Causal inference and discovery

with perspectives in Earth sciences

Prof. Dr. Jakob Runge

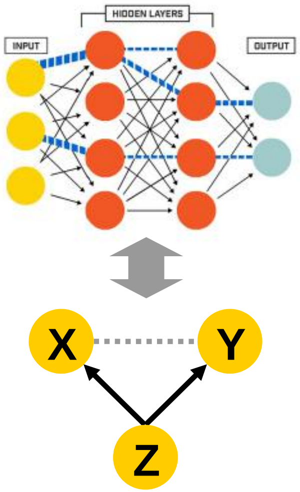
Group leader *Causal Inference* at
Institute of Data Science Jena

Chair of *Climate Informatics* at

Causal inference group at DLR and TU Berlin

Mission

Developing theory, methods, and accessible tools for causal inference in Earth system sciences and beyond



erc Causal Earth

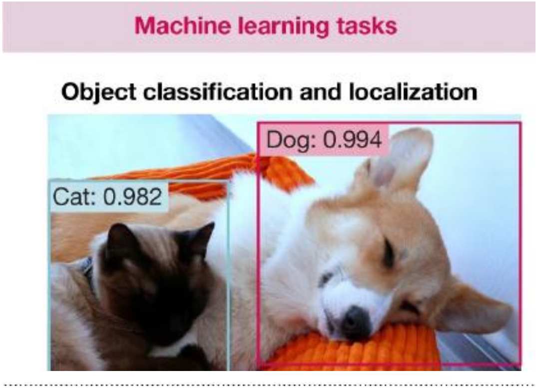
iMIRACLI

XAIDA

HELMHOLTZ AI

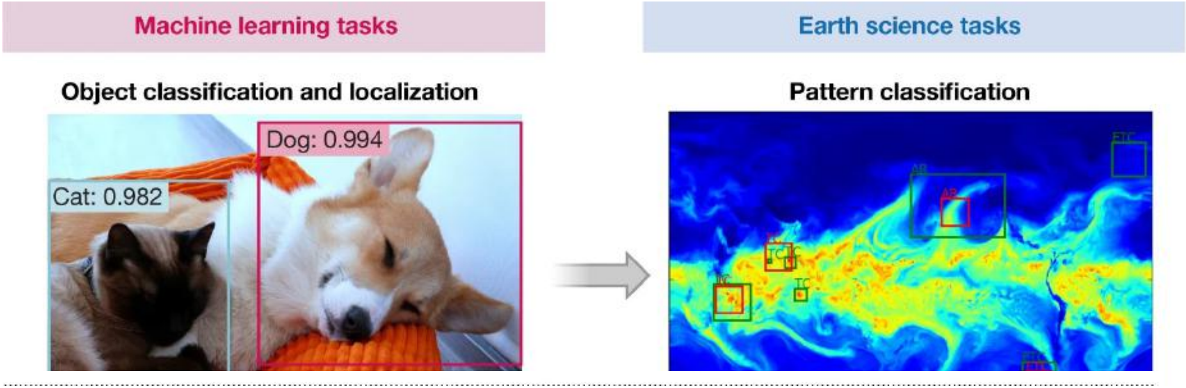
→ open positions: climateinformaticslab.com

Machine learning in science and engineering



Turing-Award 2018
LeCun, Hinton, Bengio

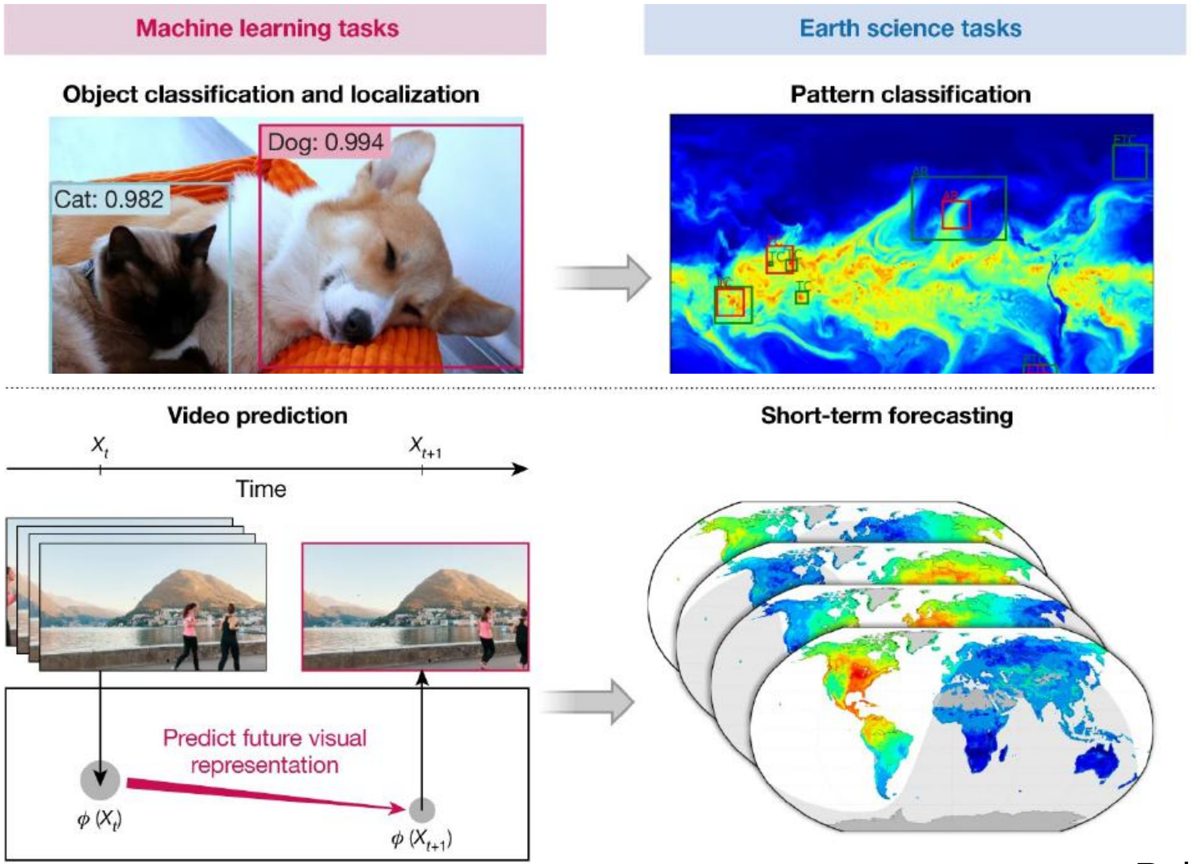
Machine learning in science and engineering



Turing-Award 2018
LeCun, Hinton, Bengio

Reichstein et al. 2019

Machine learning in science and engineering

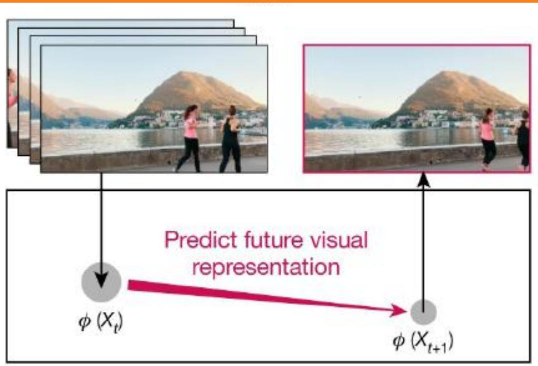


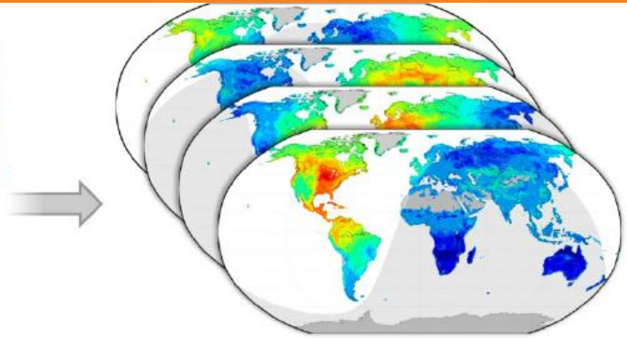
Turing-Award 2018
LeCun, Hinton, Bengio

Reichstein et al. 2019

Machine learning in science and engineering

ML can answer questions about
"What is?"
by learning statistical associations from complex data



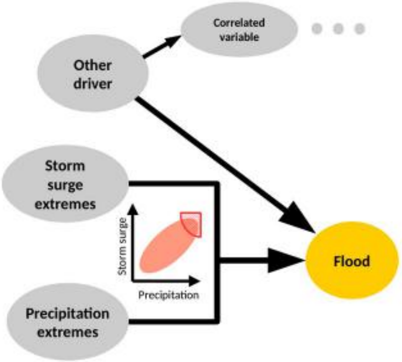


Turing-Award 2018
LeCun, Hinton, Bengio

Reichstein et al. 2019

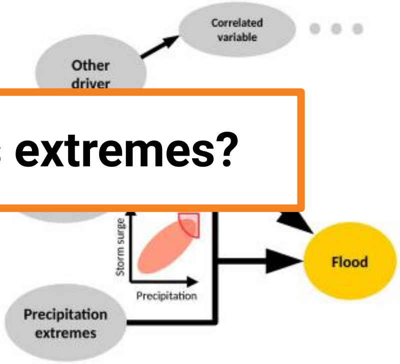
Some challenges in Earth sciences

Some challenges in Earth sciences



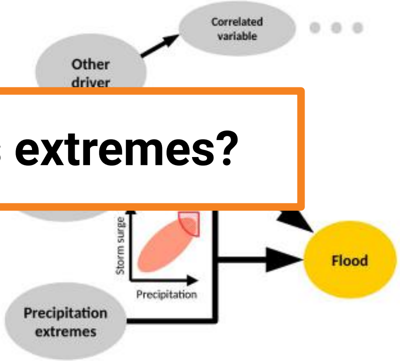
XAIDA

Some challenges in Earth sciences

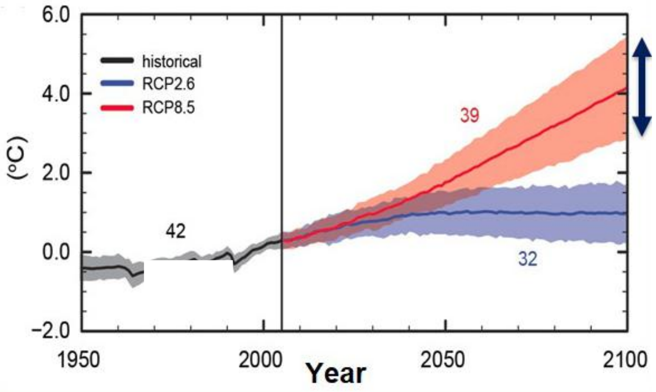


What causes extremes?

