Department of Mathematical Sciences

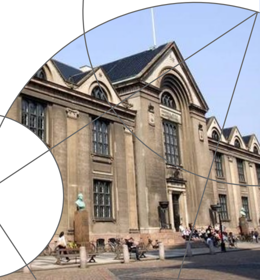
Causal Inference in Dynamical Systems

Convergent Cross Mapping and Alternative Approaches

Master's Defense

Rasmus Juhl Christensen
stud.scient.

January 15th, 2024



Program

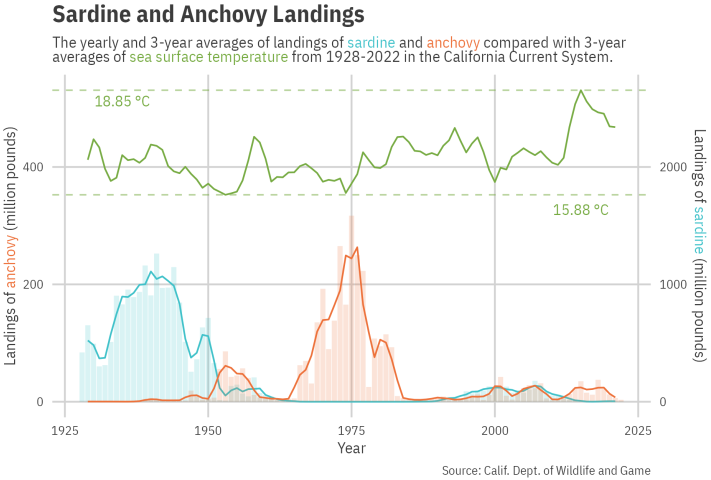
10.30-11.00: A presentation on the following topics:

1. Time Series
2. Stochastic Models
3. Granger Causality
4. Dynamical Models
5. Takens' Theorem
6. Convergent Cross Mapping
7. Extensions

11.00: Questions.

Time Series

Consider the following example:



Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \ldots, X_T$$
$$Y_1, \ldots, Y_T$$
$$Z_1, \ldots, Z_T$$

We consider thus the following features of our models:

Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \ldots, X_T$$
$$Y_1, \ldots, Y_T$$
$$Z_1, \ldots, Z_T$$

We consider thus the following features of our models:

- Dependence over time

Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \ldots, X_T$$
$$Y_1, \ldots, Y_T$$
$$Z_1, \ldots, Z_T$$

We consider thus the following features of our models:

- Dependence over time
- Randomness or unpredictability

Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \ldots, X_T$$
$$Y_1, \ldots, Y_T$$
$$Z_1, \ldots, Z_T$$

We consider thus the following features of our models:

- Dependence over time
- Randomness or unpredictability
- Interference - what happens when we „push“ a variable?

Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \dots, X_T$$
$$Y_1, \dots, Y_T$$
$$Z_1, \dots, Z_T$$

We consider thus the following features of our models:

- Dependence over time
- Randomness or unpredictability
- Interference - what happens when we „push“ a variable?

The key: building a model. In this presentation, we will introduce and compare two settings:

Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \dots, X_T$$
$$Y_1, \dots, Y_T$$
$$Z_1, \dots, Z_T$$

We consider thus the following features of our models:

- Dependence over time
- Randomness or unpredictability
- Interference - what happens when we „push" a variable?

The key: building a model. In this presentation, we will introduce and compare two settings:

- Stochastic model - randomness coming from noise,

Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \ldots, X_T$$
$$Y_1, \ldots, Y_T$$
$$Z_1, \ldots, Z_T$$

We consider thus the following features of our models:

- Dependence over time
- Randomness or unpredictability
- Interference - what happens when we „push" a variable?

The key: building a model. In this presentation, we will introduce and compare two settings:

- Stochastic model - randomness coming from noise,
- Dynamical model - randomness coming from chaos.

Time Series Characteristics

We imagine the following generical set-up in this presentation:

$$X_1, \ldots, X_T$$
$$Y_1, \ldots, Y_T$$
$$Z_1, \ldots, Z_T$$

We consider thus the following features of our models:

- Dependence over time
- Randomness or unpredictability
- Interference - what happens when we „push“ a variable?

