A Tutorial on Causal Inference in Dynamical Systems

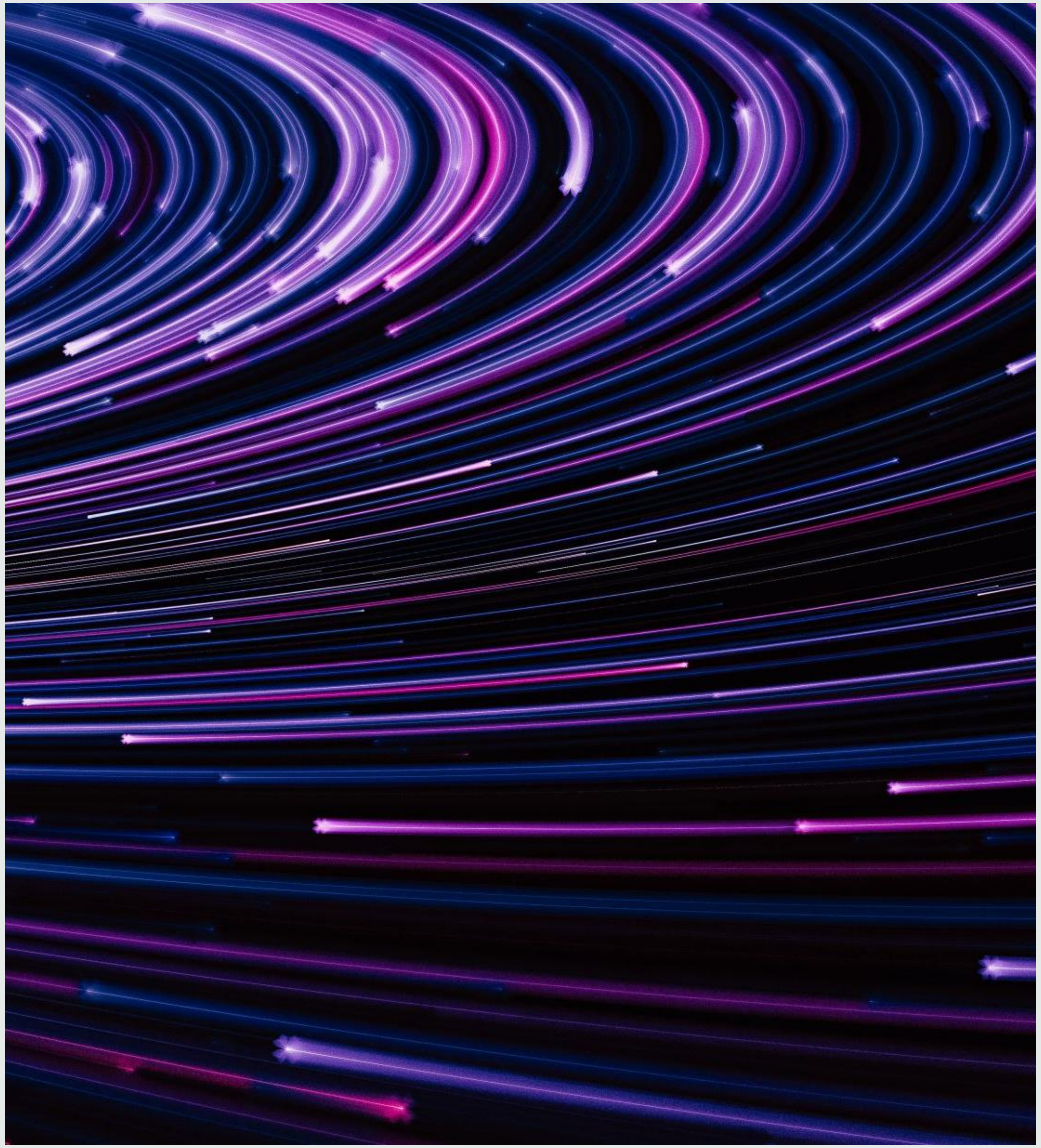
Kurt Butler
Stony Brook University

2021 Bellairs Workshop on Machine Learning and Statistical Signal Processing for Data on Graphs

This talk is based on work with Chen Cui, Sima Mofakham, Yuri Saalmann, Charles Mikell, Jessica Phillips, Guanchao Feng, Marzieh Ajirak, Yuanqing Song, Daniel Waxman, Liu Yang, Lingqing Gan, Yuhao Liu, Petar Djuric and others.

Overview

1. Some causality theory
2. Approaches to causality in dynamical systems
    1. Granger Causality
    2. Convergent Cross Mapping
    3. Mixed-State Methods
3. Some directions for future work

