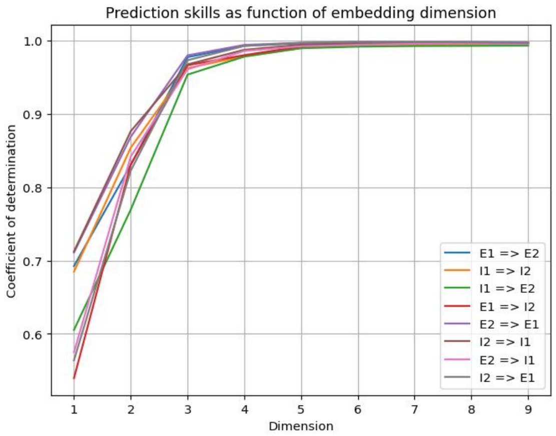


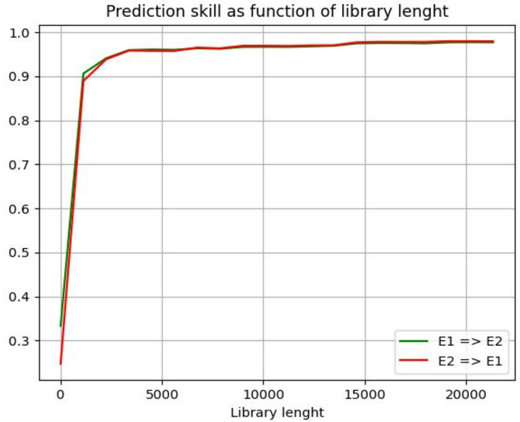


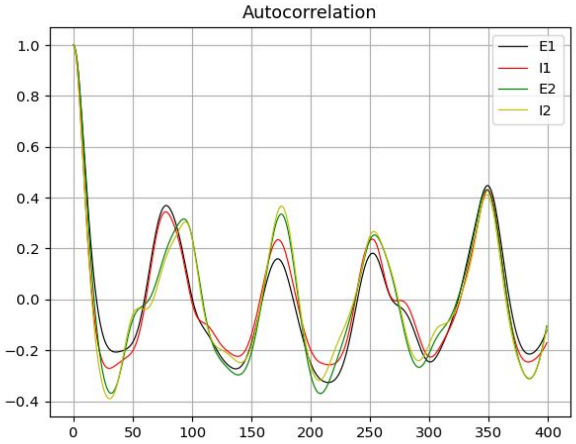


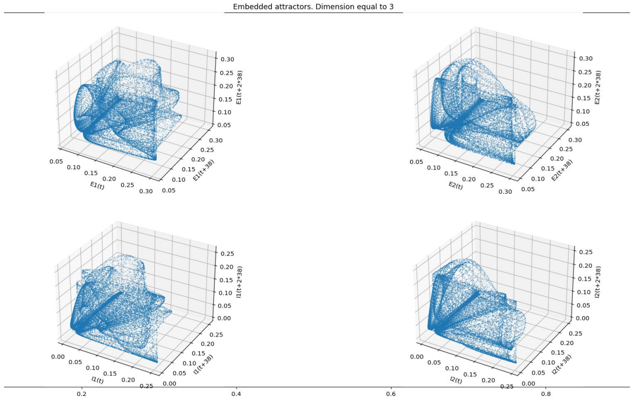
Wilson-Cowan model - Bidirectional case





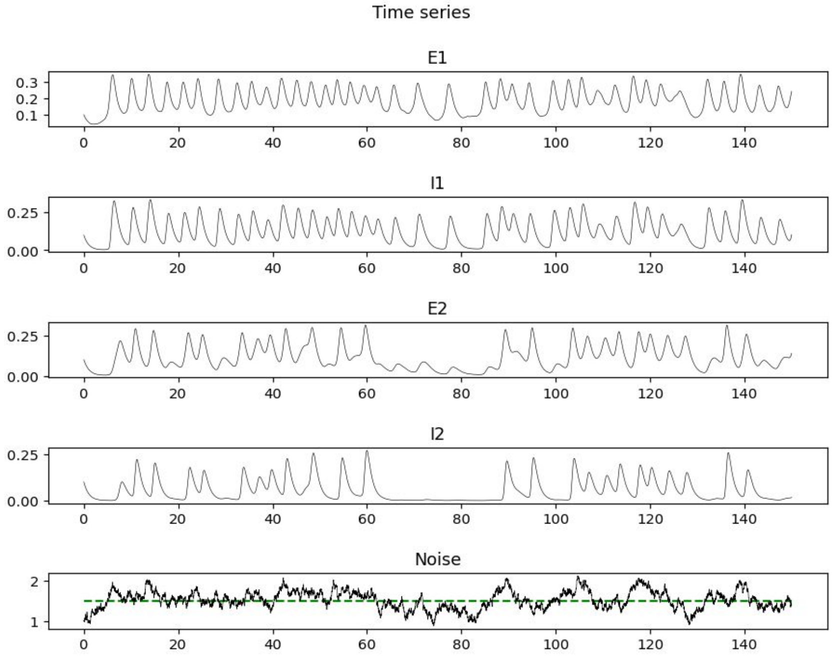


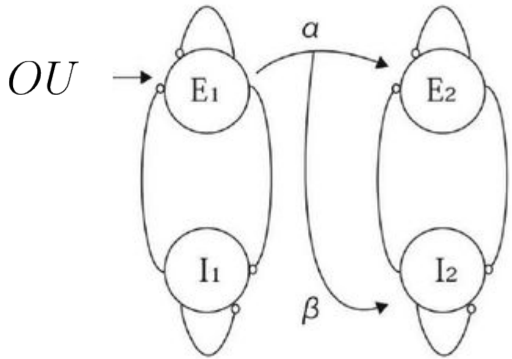




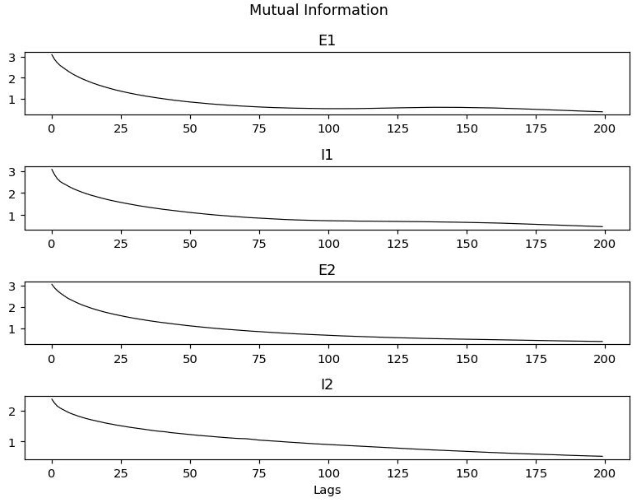
Wilson-Cowan model - Noisy input

- What about a noisy input?
- Unidirectional case in which E_1 views as input an Ornstein-Uhlenbeck process.

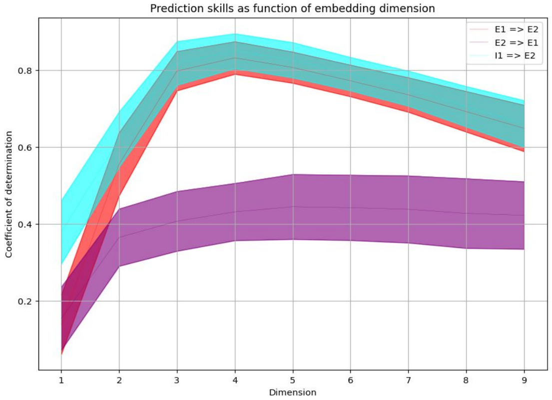


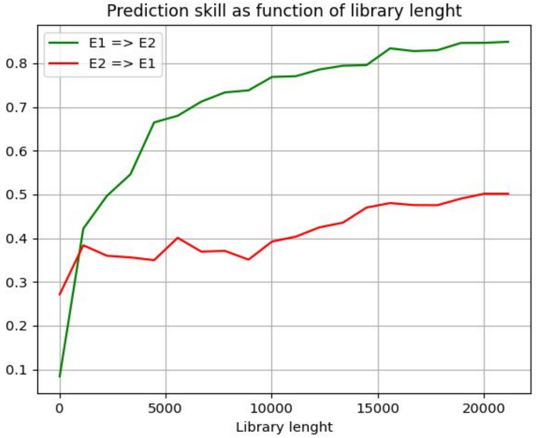


$$d\nu(t) = \theta(\mu - \nu(t))dt + \sigma dW(t)$$



Conclusion





- An overview of the gold standard of causal interaction identification was given.
- Limitations of Granger Analysis were explained.
- Non separability was exploited for causal interaction identification via CCM .
- The method was applied to a chaotic regime in two coupled Wilson-Cowan models.
- It has been shown that CCM was able to identify correctly causal links even with noisy input.