Some challenges in Earth sciences



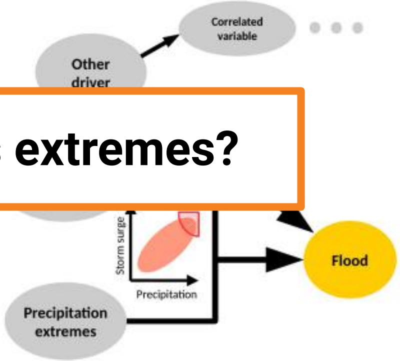
What causes extremes?

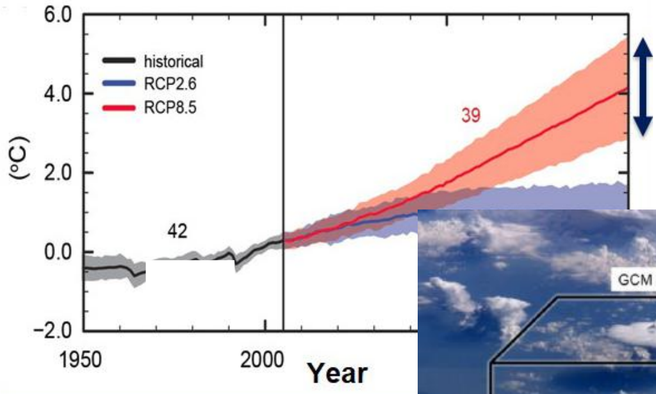


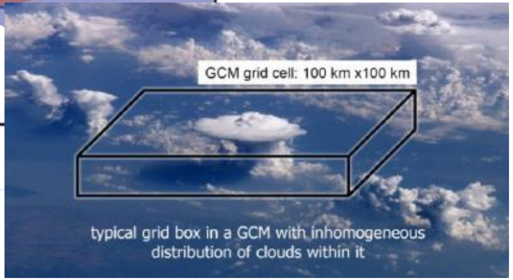
IPCC, Gentine

Some challenges in Earth sciences

What causes extremes?

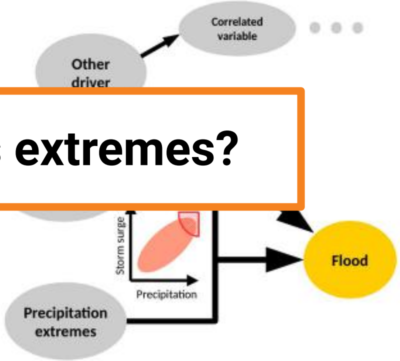




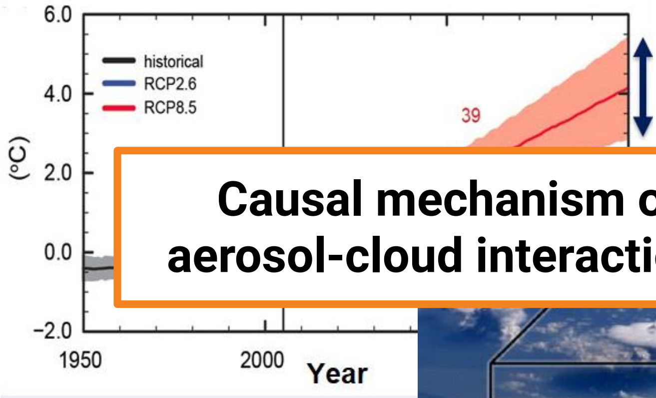


IPCC, Gentine

Some challenges in Earth sciences



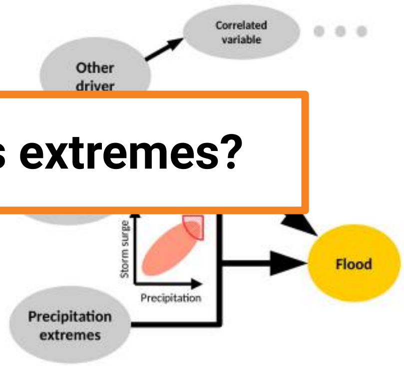
What causes extremes?



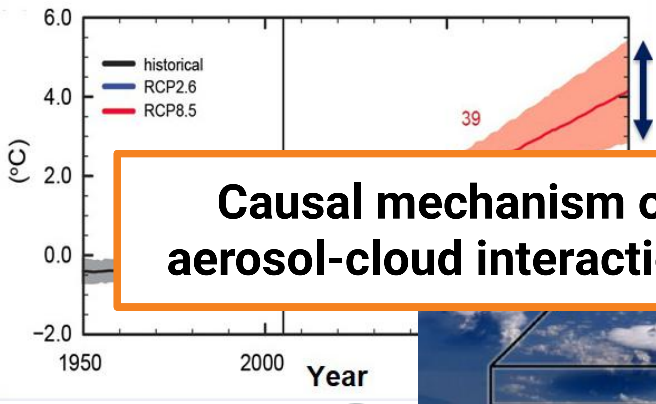
Causal mechanism of aerosol-cloud interactions

IPCC, Gentine

Some challenges in Earth sciences

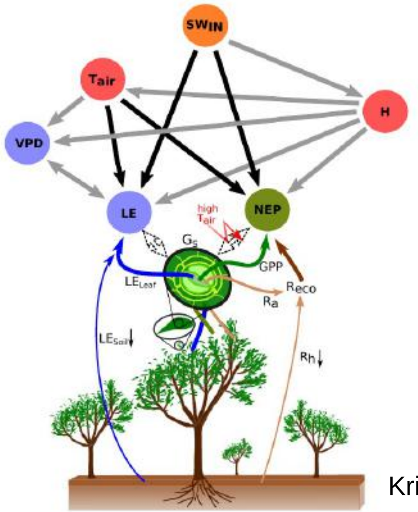


What causes extremes?



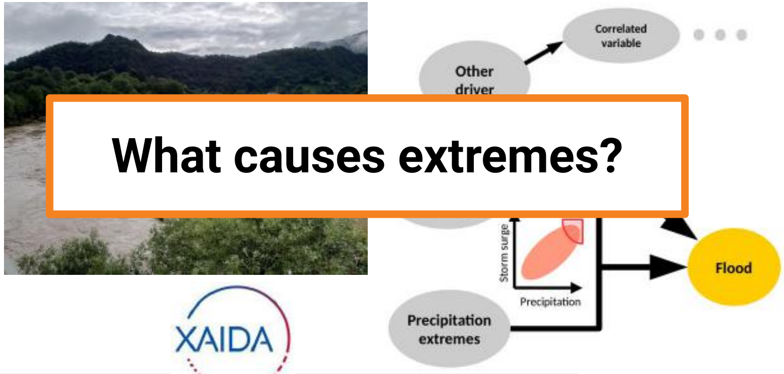
Causal mechanism of aerosol-cloud interactions

IPCC, Gentine

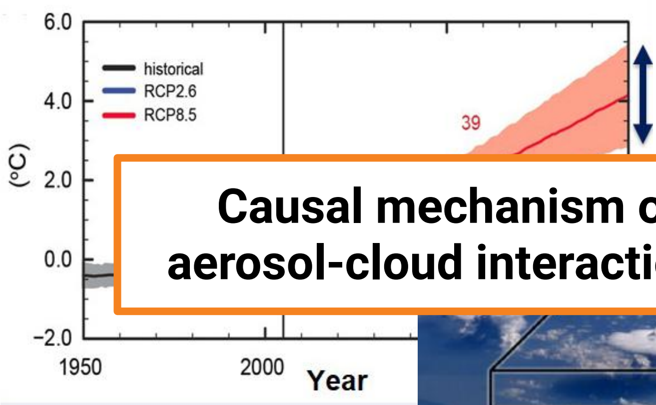


Krich, 2021

Some challenges in Earth sciences

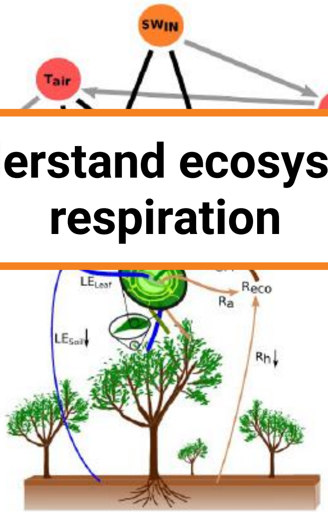


What causes extremes?



Causal mechanism of aerosol-cloud interactions

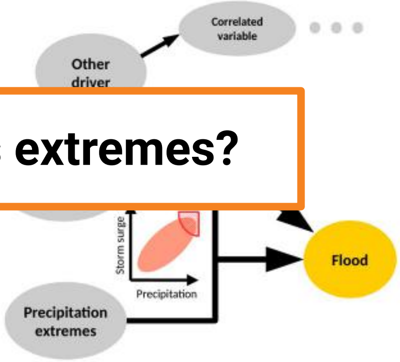
IPCC, Gentine



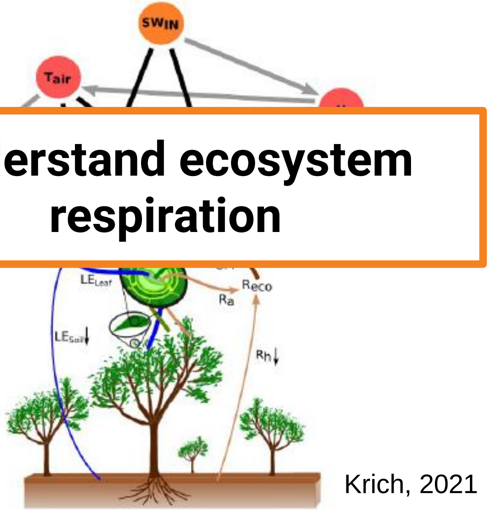
Understand ecosystem respiration

Krich, 2021

Some challenges in Earth sciences

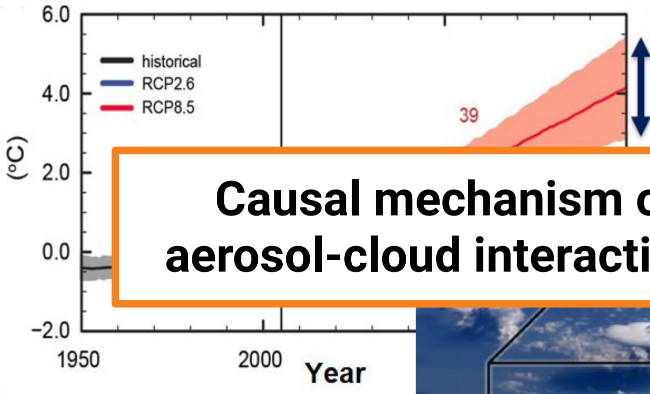


What causes extremes?