The key: building a model. In this presentation, we will introduce and compare two settings:

- Stochastic model - randomness coming from noise,
- Dynamical model - randomness coming from chaos.

$\longrightarrow$ identifiability from lagged information.

Stochastic Models

We get the randomness in stochastic models via noise terms. For each variable, we will model as follows:

$$X_t = f(\underbrace{X_{t-1}}_{\text{last state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}}, \underbrace{u_t^X}_{\text{noise}})$$

Stochastic Models

We get the randomness in stochastic models via noise terms. For each variable, we will model as follows:

$$X_t = f(\underbrace{X_{t-1}}_{\text{last state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}}, \underbrace{u_t^X}_{\text{noise}})$$

For instance, one could imagine the following situation:

$$X_t = f_X(X_{t-1}, Z_{t-1}, u_t^X)$$
$$Y_t = f_Y(Y_{t-1}, Z_{t-1}, u_t^Y)$$
$$Z_t = f_Z(Z_{t-1}, u_t^Z)$$

Stochastic Models

We get the randomness in stochastic models via noise terms. For each variable, we will model as follows:

$$X_t = f(\underbrace{X_{t-1}}_{\text{last state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}}, \underbrace{u_t^X}_{\text{noise}})$$

For instance, one could imagine the following situation:

$$\begin{aligned} X_t &= f_X(X_{t-1}, Z_{t-1}, u_t^X) \\ Y_t &= f_Y(Y_{t-1}, Z_{t-1}, u_t^Y) \\ Z_t &= f_Z(Z_{t-1}, u_t^Z) \end{aligned}$$

Causality:

- „Changes in the cause will result in changes in the effect“.

Stochastic Models

We get the randomness in stochastic models via noise terms. For each variable, we will model as follows:

$$X_t = f(\underbrace{X_{t-1}}_{\text{last state}}, \underbrace{\dots\dots\dots\dots}_{\text{additional information}}, \underbrace{u_t^X}_{\text{noise}})$$

For instance, one could imagine the following situation:

$$\begin{aligned} X_t &= f_X(X_{t-1}, Z_{t-1}, u_t^X) \\ Y_t &= f_Y(Y_{t-1}, Z_{t-1}, u_t^Y) \\ Z_t &= f_Z(Z_{t-1}, u_t^Z) \end{aligned}$$

Causality:

- „Changes in the cause will result in changes in the effect".
- In this example: Z *causes* X and Y.

Stochastic Models

We get the randomness in stochastic models via noise terms. For each variable, we will model as follows:

$$X_t = f(\underbrace{X_{t-1}}_{\text{last state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}}, \underbrace{u_t^X}_{\text{noise}})$$

For instance, one could imagine the following situation:

$$\begin{aligned} X_t &= f_X(X_{t-1}, Z_{t-1}, u_t^X) \\ Y_t &= f_Y(Y_{t-1}, Z_{t-1}, u_t^Y) \\ Z_t &= f_Z(Z_{t-1}, u_t^Z) \end{aligned}$$

Causality:

- „Changes in the cause will result in changes in the effect".
- In this example: Z *causes* X and Y.
- Reformulated in terms of Granger causality:

$$X_t \not\perp\!\!\!\perp Z_{t-1} \mid (X_{t-1}, Y_{t-1})$$

Dynamical Models

We get the randomness in dynamical models via chaos. We move to continuous setting:

$$\frac{dX_t}{dt} = f(\underbrace{X_t}_{\text{current state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}})$$

Dynamical Models

We get the randomness in dynamical models via chaos. We move to continuous setting:

$$\frac{dX_t}{dt} = f(\underbrace{X_t}_{\text{current state}}, \underbrace{\dots\dots\dots\dots\dots}_{\text{additional information}})$$

Completely deterministic!

Dynamical Models

We get the randomness in dynamical models via chaos. We move to continuous setting:

$$\frac{dX_t}{dt} = f(\underbrace{X_t}_{\text{current state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}})$$

Completely deterministic!