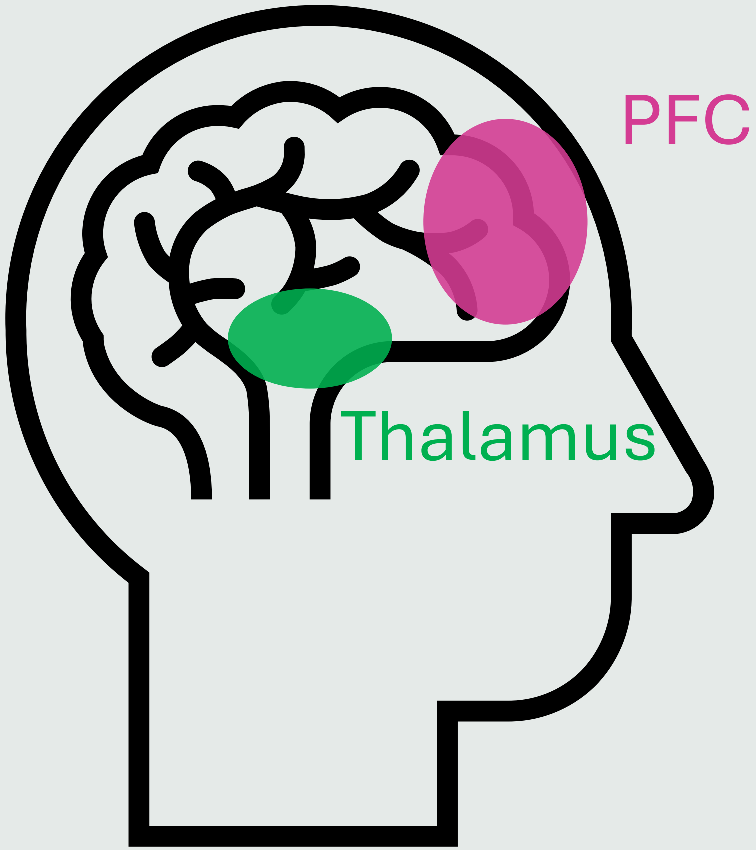


A Neuroscience Problem

Does neural activity in the thalamus cause activity in the prefrontal cortex?

The Problem: Given a dataset, it is unclear if

- Thalamus influences PFC unidirectionally.
- Thalamus and PFC are coupled.
- Thalamus and PFC are both driven by a (latent) activity in the rest of the brain.



Causal Inference

Causal inference is the subject that attempts to distinguish cause-and-effect from statistical correlation.

The most popular theory of causal inference studies **interventions**.

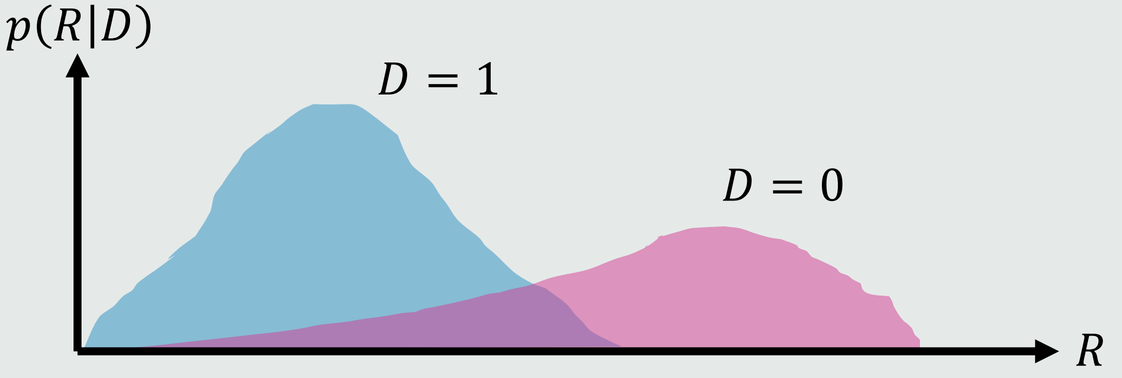
D: Boolean (0,1) equal to 1 if patient receives drug.
R: Time needed to recover from disease.

$$p(R, D) = p(R|D)p(D)$$

Example:

If a drug causes patients to recover from a disease sooner,

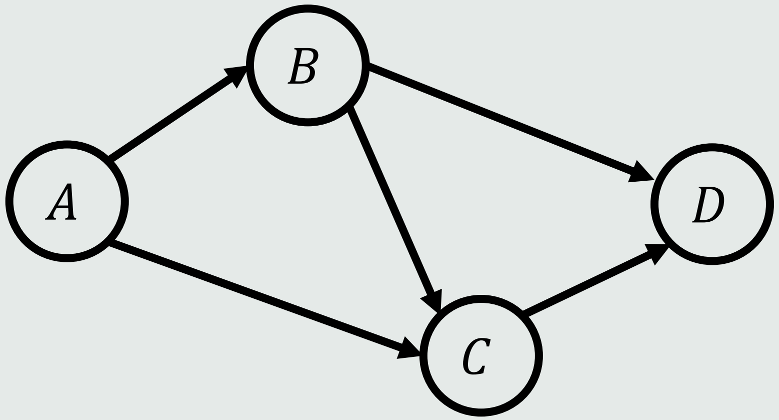
then taking the drug (an intervention) should change the probability distribution of recovery time.



do-Calculus

When we intervene on a system, the causal graph should change.

The *do* operator tells us how to modify a graph when we intervene on a system.†

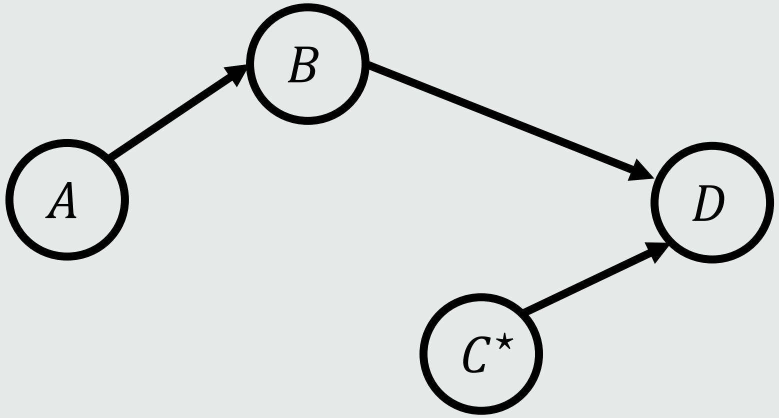


Original Graph

$$p(A, B, C, D) = p(D|B, C)p(C|A, B)p(B|A)p(A)$$

Consider an intervention where we force $C = 3$.

