European Research Council
Established by the European Commission

DLR
Deutsches Zentrum
für Luft- und Raumfahrt
German Aerospace Center

Causal inference and causal discovery with perspectives in Earth sciences

---

Jakob Runge
Lecture developed together with Andreas Gerhardus

December 4, 2020

Knowledge for Tomorrow

DLR Institute of Data Science

DLR

Table of contents

Motivation

Statistical Dependence vs Causation

Consider the following setup:

- You have a lawn with a sprinkler.
- If the sprinkler is on, the lawn will become wet.
- Each day you toss a coin and turn the sprinkler on if it shows heads.

Statistical Dependence vs Causation

Consider the following setup:

- You have a lawn with a sprinkler.
- If the sprinkler is on, the lawn will become wet.
- Each day you toss a coin and turn the sprinkler on if it shows heads.

Two easy questions:

- Does turning on the sprinkler cause the lawn to be wet?
- Does making the lawn wet turn the sprinkler on?

Statistical Dependence vs Causation

Consider the following setup:

- You have a lawn with a sprinkler.
- If the sprinkler is on, the lawn will become wet.
- Each day you toss a coin and turn the sprinkler on if it shows heads.

Two easy questions:

- Does turning on the sprinkler cause the lawn to be wet? **Yes!**
- Does making the lawn wet turn the sprinkler on? **No!**

Statistical Dependence vs Causation

How does this relate to statistical dependence?

We can model the setup with two random variables:

$$X: \quad X = 1 \text{ if the sprinkler is on, else } X = 0$$

$$Y: \quad Y = 1 \text{ if the lawn is wet, else } Y = 0$$

Statistical Dependence vs Causation

How does this relate to statistical dependence?

We can model the setup with two random variables:

$$X: \quad X = 1 \text{ if the sprinkler is on, else } X = 0$$
$$Y: \quad Y = 1 \text{ if the lawn is wet, else } Y = 0$$

Since turning the sprinkler on causes the lawn to be wet, we get

$$\begin{aligned}\mathbb{P}(\text{lawn wet}|\text{sprinkler on}) &= \mathbb{P}(Y = 1|X = 1) \\ &> \mathbb{P}(Y = 1) = \mathbb{P}(\text{lawn wet}) \ .\end{aligned}$$

This shows that X **and** Y **are dependent random variables**.

Statistical Dependence vs Causation

How does this relate to statistical dependence?

We can model the setup with two random variables:

$$X: \quad X = 1 \text{ if the sprinkler is on, else } X = 0$$
$$Y: \quad Y = 1 \text{ if the lawn is wet, else } Y = 0$$

Since turning the sprinkler on causes the lawn to be wet, we get

$$\begin{aligned}\mathbb{P}(\text{lawn wet}|\text{sprinkler on}) &= \mathbb{P}(Y = 1|X = 1)\\ &> \mathbb{P}(Y = 1) = \mathbb{P}(\text{lawn wet})\ .\end{aligned}$$

This shows that X **and** Y **are dependent random variables**.

However, the dependence also goes in the other direction. Observing the lawn to be wet increases the probability that the sprinkler is on:

$$\begin{aligned}\mathbb{P}(\text{sprinkler on}|\text{lawn wet}) &= \mathbb{P}(X = 1|Y = 1)\\ &> \mathbb{P}(X = 1) = \mathbb{P}(\text{sprinkler on})\ .\end{aligned}$$

Statistical Dependence vs Causation

How does this relate to statistical dependence?

We can model the setup with two random variables:

$$X: \quad X = 1 \text{ if the sprinkler is on, else } X = 0$$
$$Y: \quad Y = 1 \text{ if the lawn is wet, else } Y = 0$$

Since turning the sprinkler on causes the lawn to be wet, we get

$$\begin{aligned}\mathbb{P}(\text{lawn wet}|\text{sprinkler on}) &= \mathbb{P}(Y = 1|X = 1)\\ &> \mathbb{P}(Y = 1) = \mathbb{P}(\text{lawn wet})\ .\end{aligned}$$

This shows that X **and** Y **are dependent random variables**.

However, the dependence also goes in the other direction. Observing the lawn to be wet increases the probability that the sprinkler is on:

$$\begin{aligned}\mathbb{P}(\text{sprinkler on}|\text{lawn wet}) &= \mathbb{P}(X = 1|Y = 1)\\ &> \mathbb{P}(X = 1) = \mathbb{P}(\text{sprinkler on})\ .\end{aligned}$$

The probabilities alone do not allow us to decide wheter X causes Y or Y causes X

Conceptual Findings

- Statistical dependence is symmetric between cause and effect, i.e., it does not distinguish between cause and effect
- Causation is asymmetric between cause and effect, i.e., it does distinguish between cause and effect
- Statistical dependence is not causation.

Observational Causal Inference

Causal Inference

What causal inference is about:

- Provide mathematical language for causal notions such as cause and effect
- Provide methods and assumptions for answering causal questions from observational data
- Provide methods and assumptions for learning parts of the causal relationships from observational data

Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

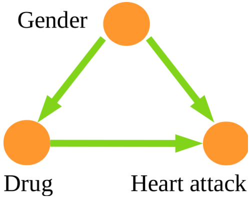
Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

Randomized control trials:

Randomly assign the subjects to the treatment and control group.



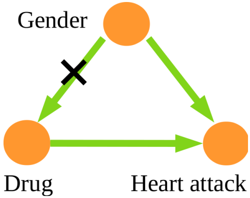
Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

Randomized control trials:

Randomly assign the subjects to the treatment and control group.



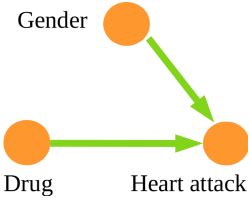
Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

Randomized control trials:

Randomly assign the subjects to the treatment and control group.



The gender of a subject does not longer influence whether a subject belongs to the treatment or control group.

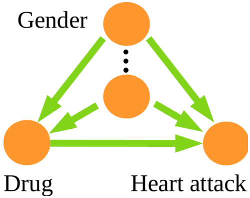
Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

Randomized control trials:

Randomly assign the subjects to the treatment and control group.



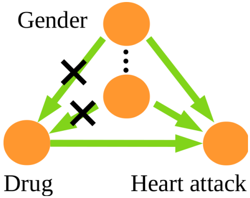
Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

Randomized control trials:

Randomly assign the subjects to the treatment and control group.



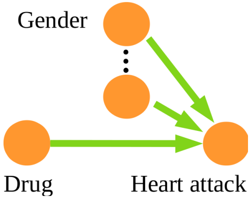
Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

Randomized control trials:

Randomly assign the subjects to the treatment and control group.



In fact, all confounders are eliminated!

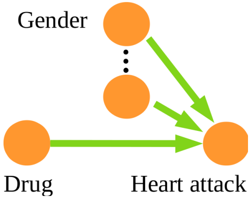
Randomized Control Trials

Question:

How can one, in general, determine whether a certain drug is helpful in reducing the risk of developing a certain desease?

Randomized control trials:

Randomly assign the subjects to the treatment and control group.



In fact, all confounders are eliminated! Therefore, in RCT data causal effects can be assessed by statistical proportions.

Experimentation

Abstraction of the question:

Consider two events or variables X and Y. How can we determine whether X causes Y?

Experimentation:

Experimentally manipulate X while keeping all other conditions exactly the same. If this results in a change of Y, X causes Y.

Such an idealization of experimental manipulation is referred to as an *intervention on* X.

Experimentation

Abstraction of the question:

Consider two events or variables X and Y. How can we determine whether X causes Y?

Experimentation:

Experimentally manipulate X while keeping all other conditions exactly the same. If this results in a change of Y, X causes Y.

Such an idealization of experimental manipulation is referred to as an *intervention on* X.

Example: Lawn with sprinkler

- Turning the sprinkler on makes the lawn wet.
- Making the lawn wet does not turn on the sprinkler.

Experimentation

Abstraction of the question:

Consider two events or variables X and Y. How can we determine whether X causes Y?

Experimentation:

Experimentally manipulate X while keeping all other conditions exactly the same. If this results in a change of Y, X causes Y.

Such an idealization of experimental manipulation is referred to as an *intervention on* X.

Example: Lawn with sprinkler

- Turning the sprinkler on makes the lawn wet.
- Making the lawn wet does not turn on the sprinkler.

We need to imagine the lawn being made wet without changing something else, in particular without turning the sprinkler on

Challenges in Context of Climate Sciences

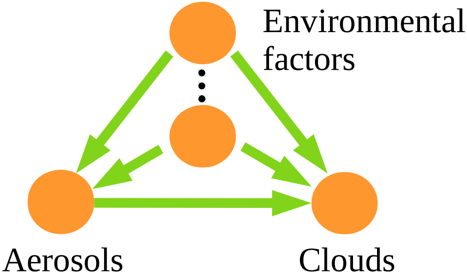
Challenges in Context of Climate Sciences

When trying to apply these concepts in the context of climate sciences, we face two fundamental challenges.

Challenges in Context of Climate Sciences

When trying to apply these concepts in the context of climate sciences, we face two fundamental challenges.

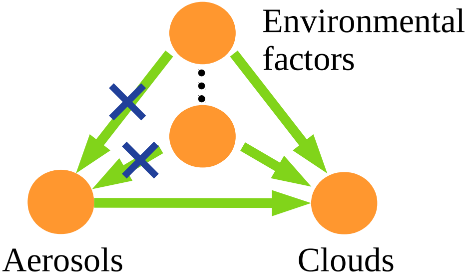
Challenge: Experimentation in real climate not possible
For most parts it is impossible, and else arguably ethically questionable, to deliberately intervene on climate variables.



Challenges in Context of Climate Sciences

When trying to apply these concepts in the context of climate sciences, we face two fundamental challenges.

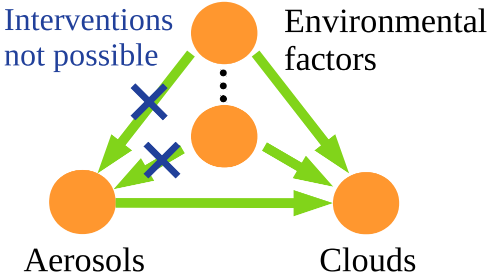
Challenge: Experimentation in real climate not possible
For most parts it is impossible, and else arguably ethically questionable, to deliberately intervene on climate variables.



Challenges in Context of Climate Sciences

When trying to apply these concepts in the context of climate sciences, we face two fundamental challenges.

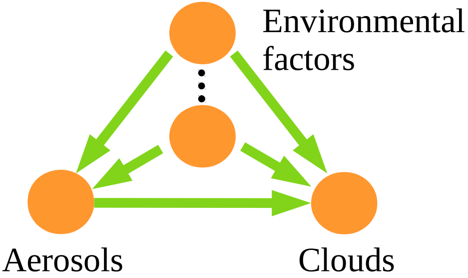
Challenge: Experimentation in real climate not possible
For most parts it is impossible, and else arguably ethically questionable, to deliberately intervene on climate variables.



Challenges in Context of Climate Sciences

When trying to apply these concepts in the context of climate sciences, we face two fundamental challenges.

Challenge: Experimentation in real climate not possible
For most parts it is impossible, and else arguably ethically questionable, to deliberately intervene on climate variables.



Challenges in Context of Climate Sciences

When trying to apply these concepts in the context of climate sciences, we face two fundamental challenges.

Challenge: Experimentation in real climate not possible
For most parts it is impossible, and else arguably ethically questionable, to deliberately intervene on climate variables.

Challenge: Ground truth not known
Causal relationships are mostly known among microscopic physical variables, but usually less so among macroscopic variables.

Challenges in Context of Climate Sciences

When trying to apply these concepts in the context of climate sciences, we face two fundamental challenges.

Challenge: Experimentation in real climate not possible
For most parts it is impossible, and else arguably ethically questionable, to deliberately intervene on climate variables.

Challenge: Ground truth not known
Causal relationships are mostly known among microscopic physical variables, but usually less so among macroscopic variables.

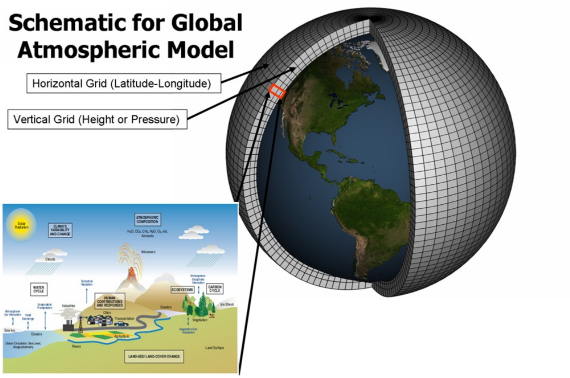
For example:

- Is there a causal relationship between arctic sea ice extent in winter and mid latitude weather in summer?
- Is there a causal relationship between ENSO and extreme weather events somewhere on Earth?

Two Approaches to Address These Challenges

Simulation:

Experimentation in climate models: Fix (intervene on) certain parameters and analyze the resulting effect on other parameters.



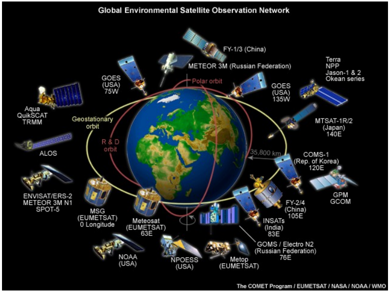
Two Approaches to Address These Challenges

Simulation:

Experimentation in climate models: Fix (intervene on) certain parameters and analyze the resulting effect on other parameters.

Observational causal inference:

Answer causal questions, for example questions about cause and effect, from observation data.



Two Approaches to Address These Challenges

Simulation:

Experimentation in climate models: Fix (intervene on) certain parameters and analyze the resulting effect on other parameters.

Observational causal inference:

Answer causal questions, for example questions about cause and effect, from observation data.

Seminal work by Judea Pearl, Peter Spirtes, Clark Glymour, Richard Scheines.




Two Approaches to Address These Challenges

Simulation:
Experimentation in climate models: Fix (intervene on) certain parameters and analyze the resulting effect on other parameters.

Observational causal inference:
Answer causal questions, for example questions about cause and effect, from observation data.

To mathematically treat causal inference, we need to formalize the idea of an underlying **structural causal model (SCM)**, *define hypothetical* **interventions**, *and establish* **identifiability** *criteria to decide if and how causal effects can be estimated from data alone (without interventions).*

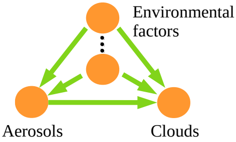
Structural Causal Models

Intuition:

A structural causal model (SCM) specifies the functional causal relationships between a set of random variables.

Example:

$$
\begin{aligned}
X_{\text{aerosols}} &:= f_{\text{aerosols}}(X_{\text{env. facs.}}, \eta_{\text{aerosols}}) \\
X_{\text{clouds}} &:= f_{\text{clouds}}(X_{\text{aerosols}}, X_{\text{env. facs.}}, \eta_{\text{clouds}}) \\
X_{\text{env. facs.}} &:= f_{\text{env. facs.}}(\eta_{\text{env. facs.}})
\end{aligned}
$$

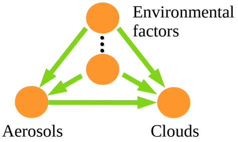


Structural Causal Models

Intuition:
A structural causal model (SCM) specifies the functional causal relationships between a set of random variables.

Example:

$$
\begin{aligned}
X_{\text{aerosols}} &:= f_{\text{aerosols}}(X_{\text{env. facs.}}, \eta_{\text{aerosols}}) \\
X_{\text{clouds}} &:= f_{\text{clouds}}(X_{\text{aerosols}}, X_{\text{env. facs.}}, \eta_{\text{clouds}}) \\
X_{\text{env. facs.}} &:= f_{\text{env. facs.}}(\eta_{\text{env. facs.}})
\end{aligned}
$$



Claims of an SCM:

- Specifies the *direct causes*, also called *parents*, of each variable
- The functions f_i are *independent mechanisms* by which nature determines the values of each variable based on the values of its direct causes plus random fluctuations

Structural Causal Models

Definition:

A structural causal model over the set of random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ consists auf n structural assignments

$$
\begin{aligned}
X_1 &:= f_1(PA_1, \eta_1) \\
&\vdots \\
X_n &:= f_n(PA_n, \eta_n)
\end{aligned}
$$

together with a specification of the probability distributions of the jointly independent 'noise' variables η_i. Here $PA_i \subseteq \mathbf{X} \setminus \{X_i\}$ are the direct causes of X_i.

Here we also assume that there are no cyclic dependencies.

Structural Causal Models

Definition:

A structural causal model over the set of random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ consists auf n structural assignments

$$\begin{aligned} X_1 &:= f_1(PA_1, \eta_1) \\ &\vdots \\ X_n &:= f_n(PA_n, \eta_n) \end{aligned}$$

together with a specification of the probability distributions of the jointly independent 'noise' variables η_i. Here $PA_i \subseteq \mathbf{X} \setminus \{X_i\}$ are the direct causes of X_i.

Here we also assume that there are no cyclic dependencies.

Interpretation of noise variables:

- Summarize all background factors outside causal model, i.e., all factors apart from $X_1, \ldots, X_n$
- Their joint independence means the model is sufficient to describe the causal relationship among the variables $X_1, \ldots, X_n$

Causal Graphs

Definition:

Consider an SCM over the variables $\mathbf{X} = \{X_1, \ldots, X_n\}$. Its causal graph $\mathcal{G}$ is the directed graph with

1. a vertex (node) for each variable X_i
2. for all i a directed edge from each variable $X_j \in PA_i$ to X_i

Causal Graphs

Definition:

Consider an SCM over the variables $\mathbf{X} = \{X_1, \ldots, X_n\}$. Its causal graph $\mathcal{G}$ is the directed graph with

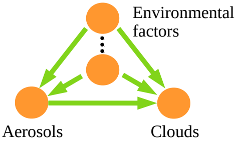
1. a vertex (node) for each variable X_i
2. for all i a directed edge from each variable $X_j \in PA_i$ to X_i

Example:

$$X_{\text{aerosols}} := f_{\text{aerosols}}(X_{\text{env. facs.}}, \eta_{\text{aerosols}})$$

$$X_{\text{clouds}} := f_{\text{clouds}}(X_{\text{aerosols}}, X_{\text{env. facs.}}, \eta_{\text{clouds}})$$

$$X_{\text{env. facs.}} := f_{\text{env. facs.}}(\eta_{\text{env. facs.}})$$



Causal Graphs

Definition:

Consider an SCM over the variables $\mathbf{X} = \{X_1, \ldots, X_n\}$. Its causal graph $\mathcal{G}$ is the directed graph with

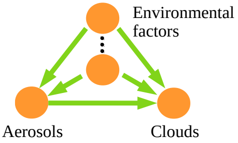
1. a vertex (node) for each variable X_i
2. for all i a directed edge from each variable $X_j \in PA_i$ to X_i

Example:

$$X_{\text{aerosols}} := f_{\text{aerosols}}(X_{\text{env. facs.}}, \eta_{\text{aerosols}})$$

$$X_{\text{clouds}} := f_{\text{clouds}}(X_{\text{aerosols}}, X_{\text{env. facs.}}, \eta_{\text{clouds}})$$

$$X_{\text{env. facs.}} := f_{\text{env. facs.}}(\eta_{\text{env. facs.}})$$



Interpretation:

- Nodes represent variables
- Edges represent causal influences

Interventions

Definition:

Consider SCM over the variables $\mathbf{X} = \{X_1, \dots, X_n\}$.

An *intervention on* X_i, denoted as $do(X_i := x_{i,0})$, defines a modified SCM obtained by replacing the old assignment of X_i with $X_i := x_{i,0}$ where $x_{i,0}$ is some value in the range of X_i:

$$
\begin{aligned}
X_1 &:= f_1(PA_1, \eta_1) \\
&\vdots \\
X_i &:= x_{i,0} \\
&\vdots \\
X_n &:= f_n(PA_n, \eta_n)
\end{aligned}
$$

Interventions

Definition:

Consider SCM over the variables $\mathbf{X} = \{X_1, \ldots, X_n\}$.