For instance, one could imagine the following situation:

$$\frac{dX_t}{dt} = f_X(X_t, Z_t), \quad \frac{dY_t}{dt} = f_Y(Y_t, Z_t), \quad \frac{dZ_t}{dt} = f_Z(Z_t)$$

Dynamical Models

We get the randomness in dynamical models via chaos. We move to continuous setting:

$$\frac{dX_t}{dt} = f(\underbrace{X_t}_{\text{current state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}})$$

Completely deterministic!

For instance, one could imagine the following situation:

$$\frac{dX_t}{dt} = f_X(X_t, Z_t), \quad \frac{dY_t}{dt} = f_Y(Y_t, Z_t), \quad \frac{dZ_t}{dt} = f_Z(Z_t)$$

Causality:

- „Changes in the cause will result in changes in the effect“.

Dynamical Models

We get the randomness in dynamical models via chaos. We move to continuous setting:

$$\frac{dX_t}{dt} = f(\underbrace{X_t}_{\text{current state}}, \underbrace{\ldots\ldots\ldots\ldots\ldots}_{\text{additional information}})$$

Completely deterministic!

For instance, one could imagine the following situation:

$$\frac{dX_t}{dt} = f_X(X_t, Z_t), \quad \frac{dY_t}{dt} = f_Y(Y_t, Z_t), \quad \frac{dZ_t}{dt} = f_Z(Z_t)$$

Causality:

- „Changes in the cause will result in changes in the effect“.
- In this example: Z *causes* X and Y.

Dynamical Models

We get the randomness in dynamical models via chaos. We move to continuous setting:

$$\frac{dX_t}{dt} = f(\underbrace{X_t}_{\text{current state}}, \underbrace{\dots\dots\dots\dots\dots}_{\text{additional information}})$$

Completely deterministic!

For instance, one could imagine the following situation:

$$\frac{dX_t}{dt} = f_X(X_t, Z_t), \quad \frac{dY_t}{dt} = f_Y(Y_t, Z_t), \quad \frac{dZ_t}{dt} = f_Z(Z_t)$$

Causality:

- „Changes in the cause will result in changes in the effect".
- In this example: Z *causes* X and Y.
- How do we detect it?

Chaotic Attractors

Meteorologist Edward Lorenz at MIT modelled atmospheric convection, and discovered the Lorenz system in 1963:

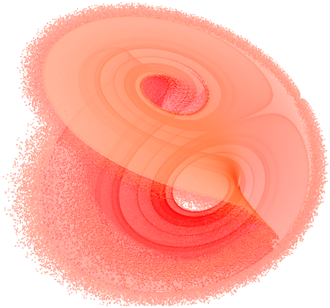
$$\frac{dX_t}{dt} = \sigma(Y_t - X_t), \quad \frac{dY_t}{dt} = -X_t Z_t + rX_t - Y_t,$$
$$\frac{dZ_t}{dt} = X_t Y_t - bZ_t$$

Chaotic Attractors

Meteorologist Edward Lorenz at MIT modelled atmospheric convection, and discovered the Lorenz system in 1963:

$$\frac{dX_t}{dt} = \sigma(Y_t - X_t), \quad \frac{dY_t}{dt} = -X_t Z_t + rX_t - Y_t,$$
$$\frac{dZ_t}{dt} = X_t Y_t - bZ_t$$

Result: a butterfly!



State Space Reconstruction

Strange behavior!

State Space Reconstruction

Strange behavior!

- Geometric object: a surface!

State Space Reconstruction

Strange behavior!

- Geometric object: a surface!
- Strong sensitivity to starting conditions.

State Space Reconstruction

Strange behavior!

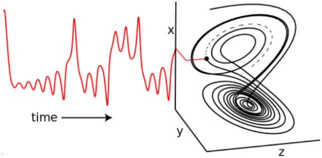
- Geometric object: a surface!
- Strong sensitivity to starting conditions.
- Dense trajectories.

State Space Reconstruction

Strange behavior!

- Geometric object: a surface!
- Strong sensitivity to starting conditions.
- Dense trajectories.