The *do* operator says that we replace $p(C|A, B)$ with $p(C^\star) = \delta(C^\star - 3)$.



Graph after Invention

$$p^{do}(A, B, C^\star, D) = p(D|B, C^\star)p(C^\star)p(B|A)p(A)$$

† Pearl, J., 2009. *Causality*. Cambridge university press.

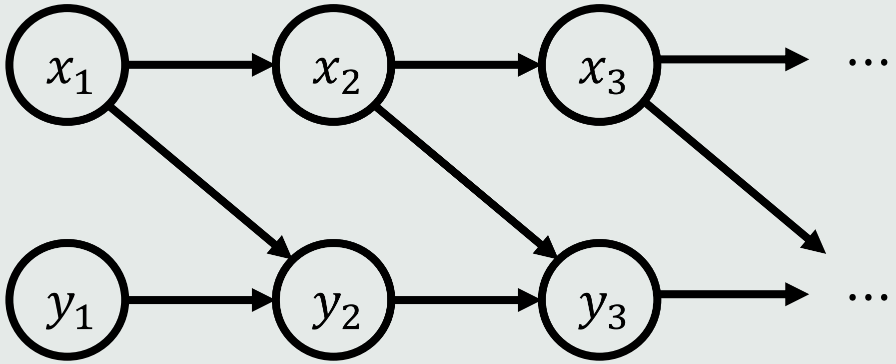
Causal Inference with Processes

Causal inference is much harder when we include time in the model.

- Interactions could be lagged or have propagation delays.
- Influences could be bidirectional.
- It's not obvious how to generalize the *do* operator to the time domain.
- It's not obvious how to draw causal graphs for continuous time.

ARX Models

$$x_{n+1} = \alpha x_n + \varepsilon_n$$
$$y_{n+1} = \beta x_n + \delta y_n + \omega_n$$
$$\varepsilon_n, \omega_n \sim N(0,1) \text{ i.i.d.}$$



† Peters, J., Janzing, D. and Schölkopf, B., 2017. *Elements of causal inference: foundations and learning algorithms* (p. 288). The MIT Press

Granger Causality

GC is one of the oldest approaches to causal inference for processes.

We say that a signal x_t **Granger-causes** a signal y_t if

$$p(y_{t+1}|x_{1:t}, y_{1:t}, z_{1:t}) \neq p(y_{t+1}|, y_{1:t}, z_{1:t})$$

i.e. the history of x_t contains unique information that can be used to predict y_{t+1}.

Testing of the Granger hypothesis can be done using linear models (linear GC) or transfer entropy (general GC). †

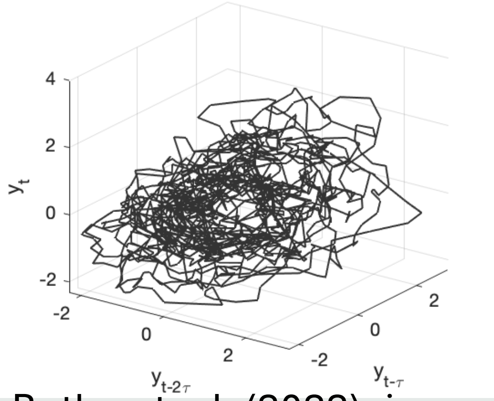
Despite it's popularity, Granger causality has some caveats:

- GC requires signals to be stochastic and stationary.
- Dynamical systems may violate the assumptions of GC (separability).‡
- It's difficult to condition on signals that we don't observe (z_t).

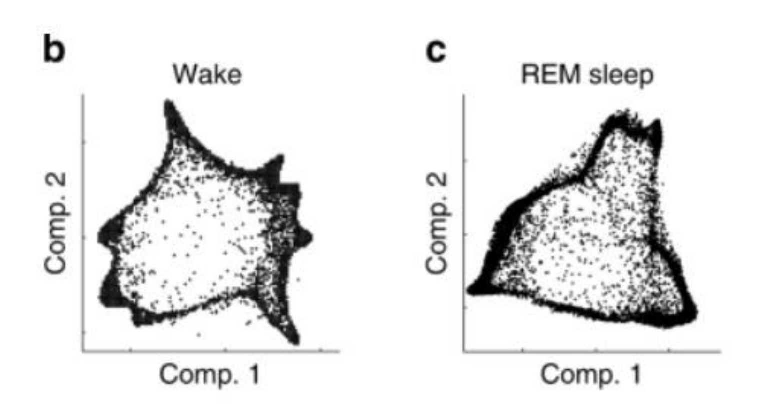
† Yuan, A.E. and Shou, W., 2021. Data-driven causal analysis of observational time series in ecology. *bioRxiv*, pp.2020-08.
‡ Sugihara, G., May, R., Ye, H., Hsieh, C.H., Deyle, E., Fogarty, M. and Munch, S., 2012. Detecting causality in complex ecosystems. *science*, *338*(6106), pp.496-500.