An *intervention on* X_i, denoted as $do(X_i := x_{i,0})$, defines a modified SCM obtained by replacing the old assignment of X_i with $X_i := x_{i,0}$ where $x_{i,0}$ is some value in the range of X_i:

$$
\begin{aligned}
X_1 &:= f_1(PA_1, \eta_1) \\
&\vdots \\
X_i &:= x_{i,0} \\
&\vdots \\
X_n &:= f_n(PA_n, \eta_n)
\end{aligned}
$$

Interpretation:

- An intervention $do(X_i := x_{i,0})$ inactivates the natural mechanism $X_i := f_i(PA_i, \eta_i)$ by forcing X_i to take the value $x_{i,0}$
- All other mechanisms remain unmodified
- Interventions can also be more complex functions than just $x_{i,0}$

Interventions

Definition:

Consider SCM over the variables $\mathbf{X} = \{X_1, \dots, X_n\}$.

An *intervention on* X_i, denoted as $do(X_i := x_{i,0})$, defines a modified SCM obtained by replacing the old assignment of X_i with $X_i := x_{i,0}$ where $x_{i,0}$ is some value in the range of X_i:

$$
\begin{aligned}
X_1 &:= f_1(PA_1, \eta_1) \\
&\vdots \\
X_i &:= x_{i,0} \\
&\vdots \\
X_n &:= f_n(PA_n, \eta_n)
\end{aligned}
$$

Interventions model the notion of 'experimentally manipulate X_i while keeping all other conditions exactly the same'

Interventions

Interventions model the notion of ‘experimentally manipulate X_i while keeping all other conditions exactly the same’

Effect on causal graphs:

The causal graph of the intervened SCM is obtained from the causal graph of the original SCM by deleting all arrows pointing into X_i.

Interventions

Interventions model the notion of ‘experimentally manipulate X_i while keeping all other conditions exactly the same’

Effect on causal graphs:

The causal graph of the intervened SCM is obtained from the causal graph of the original SCM by deleting all arrows pointing into X_i.

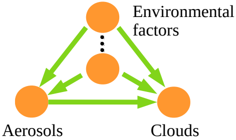
Example:

Original SCM prior to intervention:

$$X_{\text{aerosols}} := f_{\text{aerosols}}(X_{\text{env. facs.}}, \eta_{\text{aerosols}})$$

$$X_{\text{clouds}} := f_{\text{clouds}}(X_{\text{aerosols}}, X_{\text{env. facs.}}, \eta_{\text{clouds}})$$

$$X_{\text{env. facs.}} := f_{\text{env. facs.}}(\eta_{\text{env. facs.}})$$



Interventions

Interventions model the notion of 'experimentally manipulate X_i while keeping all other conditions exactly the same'

Effect on causal graphs:

The causal graph of the intervened SCM is obtained from the causal graph of the original SCM by deleting all arrows pointing into X_i.

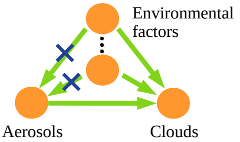
Example:

After intervention $do(X_{\text{aerosols}}) :=$ a certain aerosol concentration:

$$X_{\text{aerosols}} := \textit{a certain aerosol concentration}$$

$$X_{\text{clouds}} := f_{\text{clouds}}(X_{\text{aerosols}}, X_{\text{env. facs.}}, \eta_{\text{clouds}})$$

$$X_{\text{env. facs.}} := f_{\text{env. facs.}}(\eta_{\text{env. facs.}})$$



Identification of Causal Effects

General formalism:

- Formally define the notion of a *causal effect* where generally $\mathbb{P}(Y|do(X := x)) \quad \neq \quad \mathbb{P}(Y|X = x)$.
- Based on the causal graph, decide whether the causal effect of X on Y can be identified from the observational data.
- If the effect can be identified, express causal effect in terms of observational data

Identification of Causal Effects

General formalism:

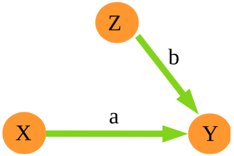
- Formally define the notion of a *causal effect* where generally $\mathbb{P}(Y|\textcolor{red}{do}(X := x)) \quad \neq \quad \mathbb{P}(Y|X = x)$.
- Based on the causal graph, decide whether the causal effect of X on Y can be identified from the observational data.
- If the effect can be identified, express causal effect in terms of observational data

Here: Convey idea through examples.

Example 1: No confounding

Consider the linear SCM:

$$
\begin{aligned}
Z &:= \eta_Z\ , && \eta_Z \sim \mathcal{N}(0, \sigma_Z^2)\\
X &:= \eta_X\ , && \eta_X \sim \mathcal{N}(0, \sigma_X^2)\\
Y &:= a \cdot X + b \cdot Z + \eta_Y\ , && \eta_Y \sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$

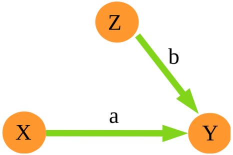


Identification of Causal Effects

Example 1: No confounding

Consider the linear SCM:

$$
\begin{aligned}
Z &:= \eta_Z \,, & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
X &:= \eta_X \,, & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Y &:= a \cdot X + b \cdot Z + \eta_Y \,, & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$

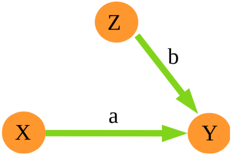


Goal: Determine a, the causal effect of X on Y

Example 1: No confounding

Consider the linear SCM:

$$
\begin{aligned}
Z &:= \eta_Z \ , & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
X &:= \eta_X \ , & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Y &:= a \cdot X + b \cdot Z + \eta_Y \ , & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$



<u>Goal:</u> Determine a, the causal effect of X on Y

<u>Idea:</u> Use covariance of Y on X

$$
\begin{aligned}
Cov(Y, X) &= Cov\big(a \cdot \eta_X + b \cdot \eta_Z + \eta_Y, \eta_X, \big) \\
&= a \cdot Cov(\eta_X, \eta_X) + b \cdot Cov(\eta_Z, \eta_X) + Cov(\eta_Y, \eta_Z) \\
&= a \cdot Var(X) + b \cdot 0 + 0
\end{aligned}
$$

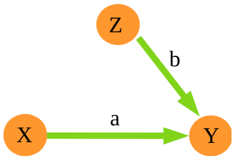
$$
\Rightarrow a = \frac{Cov(Y, X)}{Var(X)} = \rho(Y, X) \cdot \sqrt{\frac{Var(Y)}{Var(X)}}
$$

Identification of Causal Effects

Example 1: No confounding

Consider the linear SCM:

$$
\begin{aligned}
Z &:= \eta_Z \ , & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
X &:= \eta_X \ , & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Y &:= a \cdot X + b \cdot Z + \eta_Y \ , & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$



Goal: Determine a, the causal effect of X on Y

Idea: Use covariance of Y on X

$$
\begin{aligned}
Cov(Y, X) &= Cov(a \cdot \eta_X + b \cdot \eta_Z + \eta_Y, \eta_X,) \\
&= a \cdot Cov(\eta_X, \eta_X) + b \cdot Cov(\eta_Z, \eta_X) + Cov(\eta_Y, \eta_Z) \\
&= a \cdot Var(X) + b \cdot 0 + 0
\end{aligned}
$$

$$\Rightarrow a = \frac{Cov(Y, X)}{Var(X)} = \rho(Y, X) \cdot \sqrt{\frac{Var(Y)}{Var(X)}}$$

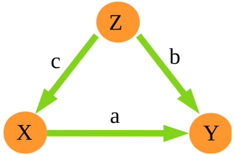
a is the coefficient of X in the linear regression of Y on X

Identification of Causal Effects

Example 2: Observed confounding

Consider the linear SCM:

$$
\begin{aligned}
Z &:= \eta_Z \ , & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
X &:= {\color{red} c \cdot Z} + \eta_X \ , & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Y &:= a \cdot X + b \cdot Z + \eta_Y \ , & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$

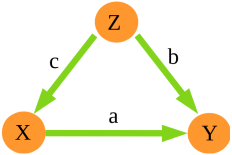


Identification of Causal Effects

Example 2: Observed confounding

Consider the linear SCM:

$$
\begin{aligned}
Z &:= \eta_Z \,, & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
X &:= c \cdot Z + \eta_X \,, & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Y &:= a \cdot X + b \cdot Z + \eta_Y \,, & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$



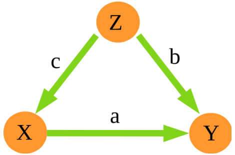
Goal: Determine a, the causal effect of X on Y

Identification of Causal Effects

Example 2: Observed confounding

Consider the linear SCM:

$$
\begin{aligned}
Z &:= \eta_Z \ , & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
X &:= {\color{red} c \cdot Z} + \eta_X \ , & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Y &:= a \cdot X + b \cdot Z + \eta_Y \ , & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$



Goal: Determine a, the causal effect of X on Y

First attempt: Linearly regress of Y on X

$$
\begin{aligned}
Cov(Y, X) &= Cov\big((a \cdot c + b) \cdot \eta_Z + a \cdot \eta_X + \eta_Y, c \cdot \eta_Z + \eta_X\big) \\
&= c \cdot (a \cdot c + b) \cdot Cov(\eta_Z, \eta_Z) + a \cdot Cov(\eta_X, \eta_X)
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= Cov(c \cdot \eta_Z + \eta_X, c \cdot \eta_Z + \eta_X) \\
&= c^2 \cdot Cov(\eta_Z, \eta_Z) + Cov(\eta_X, \eta_X)
\end{aligned}
$$

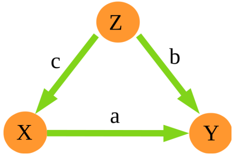
$$\Rightarrow a \neq \frac{Cov(Y, X)}{Var(X)}$$

Identification of Causal Effects

Example 2: Observed confounding

Second attempt:

- Control for Z: Regress out influence of Z on X and Y to obtain residuals ΔX and ΔY
- Linearly regress ΔY on ΔX

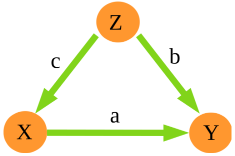


Identification of Causal Effects

Example 2: Observed confounding

Second attempt:

- Control for Z: Regress out influence of Z on X and Y to obtain residuals ΔX and ΔY
- Linearly regress ΔY on ΔX



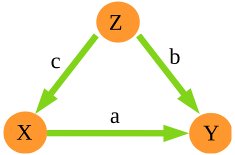
$$\begin{aligned}
\Delta X &= X - \frac{Cov(X,Z)}{Var(Z)} \cdot Z \\
&= X - \frac{c \cdot Cov(\eta_Z, \eta_Z)}{Cov(\eta_Z, \eta_Z)} \cdot Z \\
&= X - c \cdot Z \\
&= \eta_X
\end{aligned}$$

Identification of Causal Effects

Example 2: Observed confounding

Second attempt:

- Control for Z: Regress out influence of Z on X and Y to obtain residuals ΔX and ΔY
- Linearly regress ΔY on ΔX



$$\begin{aligned}
\Delta X &= X - \frac{Cov(X,Z)}{Var(Z)} \cdot Z \\
&= X - \frac{c \cdot Cov(\eta_Z, \eta_Z)}{Cov(\eta_Z, \eta_Z)} \cdot Z \\
&= X - c \cdot Z \\
&= \eta_X
\end{aligned}$$

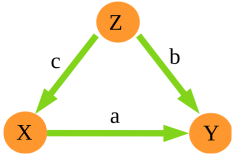
$$\begin{aligned}
\Delta Y &= Y - \frac{Cov(Y,Z)}{Var(Z)} \cdot Z \\
&= Y - \frac{(a \cdot c + b) Cov(\eta_Z, \eta_Z)}{Cov(\eta_Z, \eta_Z)} \cdot Z \\
&= X - (a \cdot c + b) \cdot Z \\
&= a \cdot \eta_X + \eta_Y
\end{aligned}$$

Identification of Causal Effects

Example 2: Observed confounding

Second attempt:

- Control for Z: Regress out influence of Z on X and Y to obtain residuals ΔX and ΔY
- Linearly regress ΔY on ΔX



$$\begin{aligned}
\Delta X &= X - \frac{Cov(X,Z)}{Var(Z)} \cdot Z \\
&= X - \frac{c \cdot Cov(\eta_Z, \eta_Z)}{Cov(\eta_Z, \eta_Z)} \cdot Z \\
&= X - c \cdot Z \\
&= \eta_X
\end{aligned}$$

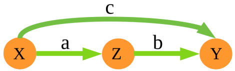
$$\begin{aligned}
\Delta Y &= Y - \frac{Cov(Y,Z)}{Var(Z)} \cdot Z \\
&= Y - \frac{(a \cdot c + b) Cov(\eta_Z, \eta_Z)}{Cov(\eta_Z, \eta_Z)} \cdot Z \\
&= X - (a \cdot c + b) \cdot Z \\
&= a \cdot \eta_X + \eta_Y
\end{aligned}$$

$$\Rightarrow a = \frac{Cov(\Delta Y, \Delta X)}{Var(\Delta X, \Delta X)} = \frac{a \cdot Cov(\eta_X, \eta_X)}{Cov(\eta_X, \eta_X)} = a$$

Example 3: Partial mediation

Consider the linear SCM:

$$
\begin{aligned}
X &:= \eta_X \,, & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Z &:= a \cdot X + \eta_Z \,, & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
Y &:= b \cdot Z + c \cdot X + \eta_Y \,, & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$

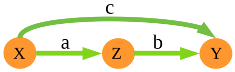


Identification of Causal Effects

Example 3: Partial mediation

Consider the linear SCM:

$$
\begin{aligned}
X &:= \eta_X \ , & \eta_X &\sim \mathcal{N}(0, \sigma_X^2) \\
Z &:= a \cdot X + \eta_Z \ , & \eta_Z &\sim \mathcal{N}(0, \sigma_Z^2) \\
Y &:= b \cdot Z + c \cdot X + \eta_Y \ , & \eta_Y &\sim \mathcal{N}(0, \sigma_Y^2)
\end{aligned}
$$



Total causal effect of X on Y: $(a \cdot b + c)$

- Linearly regress Y on X

Direct causal effect of X on Y: c

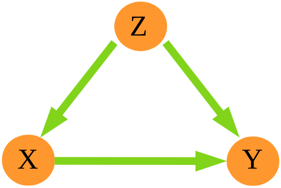
- *Multivariate regression* of Y on its parents (X, Z), then take coefficient belonging to X

Conceptual Findings

- Interventions model the notion of ‘experimentally manipulate X_i while keeping all other conditions exactly the same’.
- The causal graph allows us to decide whether a certain causal effect is identifiable.
- If the causal effect is identifiable, the causal graph further allows us to estimate it from observational data.

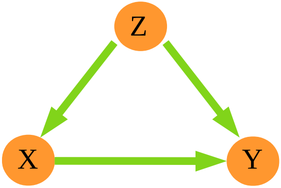
Overview: Identification of Causal Effect

Example: Observed confounder



Overview: Identification of Causal Effect

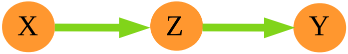
Example: Observed confounder



Control for confounder Z

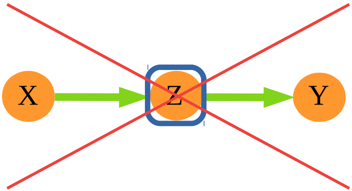
Overview: Identification of Causal Effect

Example: Mediation



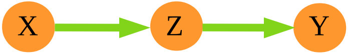
Overview: Identification of Causal Effect

Example: Mediation



Overview: Identification of Causal Effect

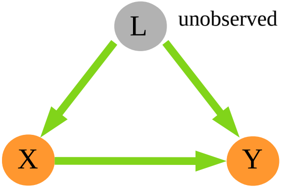
Example: Mediation



Do not control for mediator Z

Overview: Identification of Causal Effect

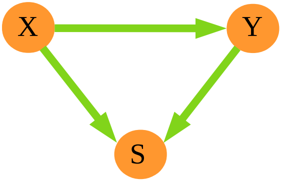
Example: Unobserved confounder



Causal effect cannot be identified

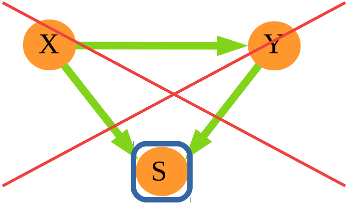
Overview: Identification of Causal Effect

Example: Common effect



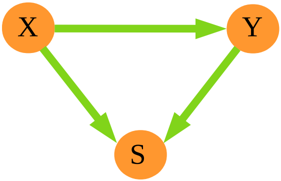
Overview: Identification of Causal Effect

Example: Common effect



Overview: Identification of Causal Effect

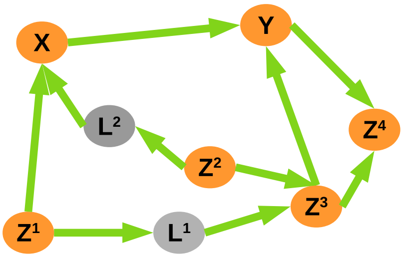
Example: Common effect



Do not control for common effects of X and Y, this would introduce bias.

However, sometimes the available dataset is already biased, e.g., if samples are missing (selection bias).

Example: Complicated case



Back-door criterion:

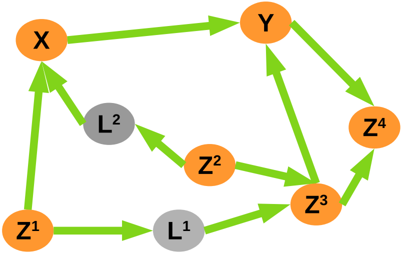
Block all non-causal dependence by controlling for an appropriate set of variables

Back-door adjustment

Back-door criterion

A set $\mathbf{Z} \subset \mathbf{X} \setminus \{X, Y\}$ satisfies the back-door criterion for identifying the causal effect of X on Y if

1. $\mathbf{Z}$ blocks all paths between X and Y with arrow into X

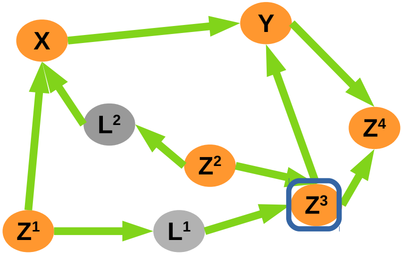


Back-door adjustment

Back-door criterion

A set $\mathbf{Z} \subset \mathbf{X} \setminus \{X, Y\}$ satisfies the back-door criterion for identifying the causal effect of X on Y if

1. $\mathbf{Z}$ blocks all paths between X and Y with arrow into X

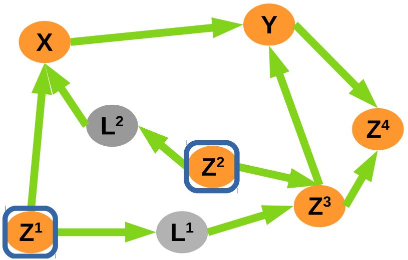


Back-door adjustment

Back-door criterion

A set $\mathbf{Z} \subset \mathbf{X} \setminus \{X, Y\}$ satisfies the back-door criterion for identifying the causal effect of X on Y if

1. $\mathbf{Z}$ blocks all paths between X and Y with arrow into X

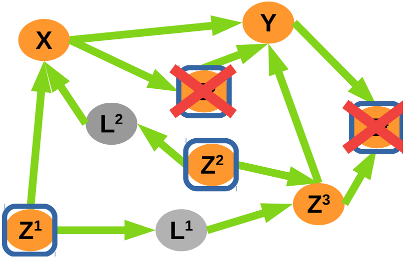


Back-door adjustment

Back-door criterion

A set $\mathbf{Z} \subset \mathbf{X} \setminus \{X, Y\}$ satisfies the back-door criterion for identifying the causal effect of X on Y if

1. $\mathbf{Z}$ blocks all paths between X and Y with arrow into X
2. No element in $\mathbf{Z}$ is a descendant of X



Back-door adjustment

Back-door criterion

A set $\mathbf{Z} \subset \mathbf{X} \setminus \{X, Y\}$ satisfies the back-door criterion for identifying the causal effect of X on Y if

1. $\mathbf{Z}$ blocks all paths between X and Y with arrow into X
2. No element in $\mathbf{Z}$ is a descendant of X

Back-door adjustment

Back-door criterion

A set $\mathbf{Z} \subset \mathbf{X} \setminus \{X, Y\}$ satisfies the back-door criterion for identifying the causal effect of X on Y if

1. $\mathbf{Z}$ blocks all paths between X and Y with arrow into X
2. No element in $\mathbf{Z}$ is a descendant of X

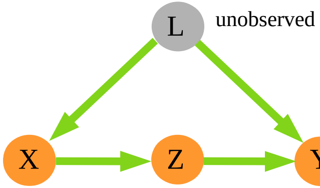
Theorem: Back-door adjustment as a sufficient criterion

If $\mathbf{Z}$ satisfy the back-door criterion for (X, Y), then

$$p(y|{\color{red}do(x)}) = \sum_{\mathbf{z}} p(y|x, \mathbf{z}) \cdot p(\mathbf{z})$$

Overview: Identification of Causal Effect

Example: Unobserved confounder plus mediation



Front-door criterion:

Allows to identify causal effect, but not as simple as controlling for a certain set of variables

Observational Causal Discovery

How to Get Causal Graphs

Relevance: Result of previous section

Knowing the causal graph of the data generating process, we can determine whether and how a causal effect can be identified from observational data

How to Get Causal Graphs

Relevance: Result of previous section

Knowing the causal graph of the data generating process, we can determine whether and how a causal effect can be identified from observational data

How to get the causal graph?