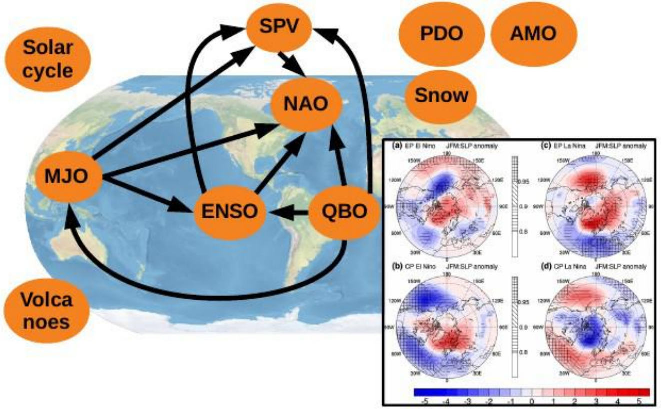
Understand ecosystem respiration

Krich, 2021

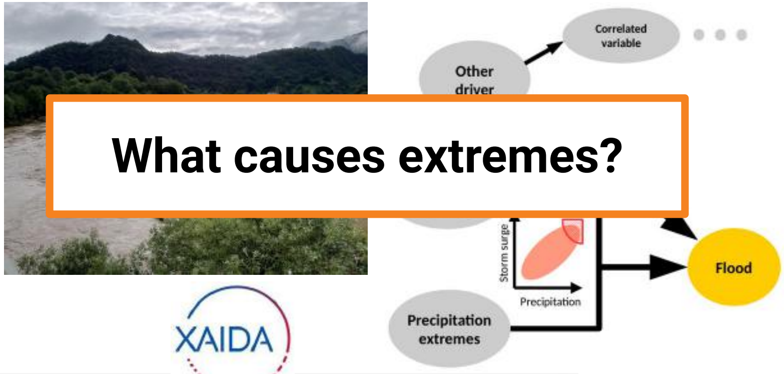


Causal mechanism of aerosol-cloud interactions

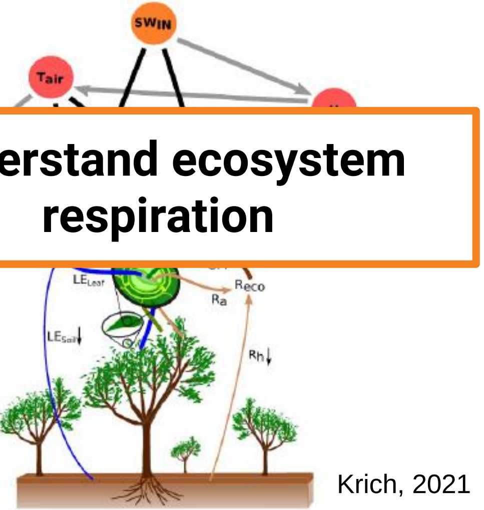
IPCC, Gentine



Some challenges in Earth sciences

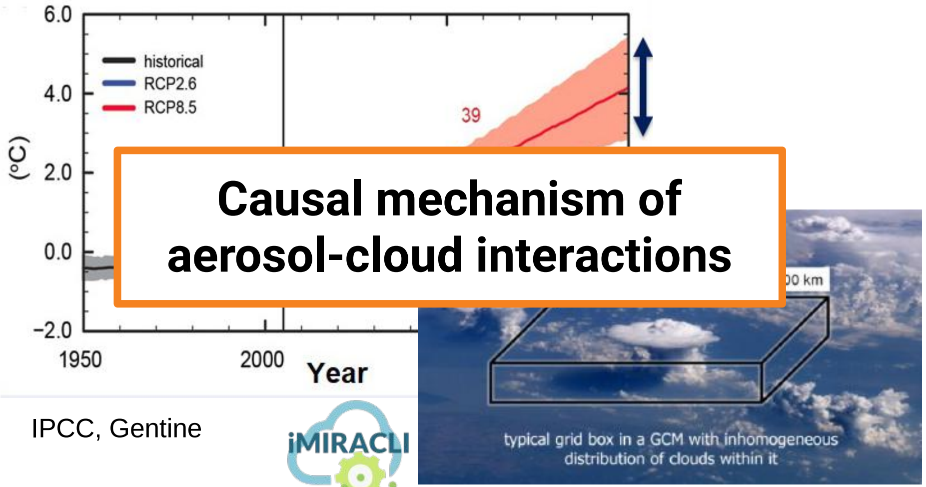


What causes extremes?



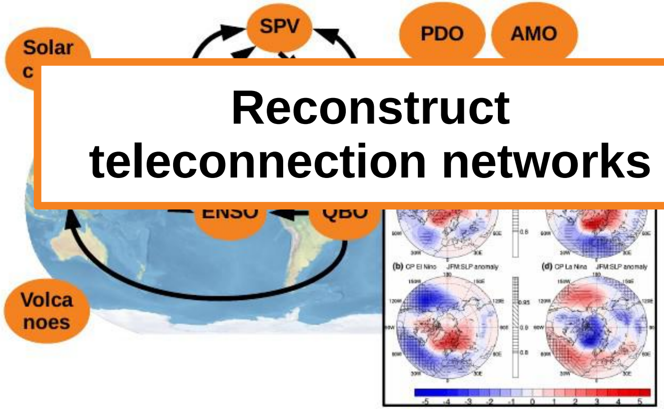
Understand ecosystem respiration

Krich, 2021



Causal mechanism of aerosol-cloud interactions

IPCC, Gentine

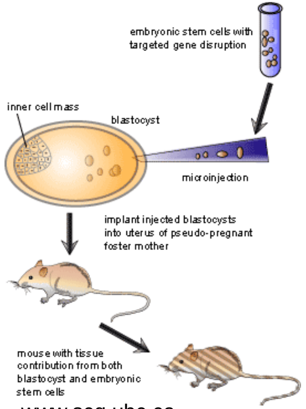


Reconstruct teleconnection networks

Answering causal questions

Answering causal questions

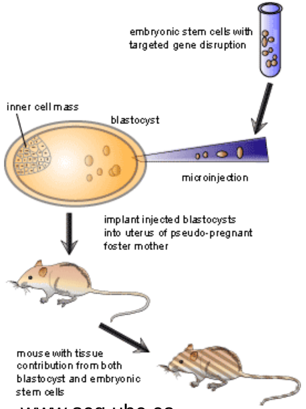
1. **Experimentation:** Randomized controlled trials, gene knockout experiments, etc



www.scq.ubc.ca

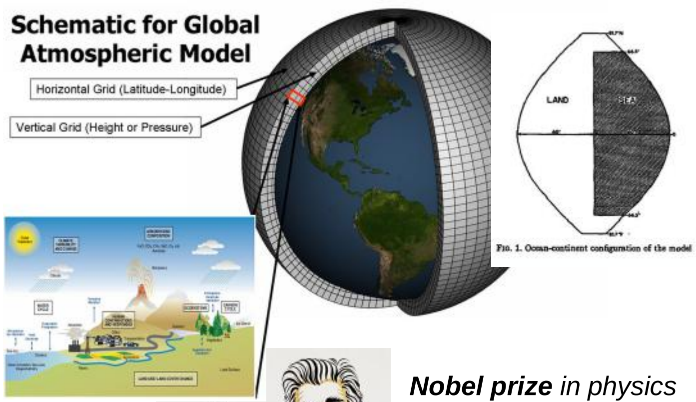
Answering causal questions

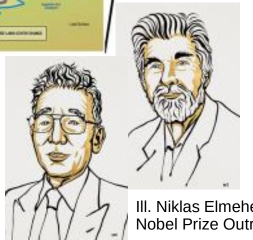
1. **Experimentation:** Randomized controlled trials, gene knockout experiments, etc.



www.scq.ubc.ca

2. **Simulation models:** Based on underlying physics (if available)

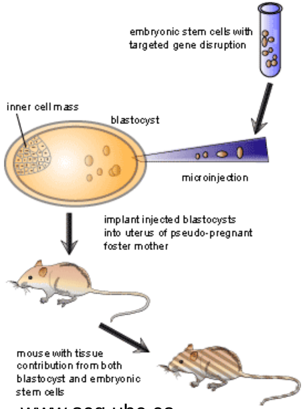




Ill. Niklas Elmehed © Nobel Prize Outreach

***Nobel prize** in physics 2021 for Klaus Hasselmann and Syukuro Manabe*

Answering causal questions



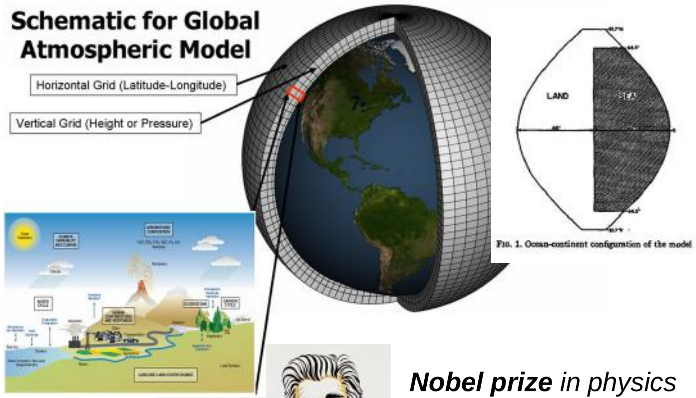
www.scq.ubc.ca

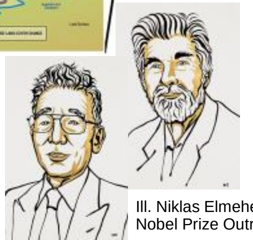
1. **Experimentation:** Randomized controlled trials, gene knockout experiments, etc.

Ethical issues, expensive, compliance, not possible, ...

2. **Simulation models:** Based on underlying physics (if available)

Expensive, time consuming, and only approximation of reality, ...

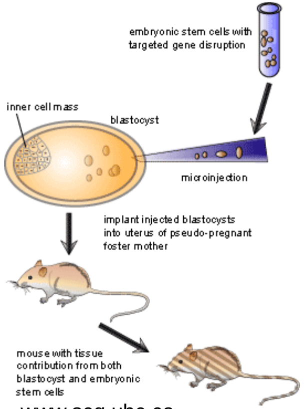




***Nobel prize** in physics 2021 for Klaus Hasselmann and Syukuro Manabe*

Ill. Niklas Elmehed © Nobel Prize Outreach

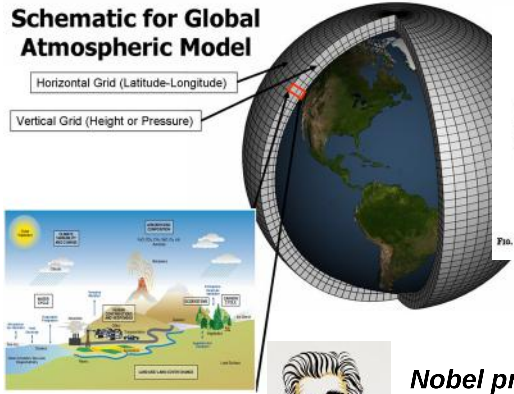
Answering causal questions



www.scq.ubc.ca

1. **Experimentation:** Randomized controlled trials, gene knockout experiments, etc.

Ethical issues, expensive, compliance, not possible, ...



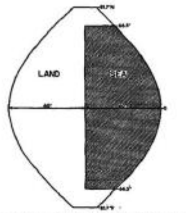
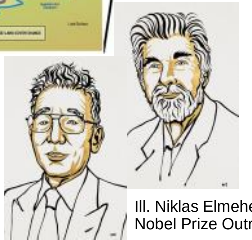


FIG. 1. Ocean-continent configuration of the model

2. **Simulation models:** Based on underlying physics (if available)

Expensive, time consuming, and only approximation of reality, ...