Low Dimensional Dynamics in Neuroscience

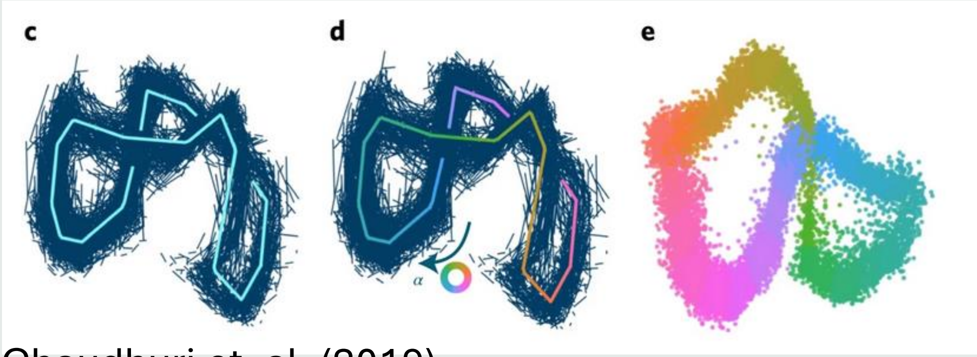
- Real-life neurological systems are clearly high-dimensional.
- However, in some cases the dynamics of the population may reside in a lower dimensional manifold.



Butler et. al. (2022), in review.



Rubin et. al. (2019).



Chaudhuri et. al. (2019)

† Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A. and Fiete, I., 2019. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, *22*(9), pp.1512-1520.

‡ Rubin, A., Sheintuch, L., Brande-Eilat, N., Pinchasof, O., Rechavi, Y., Geva, N. and Ziv, Y., 2019. Revealing neural correlates of behavior without behavioral measurements. *Nature communications*, *10*(1), pp.1-14.

* Butler K., Feng, G., Mikell, C., Mofakham, S., and Djurić, P.M., 2022, May. Predicting Latent States of Dynamical Systems with State-Space Reconstruction and Gaussian Processes. In review to be published in *ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.

State Space Models

Linear state model

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$$

For some matrix $\mathbf{A}$.

- The possible dynamics are determined by the eigenvalues of $\mathbf{A}$.
- The only set that may attract trajectories is the origin.
- Solution curves can be computed directly by the matrix exponential.

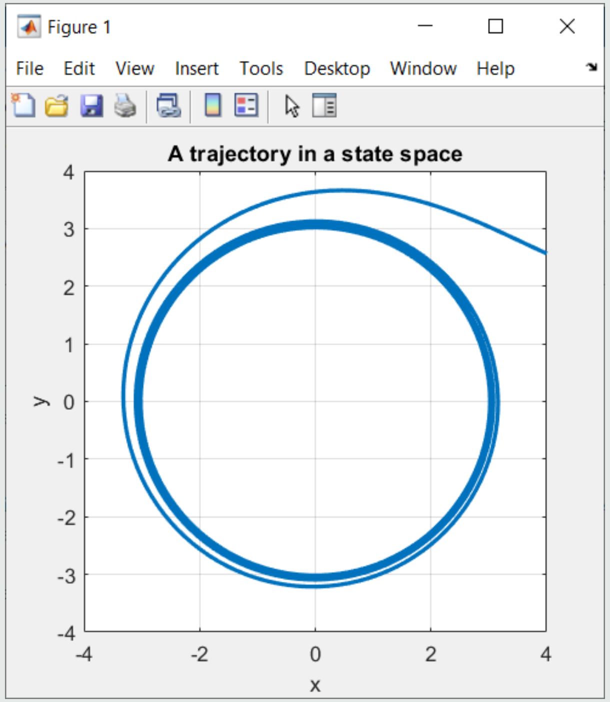
Nonlinear state model

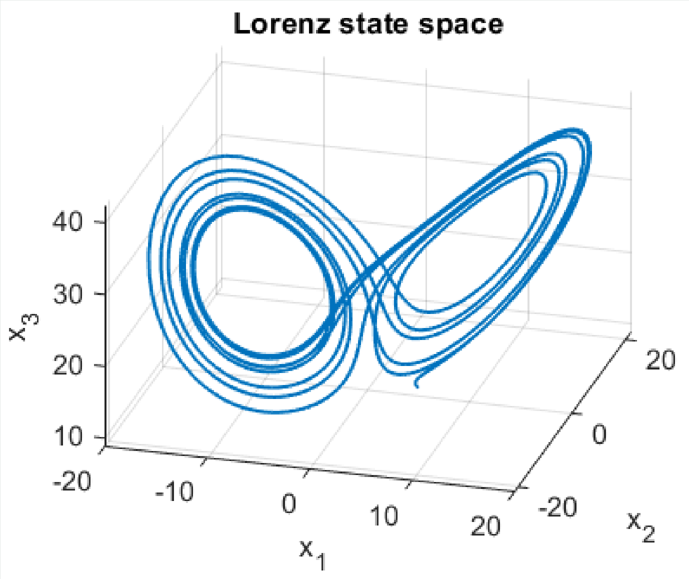
$$\dot{\mathbf{x}} = F(\mathbf{x})$$

For some smooth function F.

- Even simple functions can produce chaotic dynamics (e.g. logistic function).
- Attracting sets can be arbitrarily complicated or even fractal.
- Solution curves can only be computed if the ODE is tractable.