How to Get Causal Graphs

Relevance: Result of previous section

Knowing the causal graph of the data generating process, we can determine whether and how a causal effect can be identified from observational data

How to get the causal graph?

Option 1: Scientific knowledge

Talk to domain experts, general reasoning

Option 2: Observational causal discovery

Learn from observational data

Causal Graphs and (Conditional) Independencies

Fact:

The structure of the causal graph often has observable implications in terms of (conditional) independencies in the observed data.

Intuition:

- Statistical dependencies derive from causal relationships
- Conditioning can block and open the ‘flow of information’

Causal Graphs and (Conditional) Independencies

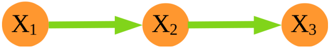
Fact:

The structure of the causal graph often has observable implications in terms of (conditional) independencies in the observed data.

Intuition:

- Statistical dependencies derive from causal relationships
- Conditioning can block and open the ‘flow of information’

Example 1: Chain



- X_1 influences X_2: $X_1 \not\perp\!\!\!\perp X_2$
- X_2 influences X_3: $X_2 \not\perp\!\!\!\perp X_3$
- X_1 influences X_3 through X_2: $X_1 \not\perp\!\!\!\perp X_3$
- Knowing X_2, X_1 does not say more about X_3: $X_1 \perp\!\!\!\perp X_3 | X_2$

Causal Graphs and (Conditional) Independencies

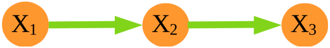
Fact:

The structure of the causal graph often has observable implications in terms of (conditional) independencies in the observed data.

Intuition:

- Statistical dependencies derive from causal relationships
- Conditioning can block and open the 'flow of information'

Example 1: Chain



- X_1 influences X_2: $X_1 \not\perp\!\!\!\perp X_2$
- X_2 influences X_3: $X_2 \not\perp\!\!\!\perp X_3$
- X_1 influences X_3 through X_2: $X_1 \not\perp\!\!\!\perp X_3$
- Knowing X_2, X_1 does not say more about X_3: $X_1 \perp\!\!\!\perp X_3 | X_2$

Causal Graphs and (Conditional) Independencies

Example 2: Fork

- X_2 influences X_1: $X_1 \not\perp\!\!\!\perp X_2$
- X_2 influences X_3: $X_2 \not\perp\!\!\!\perp X_3$
- Observing X_1 says something about X_2 and hence about X_3: $X_1 \not\perp\!\!\!\perp X_3$
- Knowing X_2, X_1 does not say more about X_3: $X_1 \perp\!\!\!\perp X_3 | X_2$

Causal Graphs and (Conditional) Independencies

Example 2: Fork

- X_2 influences X_1: $X_1 \not\!\perp\!\!\!\perp X_2$
- X_2 influences X_3: $X_2 \not\!\perp\!\!\!\perp X_3$
- Observing X_1 says something about X_2 and hence about X_3: $X_1 \not\!\perp\!\!\!\perp X_3$
- Knowing X_2, X_1 does not say more about X_3: $X_1 \perp\!\!\!\perp X_3 | X_2$

Example 3: Collider

- X_1 influences X_2: $X_1 \not\perp\!\!\!\perp X_2$
- X_3 influences X_2: $X_2 \not\perp\!\!\!\perp X_3$
- No influence between X_1 and X_3: $X_1 \perp\!\!\!\perp X_3$
- Observing X_2 introduces selection bias between X_1 and X_3: $X_1 \not\perp\!\!\!\perp X_3 | X_2$

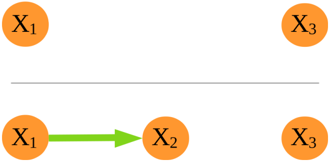
Causal Graphs and (Conditional) Independencies

Example 3: Collider

- X_1 influences X_2: $X_1 \not\perp\!\!\!\perp X_2$
- X_3 influences X_2: $X_2 \not\perp\!\!\!\perp X_3$
- No influence between X_1 and X_3: $X_1 \perp\!\!\!\perp X_3$
- Observing X_2 introduces selection bias between X_1 and X_3: $X_1 \not\perp\!\!\!\perp X_3 | X_2$

Causal Graphs and (Conditional) Independencies

Example 4: Disconnected variables



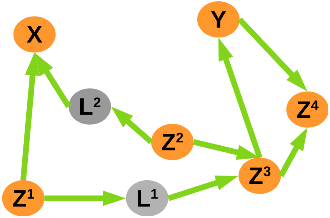
- No influence between X_1 and X_3: $X_1 \perp\!\!\!\perp X_3$

Causal Graphs and (Conditional) Independencies

General rule: d-separation

The graphical criterion of *d-separation* allows to read off all (conditional) independencies implied by the structure of a particular causal graph.

Example 5: Complicated case



Some independencies implied by d-separation

- $X \perp\!\!\!\perp Y | Z^1, Z^2$
- $X \perp\!\!\!\perp Y | Z^3$

Constraint Based Causal Discovery

Idea:

- Perform statistical tests of (conditional) independence in observational data
- Use test results to constrain the structure of the causal graph

Constraint Based Causal Discovery

Idea:

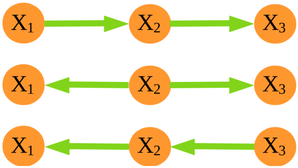
- Perform statistical tests of (conditional) independence in observational data
- Use test results to constrain the structure of the causal graph

Example 1:

Test decisions:

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- $X_1 \not\perp\!\!\!\perp X_3$
- $X_1 \perp\!\!\!\perp X_3 | X_2$

Possible causal graphs:



Constraint Based Causal Discovery

Idea:

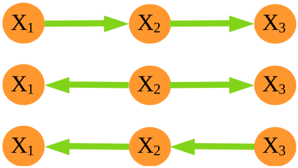
- Perform statistical tests of (conditional) independence in observational data
- Use test results to constrain the structure of the causal graph

Example 1:

Test decisions:

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- $X_1 \not\perp\!\!\!\perp X_3$
- $X_1 \perp\!\!\!\perp X_3 | X_2$

Possible causal graphs:



observationally equivalent graphs

Constraint Based Causal Discovery

Idea:

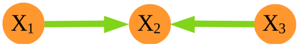
- Perform statistical tests of (conditional) independence in observational data
- Use test results to constrain the structure of the causal graph

Example 2:

Test decisions:

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- $X_1 \perp\!\!\!\perp X_3$

Possible causal graphs:



Constraint Based Causal Discovery

Idea:

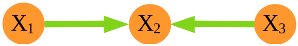
- Perform statistical tests of (conditional) independence in observational data
- Use test results to constrain the structure of the causal graph

Example 2:

Test decisions:

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- $X_1 \perp\!\!\!\perp X_3$

Possible causal graphs:



observationally equivalent graphs

Required Assumptions

Recall:

Causal questions cannot be answered from observational data without making any additional assumptions.

Required Assumptions

Recall:

Causal questions cannot be answered from observational data without making any additional assumptions.

Assumption 1: Existence of SCM

The data generating process can be described as an SCM. This implies the so called *causal Markov condition*:

d-separation in causal graph $\Rightarrow$ (conditional) independence

Assumption 2: Faithfulness

(conditional) independence $\Rightarrow$ d-separation in causal graph

Assumption 3: Causal sufficiency

There are no unobserved confounders and there are no selection variables.

Required Assumptions

Recall:

Causal questions cannot be answered from observational data without making any additional assumptions.

Assumption 1: Existence of SCM

The data generating process can be described as an SCM. This implies the so called *causal Markov condition*:

$$\text{d-separation in causal graph} \quad \Rightarrow \quad \text{(conditional) independence}$$

Assumption 2: Faithfulness

$$\text{(conditional) independence} \quad \Rightarrow \quad \text{d-separation in causal graph}$$

Assumption 3: Causal sufficiency

There are no unobserved confounders and there are no selection variables.

Assumption 3 can be dropped (but much more complicated)

Assumption 2 can be relaxed

Some Famous Algorithms

SGS-Algorithm:
Tests all possible (conditional) independence statements, use results to constrain the causal graph as discussed in the examples.

SGS stands for *S*pirtes, *G*lymour, and *S*cheines.

PC-Algorithm:
Based on SGS-Algorithm, but using much fewer conditional independence tests.

PC stands for *P*eter Spirtes and *C*lark Glymour.

FCI-Algorithm:
Similar to the PC-Algorithm, but without requiring the assumption of causal sufficiency, i.e., allowing for hidden confounders and selection variabes.

FCI stands for *f*ast *c*ausal *i*nference.

Two Main Approaches of Causal Discovery

Constraint-based causal discovery:
Constrain causal graph by using results of conditional independence tests in observational data.

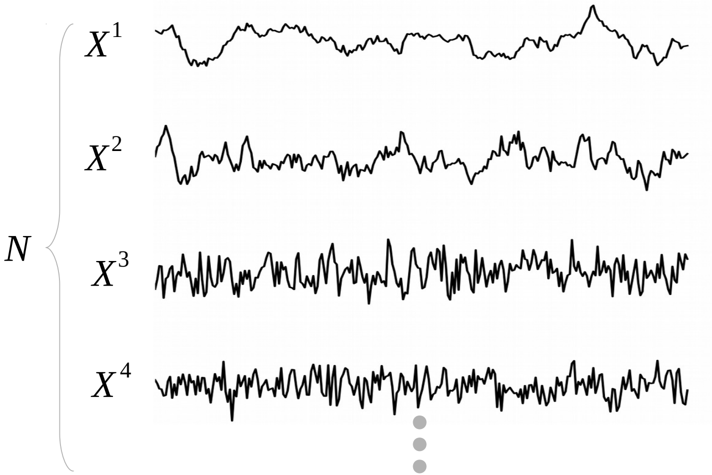
$\Rightarrow$ discussed today

SCM-based causal discovery:
Make assumption on functional causal relationships (e.g., linear or non-linear) and noise distributions (e.g., Gaussian or non-Gaussian) of data generating SCM.

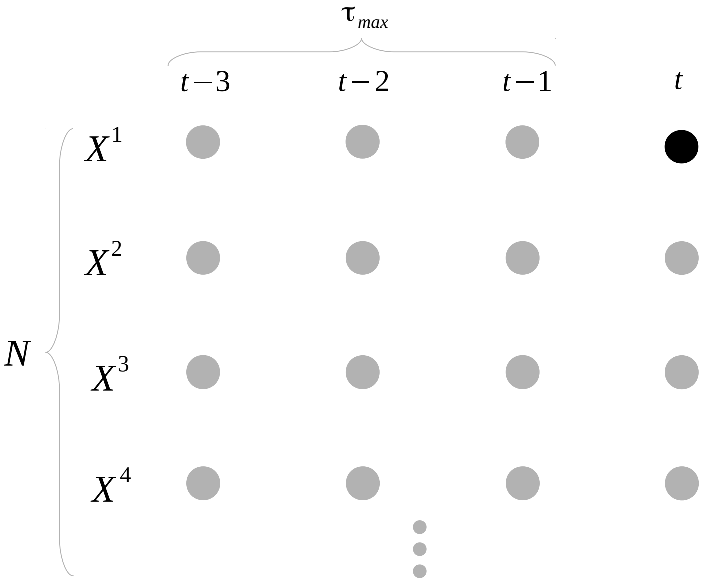
Generically, model can fit in one direction only. This allows to identify direction of causal influence.

$\Rightarrow$ not discussed today

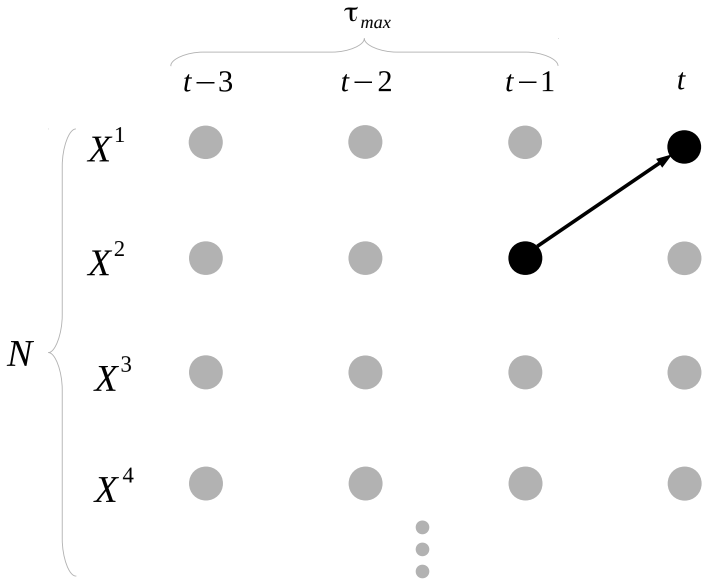
Causal Discovery in Time Series



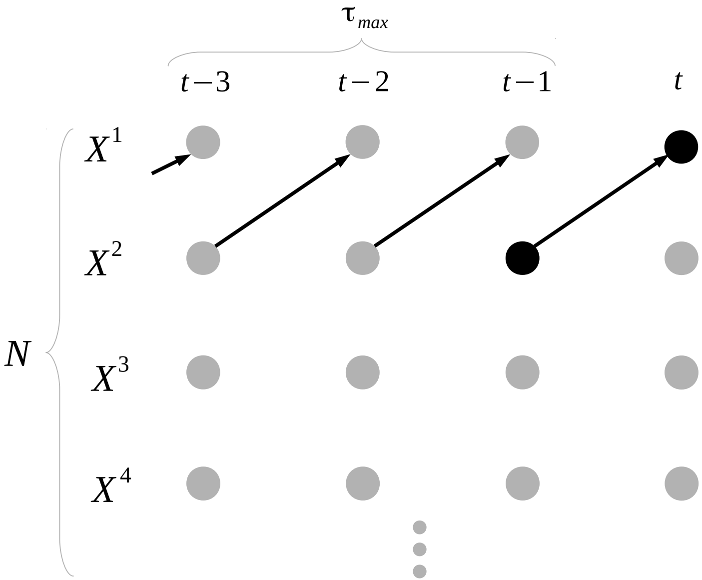
Causal Discovery in Time Series

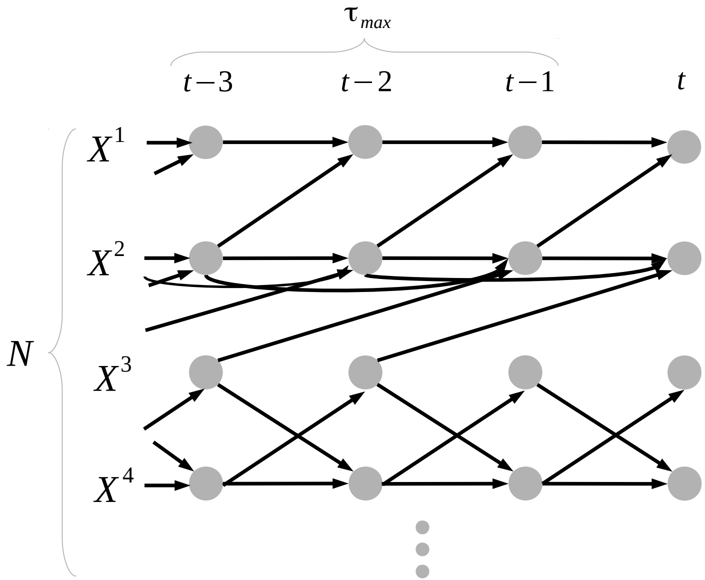


Causal Discovery in Time Series



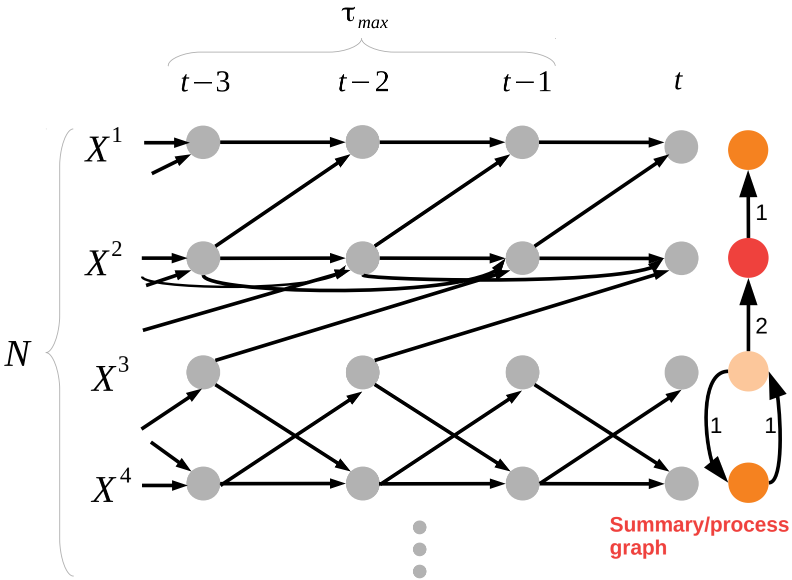
Causal Discovery in Time Series



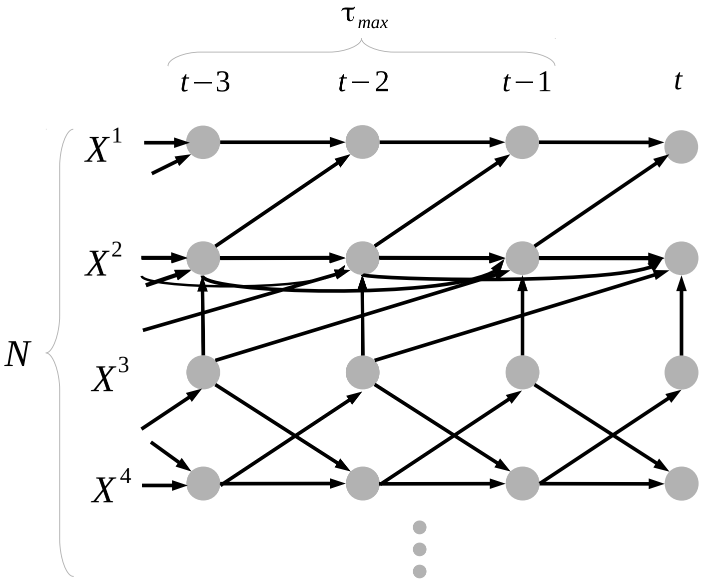


τ_{max}
$t-3$
$t-2$
$t-1$
t
X^1
X^2
X^3
X^4
N

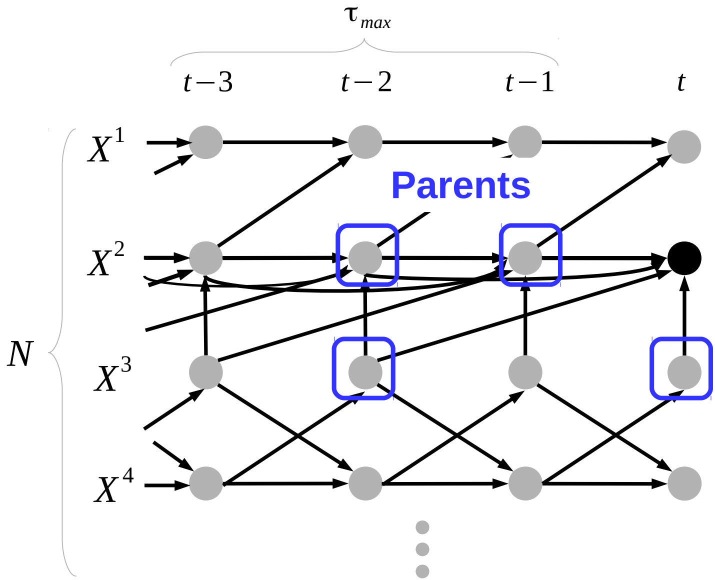
Causal Discovery in Time Series



Causal Discovery in Time Series



Causal Discovery in Time Series



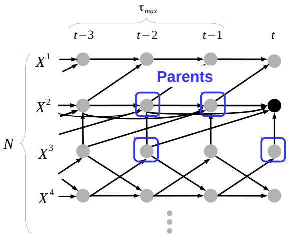
Causal Discovery in Time Series

Observations in time make things easier:

- Additional constraint: Causation cannot go back in time.