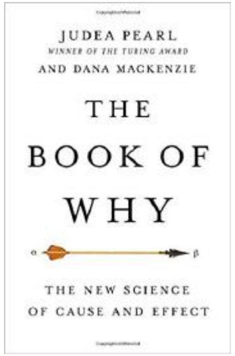
***Nobel prize** in physics 2021 for Klaus Hasselmann and Syukuro Manabe*

Ill. Niklas Elmehed © Nobel Prize Outreach

3. **Causal inference:** Based on observational <u>and/or</u> experimental data

Causal inference

Causal inference is a framework to answer causal questions from observational <u>and/or</u> experimental data.



J Pearl
Turing-Award 2011

Spirtes, Glymour, Scheines

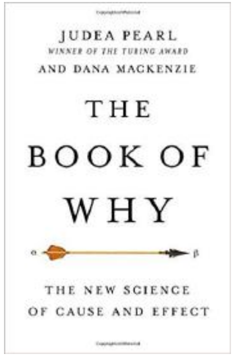
JD Angrist and GW Imbens
Nobel prize 2021

Ill. Niklas Elmehed © Nobel Prize Outreach

Causal inference

Causal inference is a framework to answer causal questions from observational <u>and/or</u> experimental data.

Pearl's hierarchy	**Activity**	**Questions**	**Examples**
3. Counterfactuals $P(y'_{x'} \mid x, y)$	Retrospection, Imagining, Understanding	Why? What If I had done...	Did climate change cause this extreme event? Was it aspirin that stopped my headache?
2. Intervention $P(y \mid do(x))$	Intervening	What if I do…?	What causes what in the data? Infer underlying mechanisms, future impacts of greenhouse gas emissions? What if I take aspirin?
1. Association $P(y \mid x)$	Seeing	What is? How are variables related?	Pattern recognition, correlation networks, statistical weather forecast (if the system is not changing)



J Pearl
Turing-Award 2011

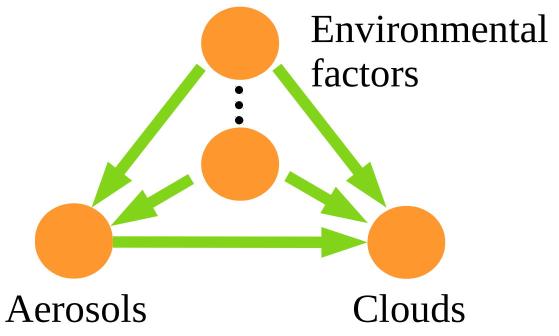
Spirtes,
Glymour,
Scheines

JD Angrist and GW Imbens
Nobel prize 2021

Ill. Niklas Elmehed © Nobel Prize Outreach

Causal inference framework

Pearl's causal inference framework assumes an underlying **structural causal model (SCM)** with an associated **graph**

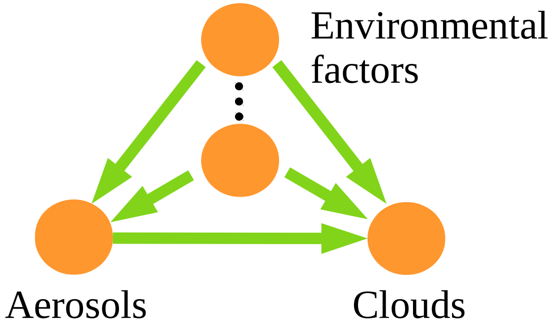


Causal inference framework

Pearl's causal inference framework assumes an underlying **structural causal model (SCM)** with an associated **graph**, for example,

$$X_{\mathrm{A}} := f_{\mathrm{A}}(X_{\mathrm{E}}, \eta_{\mathrm{A}})$$
$$X_{\mathrm{C}} := f_{\mathrm{C}}(X_{\mathrm{A}}, X_{\mathrm{E}}, \eta_{\mathrm{C}})$$
$$X_{\mathrm{E}} := f_{\mathrm{E}}(\eta_{\mathrm{E}})$$

where the graph is *acyclic* and noise terms are *independent*. This SCM entails a factorized distribution $P(\mathbf{X}) = P(X_{\mathrm{C}}|X_{\mathrm{A}}, X_{\mathrm{E}})P(X_{\mathrm{A}}|X_{\mathrm{E}})P(X_{\mathrm{E}})$

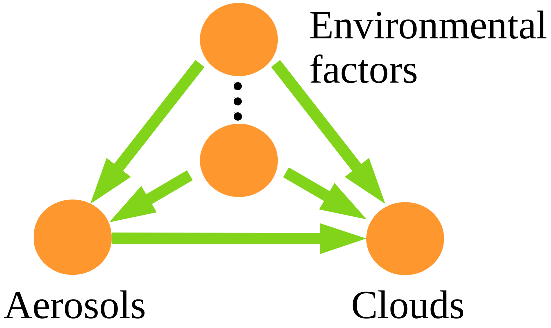


Causal inference framework

Pearl's causal inference framework assumes an underlying **structural causal model (SCM)** with an associated **graph**, for example,

$$X_{\mathrm{A}} := f_{\mathrm{A}}(X_{\mathrm{E}}, \eta_{\mathrm{A}})$$
$$X_{\mathrm{C}} := f_{\mathrm{C}}(X_{\mathrm{A}}, X_{\mathrm{E}}, \eta_{\mathrm{C}})$$
$$X_{\mathrm{E}} := f_{\mathrm{E}}(\eta_{\mathrm{E}})$$

where the graph is *acyclic* and noise terms are *independent*. This SCM entails a factorized distribution $P(\mathbf{X}) = P(X_{\mathrm{C}}|X_{\mathrm{A}}, X_{\mathrm{E}})P(X_{\mathrm{A}}|X_{\mathrm{E}})P(X_{\mathrm{E}})$

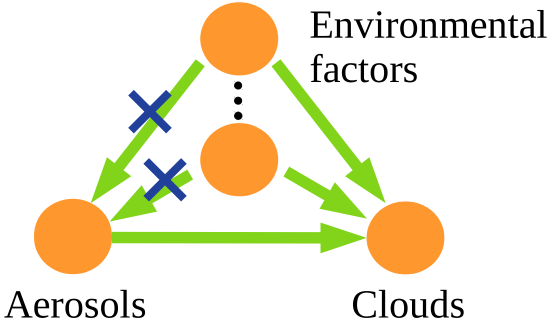


An **experiment / intervention** is then represented by the intervened SCM

$$X_{\mathrm{A}} := x'$$
$$X_{\mathrm{C}} := f_{\mathrm{C}}(X_{\mathrm{A}}, X_{\mathrm{E}}, \eta_{\mathrm{C}})$$
$$X_{\mathrm{E}} := f_{\mathrm{E}}(\eta_{\mathrm{E}})$$

Which results in the interventional distribution that defines **causal effects**:

$$P(X_{\mathrm{C}}|do(X_{\mathrm{A}} = x')) \neq P(X_{\mathrm{C}}|X_{\mathrm{A}} = x')$$



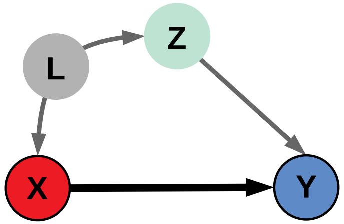
Causal inference framework

Two types of tasks:

Causal inference framework

<u>Two types of tasks:</u>

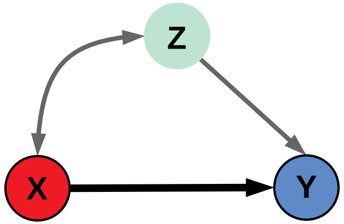
1. **Utilize qualitative causal knowledge** in form of directed acyclic graphs including observed and *unobserved / latent* variables



Causal inference framework

Two types of tasks:

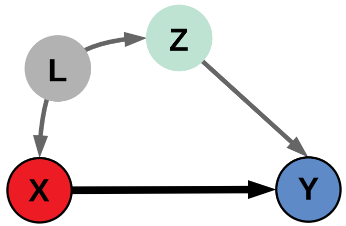
1. **Utilize qualitative causal knowledge** in form of directed acyclic graphs including observed and *unobserved / latent* variables

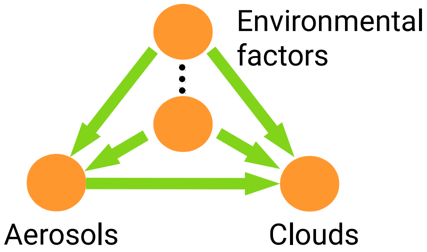


Causal inference framework

<u>Two types of tasks:</u>

1. **Utilize qualitative causal knowledge** in form of directed acyclic graphs including observed and *unobserved / latent* variables

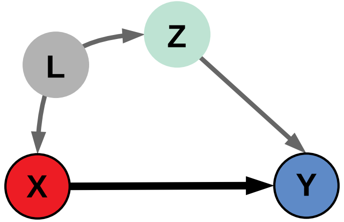


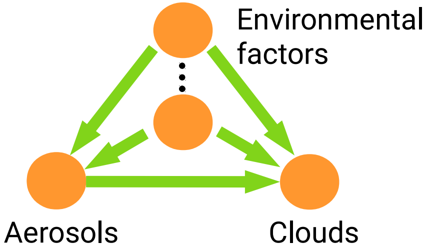


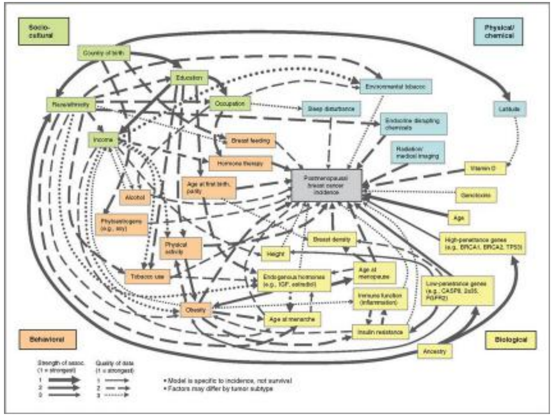
Causal inference framework

Two types of tasks:

1. **Utilize qualitative causal knowledge** in form of directed acyclic graphs including observed and *unobserved / latent* variables