Attractors and Invariant Sets





- An **attractor** is a subset of the state space that attracts and traps state trajectories.
- Even simple nonlinear systems can have attractors with complicated geometry.
- We can exploit the existence of low-dimensional attractors to do scientific tasks like causal inference.

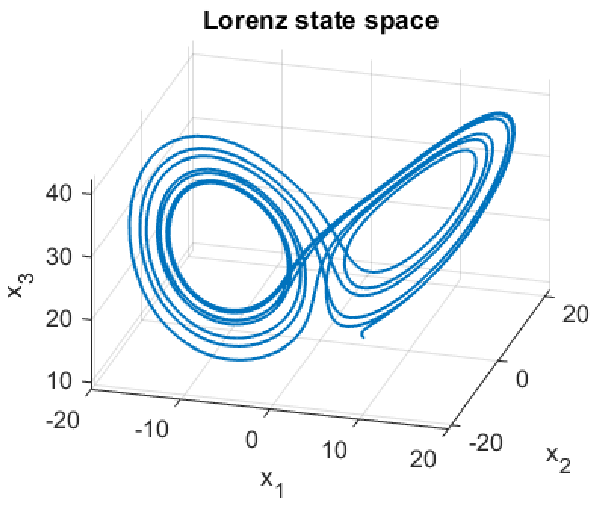
State Space Reconstruction

Given a (scalar) signal $y(t)$, we define the Q-dimensional **delay embedding** $\mathbf{m}_y(t)$ as a vector-valued signal:

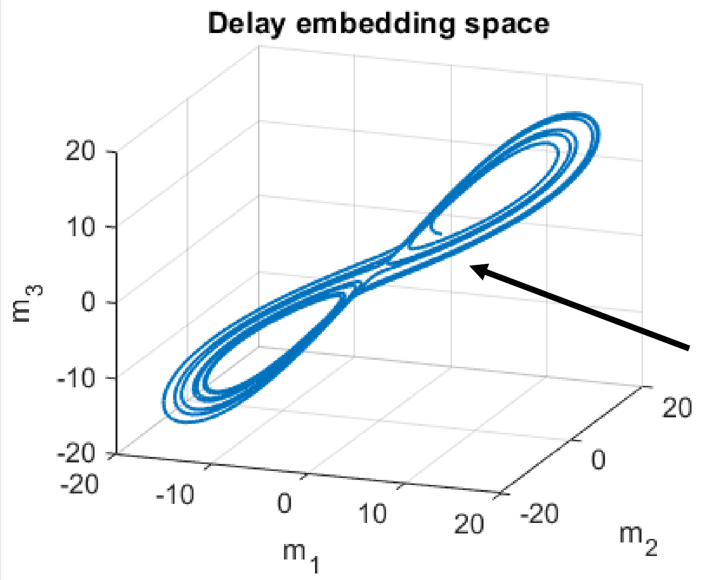
$$\mathbf{m}_y(t) = [y(t-(Q-1)\tau) \quad \cdots \quad y(t-\tau) \quad y(t)]$$

Taken's theorem† says that the delay embedding an observation signal will embed the latent attractor whenever $Q > 2d_A$ (almost surely).

$$\dot{\mathbf{x}} = F(\mathbf{x})$$
$$y = G(\mathbf{x})$$



Delay embedding gives a diffeomorphism between the latent attractor A and the shadow manifold M_x.



M_x
“shadow manifold”

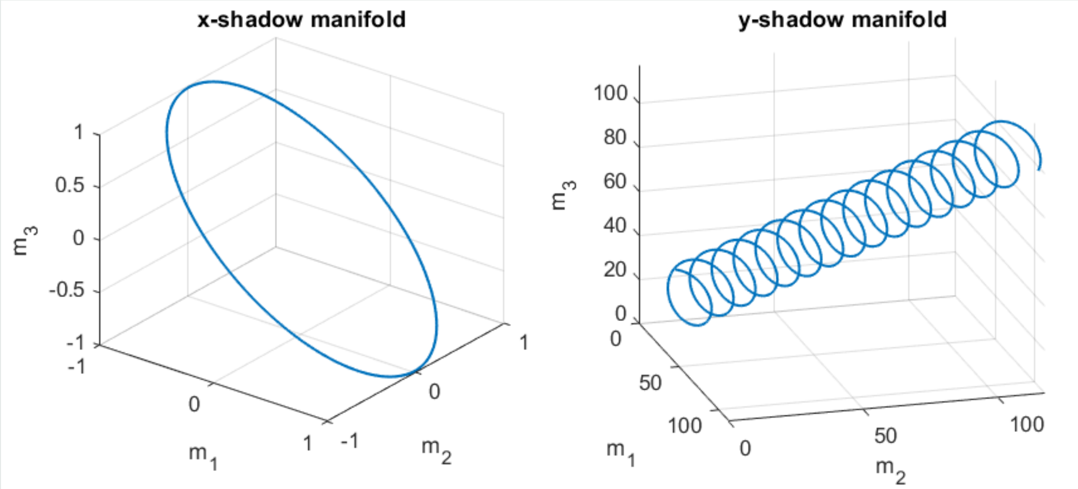
† Sauer, T., Yorke, J.A. and Casdagli, M., 1991. Embedology. Journal of statistical Physics, 65(3), pp.579-616.