Observations in time make things harder:

- High dimensionality: Resolving in time increases the number of variables
- Statistical issues: Autocorrelation makes conditional independence tests statistically harder



Causal Discovery in Time Series

Observations in time make things harder:

- High dimensionality: Resolving in time increases the number of variables
- Statistical issues: Autocorrelation makes conditional independence tests statistically harder

PCMCI-Algorithm:

Adaption of the PC-algorithm to time series in which these challenges are addressed.

Causal Discovery in Time Series

Observations in time make things harder:

- High dimensionality: Resolving in time increases the number of variables
- Statistical issues: Autocorrelation makes conditional independence tests statistically harder

PCMCI-Algorithm:

Adaption of the PC-algorithm to time series in which these challenges are addressed.

LPCMCI-Algorithm:

Generalization of PCMCI that allows for hidden confounders.

Alternative Concept: Granger Causality

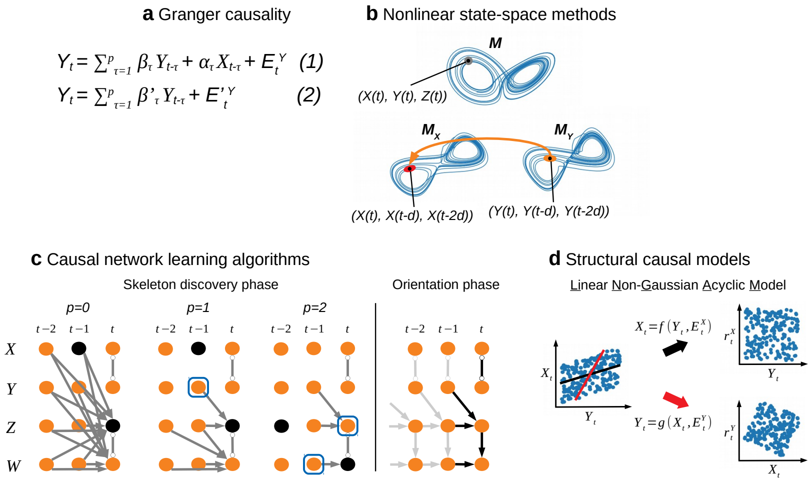
Idea:

X causes Y if the past of X helps in predicting the Y from its own past.

Limitations:

- Requires causal sufficiency
- Does not allow contemporaneous interactions

State of the art: see Runge et al. NatComm 2019



Real world challenges: see Runge et al. NatComm 2019

Challenges

Process:

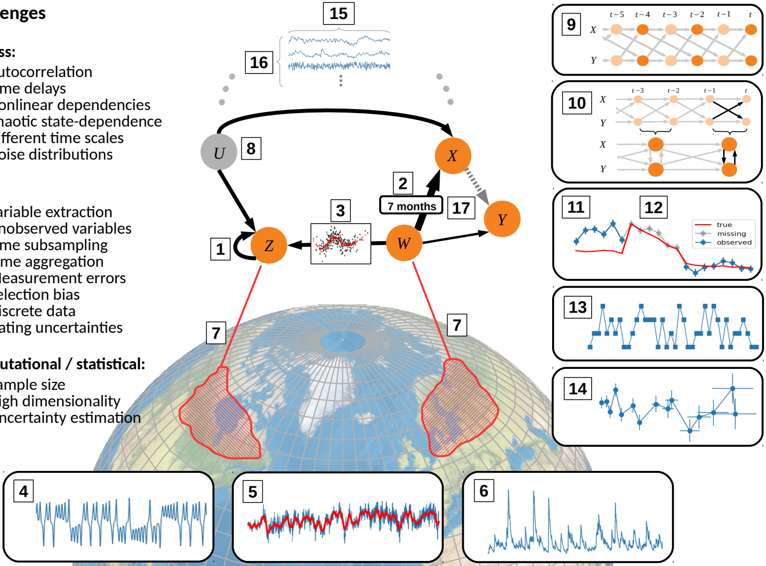
1 Autocorrelation
2 Time delays
3 Nonlinear dependencies
4 Chaotic state-dependence
5 Different time scales
6 Noise distributions

Data:

7 Variable extraction
8 Unobserved variables
9 Time subsampling
10 Time aggregation
11 Measurement errors
12 Selection bias
13 Discrete data
14 Dating uncertainties

Computational / statistical:

15 Sample size
16 High dimensionality
17 Uncertainty estimation



Application examples

Application cases

- Testing causal hypotheses
  [Runge et al., 2014, Runge et al., 2015b, Kretschmer et al., 2016, Runge et al., 2019b, Kretschmer et al., 2018, Runge et al., 2018, Runge et al., 2019a, Krich et al., 2020]

Application cases

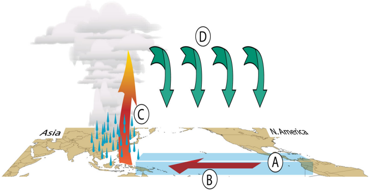
- Testing causal hypotheses
  [Runge et al., 2014, Runge et al., 2015b, Kretschmer et al., 2016, Runge et al., 2019b, Kretschmer et al., 2018, Runge et al., 2018, Runge et al., 2019a, Krich et al., 2020]
- Optimal statistical prediction schemes
  [Runge et al., 2015a, Kretschmer et al., 2017, Di Capua et al., 2019]

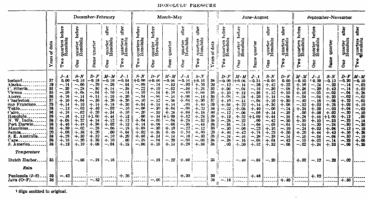
Application cases

- Testing causal hypotheses
  [Runge et al., 2014, Runge et al., 2015b, Kretschmer et al., 2016, Runge et al., 2019b, Kretschmer et al., 2018, Runge et al., 2018, Runge et al., 2019a, Krich et al., 2020]
- Optimal statistical prediction schemes
  [Runge et al., 2015a, Kretschmer et al., 2017, Di Capua et al., 2019]
- Evaluating climate/physical models
  [Schleussner et al., 2014, Nowack et al., 2019]

Reconstructing Walker Circulation

- Monthly surface pressure anomalies in the West Pacific (WPAC), surface air temperature anomalies in the Central Pacific (CPAC) and East Pacific (EPAC)





HONOLULU PRESSURE

	Years of data	December-February: Two quarters before Honolulu	One quarter before Honolulu	Same quarter	One quarter after Honolulu	Two quarters after Honolulu	March-May: Two quarters before Honolulu	One quarter before Honolulu	Same quarter	One quarter after Honolulu	Two quarters after Honolulu
		J-A	S-N	D-F	M-M	J-A	S-N	D-F	M-M	J-A	S-N
Iceland	37	0.00	−0.16	−0.18	−0.10	−0.04	+0.06	+0.06	−0.06	−0.16	+0.16
Alaska	19	−.14	−.06	−.70	−.26	−.04	−.08	−.14	−.52	−.30	+.50
C. Siberia	33	−.20	−.28	−.30	+.14	−.38	−.22	+.10	−.02	−.34	+.24
Vienna	35	−.20	+.08	−.04	+.10	−.14	−.10	−.14	+.18	−.40	+.08
Azores	38	+.16	−.14	+.14	+.16	+.20	−.34	−.06	−.14	+.06	−.14
Charleston	37	+.10	+.04	−.20	+.20	+.20	−.10	−.12	−.08	+.04	+.30
San Francisco	39	+.14	+.10	+.14	+.18	+.20	+.04	+.14	+.14	.00	+.40
Tokio	38	−.12	−.02	+.40	−.02	−.10	−.06	−.20	+.04	−.06	+.22
Cairo	36	+.18	−.02	+.20	+.34	+.10	−.20	−.10	+.28	+.08	+.18
Honolulu	39	−.16	+.12	+1.00	+.44	+.12	.00	+.44	+1.00	+.52	+.24
N. W. India	38	+.08	+.22	+.14	+.12	+.12	+.04	−.06	−.04	+.24	−.32
Port Darwin	38	+.46	+.48	+.38	+.02	+.12	+.14	+.08	−.06	−.38	−.44
Mauritius	38	+.30	−.02	−.26	−.06	−.18	+.08	−.30	+.10	−.22	−.12
Samoa	20	+.08	+.38	+.20	.00	+.04	+.02	+.36	+.48	+.10	+.60
S. E. Australia	38	+.28	+.30	+.32	−.10	+.20	+.06	+.14	−.14	−.16	−.30
Cape	38	+.16	−.20	+.30	+.10	.00	−.38	−.04	−.20	−.06	−.06
S. America	38	+.12	+.10	+.06	−.01	+.18	−.06	+.18	+.14	+.24	+.34
Temperature											
Dutch Harbor	35	......	−.06	−.24	−.16	......	......	−.18	−.22	+.04	......
Rain											
Peninsula (J-S)	39	−.42	......	......	......	+.26	......	......	......	+.30	......
Java (O-F)	36	......	......	−.32	......	......	......	−.06	......	......	......

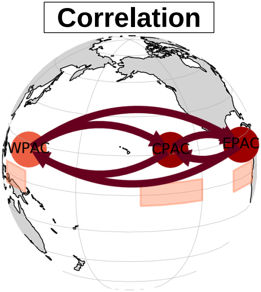
	Years of data	June-August: Two quarters before Honolulu	One quarter before Honolulu	Same quarter	One quarter after Honolulu	Two quarters after Honolulu	September-November: Two quarters before Honolulu	One quarter before Honolulu	Same quarter	One quarter after Honolulu	Two quarters after Honolulu
		D-F	M-M	J-A	S-N	D-F	M-M	J-A	S-N	D-F	M-M
Iceland	38	−0.06	+0.06	−0.34	−0.06	0.00	−0.10	[1]0.20	−0.12	−0.20	−0.10
Alaska	20	−.12	−.14	−.28	+.42	+.22	+.16	−.24	−.22	−.14	−.02
C. Siberia	33	+.06	−.04	−.16	−.30	−.08	+.26	−.20	+.42	−.14	+.02
Vienna	35	+.10	+.12	−.26	+.12	+.10	+.16	−.03	−.02	−.04	+.26
Azores	38	+.78	−.24	+.20	+.14	−.08	−.04	−.04	+.20	−.02	−.02
Charleston	37	+.10	−.14	−.08	+.10	+.30	−.40	−.18	−.08	+.22	−.02
San Francisco	39	+.04	+.22	+.14	+.36	+.08	−.32	+.02	+.26	+.02	+.10
Tokio	38	−.22	+.08	+.40	−.06	−.24	+.24	+.30	−.02	+.04	−.02
Cairo	36	−.12	−.08	−.06	+.14	−.28	.00	+.08	+.02	−.20	−.10
Honolulu	39	+.12	+.52	+1.00	+.44	−.16	+.24	+.44	+1.00	+.12	.00
N. W. India	38	−.28	−.22	.00	−.42	−.30	+.30	+.10	−.32	−.22	−.14
Port Darwin	38	−.34	−.54	−.66	−.68	−.64	−.04	−.20	−.24	−.30	−.26
Mauritius	38	−.30	−.08	−.22	+.16	−.10	−.24	+.02	+.04	−.12	−.18
Samoa	20	+.46	+.42	+.74	+.22	+.50	+.20	+.44	+.42	+.56	+.28
S. E. Australia	38	−.24	−.46	−.28	−.38	−.54	−.26	−.04	−.04	−.30	−.18
Cape	38	−.28	−.16	−.08	+.04	−.42	−.10	+.02	−.14	−.22	+.06
S. America	38	.00	+.30	+.52	+.32	−.06	−.02	+.24	+.32	−.08	+.28
Temperature											
Dutch Harbor	35	......	−.40	−.04	−.20	......	+.02	−.12	−.22	−.02	......
Rain											
Peninsula (J-S)	39	......	......	+.46	......	......	......	+.02	......	......	......
Java (O-F)	36	−.14	......	......	......	+.40	......	......	......	+.30	......

[1] Sign omitted in original.

Runge et al. *Nat. Comm.* (2019)

Reconstructing Walker Circulation

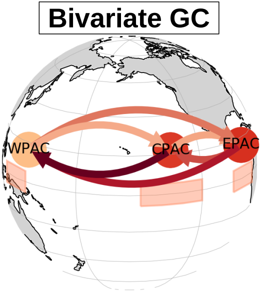
- Monthly surface pressure anomalies in the West Pacific (WPAC), surface air temperature anomalies in the Central Pacific (CPAC) and East Pacific (EPAC)
- Correlation analysis gives a completely connected graph



Runge et al. *Nat. Comm.* (2019)

Reconstructing Walker Circulation

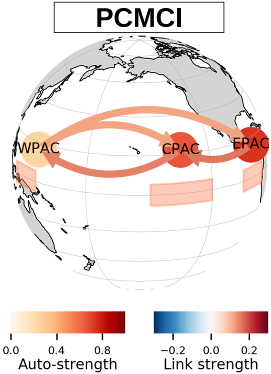
- Monthly surface pressure anomalies in the West Pacific (WPAC), surface air temperature anomalies in the Central Pacific (CPAC) and East Pacific (EPAC)
- Correlation analysis gives a completely connected graph
- Also bivariate Granger Causality cannot remove indirect and common driver links



Runge et al. *Nat. Comm.* (2019)

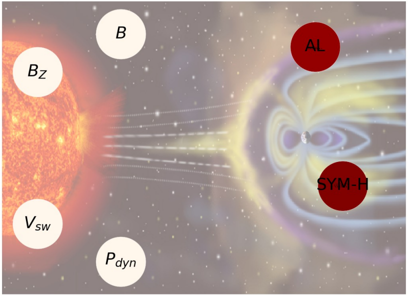
Reconstructing Walker Circulation

- Monthly surface pressure anomalies in the West Pacific (WPAC), surface air temperature anomalies in the Central Pacific (CPAC) and East Pacific (EPAC)
- Correlation analysis gives a completely connected graph
- Also bivariate Granger Causality cannot remove indirect and common driver links
- PCMCI [Runge et al., 2019b] better identifies the Walker circulation



Runge et al. *Nat. Comm.* (2019)

- Hypothesis on interaction between magnetospheric Auroral Electrojet index (AL), magnetospheric ring current strength (SYM-H), and solar wind parameters

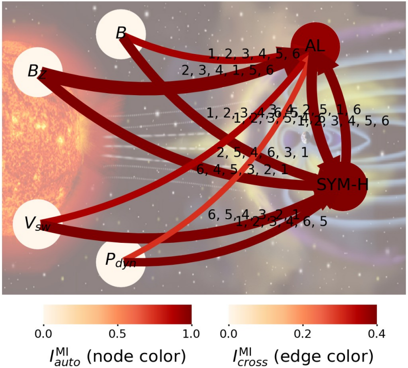


Runge et al. *Sci. Rep.* (2018), $\Delta t = 20$min resolution

Space physics

- Hypothesis on interaction between magnetospheric Auroral Electrojet index (AL), magnetospheric ring current strength (SYM-H), and solar wind parameters
- Mutual information analysis gives many dependencies

Mutual information

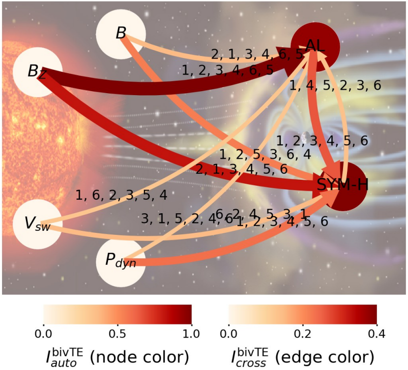


Runge et al. *Sci. Rep.* (2018), $\Delta t = 20$min resolution

Space physics

- Hypothesis on interaction between magnetospheric Auroral Electrojet index (AL), magnetospheric ring current strength (SYM-H), and solar wind parameters
- Mutual information analysis gives many dependencies
- Transfer Entropy cannot remove indirect and common driver links

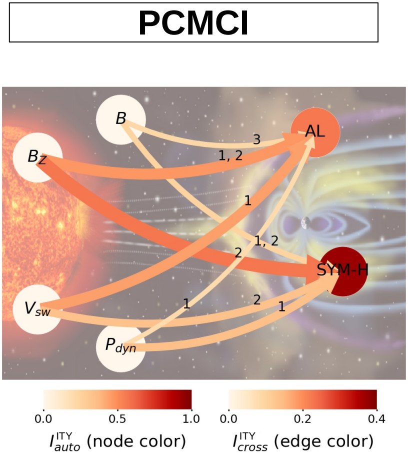
Transfer Entropy



Runge et al. *Sci. Rep.* (2018), $\Delta t = 20$min resolution

Space physics

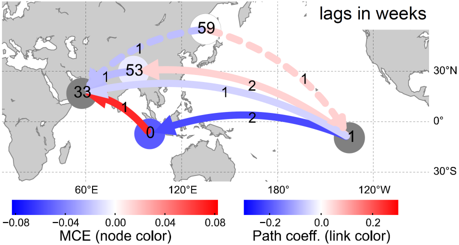
- Hypothesis on interaction between magnetospheric Auroral Electrojet index (AL), magnetospheric ring current strength (SYM-H), and solar wind parameters
- Mutual information analysis gives many dependencies
- Transfer Entropy cannot remove indirect and common driver links
- PCMCI yields novel insight that solar wind is common driver of magnetospheric indices



Runge et al. *Sci. Rep.* (2018), $\Delta t = 20$min resolution

Causal mediation analysis

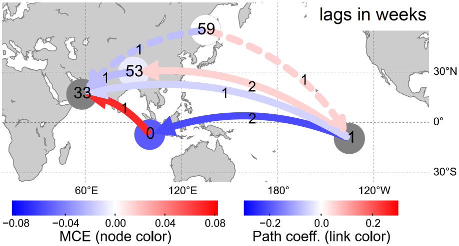
- Pathway mechanisms by which El Nino influences Indian monsoon through sea-level pressure system



Runge et al. *Nat. Comm.* (2015)

Causal mediation analysis

- Pathway mechanisms by which El Nino influences Indian monsoon through sea-level pressure system
- *Mediated Causal Effect* (MCE) quantifies how much an intermediate variable (node) contributes to a causal effect