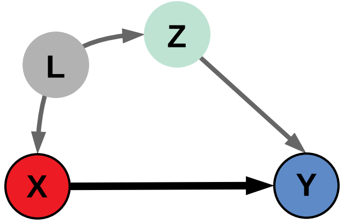




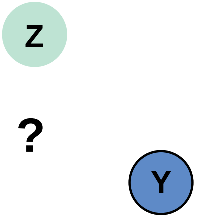
Causal inference framework

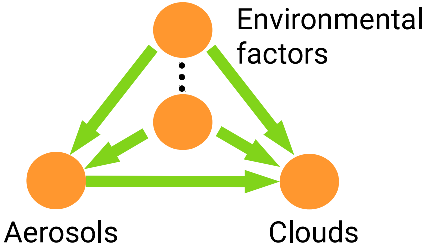
<u>Two types of tasks:</u>

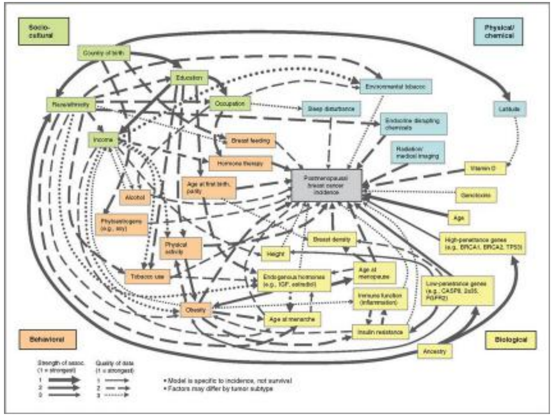
1. **Utilize qualitative causal knowledge** in form of directed acyclic graphs including observed and *unobserved / latent* variables



2. **Learn causal graphs** based on general assumptions



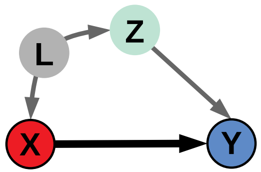




Causal inference: 1. Utilizing causal graphs

Causal effect estimation: Given causal graph and data, compute causal effect of intervention from observational distribution *P(**V**)*

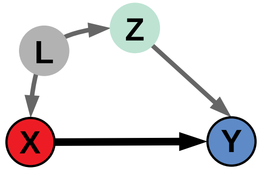
$$P\left(Y|do(X{=}x)\right) = \text{function of } P(\mathbf{V})$$

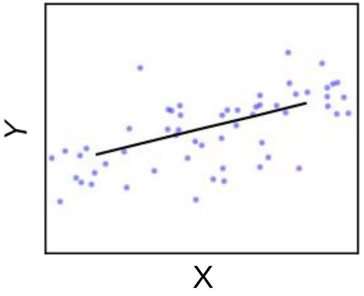


Causal inference: 1. Utilizing causal graphs

<u>Causal effect estimation:</u> Given causal graph and data, compute causal effect of intervention from observational distribution *P(**V**)*

$$P\left(Y|do(X{=}x)\right) = \text{function of } P(\mathbf{V})$$

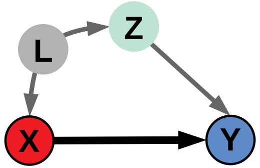




Causal inference: 1. Utilizing causal graphs

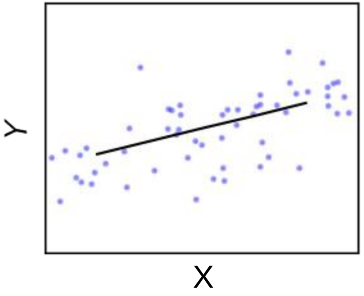
<u>Causal effect estimation:</u> Given causal graph and data, compute causal effect of intervention from observational distribution ***P*(*V*)**

$$P\left(Y|do(X{=}x)\right) = \text{function of } P(\mathbf{V})$$



'Correlation' regression

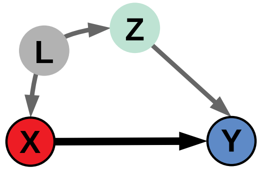
$$Y = \boxed{\beta_{YX}} X$$



Causal inference: 1. Utilizing causal graphs

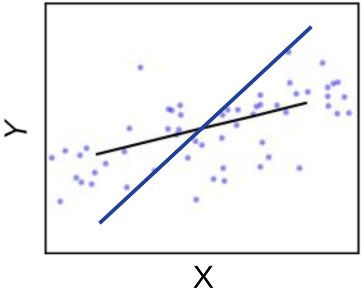
<u>Causal effect estimation:</u> Given causal graph and data, compute causal effect of intervention from observational distribution *P(**V**)*

$$P\left(Y | do(X{=}x)\right) = \text{function of } P(\mathbf{V})$$



‘Correlation’ regression

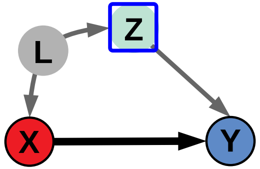
$$Y = \boxed{\beta_{YX}} X$$



Causal inference: 1. Utilizing causal graphs

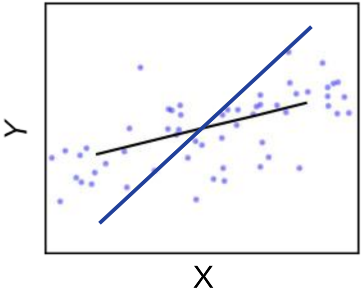
Causal effect estimation: Given causal graph and data, compute causal effect of intervention from observational distribution *P(**V**)*

$$P\left(Y|do(X{=}x)\right) = \text{function of } P(\mathbf{V})$$



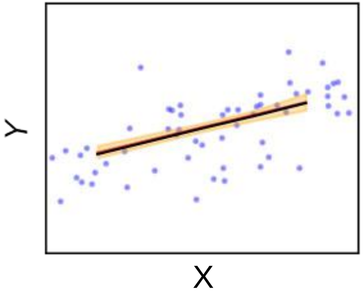
‘Correlation’ regression

$$Y = \boxed{\beta_{YX}} X$$



Causal regression

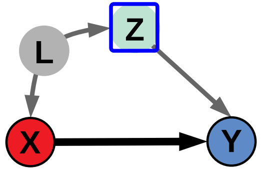
$$Y = \boxed{\beta_{YX\cdot Z}} X + \beta_{YZ\cdot X} Z$$



Causal inference: 1. Utilizing causal graphs

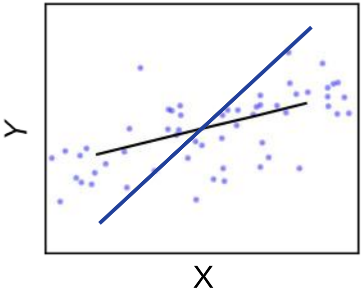
<u>Causal effect estimation:</u> Given causal graph and data, compute causal effect of intervention from observational distribution *P(**V**)*

$$P\,(Y|do(X{=}x)) = \text{function of } P(\mathbf{V})$$



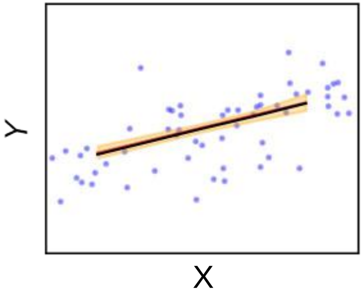
‘Correlation’ regression

$$Y = \boxed{\beta_{YX}}\,X$$



Causal regression

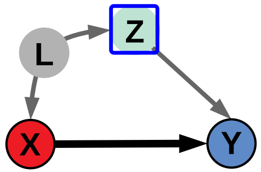
$$\mathbb{E}[Y|do(x=1)] - \mathbb{E}[Y|do(x=0)] = \beta_{YX\cdot Z}$$



Causal inference: 1. Utilizing causal graphs

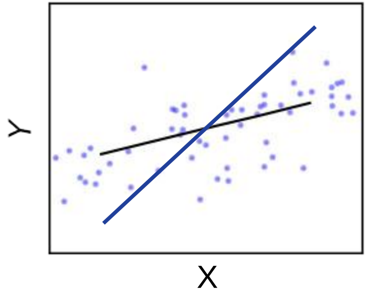
<u>Causal effect estimation:</u> Given causal graph and data, compute causal effect of intervention from observational distribution *P(V)*

$$P\left(Y|do(X{=}x)\right) = \text{function of } P(\mathbf{V})$$



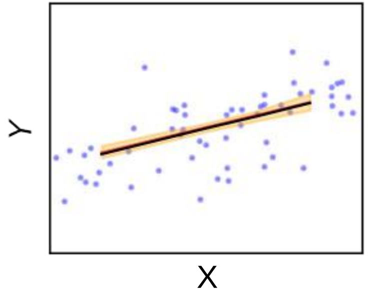
'Correlation' regression

$$Y = \boxed{\beta_{YX}} X$$



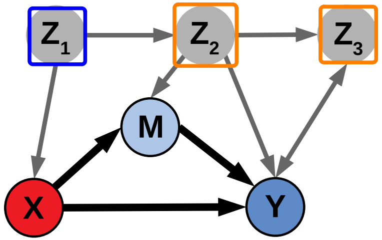
Causal regression

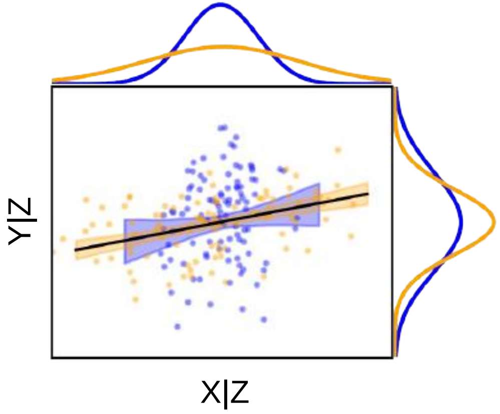
$$Y = \boxed{\beta_{YX\cdot Z}} X + \beta_{YZ\cdot X} Z$$



Optimal causal effect estimators

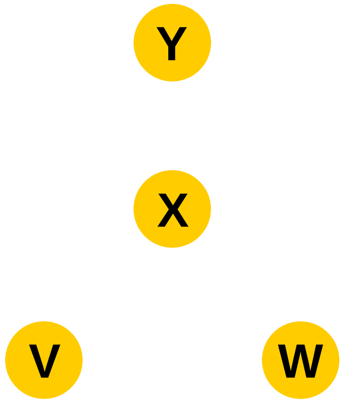
(Runge NeurIPS 2021)





Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution $P(\boldsymbol{V})$

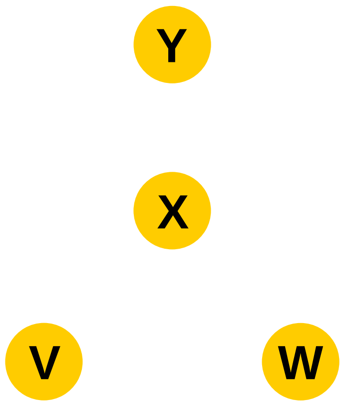


Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution $P(\mathbf{V})$

Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

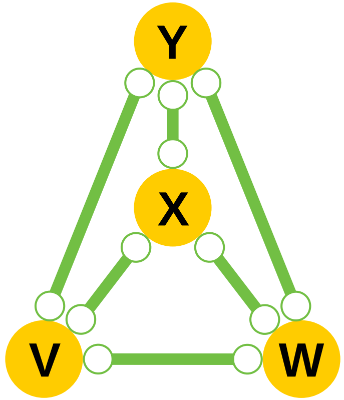


Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

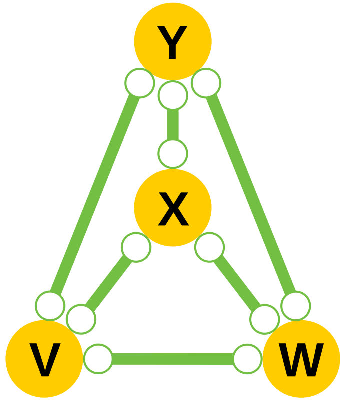


Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$V \perp\!\!\!\perp W$ and $(V, W) \perp\!\!\!\perp Y | X$

all others are dependent



Example assumptions:

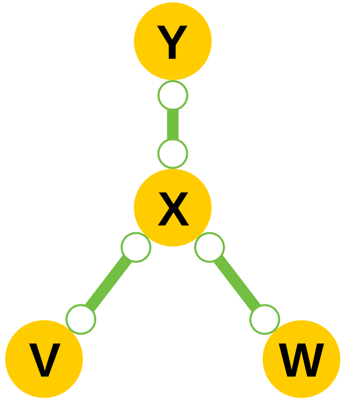
- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$V \perp\!\!\!\perp W \text{ and } (V,W) \perp\!\!\!\perp Y|X$$

all others are dependent



Example assumptions:

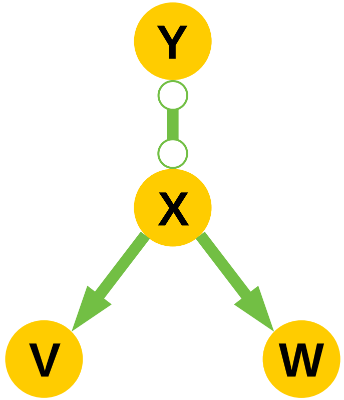
- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$V \perp\!\!\!\perp W \text{ and } (V,W) \perp\!\!\!\perp Y|X$$

all others are dependent



Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

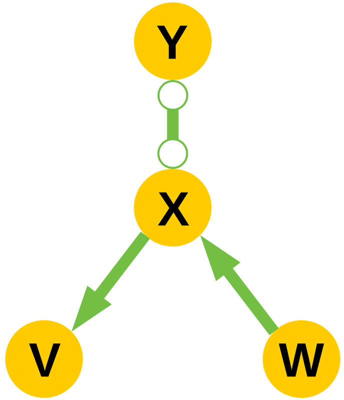
Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

$$V \perp\!\!\!\perp W \text{ and } (V, W) \perp\!\!\!\perp Y | X$$

all others are dependent

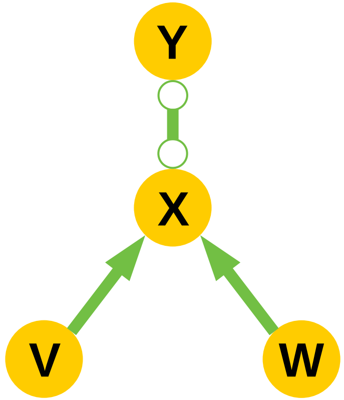


Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$V \perp\!\!\!\perp W \text{ and } (V,W) \perp\!\!\!\perp Y|X$$

all others are dependent



Example assumptions:

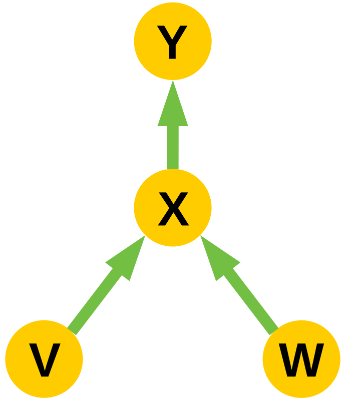
- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$V \perp\!\!\!\perp W \text{ and } (V,W) \perp\!\!\!\perp Y|X$$

all others are dependent



Example assumptions:

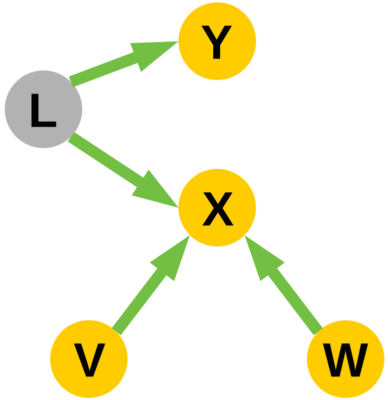
- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$\mathrm{V} \perp\!\!\!\perp W \text{ and } (V,W) \perp\!\!\!\perp Y|X$$

all others are dependent



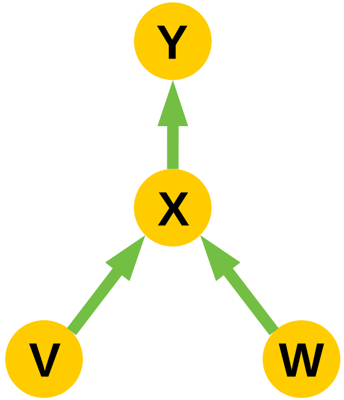
Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$V \perp\!\!\!\perp W$ and $(V,W) \perp\!\!\!\perp Y|X$
all others are dependent



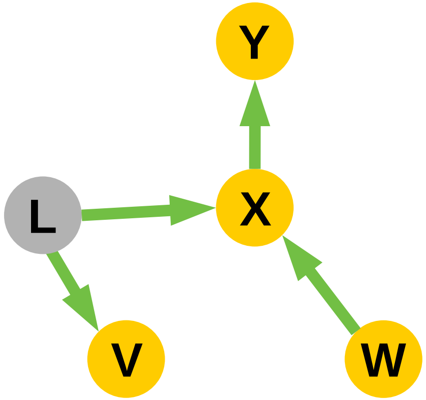
Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$V \perp\!\!\!\perp W$ and $(V, W) \perp\!\!\!\perp Y | X$
all others are dependent



Example assumptions:

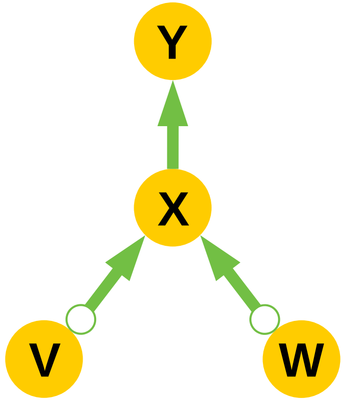
- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$V \perp\!\!\!\perp W \text{ and } (V, W) \perp\!\!\!\perp Y | X$$

all others are dependent



Example assumptions:

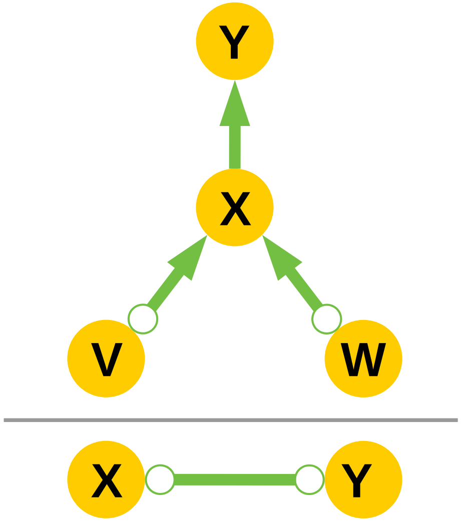
- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$V \perp\!\!\!\perp W \text{ and } (V,W) \perp\!\!\!\perp Y|X$$

all others are dependent



Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

- Assumptions on functional dependencies and noise distributions
  → ***Restricted structural causal modeling*** (Peters et al. 2018)

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

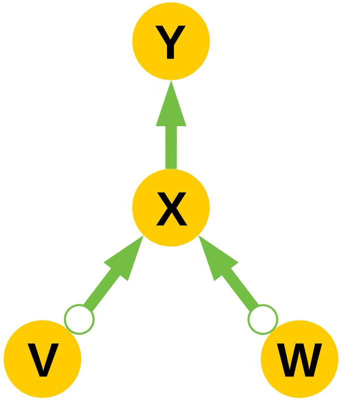
Example assumptions:

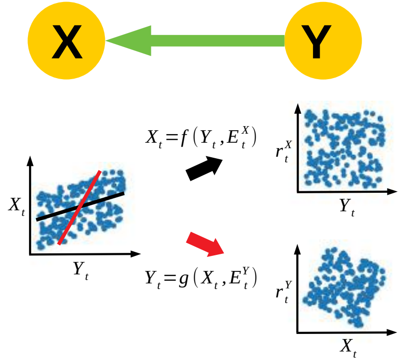
- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)

- Assumptions on functional dependencies and noise distributions
  → ***Restricted structural causal modeling*** (Peters et al. 2018)

$V \perp\!\!\!\perp W$ and $(V,W) \perp\!\!\!\perp Y|X$

all others are dependent



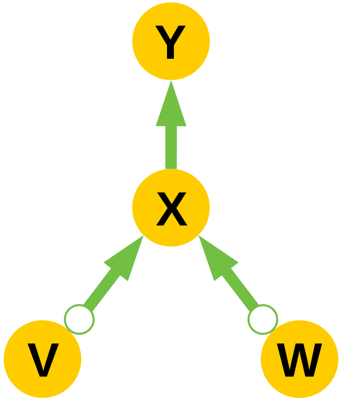


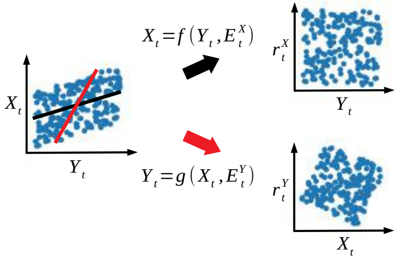
Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

$$V \perp\!\!\!\perp W \text{ and } (V,W) \perp\!\!\!\perp Y|X$$

all others are dependent



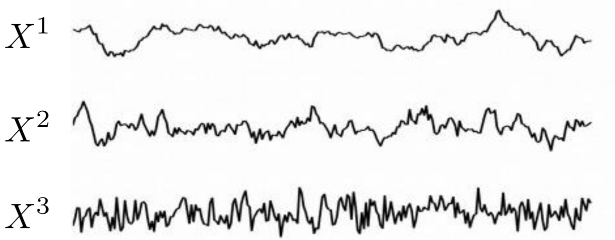


Example assumptions:

- Statistical dependencies imply causal relations (*Markov condition*) and independencies an absence thereof (*Faithfulness*)
  → ***Constraint-based causal discovery*** (Spirtes et al. 2000)
- Assumptions on functional dependencies and noise distributions
  → ***Restricted structural causal modeling*** (Peters et al. 2018)
- Score-based Bayesian network learning
- ...