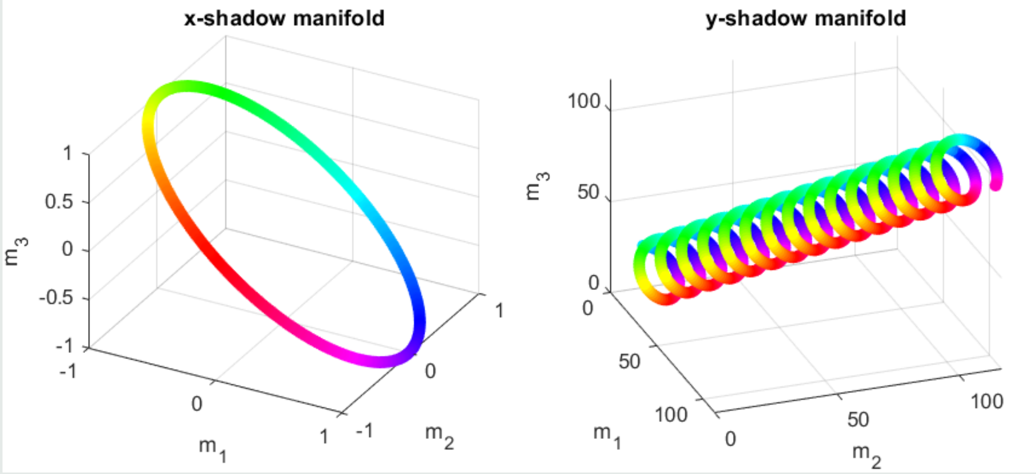
Convergent Cross Mapping

If **one signal drives the other**, then there exists a cross-map between the shadow manifolds that goes in the **opposite direction.** †

Causal graph:

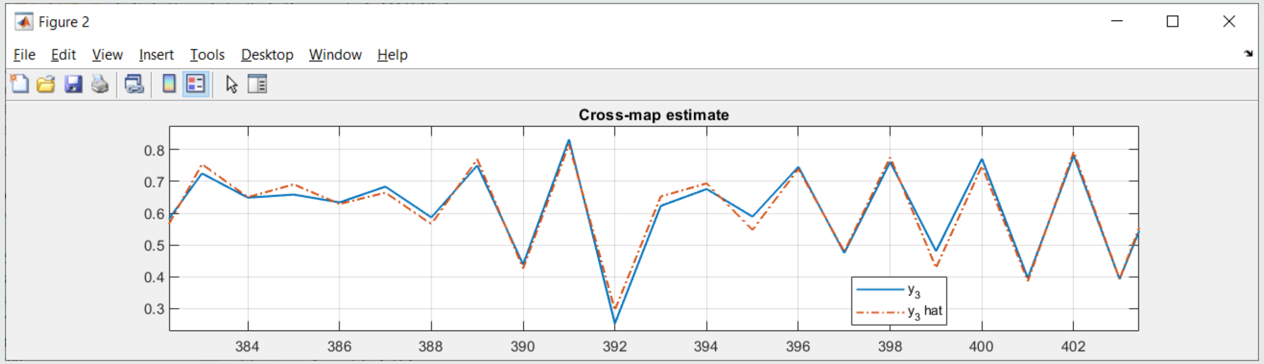
Shadow manifolds:

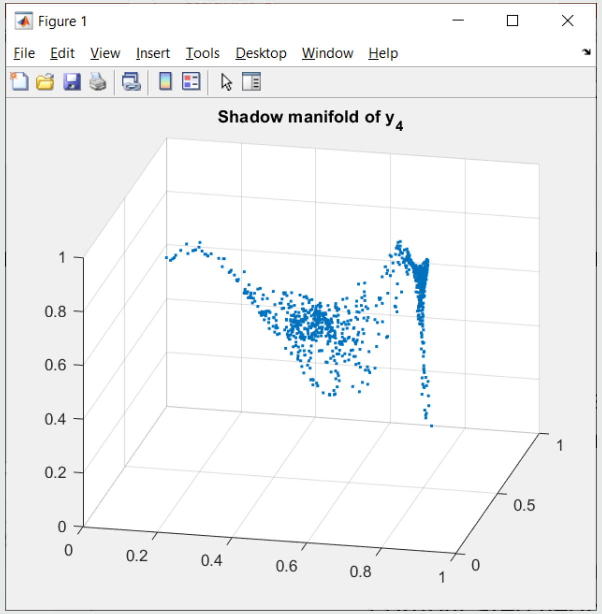


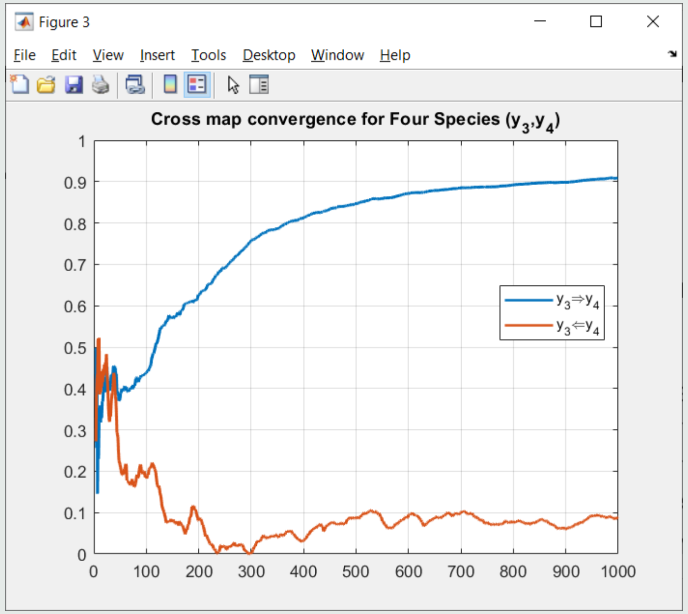


† Sugihara, G., May, R., Ye, H., Hsieh, C.H., Deyle, E., Fogarty, M. and Munch, S., 2012. Detecting causality in complex ecosystems. *science*, *338*(6106), pp.496-500.
‡ Feng, G., Yu, K., Wang, Y., Yuan, Y. and Djurić, P.M., 2020, May. Improving Convergent Cross Mapping for Causal Discovery with Gaussian Processes. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3692-3696). IEEE.

CCM Example





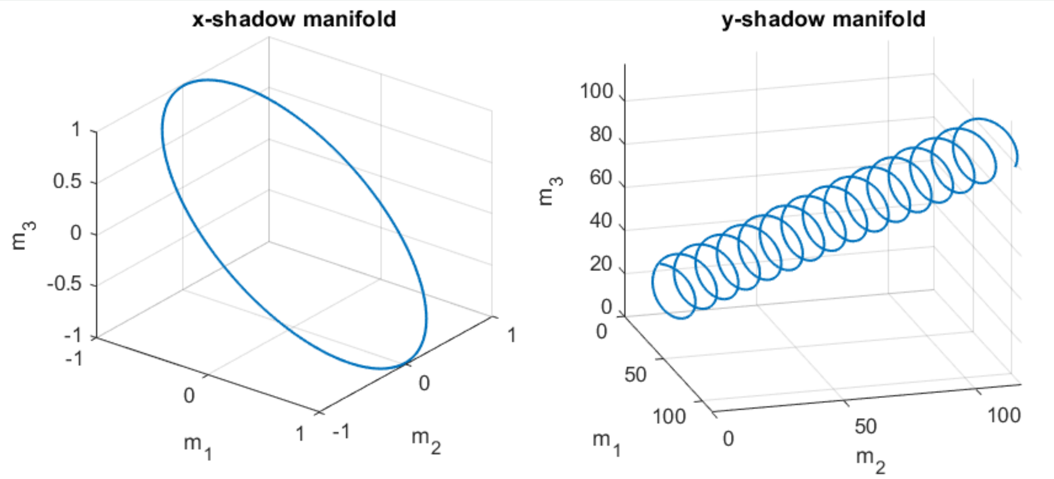


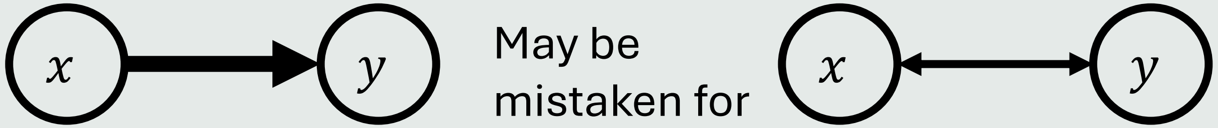
To test if there is an arrow $x \to y$:

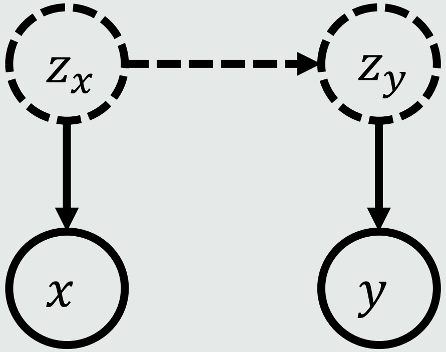
1. Produce the shadow manifold M_y.
2. Learn the cross-map and make predictions $\hat{x} = f(\mathbf{m}_y)$.
3. Repeat this procedure in an online manner, and check if the cross-map is **converging** to an accurate estimator.

Caveats of CCM

- If there are not attractors or invariant sets in the latent space, then learning the cross-map from data is infeasible.
- CCM does not work for largely stochastic systems.
- CCM can mistake strong unidirectional forcing for bidirectional coupling.
- CCM tests for causality *in the latent space*.







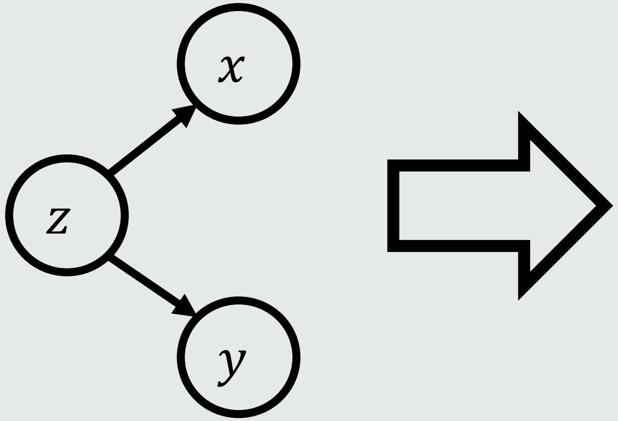
Mixed-State Methods

Mixed-state SSR is when we concatenate the delay-embedding vectors of various measurements.