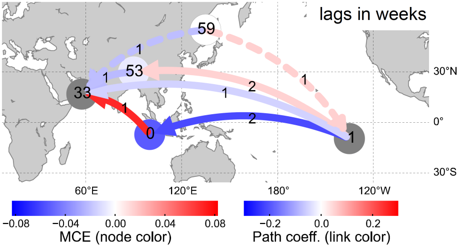


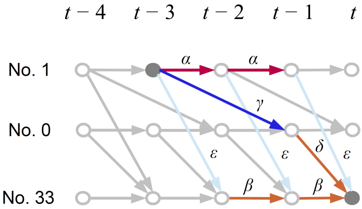
Runge et al. *Nat. Comm.* (2015)

Causal mediation analysis

- Pathway mechanisms by which El Nino influences Indian monsoon through sea-level pressure system
- *Mediated Causal Effect* (MCE) quantifies how much an intermediate variable (node) contributes to a causal effect
- Linear path analysis (early approach due to Sewall Wright in 1920s)

Runge et al. *Nat. Comm.* (2015)

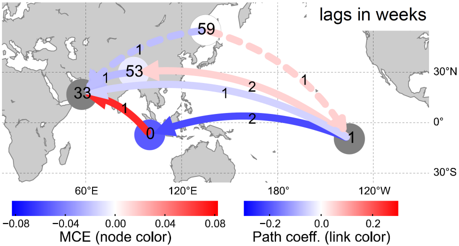


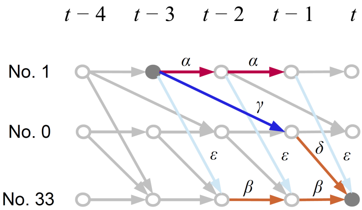


Causal mediation analysis

- Pathway mechanisms by which El Nino influences Indian monsoon through sea-level pressure system
- *Mediated Causal Effect* (MCE) quantifies how much an intermediate variable (node) contributes to a causal effect
- Linear path analysis (early approach due to Sewall Wright in 1920s)
- Nonlinear extension in Runge *Physical Review E* (2015)

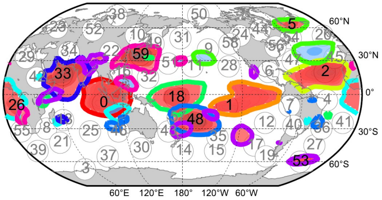
Runge et al. *Nat. Comm.* (2015)





Causal complex network analysis

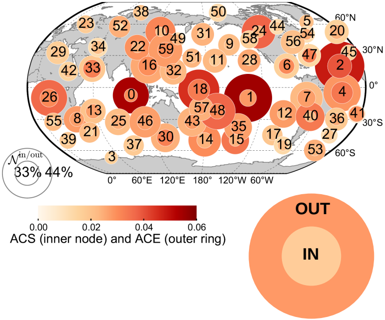
- Complex network measures based on extracted causal network from sea-level pressure system



Runge et al. *Nat. Comm.* (2015)

Causal complex network analysis

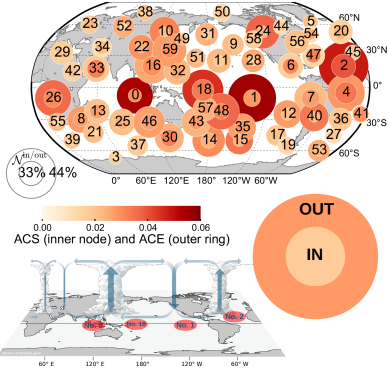
- Complex network measures based on extracted causal network from sea-level pressure system
- Global *causal gateways* based on *Average Causal Effect* (ACE)



Runge et al. *Nat. Comm.* (2015)

Causal complex network analysis

- Complex network measures based on extracted causal network from sea-level pressure system
- Global *causal gateways* based on *Average Causal Effect* (ACE)
- Here well represents tropical atmospheric uplift regions

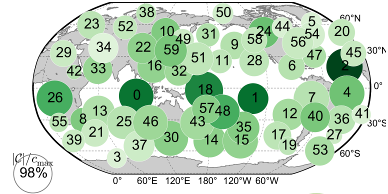


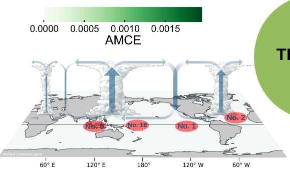
Runge et al. *Nat. Comm.* (2015)

Causal complex network analysis

- Complex network measures based on extracted causal network from sea-level pressure system
- Global *causal gateways* based on *Average Causal Effect* (ACE)
- Here well represents tropical atmospheric uplift regions
- Global *causal mediators* based on *Mediated Causal Effect* (MCE)

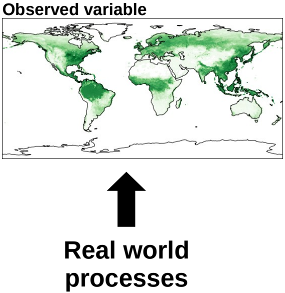
Runge et al. *Nat. Comm.* (2015)

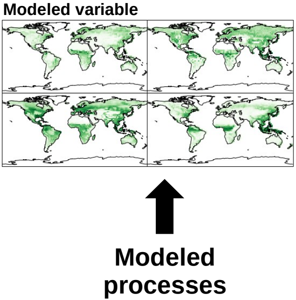




Causal model evaluation (Nowack et al., 2020)

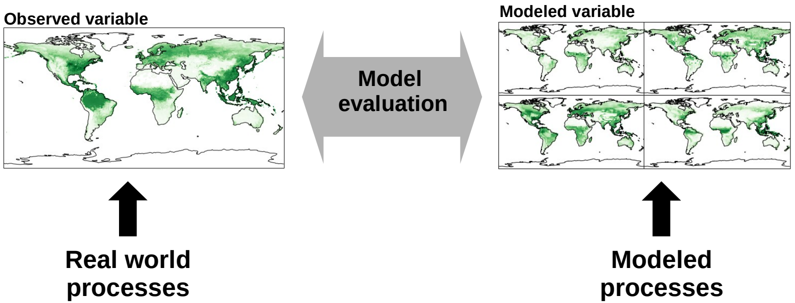
Motivation: Simple statistics (e.g. mean, variance, trends) can be right for the wrong reasons





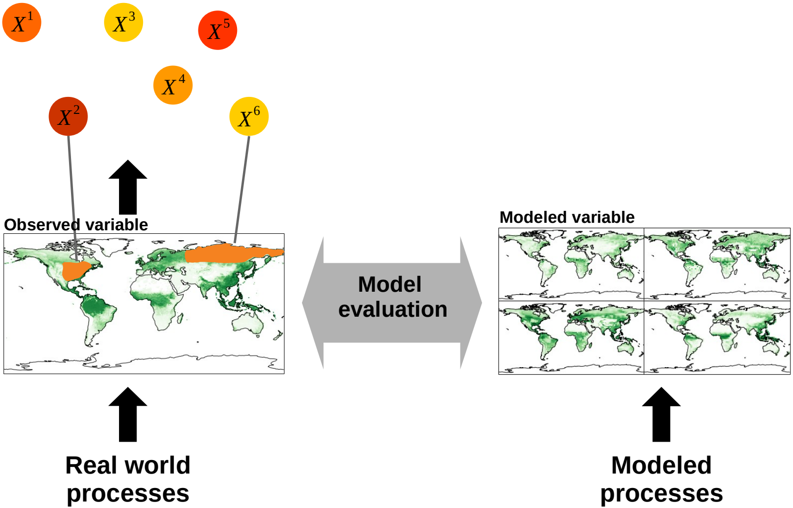
Causal model evaluation (Nowack et al., 2020)

Motivation: Simple statistics (e.g. mean, variance, trends) can be right for the wrong reasons



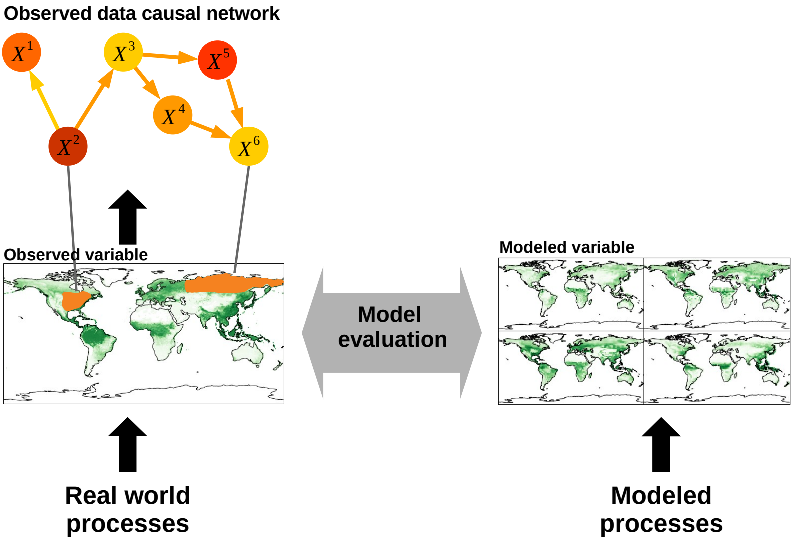
Causal model evaluation (Nowack et al., 2020)

Idea: Compare climate models and observations in terms of causal relationships



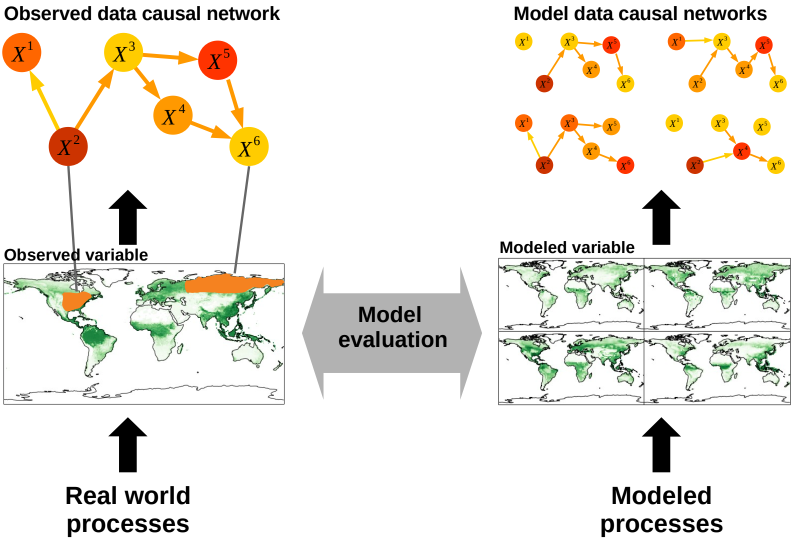
Causal model evaluation (Nowack et al., 2020)

Idea: Compare climate models and observations in terms of causal relationships



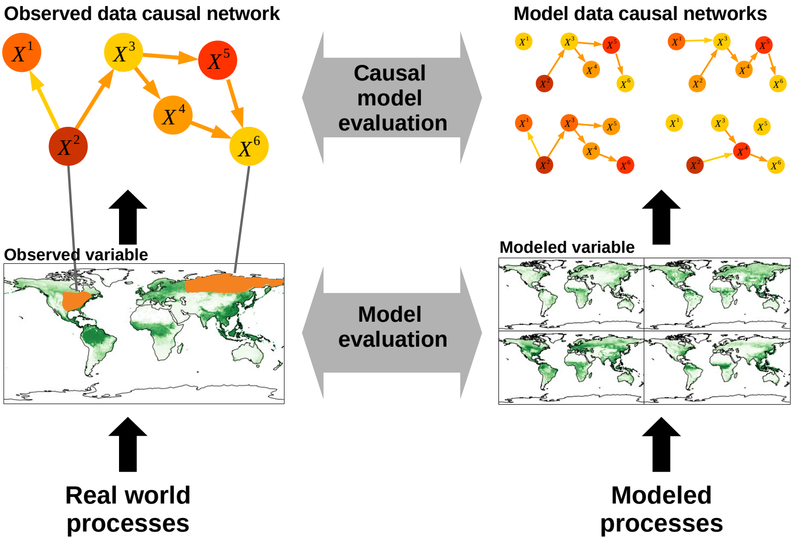
Causal model evaluation (Nowack et al., 2020)

Idea: Compare climate models and observations in terms of causal relationships



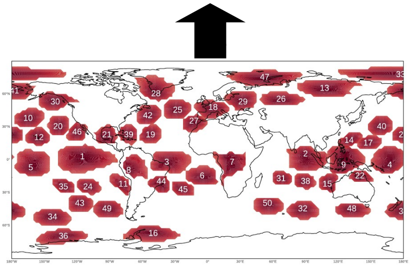
Causal model evaluation (Nowack et al., 2020)

Idea: Compare climate models and observations in terms of causal relationships



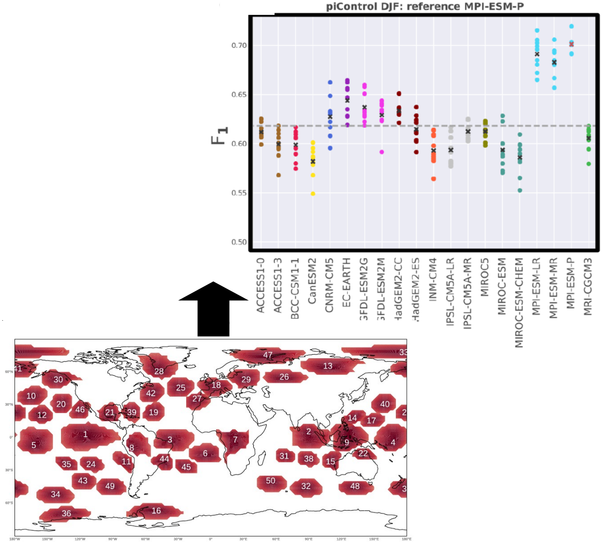
Causal model evaluation (Nowack et al., 2020)

First results: CMIP5 simulations (historical and preindustrial) vs NCEP/NCAR reanalysis data of regional 3-day-mean sea level pressure



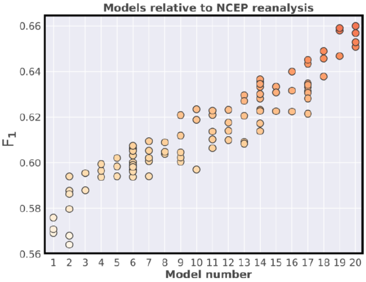
Causal model evaluation (Nowack et al., 2020)

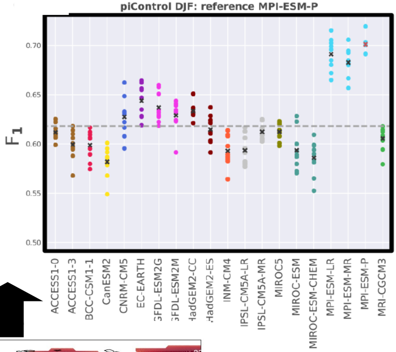
Validation: Similar climate models have similar causal networks; F-score as network comparison metric

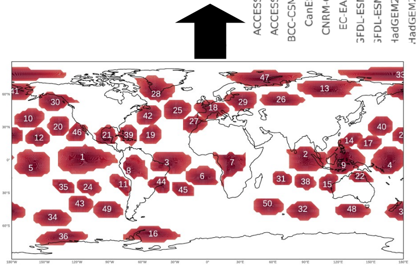


Causal model evaluation (Nowack et al., 2020)

Model evaluation: Significant differences in comparison to reanalysis



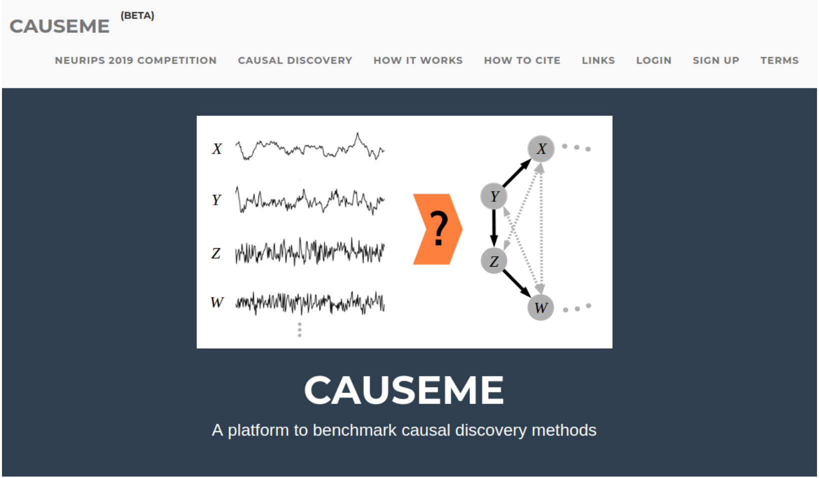




Causality benchmark platform

Causality benchmark platform `CauseMe.net`

Joint work with Jordi Munoz-Mari and Gustau Camps-Valls (U Valencia)



Causality benchmark platform CauseMe.net

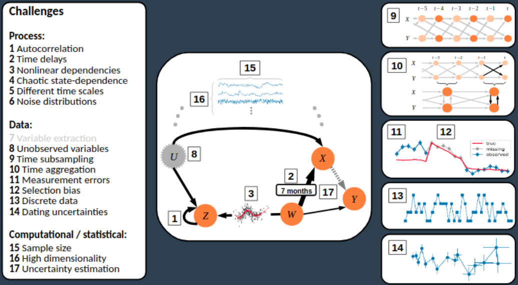
Joint work with Jordi Munoz-Mari and Gustau Camps-Valls (U Valencia)

CAUSEME (BETA)

NEURIPS 2019 COMPETITION | CAUSAL DISCOVERY | HOW IT WORKS | HOW TO CITE | LINKS | LOGIN | SIGN UP | TERMS

HOW IT WORKS

Causeme currently covers a wide range of synthetic model data mimicking a number of real world challenges. These cover time delays, autocorrelation, nonlinearity, chaotic dynamics, extreme events, measurement error, and will be extended by many more. Method developers can upload their predictions (matrices of causal connections) and the platform evaluates and ranks the methods according to different metrics of performance. After registering and logging in, more information, datasets, and example code snippets are given.

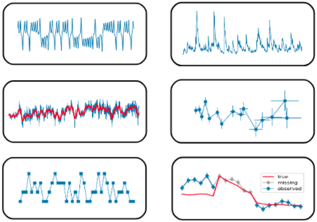


Causality benchmark platform CauseMe.net

Joint work with Jordi Munoz-Mari and Gustau Camps-Valls (U Valencia)

JAKOB RUNGE DATA AND MODELS METHODS ↓⁼RANKING HOWTO MY RESULTS LOGOUT

DATA AND MODELS



Below you find a list of available datasets. Currently, they come from dynamical model systems featuring different challenges for causal discovery from time series as discussed in the accompanying Nature Communications Perspective paper. At the end of this page you find information on how to contribute real world datasets or model systems. Clicking on the model name will bring you to a description of the model and a list of experimental datasets. Please see the CauseMe workflow description in HowTo on how to upload your results for these experiments.

You can search through the database by name, description or tags.

Filter models:

Name	Long name	Type	Tags

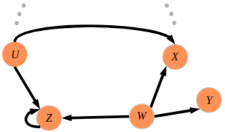
DLR

Causality benchmark platform CauseMe.net

Joint work with Jordi Munoz-Mari and Gustau Camps-Valls (U Valencia)

JAKOB RUNGE   DATA AND MODELS   METHODS   RANKING   HOWTO   MY RESULTS   LOGOUT

METHODS



Below you find a list of methods applied by users of this platform. Clicking on the name will bring you to a description of the method. You can search through the database by name, user, and tags. Register your own methods on My Results!