A resulting time series becomes:

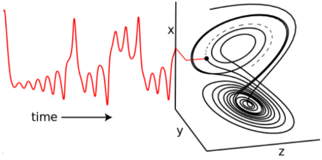


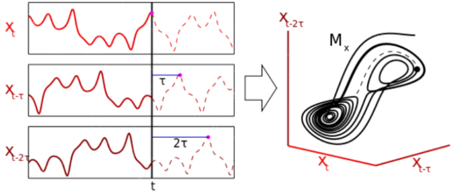
State Space Reconstruction

Strange behavior!

- Geometric object: a surface!
- Strong sensitivity to starting conditions.
- Dense trajectories.

A resulting time series becomes:



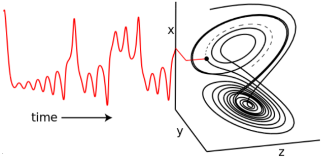


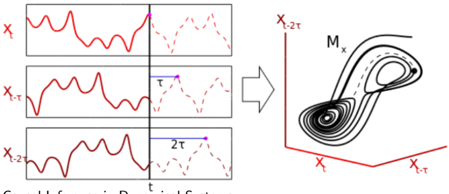
State Space Reconstruction

Strange behavior!

- Geometric object: a surface!
- Strong sensitivity to starting conditions.
- Dense trajectories.

A resulting time series becomes:





$$\frac{dX_t}{dt} = f_X(X_t, Z_t)$$
$$\frac{dY_t}{dt} = f_Y(Y_t, Z_t)$$
$$\frac{dZ_t}{dt} = f_Z(Z_t)$$

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \ldots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \ldots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

- if there is no causal link from Z or Y to X, then Takens' theorem does *not* hold!

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \dots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

- if there is no causal link from Z or Y to X, then Takens' theorem does *not* hold!
- if there is a causal link from from Z to X and Y to X, then we can *hope* Takens' theorem does hold.

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \dots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

- if there is no causal link from Z or Y to X, then Takens' theorem does *not* hold!
- if there is a causal link from from Z to X and Y to X, then we can *hope* Takens' theorem does hold.
- we can hope to recover a system consisting of *all* causes of a variable *in principle*.

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \ldots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

- if there is no causal link from Z or Y to X, then Takens' theorem does *not* hold!
- if there is a causal link from from Z to X and Y to X, then we can *hope* Takens' theorem does hold.
- we can hope to recover a system consisting of *all* causes of a variable *in principle*.

Key ingredients:

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \ldots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

- if there is no causal link from Z or Y to X, then Takens' theorem does *not* hold!
- if there is a causal link from from Z to X and Y to X, then we can *hope* Takens' theorem does hold.
- we can hope to recover a system consisting of *all* causes of a variable *in principle*.

Key ingredients:

- Causal model consisting of functional assignments in ODE's.

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \ldots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

- if there is no causal link from Z or Y to X, then Takens' theorem does *not* hold!
- if there is a causal link from from Z to X and Y to X, then we can *hope* Takens' theorem does hold.
- we can hope to recover a system consisting of *all* causes of a variable *in principle*.

Key ingredients:

- Causal model consisting of functional assignments in ODE's.
- Emergence of geometric structure from solutions.

Causal Inference

Takens' theorem:

$$(X_t, Y_t, Z_t) \mapsto (X_t, X_{t-\tau}, \dots, X_{t-4\tau})$$

is an *embedding* of the attractor *generically*. In particular,

- if there is no causal link from Z or Y to X, then Takens' theorem does *not* hold!
- if there is a causal link from from Z to X and Y to X, then we can *hope* Takens' theorem does hold.
- we can hope to recover a system consisting of *all* causes of a variable *in principle*.

Key ingredients:

- Causal model consisting of functional assignments in ODE's.
- Emergence of geometric structure from solutions.
- Decomposition of geometry according to causal structure.

Convergent Cross Mapping

In practice, we start for a fixed number of points by computing the state space reconstructions or *shadow manifolds* M_X, M_Y and M_Z.

Convergent Cross Mapping

In practice, we start for a fixed number of points by computing the state space reconstructions or *shadow manifolds* M_X, M_Y and M_Z.