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution $P(\boldsymbol{V})$



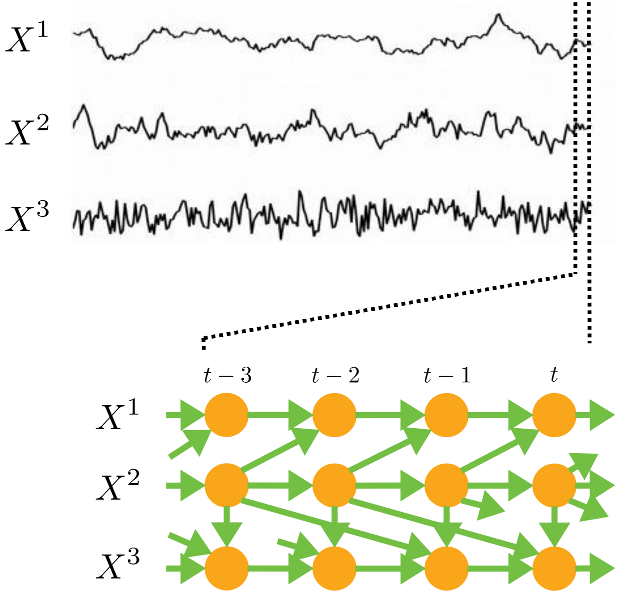
Time series case:

Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

Time series case:

- ***PCMCI causal discovery framework*** (Runge et al. SciAdv 2019, UAI 2020, Gerhardus and Runge NeurIPS 2020)

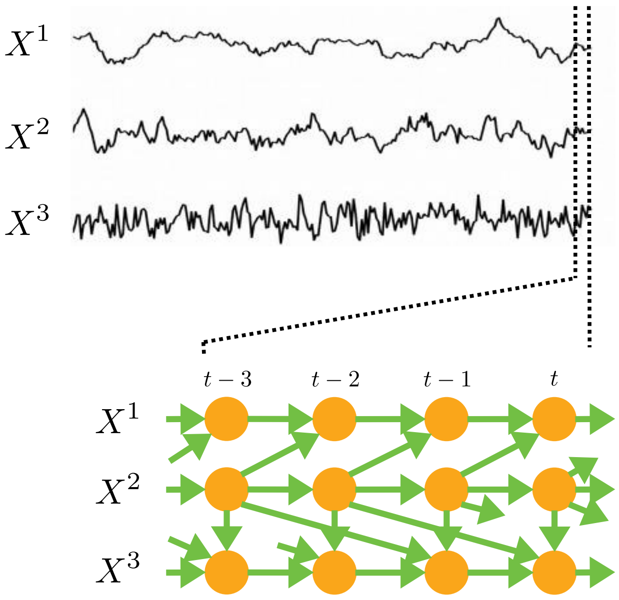


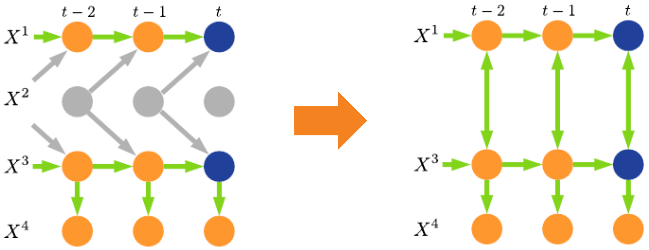
Causal inference: 2. Learning causal graphs

Task: Given data and *general assumptions*, estimate causal graph from observational distribution *P(**V**)*

Time series case:

- ***PCMCI causal discovery framework***
  (Runge et al. SciAdv 2019, UAI 2020, Gerhardus and Runge NeurIPS 2020)





Causal inference: 2. Learning causal graphs

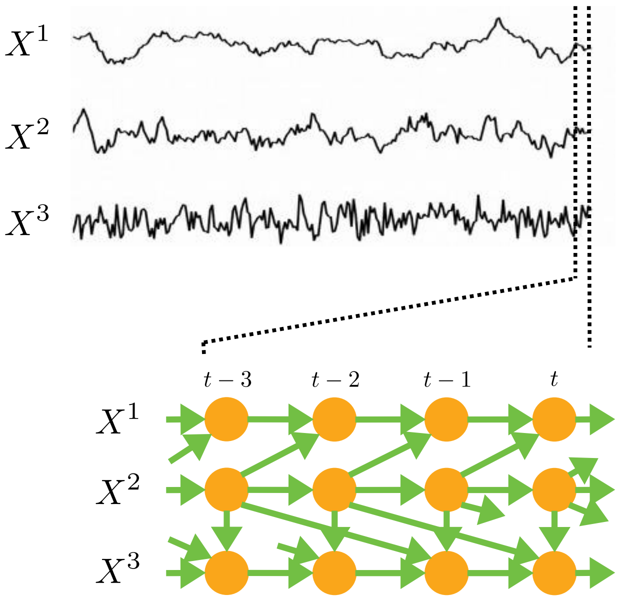
Task: Given data and *general assumptions*, estimate causal graph from observational distribution $P(\mathbf{V})$

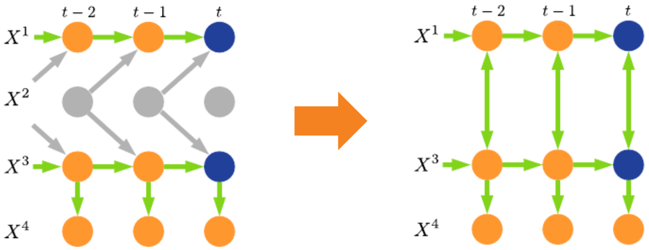
Time series case:

- ***PCMCI causal discovery framework*** (Runge et al. SciAdv 2019, UAI 2020, Gerhardus and Runge NeurIPS 2020)

- Assuming no instantaneous effects
  → *Granger causality* (Granger 1969)

- **… → *see Runge et al. NatComm Perspective (2019)***

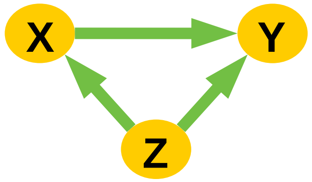




Causal inference and machine learning

Causal inference engine (Pearl): Given a query, a causal graph and data, output estimand in terms of observational distribution *P(**V**)*

$$P\left(Y|do(X{=}x)\right) = \int P(y|x,z)P(z)dz$$

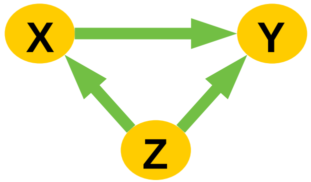


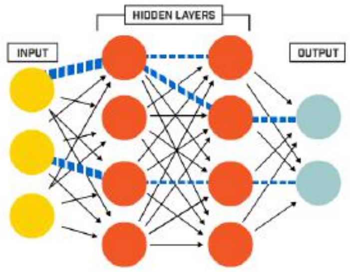
Causal inference and machine learning

Causal inference engine (Pearl): Given a query, a causal graph and data, output estimand in terms of observational distribution *P(**V**)* → ***estimand can be estimated with (deep) ML***

$$P\left(Y|do(X{=}x)\right) = \int P(y|x,z)P(z)dz$$

$$\implies \hat{Y} = \int \hat{f}(X=x, Z=z)p(z)dz$$



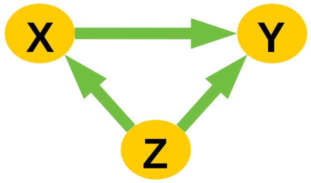


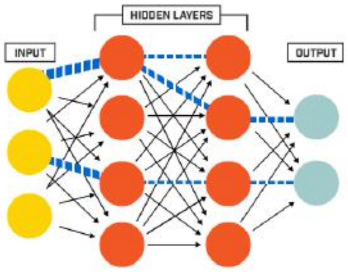
Causal inference and machine learning

Causal inference engine (Pearl): Given a query, a causal graph and data, output estimand in terms of observational distribution *P(**V**)* → ***estimand can be estimated with (deep) ML***

$$P\left(Y|do(X{=}x)\right) = \int P(y|x,z)P(z)dz$$

$$\implies \hat{Y} = \int \hat{f}(X=x, Z=z)p(z)dz$$





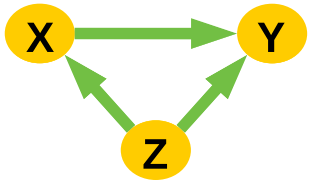
Causal graph learning:

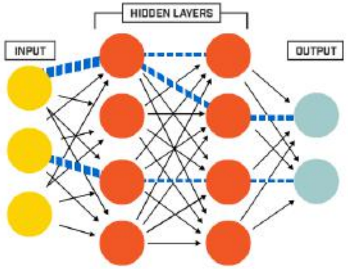
Causal inference and machine learning

Causal inference engine (Pearl): Given a query, a causal graph and data, output estimand in terms of observational distribution *P(**V**)*
→ ***estimand can be estimated with (deep) ML***

$$P\,(Y|do(X{=}x)) = \int P(y|x,z)P(z)dz$$

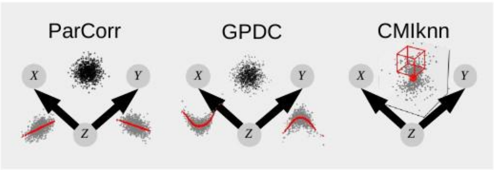
$$\implies \hat{Y} = \int \hat{f}(X=x, Z=z)p(z)dz$$





Causal graph learning:

- ML for non-parametric conditional independence testing (Runge AISTATS 2018, Strobl et al., 2017, ...)

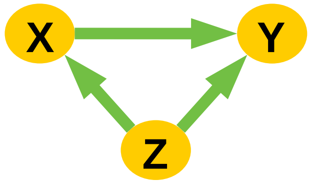


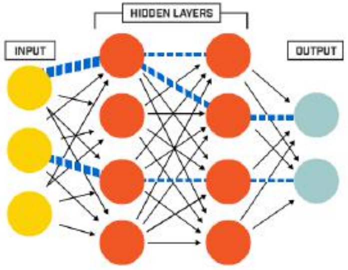
Causal inference and machine learning

Causal inference engine (Pearl): Given a query, a causal graph and data, output estimand in terms of observational distribution *P(**V**)*
→ ***estimand can be estimated with (deep) ML***

$$P\,(Y|do(X{=}x)) = \int P(y|x,z)P(z)dz$$

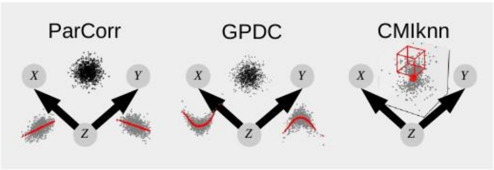
$$\Longrightarrow \hat{Y} = \int \hat{f}(X = x, Z = z)p(z)dz$$





Causal graph learning:

- ML for non-parametric conditional independence testing (Runge AISTATS 2018, Strobl et al., 2017, ...)
- Structure learning as continuous optimization problem
  → *Brouillard et al., 2020, and many more recently*

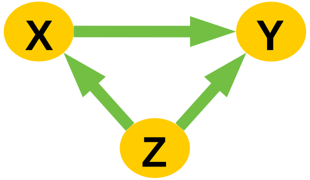


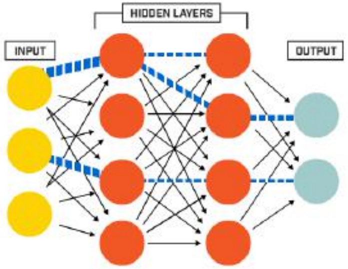
Causal inference and machine learning

Causal inference engine (Pearl): Given a query, a causal graph and data, output estimand in terms of observational distribution *P(**V**)*
→ ***estimand can be estimated with (deep) ML***

$$P\,(Y|do(X{=}x)) = \int P(y|x,z)P(z)dz$$

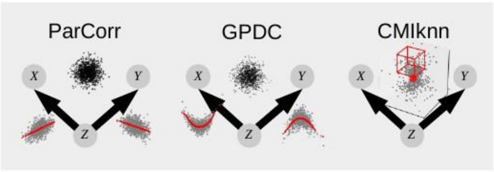
$$\Longrightarrow\ \hat{Y} = \int \hat{f}(X=x, Z=z)p(z)dz$$