Show 10 entries

Filter methods:

Name	User	Tags
adaptive-lasso	Jakob runge	Linear, time delays, high-dimensional
correlation	Jakob runge	Linear, time series, non-conditional
distance-correlation	Jakob runge	Time delays, nonlinear, non-conditional
FullCI-CMIknn	Jakob runge	Time delays, nonlinear
FullCI-GPDC	Jakob runge	Time delays, nonlinear
FullCI-ParCorr	Jakob runge	Linear, time delays

Causality benchmark platform CauseMe.net

Joint work with Jordi Munoz-Mari and Gustau Camps-Valls (U Valencia)

JAKOB RUNGE — DATA AND MODELS — METHODS — RANKING — HOWTO — MY RESULTS — LOGOUT

RANKING

The table below presents a ranking of methods for different experiments and can be sorted according to the different metrics in columns. Optionally, the table can be filtered by metric values above or below a certain threshold. For example, one can display only methods with a FPR below 6% and sort these by TPR in decreasing order. In addition, the search field can be used on the whole table to select only particular experiments or particular methods (or both). For example, "varmodel N-10 T-150" will list all methods with 'varmodel' in the string and all experiments with N=10 variables and sample length T=150. See here for a description of metrics: AUC is based on scores, while F-measure, FPR, and TPR are based on binary link predictions by thresholding uploaded p-values at 0.05 (only available if p-values were uploaded). TLR requires lag predictions.

Filter: FPR < 0.06 Go ☐ Paper ☐ Code ☑ Validated

Show 100 entries Search: linear-VAR_multirealizations_N-

Id	User	Experiment	Method (params)	Paper	Code	Valid.	Time	AUC	AUC-PR	F-measure	FPR	TPR	TLR	Boxplot FPR	Boxplot TPR
37	Jakob Runge	linear-VAR_multirealiz	PCMCI-ParCorr (tau_max=5,pc_	✔	✔	✔	2.97	0.98	0.89	0.56	0.05	0.92	0.98		
145	Jakob Runge	linear-VAR_multirealiz	adaptive-lasso (tau_max=5,)	✔	✖	✔	18.99	0.96	0.86	0.75	0.02	0.92	0.99		
215	Jakob Runge	linear-VAR_multirealiz	varmodel (maxlags=5,)	✔	✔	✔	0.48	0.95	0.69	0.50	0.05	0.76	0.98		
249	Jakob Runge	linear-VAR_multirealiz	FullCI-ParCorr (tau_max=5,)	✔	✔	✔	11.24	0.94	0.70	0.51	0.05	0.74	0.98		

Showing 1 to 4 of 4 entries (filtered from 1,604 total entries)

Previous 1 Next

Summary

- **Causal inference:**
  Answering causal questions from empirical data

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)
  - Potentially causal sufficiency

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)
  - Potentially causal sufficiency
  - Assumptions on dependency types (linearity, etc) and distributions

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known → **estimate causal effects**
  2. Causal graph unknown → **causal discovery** (→ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)
  - Potentially causal sufficiency
  - Assumptions on dependency types (linearity, etc) and distributions
  - Time series data: time order, stationarity, ...

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)
  - Potentially causal sufficiency
  - Assumptions on dependency types (linearity, etc) and distributions
  - Time series data: time order, stationarity, ...
- These assumptions can sometimes not be **tested** from the same data or even any empirical data

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known → **estimate causal effects**
  2. Causal graph unknown → **causal discovery** (→ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)
  - Potentially causal sufficiency
  - Assumptions on dependency types (linearity, etc) and distributions
  - Time series data: time order, stationarity, ...
- These assumptions can sometimes not be **tested** from the same data or even any empirical data
- Causal conclusions require to

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)
  - Potentially causal sufficiency
  - Assumptions on dependency types (linearity, etc) and distributions
  - Time series data: time order, stationarity, ...
- These assumptions can sometimes not be **tested** from the same data or even any empirical data
- Causal conclusions require to
  - **state assumptions** and explain reasons for believing them

Summary

- **Causal inference:**
  Answering causal questions from empirical data
- Two settings:
  1. Causal graph assumed known $\rightarrow$ **estimate causal effects**
  2. Causal graph unknown $\rightarrow$ **causal discovery** ($\rightarrow$ then causal effects)
- **Causal inference** requires **assumptions**
  - Underlying SCM / Causal Markov Condition
  - Faithfulness (for causal discovery)
  - Potentially causal sufficiency
  - Assumptions on dependency types (linearity, etc) and distributions
  - Time series data: time order, stationarity, ...
- These assumptions can sometimes not be **tested** from the same data or even any empirical data
- Causal conclusions require to
  - **state assumptions** and explain reasons for believing them
  - and to indicate how conclusions are altered for different assumptions

Thank you! Questions?

- *Nature Comm.* Perspective on causal discovery in time series [Runge et al., 2019a]
- Causal inference: full theory [Pearl, 2000], primer [Pearl et al., 2016], linear models [Pearl, 2013], popular science book [Pearl and Mackenzie, 2018]
- Causal discovery: general [Spirtes et al., 2000], for time series [Runge, 2018a, Runge et al., 2019a]
- Restricted SCMs [Peters et al., 2017]
- PCMCI [Runge et al., 2019b] in *Science Advances*
- PCMCI^{+} [Runge, 2020] in *UAI*
- LPCMCI [Gerhardus and Runge, 2020] in *NeurIPS*
- My software: `jakobrunge.github.io/tigramite`

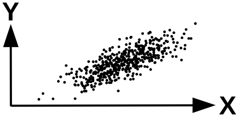
Causal discovery in a nutshell

Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?

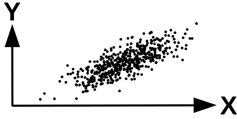
Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?



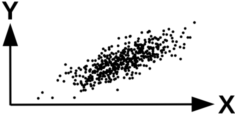
Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?



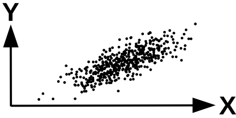
Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?



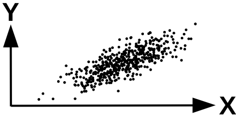
Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?



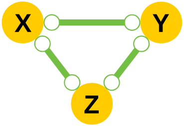
Causal discovery in a nutshell [Spirtes et al., 2000]

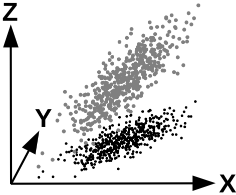
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness



Causal discovery in a nutshell [Spirtes et al., 2000]

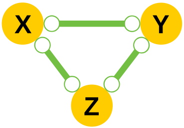
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness

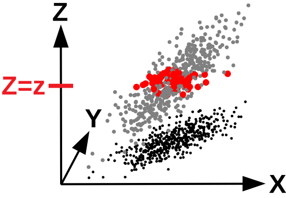




Causal discovery in a nutshell [Spirtes et al., 2000]

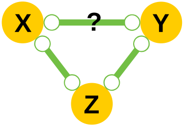
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
  $\Leftrightarrow p(X, Y \mid Z) = p(X \mid Z)p(Y \mid Z)$

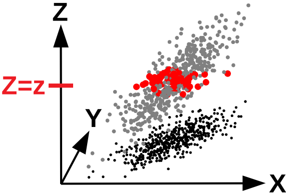




Causal discovery in a nutshell [Spirtes et al., 2000]

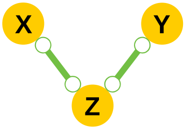
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$

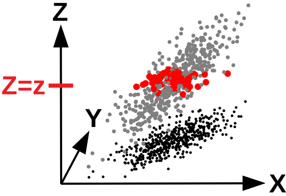




Causal discovery in a nutshell [Spirtes et al., 2000]

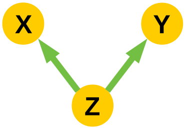
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\implies$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence
  $\implies$ no causal connectedness

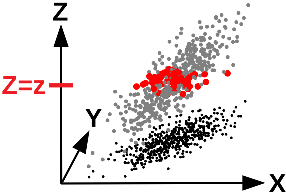




Causal discovery in a nutshell [Spirtes et al., 2000]

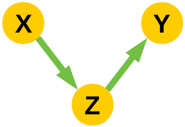
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence
  $\Longrightarrow$ <u>no</u> causal connectedness

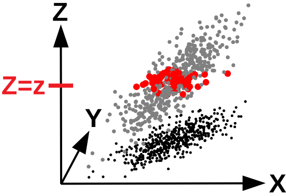




Causal discovery in a nutshell [Spirtes et al., 2000]

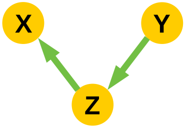
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence
  $\Longrightarrow$ no causal connectedness

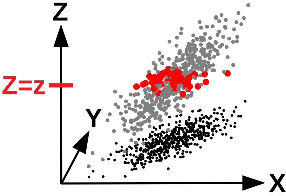




Causal discovery in a nutshell [Spirtes et al., 2000]

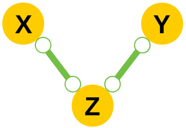
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence
  $\Longrightarrow$ no causal connectedness

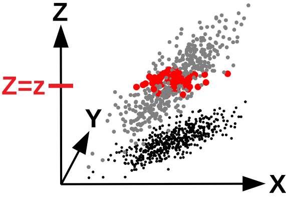




Causal discovery in a nutshell [Spirtes et al., 2000]

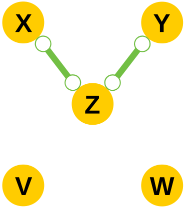
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence
  $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs

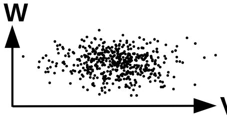




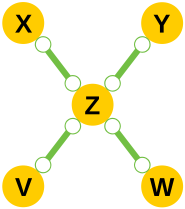
Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)



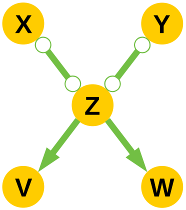


Causal discovery in a nutshell [Spirtes et al., 2000]



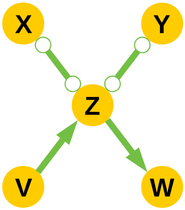
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
  - Which causal models explain these? (assuming no unobserved variables = latents)

Causal discovery in a nutshell [Spirtes et al., 2000]



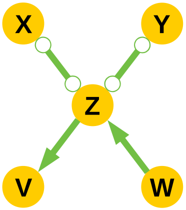
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)

Causal discovery in a nutshell [Spirtes et al., 2000]



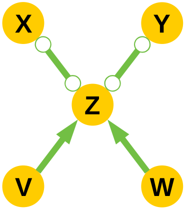
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence
  $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence
  $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
  - Which causal models explain these? (assuming no unobserved variables = latents)

Causal discovery in a nutshell [Spirtes et al., 2000]



- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)

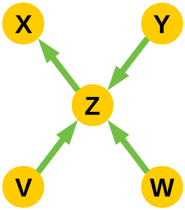
Causal discovery in a nutshell [Spirtes et al., 2000]



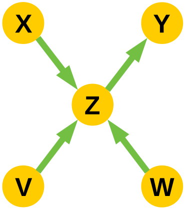
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)

Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
  - Which causal models explain these? (assuming no unobserved variables = latents)



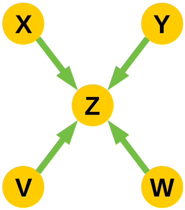
Causal discovery in a nutshell [Spirtes et al., 2000]



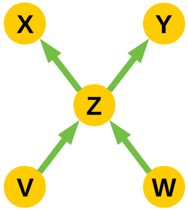
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\implies$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\implies$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)

Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)



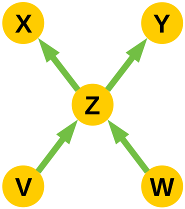
Causal discovery in a nutshell [Spirtes et al., 2000]



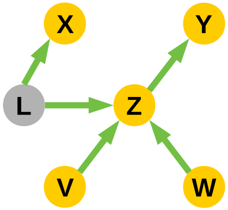
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)

Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
  - Which causal models explain these? (assuming no unobserved variables = latents)
  - Only possible model



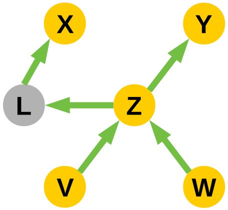
Causal discovery in a nutshell [Spirtes et al., 2000]



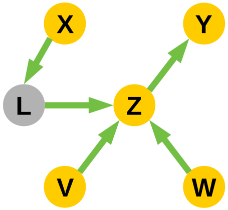
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model

Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
  - Which causal models explain these? (assuming no unobserved variables = latents)
  - Only possible model



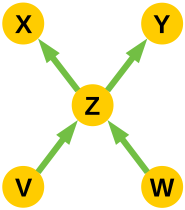
Causal discovery in a nutshell [Spirtes et al., 2000]



- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\!\perp\!\!\!\perp Y$, $X \not\!\perp\!\!\!\perp W$, $Y \not\!\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model

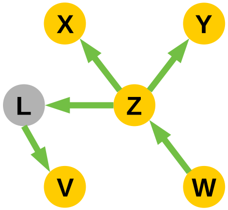
Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\!\perp\!\!\!\perp Y$, $X \not\!\perp\!\!\!\perp W$, $Y \not\!\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model

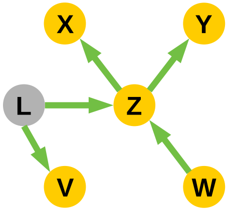


Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
  - Which causal models explain these? (assuming no unobserved variables = latents)
  - Only possible model



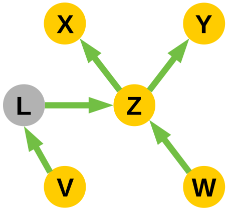
Causal discovery in a nutshell [Spirtes et al., 2000]



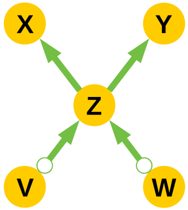
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model

Causal discovery in a nutshell [Spirtes et al., 2000]

- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\implies$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\implies$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model

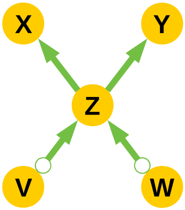


Causal discovery in a nutshell [Spirtes et al., 2000]



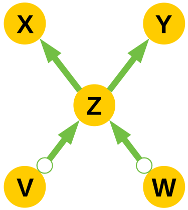
- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\!\perp\!\!\!\perp Y$, $X \not\!\perp\!\!\!\perp W$, $Y \not\!\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model (accounting for latents)

Causal discovery in a nutshell [Spirtes et al., 2000]



- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ <u>no</u> causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y | Z$, $X \perp\!\!\!\perp V | Z$, $X \perp\!\!\!\perp W | Z$, $Y \perp\!\!\!\perp V | Z$, $Y \perp\!\!\!\perp W | Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model (accounting for latents)
    - Causal inference methods use (some of) these assumptions

Causal discovery in a nutshell [Spirtes et al., 2000]

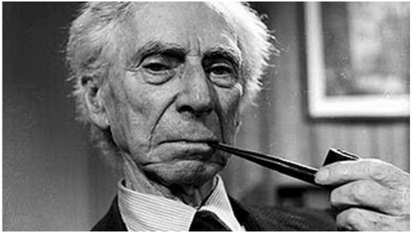


- Observed dependence: Possible causal models?
- **Causal Markov Condition**: statistical dependence $\Longrightarrow$ causal connectedness
- Observed *conditional independence:* $X \perp\!\!\!\perp Y \mid Z$
- **Faithfulness assumption**: statistical independence $\Longrightarrow$ no causal connectedness
- *Markov equivalence*: Cannot distinguish graphs
- Suppose $V \perp\!\!\!\perp W$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp V|Z$, $X \perp\!\!\!\perp W|Z$, $Y \perp\!\!\!\perp V|Z$, $Y \perp\!\!\!\perp W|Z$, but all others are dependent ($X \not\perp\!\!\!\perp Y$, $X \not\perp\!\!\!\perp W$, $Y \not\perp\!\!\!\perp W$, etc.)
    - Which causal models explain these? (assuming no unobserved variables = latents)
    - Only possible model (accounting for latents)
    - Causal inference methods use (some of) these assumptions, and others (e.g., time order)

Assumptions for causal inference

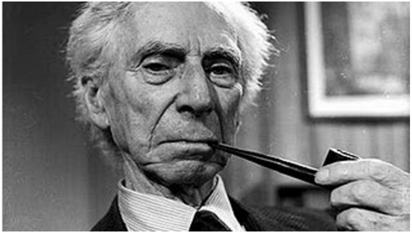
Assumptions (make causality great again)

- Bertrand Russell (1913): "The law of causality, I believe, is a relic of bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."



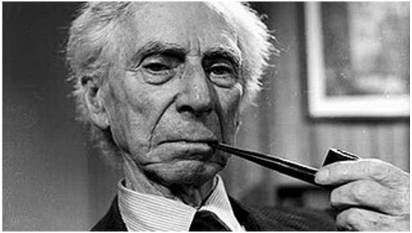
Assumptions (make causality great again)

- Bertrand Russell (1913): "The law of causality, I believe, is a relic of bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."
- Assumptions are not a limitation, but a *tool* to arrive at causal conclusions



Assumptions (make causality great again)

- Bertrand Russell (1913): "The law of causality, I believe, is a relic of bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."
- Assumptions are not a limitation, but a *tool* to arrive at causal conclusions
- Stating scientific results based on a clearly defined set of assumptions provides a basis for arguing whether a particular assumption is justified or not



Assumptions for causal inference

- Causal Sufficiency

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles
- Selection bias

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles
- Selection bias
- Time series:

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles
- Selection bias
- Time series:
    - Time order

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles
- Selection bias
- Time series:
    - Time order
    - Stationarity

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles
- Selection bias
- Time series:
  - Time order
  - Stationarity
  - Instantaneous effects

Assumptions for causal inference

- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles
- Selection bias
- Time series:
  - Time order
  - Stationarity
  - Instantaneous effects
- Assumptions on dependency types (linearity etc) and distributions ($\rightarrow$ Structural Causal Models)

Assumptions for causal inference

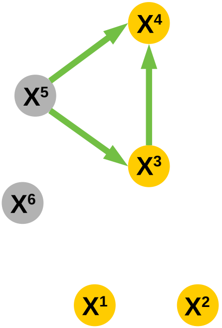
- Causal Sufficiency
- Causal Markov Condition
- Faithfulness
- Causal cycles
- Selection bias
- Time series:
  - Time order
  - Stationarity
  - Instantaneous effects
- Assumptions on dependency types (linearity etc) and distributions ($\rightarrow$ Structural Causal Models)
- Causal approaches typically utilize a subset (not all!) of these assumptions

Causal Sufficiency

Causal Sufficiency

Causal Sufficiency

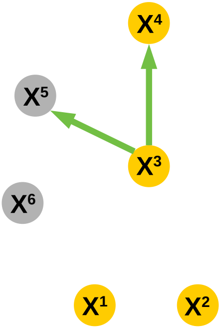
A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



Causal Sufficiency

Causal Sufficiency

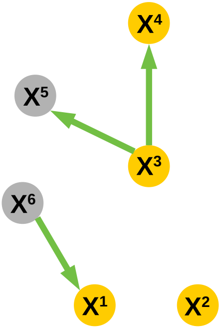
A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



Causal Sufficiency

Causal Sufficiency

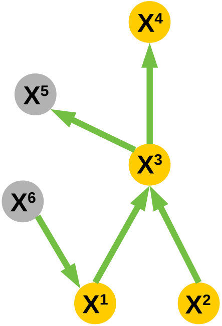
A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



Causal Sufficiency

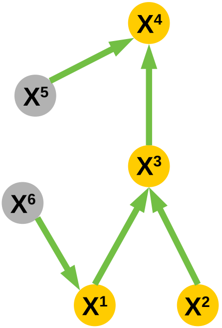
Causal Sufficiency

A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



Causal Sufficiency

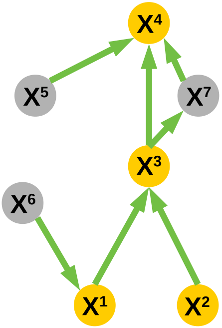
A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



Causal Sufficiency

Causal Sufficiency

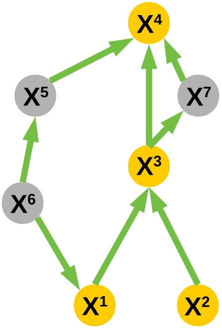
A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



Causal Sufficiency

Causal Sufficiency

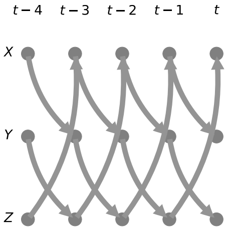
A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



Causal Sufficiency

Causal Sufficiency

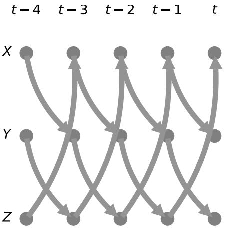
A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).