To compute the cross-mapping skill from say X to Y, we proceed as follows:

Convergent Cross Mapping

In practice, we start for a fixed number of points by computing the state space reconstructions or *shadow manifolds* M_X, M_Y and M_Z.

To compute the cross-mapping skill from say X to Y, we proceed as follows:

1. At time t, we consider $x_t = (X_t, X_{t-\tau}, \ldots, X_{t-E\tau}) \in M_X$.

Convergent Cross Mapping

In practice, we start for a fixed number of points by computing the state space reconstructions or *shadow manifolds* M_X, M_Y and M_Z.

To compute the cross-mapping skill from say X to Y, we proceed as follows:

1. At time t, we consider $x_t = (X_t, X_{t-\tau}, \dots, X_{t-E\tau}) \in M_X$.
2. We locate the closest points on M_X of x_t, and denote their time stamps $t_1, \dots, t_{E+1}$.

Convergent Cross Mapping

In practice, we start for a fixed number of points by computing the state space reconstructions or *shadow manifolds* M_X, M_Y and M_Z.

To compute the cross-mapping skill from say X to Y, we proceed as follows:

1. At time t, we consider $x_t = (X_t, X_{t-\tau}, \ldots, X_{t-E\tau}) \in M_X$.
2. We locate the closest points on M_X of x_t, and denote their time stamps $t_1, \ldots, t_{E+1}$.
3. We compute the estimate

$$\hat{Y}_t | M_X = \sum_{i=1}^{E+1} \frac{u_i}{\sum_{j=1}^{E+1} u_j} Y_{t_i}$$

with $u_i = \exp\left(-\frac{\|x(t)-x(t_i)\|}{\|x(t)-x(t_1)\|}\right)$.

Convergent Cross Mapping

In practice, we start for a fixed number of points by computing the state space reconstructions or *shadow manifolds* M_X, M_Y and M_Z.

To compute the cross-mapping skill from say X to Y, we proceed as follows:

1. At time t, we consider $x_t = (X_t, X_{t-\tau}, \ldots, X_{t-E\tau}) \in M_X$.
2. We locate the closest points on M_X of x_t, and denote their time stamps $t_1, \ldots, t_{E+1}$.
3. We compute the estimate

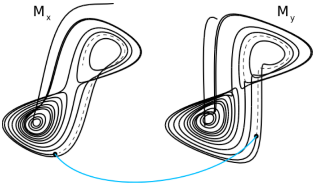
$$\hat{Y}_t | M_X = \sum_{i=1}^{E+1} \frac{u_i}{\sum_{i=1}^{E+1} u_j} Y_{t_i}$$

with $u_i = \exp\left(-\frac{\|x(t)-x(t_i)\|}{\|x(t)-x(t_1)\|}\right)$.

We compute the cross-mapping skill as the correlation between $\hat{Y}_t | M_X$ and Y_t.

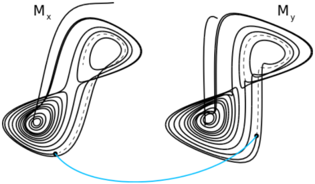
Convergent Cross Mapping pt. 2

Visualization:

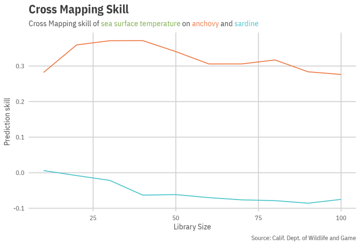


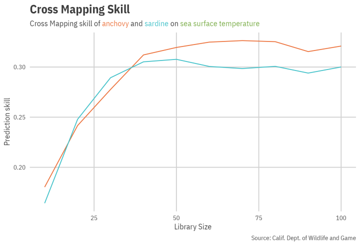
Convergent Cross Mapping pt. 2

Visualization:



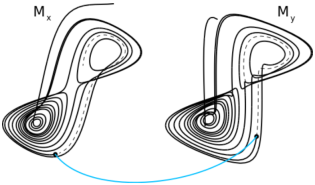
In practice (no transformation!) with $\tau = 1$, $E = 4$:



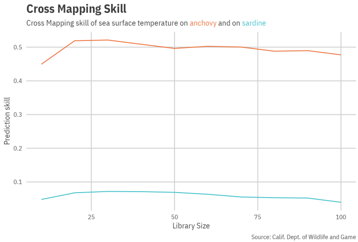


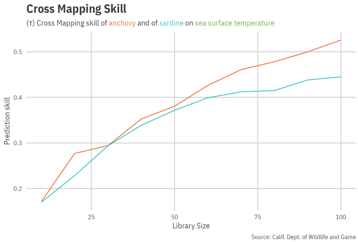
Convergent Cross Mapping pt. 2

Visualization:



In practice (no transformation!) with $\tau = 2$, $E = 4$:





Comparison of Models

In both settings, we share the modelling approach:

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Advantages of convergent cross mapping:

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Advantages of convergent cross mapping:

- We do not need to observe all variables of interest.

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Advantages of convergent cross mapping:

- We do not need to observe all variables of interest.
- We do not need to assume much about the functional expressions.*

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Advantages of convergent cross mapping:

- We do not need to observe all variables of interest.
- We do not need to assume much about the functional expressions.*

Disadvantages of convergent cross mapping:

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Advantages of convergent cross mapping:

- We do not need to observe all variables of interest.
- We do not need to assume much about the functional expressions.*

Disadvantages of convergent cross mapping:

- No model formulated that explicitly satisfies all technical assumptions.

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Advantages of convergent cross mapping:

- We do not need to observe all variables of interest.
- We do not need to assume much about the functional expressions.*

Disadvantages of convergent cross mapping:

- No model formulated that explicitly satisfies all technical assumptions.
- Problems in face of synchronicity and noise.

Comparison of Models

In both settings, we share the modelling approach:

- Functional assignments based on graph structure.
- Modularity assumption on interventions.

However, there are two properties that distinguish these models:

- Non-separability of information.
- Deterministic non-linear dynamics.

Advantages of convergent cross mapping:

- We do not need to observe all variables of interest.
- We do not need to assume much about the functional expressions.*

Disadvantages of convergent cross mapping:

- No model formulated that explicitly satisfies all technical assumptions.
- Problems in face of synchronicity and noise.
- Regime dependence.

Future Work

The following ideas was considered, but not fully developed:

Future Work

The following ideas was considered, but not fully developed:

1. Give a full description of a model satisfying the theoretical guarantees.

Future Work

The following ideas was considered, but not fully developed:

1. Give a full description of a model satisfying the theoretical guarantees.
2. We can imagine that we observe

$$X_t + \varepsilon_t^X$$

with some observation noise X_t.

Future Work

The following ideas was considered, but not fully developed:

1. Give a full description of a model satisfying the theoretical guarantees.
2. We can imagine that we observe

$$X_t + \varepsilon_t^X$$

with some observation noise X_t.

3. We can imagine

$$\frac{dX_t}{dt} = f(X_t, X_{t-\tau}, Y_t, \ldots)$$

Future Work

The following ideas was considered, but not fully developed:

1. Give a full description of a model satisfying the theoretical guarantees.
2. We can imagine that we observe

$$X_t + \varepsilon_t^X$$

   with some observation noise X_t.
3. We can imagine

$$\frac{dX_t}{dt} = f(X_t, X_{t-\tau}, Y_t, \dots)$$

4. Alternative, we can imagine an SDE

$$dX_t = f(X_{t-}, Y_{t-}, \dots) dZ_t$$

Future Work

The following ideas was considered, but not fully developed:

1. Give a full description of a model satisfying the theoretical guarantees.
2. We can imagine that we observe
$$X_t + \varepsilon_t^X$$
with some observation noise X_t.
3. We can imagine
$$\frac{dX_t}{dt} = f(X_t, X_{t-\tau}, Y_t, \ldots)$$
4. Alternative, we can imagine an SDE
$$dX_t = f(X_{t-}, Y_{t-}, \ldots)dZ_t$$
5. The link between limiting distributions and attractor sets.

Questions