$$\mathbf{m}_{xy}(t) = \left(\mathbf{m}_x(t), \mathbf{m}_y(t)\right)$$

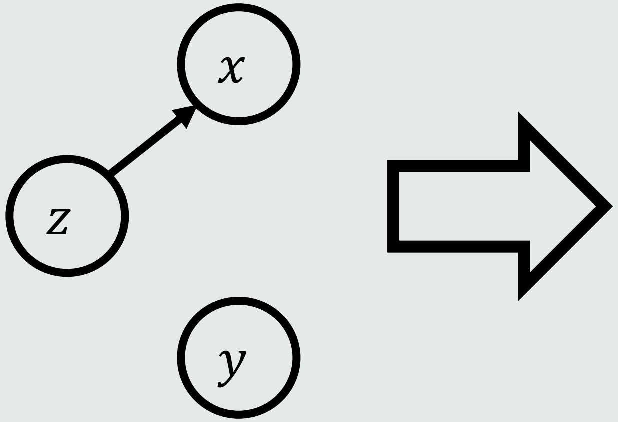
The mixed-state embedding does **not duplicate topological information**.

Therefore, the dimensions of the shadow manifolds can identify **common causes**. †



$$M_{xy} \cong M_x \cong M_y$$

$$\dim(M_{xy}) = \dim(M_x) = \dim(M_y)$$

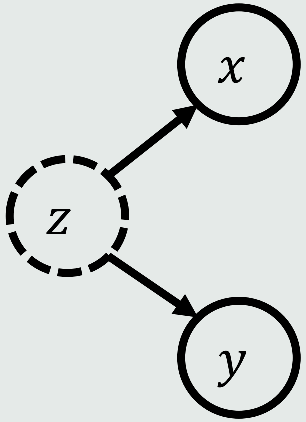


$$M_{xy} \cong M_x \times M_y$$

$$\dim(M_{xy}) = \dim(M_x) + \dim(M_y)$$

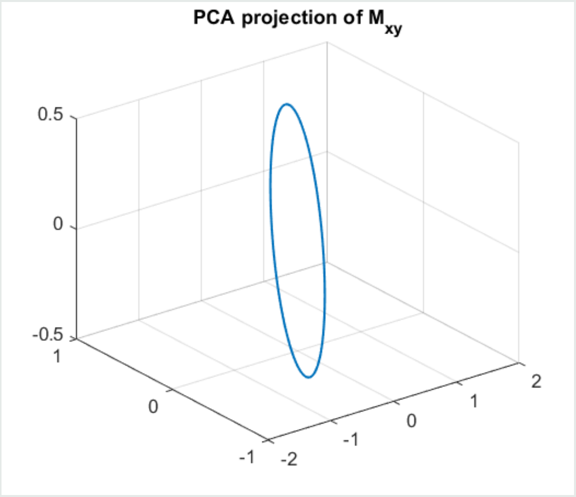
† Benko, Z., Zlatniczki, A., Fabó, D., Sólyom, A., Eross, L., Telcs, A. and Somogyvári, Z., 2018. Exact inference of causal relations in dynamical systems. *arXiv preprint arXiv:1808.10806*.

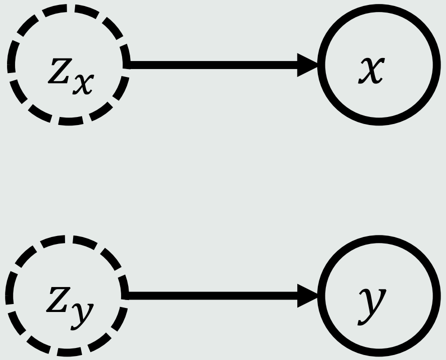
Mixed-State Methods



$$x(t) = \sin(ft)$$
$$y(t) = \sin(ft + \phi)$$

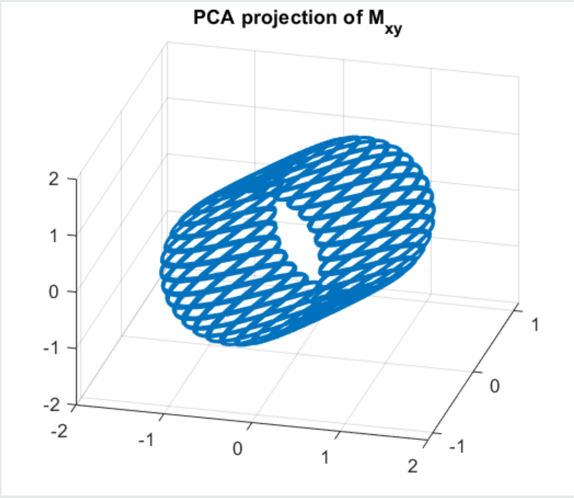
$$M_x \cong \mathcal{C}$$
$$M_y \cong \mathcal{C}$$
$$M_{xy} \cong \mathcal{C}$$





$$x(t) = \sin(f_1 t)$$
$$y(t) = \sin(f_2 t)$$
$$f_1/f_2 \notin \mathbb{Q}$$

$$M_x \cong \mathcal{C}$$
$$M_y \cong \mathcal{C}$$
$$M_{xy} \cong \mathcal{C} \times \mathcal{C}$$



Directions for future work

- Methods for analyzing systems with both stochastic and dynamical elements.
- Applications of state-space reconstruction to systems without attractors.
    - Modeling latent dynamics and properties (e.g., dissipation)
    - Static causal inference
    - Time-varying causal inference
- Generalization of CCM and other methods to more than pairs of variables.