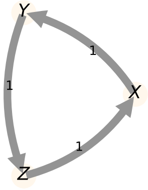


Causal Sufficiency

Causal Sufficiency

A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).

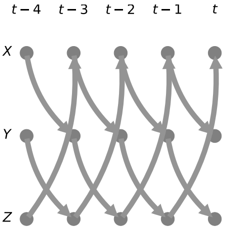


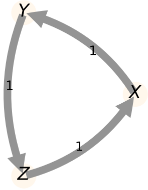


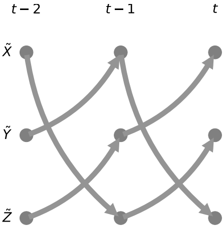
Causal Sufficiency

Causal Sufficiency

A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).



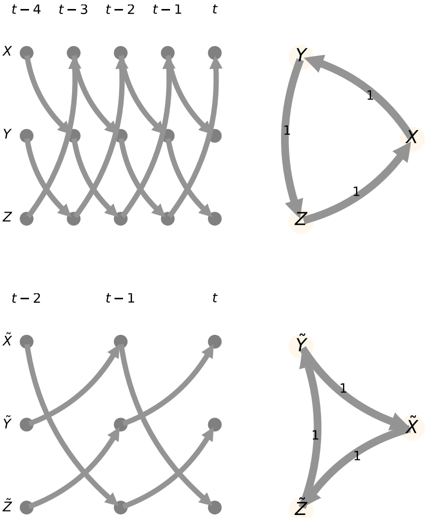




Causal Sufficiency

Causal Sufficiency

A set $W \subset \mathbf{X}$ of variables is *causally sufficient* for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W (or has the same value for all units in the population).

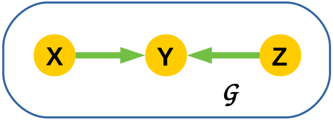


Causal Markov Condition

Pseudo-indeterministic causal graphs

A causal graph $\mathcal{G} = (\mathbf{X}, \mathcal{E})$ is *pseudo-indeterministic* for a population, if and only if $\mathcal{G}$ is not a deterministic causal structure for the population and there exists a causal graph $\mathcal{G}'$ for the population over a set of variables $\mathbf{X}'$ that properly includes $\mathbf{X}$ such that

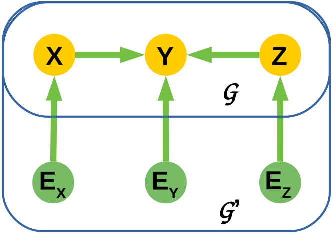
1. $\mathcal{G}'$ is a deterministic causal structure for the population
2. If X and Y are in $\mathbf{X}$, then $X \to Y$ is in $\mathcal{E}$ if and only if $X \to Y$ is in $\mathcal{E}'$
3. No variable in $\mathbf{X}$ is a cause of a variable in $\mathbf{X}' \setminus \mathbf{X}$
4. No variable in $\mathbf{X}' \setminus \mathbf{X}$ is a common cause of two variable in $\mathbf{X}$



Pseudo-indeterministic causal graphs

A causal graph $\mathcal{G} = (\mathbf{X}, \mathcal{E})$ is *pseudo-indeterministic* for a population, if and only if $\mathcal{G}$ is not a deterministic causal structure for the population and there exists a causal graph $\mathcal{G}'$ for the population over a set of variables $\mathbf{X}'$ that properly includes $\mathbf{X}$ such that

1. $\mathcal{G}'$ is a deterministic causal structure for the population
2. If X and Y are in $\mathbf{X}$, then $X \rightarrow Y$ is in $\mathcal{E}$ if and only if $X \rightarrow Y$ is in $\mathcal{E}'$
3. No variable in $\mathbf{X}$ is a cause of a variable in $\mathbf{X}' \setminus \mathbf{X}$
4. No variable in $\mathbf{X}' \setminus \mathbf{X}$ is a common cause of two variable in $\mathbf{X}$



Causal Markov Condition

The joint distribution $P(\mathbf{X})$ obeys *Global Markov condition* on $\mathcal{G}$ iff for all disjoint subsets X^i, X^j, $S \subset \mathbf{X}$,

$$X^i \bowtie X^j \mid S \implies X^i \perp\!\!\!\perp X^j \mid S \tag{1}$$

$$X^i \not\perp\!\!\!\perp X^j \mid S \implies X^i \not\bowtie X^j \mid S \quad \text{(contraposition)}$$

Separation implies independence and *dependence implies connectedness*.

D-separation (for DAGs) characterizes all and only the conditional independence relations that follow from satisfying the Markov condition.

- Probability distributions for pseudo-indeterministic systems (in which the exogenous variables are independently distributed) satisfy the Markov Condition

Causal Markov Condition

The joint distribution $P(\mathbf{X})$ obeys *Global Markov condition* on $\mathcal{G}$ iff for all disjoint subsets X^i, X^j, $S \subset \mathbf{X}$,

$$X^i \bowtie X^j \mid S \quad \Longrightarrow \quad X^i \perp\!\!\!\perp X^j \mid S \tag{1}$$

$$X^i \not\!\perp\!\!\!\perp X^j \mid S \quad \Longrightarrow \quad X^i \not\bowtie X^j \mid S \quad \text{(contraposition)}$$

Separation implies independence and *dependence implies connectedness*.

D-separation (for DAGs) characterizes all and only the conditional independence relations that follow from satisfying the Markov condition.

- Probability distributions for pseudo-indeterministic systems (in which the exogenous variables are independently distributed) satisfy the Markov Condition
- Macroscopic systems are (mostly) deterministic, but in practice we never have access to all of the causally relevant variables affecting a macroscopic system $\rightarrow$ if we include enough variables so that the excluded variables are probabilistically independent of one another, then our model will satisfy the Markov Condition

Causal Markov Condition

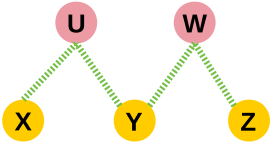
The Markov Condition holds for most systems, some arguable exeptions (`plato.stanford.edu`):

Artificial variables: Suppose X, Y, Z are independent and causally unrelated. Now define $U = X + Y$ and $W = Y + Z$. Then U and W will be dependent, even though there is no causal relation between them.

Causal Markov Condition

The Markov Condition holds for most systems, some arguable exeptions (`plato.stanford.edu`):

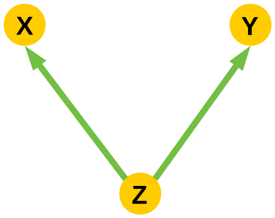
Artificial variables: Suppose X, Y, Z are independent and causally unrelated. Now define $U = X + Y$ and $W = Y + Z$. Then U and W will be dependent, even though there is no causal relation between them.



Causal Markov Condition

The Markov Condition holds for most systems, some arguable exeptions (`plato.stanford.edu`):

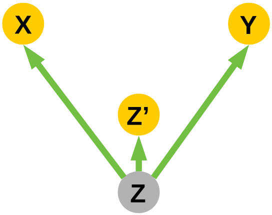
Coarse grained variables: Suppose X, Y, Z are continuous variables, Z is a common cause of X and Y, and neither X nor Y causes the other. Suppose we replace Z with a coarser variable, Z'. Then we would not expect Z' to screen X off from Y. The value of X may well contain information about the value of Z beyond what is given by Z', and this may affect the probability of Y.



Causal Markov Condition

The Markov Condition holds for most systems, some arguable exeptions (`plato.stanford.edu`):

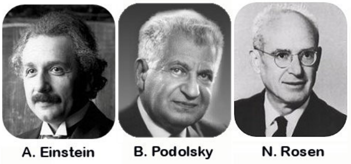
Coarse grained variables: Suppose X, Y, Z are continuous variables, Z is a common cause of X and Y, and neither X nor Y causes the other. Suppose we replace Z with a coarser variable, Z'. Then we would not expect Z' to screen X off from Y. The value of X may well contain information about the value of Z beyond what is given by Z', and this may affect the probability of Y.



Causal Markov Condition

The Markov Condition holds for most systems, some arguable exeptions (`plato.stanford.edu`):

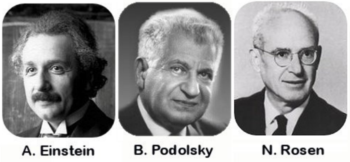
Quantum systems: EPR (Einstein-Podolsky-Rosen) set-up: two singlet particles are perfectly anti-correlated even if they are sufficiently far away from each other that it is impossible for one outcome to causally influence the other. Either the Causal Markov condition or locality principle must be wrong.



Causal Markov Condition

The Markov Condition holds for most systems, some arguable exeptions (`plato.stanford.edu`):

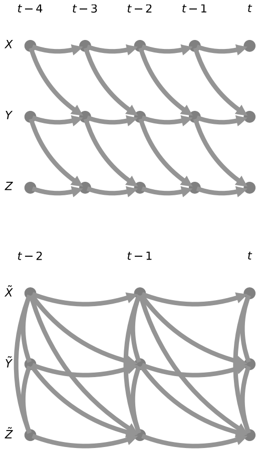
In most cases is the Markov Condition violated only for the distribution of the *observed* variables → more a violation of *Causal Sufficiency*



Causal Markov Condition

The Markov Condition holds for most systems, some arguable exeptions (`plato.stanford.edu`):

For example, also if time series are aggregated



Causal Markov Condition

Spirtes, P., C. Glymour, and R. Scheines. 2000. Causation, Prediction, and Search:

The basis for the Causal Markov Condition is, first, that it is necessarily true of populations of structurally alike pseudo-indeterministic systems whose exogenous variables are distributed independently, and second, it is supported by almost all of our experience with systems that can be put through repetitive processes and whose fundamental propensities can be tested. Any persuasive case against the Condition would have to exhibit macroscopic systems for which it fails and give some powerful reason why we should think the macroscopic natural and social systems for which we wish causal explanations also fail to satisfy the condition. It seems to us that no such case has been made.

Faithfulness

Faithfulness Condition

Definition: The joint distribution $P(\mathbf{X})$ is *faithful* to $\mathcal{G}$ iff every conditional independence relation true in $P(\mathbf{X})$ is entailed by the Causal Markov Condition applied to $\mathcal{G}$.

Theorem: Faithfulness holds iff for all disjoint subsets X^i, X^j, S,

$$X^i \perp\!\!\!\perp X^j \mid S \implies X^i \bowtie X^j \mid S \tag{2}$$

$$X^i \not\bowtie X^j \mid S \implies X^i \not\perp\!\!\!\perp X^j \mid S \quad \text{(contraposition)}$$

Conditional independence implies separation and *connectedness implies dependence*.

- Faithfulness is violated if we observe a statistical independence relation that are not entailed by the Markov condition

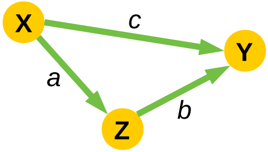
Faithfulness Violations

Counteracting mechanisms: For $c = -ab$ we have $X \perp\!\!\!\perp Y$ even though they are connected in the graph:

$$\begin{aligned} X &= \eta^X \\ Z &= aX + \eta^Y \\ Y &= bZ + cX + \eta^Z\,. \end{aligned} \tag{3}$$

In linear models coefficient values form real space and the set of points in this space that create vanishing partial correlations not implied by the Causal Markov Condition have Lebesgue measure zero

But: For *finite* sample sizes such cases may be more common!



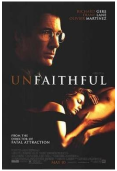
Faithfulness Violations

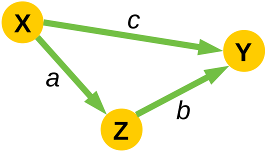
Counteracting mechanisms: For $c = -ab$ we have $X \perp\!\!\!\perp Y$ even though they are connected in the graph:

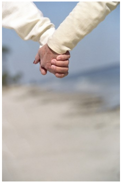
$$\begin{aligned} X &= \eta^X \\ Z &= aX + \eta^Y \\ Y &= bZ + cX + \eta^Z\,. \end{aligned} \tag{3}$$

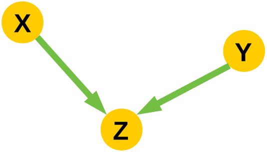
In linear models coefficient values form real space and the set of points in this space that create vanishing partial correlations not implied by the Causal Markov Condition have Lebesgue measure zero

But: For *finite* sample sizes such cases may be more common!







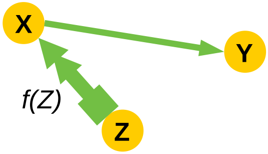


Faithfulness Violations

Determinism: Here we have $I(X;Y|Z) = I(f(Z);Y|Z) = 0$ (since $H(f(Z)|Z) = 0$) implying $X \perp\!\!\!\perp Y \mid Z$ even though Y depends on X in the model:

$$Z = \eta^Z$$
$$X = f(Z)$$
$$Y = h(X) + \eta^Y$$

One can argue that the complexity of the underlying processes will almost always imply that variables do not deterministically depend on their parents, but some unresolved processes constitute 'intrinsic' or 'dynamical' noise.

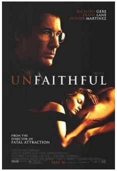


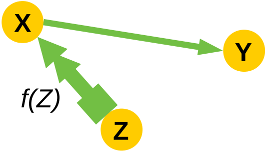
Faithfulness Violations

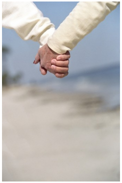
Determinism: Here we have $I(X;Y|Z) = I(f(Z);Y|Z) = 0$ (since $H(f(Z)|Z) = 0$) implying $X \perp\!\!\!\perp Y \mid Z$ even though Y depends on X in the model:

$$Z = \eta^Z$$
$$X = f(Z)$$
$$Y = h(X) + \eta^Y$$

One can argue that the complexity of the underlying processes will almost always imply that variables do not deterministically depend on their parents, but some unresolved processes constitute 'intrinsic' or 'dynamical' noise.







Faithfulness Violations

Purely non-pairwise dependencies:

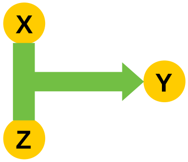
Here $I(X;Y) = I(Z;Y) = 0$ even though both are connected to Y:

$$Y = X \oplus Z + \eta^Y$$

X, Z binary random variables with

$$P(X = 0,1) = P(Z = 0,1) = 0.5$$

But if $P(X) \neq P(Z)$, Faithfulness is not violated anymore! Another way out: Hypergraphs.



Faithfulness Violations

Purely non-pairwise dependencies:

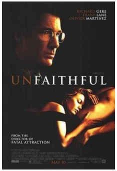
Here $I(X;Y) = I(Z;Y) = 0$ even though both are connected to Y:

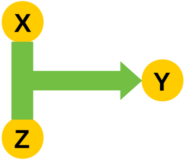
$$Y = X \oplus Z + \eta^Y$$

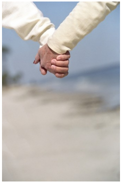
X, Z binary random variables with

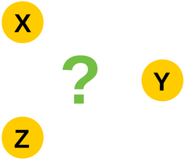
$$P(X = 0,1) = P(Z = 0,1) = 0.5$$

But if $P(X) \neq P(Z)$, Faithfulness is not violated anymore! Another way out: Hypergraphs.







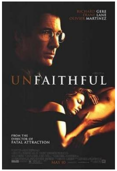


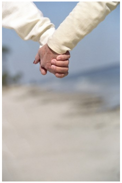
Faithfulness Violations

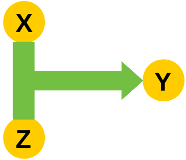
→ **Non-faithful distributions arise from a pathological fine-tuning of dependence parameters**

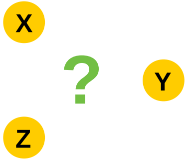
But (`plato.stanford.edu`):

- 'no fine-tuning' condition seems implausible as a metaphysical or conceptual constraint upon the connection between causation and probabilities (e.g., genes suppress each other)
- Faithfulness is a *methodological principle rather than a metaphysical principle*: it is preferable to postulate a causal structure that implies the independencies rather than one that is merely consistent with independence









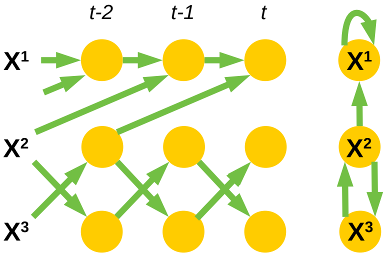
Causal cycles

Causal cycles

- Here we assume *directed acyclic graphs* (DAGs) as underlying causal structure

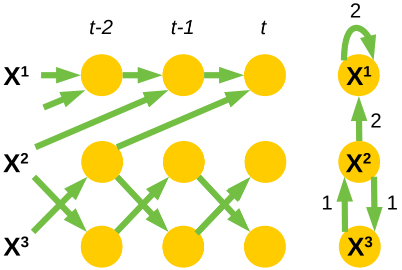
Causal cycles

- Here we assume *directed acyclic graphs* (DAGs) as underlying causal structure
- Cycles can often be resolved in time-series graphs



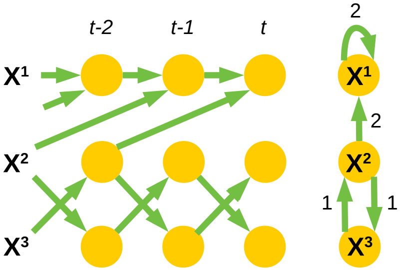
Causal cycles

- Here we assume *directed acyclic graphs* (DAGs) as underlying causal structure
- Cycles can often be resolved in time-series graphs



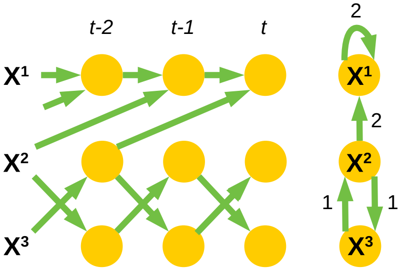
Causal cycles

- Here we assume *directed acyclic graphs* (DAGs) as underlying causal structure
- Cycles can often be resolved in time-series graphs
- One can argue that since *the speed of light is finite*, effects cannot cause causes and hence causal graphs among time-resolved variables are always acyclic



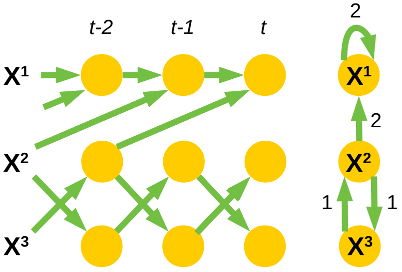
Causal cycles

- Here we assume *directed acyclic graphs* (DAGs) as underlying causal structure
- Cycles can often be resolved in time-series graphs
- One can argue that since *the speed of light is finite*, effects cannot cause causes and hence causal graphs among time-resolved variables are always acyclic
- But what if cannot resolve time? One can think of causal graphs for systems in *equilibrium* and use the equilibrium distribution for causal discovery



Causal cycles

- Here we assume *directed acyclic graphs* (DAGs) as underlying causal structure
- Cycles can often be resolved in time-series graphs
- One can argue that since *the speed of light is finite*, effects cannot cause causes and hence causal graphs among time-resolved variables are always acyclic
- But what if cannot resolve time? One can think of causal graphs for systems in *equilibrium* and use the equilibrium distribution for causal discovery
- Difficult topic not further discussed here



Selection bias

Selection bias

$\rightarrow$ covered in next lecture

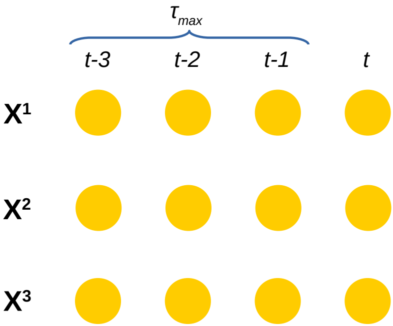
Further assumptions for time series graphs

Further assumptions for time series graphs

- Time order

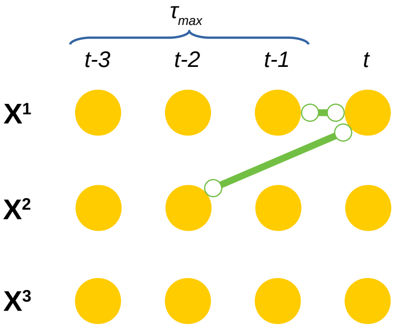
Further assumptions for time series graphs

- Time order



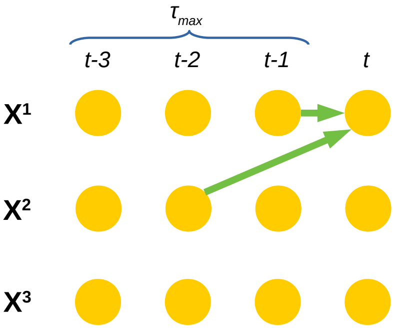
Further assumptions for time series graphs

- Time order



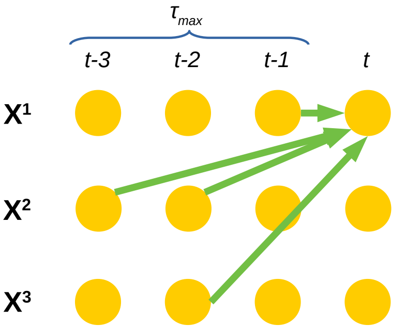
Further assumptions for time series graphs

- Time order



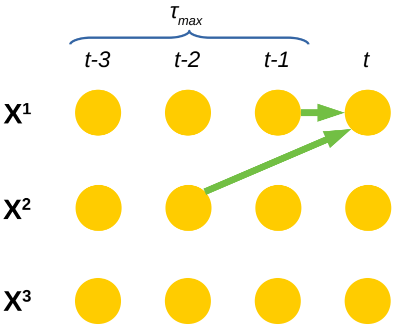
Further assumptions for time series graphs

- Time order



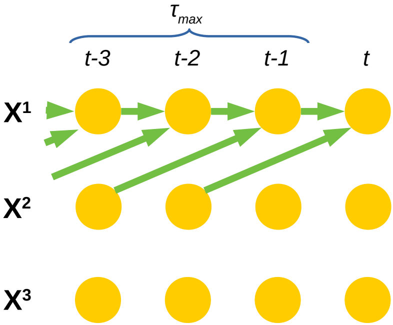
Further assumptions for time series graphs

- Time order



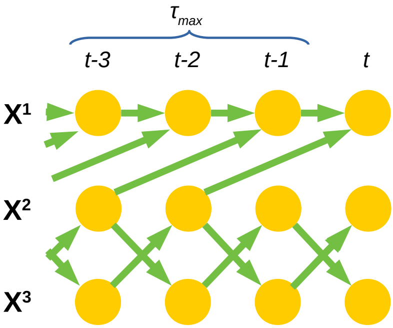
Further assumptions for time series graphs

- Time order



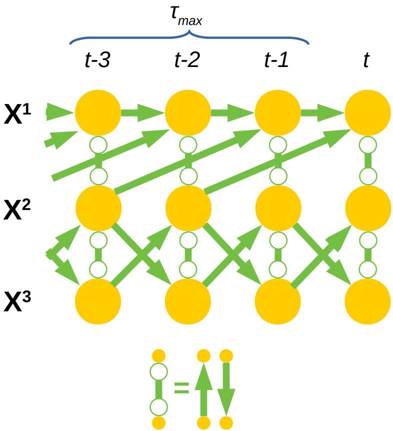
Further assumptions for time series graphs

- Time order



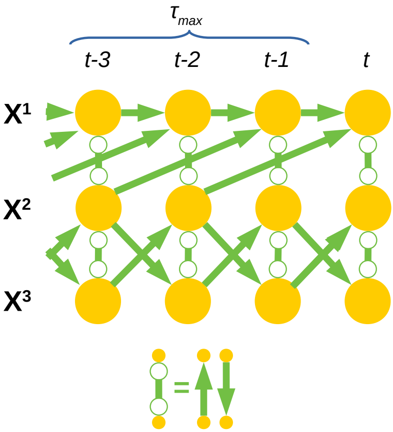
Further assumptions for time series graphs

- Time order
- No *contemporaneous effects*: Then causal graphs can be efficiently estimated using simplifications
  $\rightarrow$ lecture on PCMCI and Granger Causality



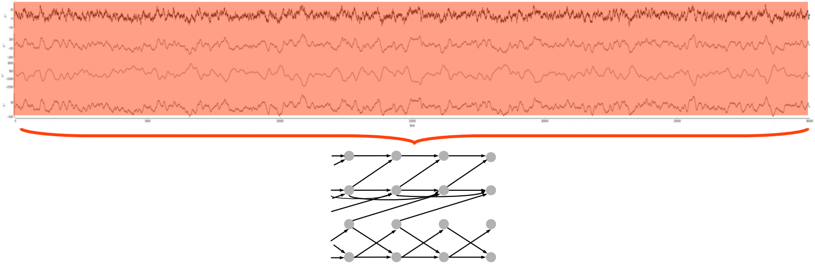
Further assumptions for time series graphs

- Time order
- No *contemporaneous effects*: Then causal graphs can be efficiently estimated using simplifications
  $\rightarrow$ lecture on PCMCI and Granger Causality
- Stationarity



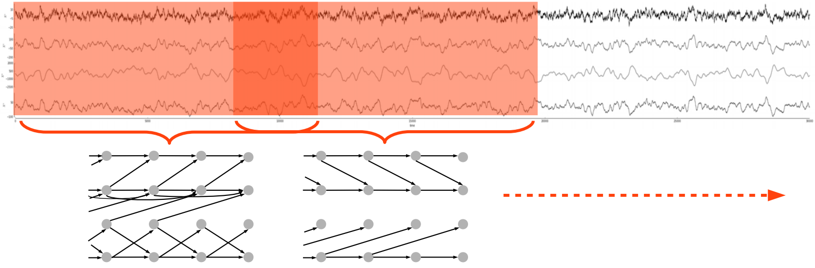
Stationarity

Definition: The time series process $\mathbf{X}_t$ with graph(s) $\mathcal{G}_t$ is called *causally stationary over a time index set* $\mathcal{T}$ iff $\mathcal{G}_t = \mathcal{G}_s$ for all $t, s \in \mathcal{T}$.



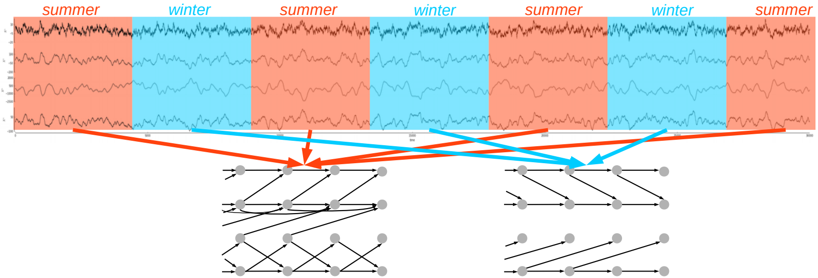
Stationarity

Definition: The time series process $\mathbf{X}_t$ with graph(s) $\mathcal{G}_t$ is called *causally stationary over a time index set* $\mathcal{T}$ iff $\mathcal{G}_t = \mathcal{G}_s$ for all $t, s \in \mathcal{T}$.



Stationarity

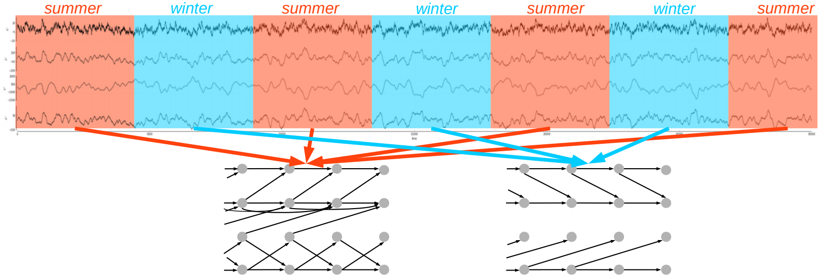
Definition: The time series process $\mathbf{X}_t$ with graph(s) $\mathcal{G}_t$ is called *causally stationary over a time index set* $\mathcal{T}$ iff $\mathcal{G}_t = \mathcal{G}_s$ for all $t, s \in \mathcal{T}$.



Stationarity

Definition: The time series process $\mathbf{X}_t$ with graph(s) $\mathcal{G}_t$ is called *causally stationary over a time index set* $\mathcal{T}$ iff $\mathcal{G}_t = \mathcal{G}_s$ for all $t, s \in \mathcal{T}$.

Weaker form than the common definition of stationarity in mean, variance, spectral properties, or of the value of individual coefficients in a linear model.



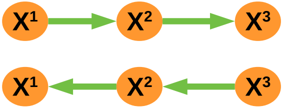
Assumptions on dependencies and distributions

Assumptions on dependencies and distributions

- Conditional-independence alone can only recover causal structures up to a Markov equivalence class

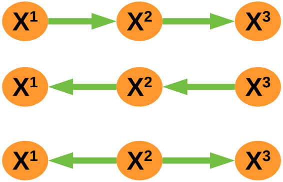
Assumptions on dependencies and distributions

- Conditional-independence alone can only recover causal structures up to a Markov equivalence class



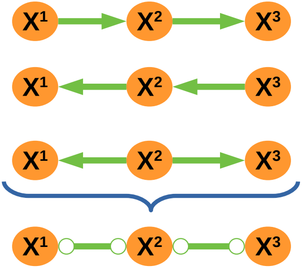
Assumptions on dependencies and distributions

- Conditional-independence alone can only recover causal structures up to a Markov equivalence class



Assumptions on dependencies and distributions

- Conditional-independence alone can only recover causal structures up to a Markov equivalence class



Assumptions on dependencies and distributions

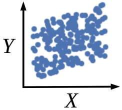
- Conditional-independence alone can only recover causal structures up to a Markov equivalence class
- Especially useless for problems with just two variables

Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:

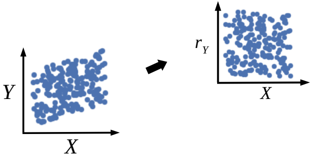
Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:
- Regression of Y on X yields residual r_Y. Suppose we measure that r_Y and X are uncorrelated, but r_X and Y are correlated



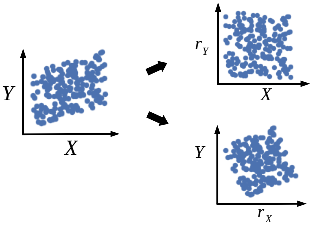
Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:
- Regression of Y on X yields residual r_Y. Suppose we measure that r_Y and X are uncorrelated, but r_X and Y are correlated



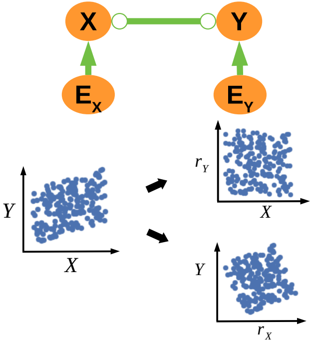
Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:
- Regression of Y on X yields residual r_Y. Suppose we measure that r_Y and X are uncorrelated, but r_X and Y are correlated



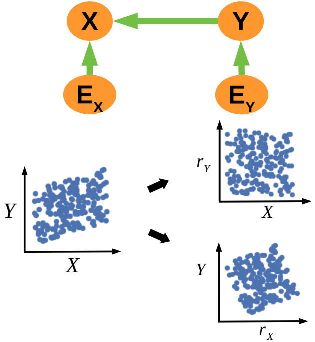
Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:
- Regression of Y on X yields residual r_Y. Suppose we measure that r_Y and X are uncorrelated, but r_X and Y are correlated
- Exploit the conditional independence constraints on the graph including error terms (as estimated from regression)



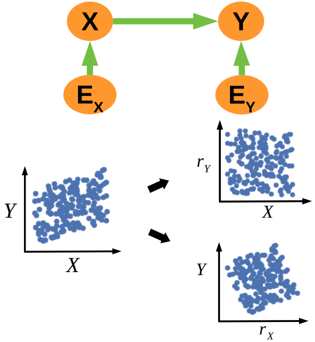
Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:
- Regression of Y on X yields residual r_Y. Suppose we measure that r_Y and X are uncorrelated, but r_X and Y are correlated
- Exploit the conditional independence constraints on the graph including error terms (as estimated from regression)
- Asymmetry allows to conclude on causal direction



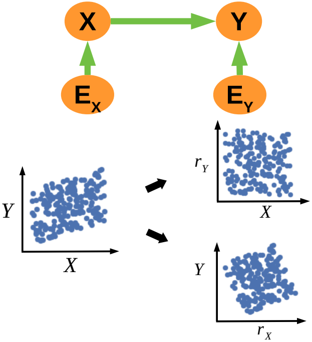
Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:
- Regression of Y on X yields residual r_Y. Suppose we measure that r_Y and X are uncorrelated, but r_X and Y are correlated
- Exploit the conditional independence constraints on the graph including error terms (as estimated from regression)
- Asymmetry allows to conclude on causal direction



Assumptions on dependencies and distributions

- Causal inference can also utilize assumptions on dependencies and distributions: E.g., linear additive models with non-gaussian noise:
- Regression of Y on X yields residual r_Y. Suppose we measure that r_Y and X are uncorrelated, but r_X and Y are correlated
- Exploit the conditional independence constraints on the graph including error terms (as estimated from regression)
- Asymmetry allows to conclude on causal direction
- Covered in lecture on *Structural Causal Models*



References i

[Runge et al., 2019b, Runge et al., 2019a, Camps-Valls et al., 2019, Di Capua et al., 2019, Krich et al., 2020, Trifunov et al., 2019a, Trifunov et al., 2019b, Reimers et al., 2019, Runge, 2018b, Bollt et al., 2018, Runge, 2018a, Kretschmer et al., 2018, Tibau et al., 2018, Runge et al., 2018, Kretschmer et al., 2017, Kretschmer et al., 2016, Runge, 2015, Runge et al., 2015a, Runge et al., 2015b, Runge et al., 2014, Schleussner et al., 2014, Runge et al., 2012b, Runge et al., 2012a, Pompe and Runge, 2011]

Bollt, E. M., Sun, J., and Runge, J. (2018).
Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications.
Chaos An Interdiscip. J. Nonlinear Sci., 28(7):075201.

References ii

Camps-Valls, G., Sejdinovic, D., Runge, J., and Reichstein, M. (2019).
A perspective on gaussian processes for earth observation.
National Science Review, 6(4):616–618.

Di Capua, G., Kretschmer, M., Runge, J., Alessandri, A., Donner, R., van den Hurk, B., Vellore, R., Krishnan, R., and Coumou, D. (2019).
Long-lead statistical forecasts of the indian summer monsoon rainfall based on causal precursors.
Weather and Forecasting, 34(5):1377–1394.

References iii

Kretschmer, M., Cohen, J., Matthias, V., Runge, J., and Coumou, D. (2018).
The different stratospheric influence on cold-extremes in Eurasia and North America.
npj Climate and Atmospheric Sciences, 1(1):44.

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J. (2016).
Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation.
Journal of Climate, 29(11):4069–4081.

Kretschmer, M., Runge, J., and Coumou, D. (2017).
Early prediction of weak stratospheric polar vortex states using causal precursors.
Geophysical Research Letters, 44(16):8592–8600.

References iv

Krich, C., Runge, J., Miralles, D., Migliavacca, M., Perez-Priego, O., El-Madany, T., Carrara, A., and Mahecha, M. D. (2020).
Estimating causal networks in biosphere–atmosphere interaction with the pcmci approach.
Biogeosciences, pages 1033–1061.

Nowack, P. J., Runge, J., Eyring, V., and Haigh, H. D. (2019).
Causal teleconnection fingerprints for climate model evaluation.
Nature Communications, minor revision.

Pearl, J. (2000).
Causality: Models, Reasoning, and Inference.
Cambridge University Press, New York, NY.

Pearl, J. (2013).
Linear models: A useful microscope for causal analysis.
J. Causal Inference, 1(1):155–170.

Pearl, J., Glymour, M., and Jewell, N. P. (2016).
Causal inference in statistics: A Primer.
John Wiley & Sons.

Pearl, J. and Mackenzie, D. (2018).
The Book of Why: The New Science of Cause and Effect.
Basic books, New York.

Peters, J., Janzing, D., and Schölkopf, B. (2017).
Elements of causal inference: foundations and learning algorithms.
MIT Press, Cambridge, MA.

Pompe, B. and Runge, J. (2011).
Momentary information transfer as a coupling measure of time series.
Physical Review E, 83(5):1–12.

Reimers, C., Runge, J., and Denzler, J. (2019).
Using causal inference to globally understand black box predictors beyond saliency maps.
In *9th Int. Work. Clim. Informatics*, pages 1–4.

Runge, J. (2015).
Quantifying information transfer and mediation along causal pathways in complex systems.
Physical Review E, 92(6):062829.

Runge, J. (2018a).
Causal network reconstruction from time series: From theoretical assumptions to practical estimation.
Chaos An Interdiscip. J. Nonlinear Sci., 28(7):075310.

Runge, J. (2018b).
Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information.
In Storkey, A. & Perez-Cruz, F., editor, *Proc. 21st Int. Conf. Artif. Intell. Stat.* Playa Blanca, Lanzarote, Canary Islands: PMLR.

Runge, J. (2020).
Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets.
In Sontag, D. and Peters, J., editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence, UAI 2020, Toronto, Canada, 2019*. AUAI Press.

Runge, J., Balasis, G., Daglis, I. A., Papadimitriou, C., and Donner, R. V. (2018).
Common solar wind drivers behind magnetic storm–magnetospheric substorm dependency.
Scientific Reports, 8(1):16987.

References ix

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J. (2019a).
Inferring causation from time series in earth system sciences.
Nature Communications, 10(1):2553.

Runge, J., Donner, R. V., and Kurths, J. (2015a).
Optimal model-free prediction from multivariate time series.
Physical Review E, 91(5):052909.

References x

Runge, J., Heitzig, J., Marwan, N., and Kurths, J. (2012a).
Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy.
Physical Review E, 86(6):061121.

Runge, J., Heitzig, J., Petoukhov, V., and Kurths, J. (2012b).
Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy.
Phys. Rev. Lett., 108(25):258701.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. (2019b).
Detecting and quantifying causal associations in large nonlinear time series datasets.
Science Advances, eaau4996(5).

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M., and Kurths, J. (2015b).
Identifying causal gateways and mediators in complex spatio-temporal systems.
Nature Communications, 6:8502.

Runge, J., Petoukhov, V., and Kurths, J. (2014).
Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models.
Journal of Climate, 27(2):720–739.

Schleussner, C.-F., Runge, J., Lehmann, J., and Levermann, A. (2014).
The role of the North Atlantic overturning and deep ocean for multi-decadal global-mean-temperature variability.
Earth Syst. Dyn., 5(1):103–115.

Spirtes, P., Glymour, C., and Scheines, R. (2000).
Causation, Prediction, and Search.
MIT Press, Boston.

Tibau, X.-a., Requena-Mesa, C., Reimers, C., Denzler, J., Eyring, V., Reichstein, M., and Runge, J. (2018).
SupernoVAE : VAE based kernel PCA for analysis of spatio-temporal Earth data.
In *8th Int. Work. Clim. Informatics*, pages 1–4.

Trifunov, V. T., Shadaydeh, M., Runge, J., Eyring, V., Reichstein, M., and Denzler, J. (2019a).
Causal link estimation under hidden confounding in ecological time series.
In *9th Int. Work. Clim. Informatics*, pages 1–4.

Trifunov, V. T., Shadaydeh, M., Runge, J., Eyring, V., Reichstein, M., and Denzler, J. (2019b).
Nonlinear causal link estimation under hidden confounding with an application to time-series anomaly detection.
In *German Conference on Pattern Recognition (GCPR)*, pages 1–4.