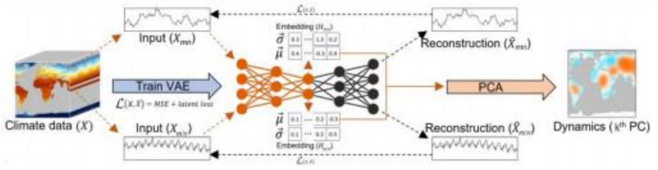




Causal graph learning:

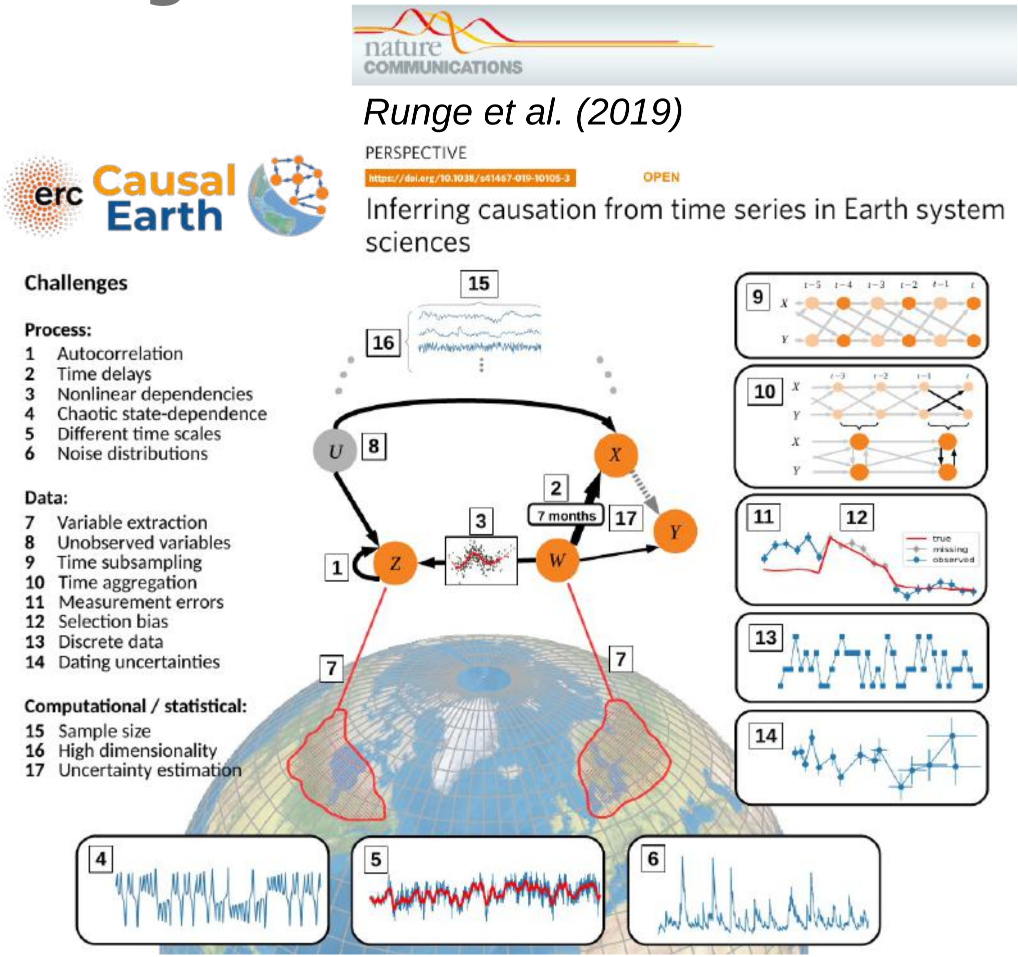
- ML for non-parametric conditional independence testing (Runge AISTATS 2018, Strobl et al., 2017, ...)
- Structure learning as continuous optimization problem
  → *Brouillard et al., 2020, and many more recently*
- Dimension-reduction to reconstruct variables
- ...





Challenges

Runge et al. (2019)



Challenges

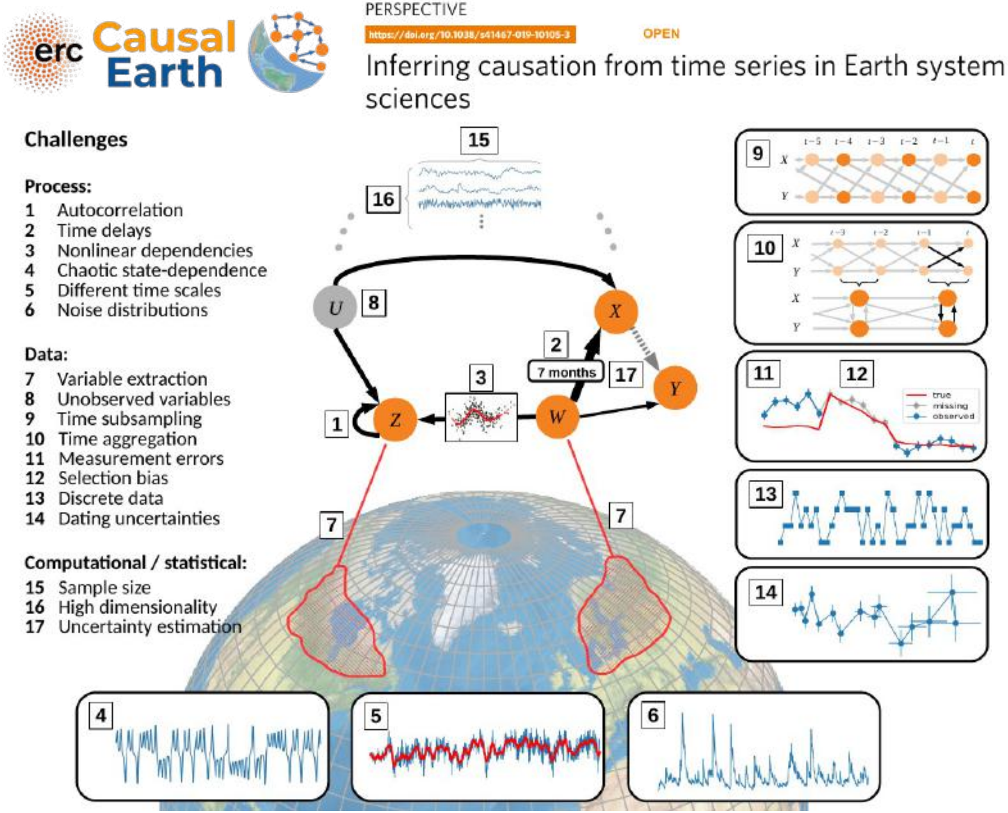
Runge et al. (2019)

nature COMMUNICATIONS

PERSPECTIVE

https://doi.org/10.1038/s41467-019-10105-3 OPEN

Inferring causation from time series in Earth system sciences



Schölkopf et al. (2020)

Toward Causal Representation Learning

This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalization, thereby assaying how causality can contribute to modern machine learning research.

By Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio

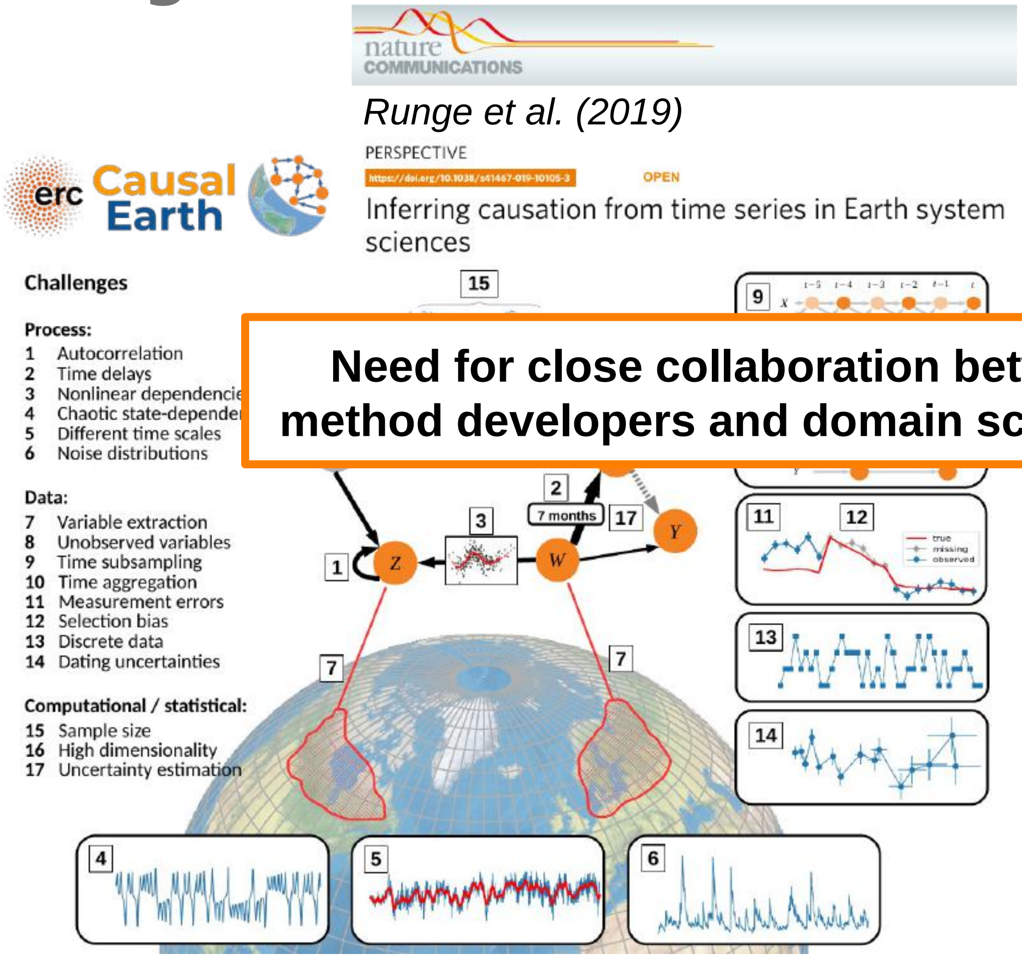
ABSTRACT | The two fields of machine learning and graphical causality arose and are developed separately. However, there is, now, cross-pollination and increasing interest in both fields to benefit from the advances of the other. In this article, we review fundamental concepts of causal inference and relate them to crucial open problems of machine learning, including transfer and generalization, thereby assaying how causality can contribute to modern machine learning research. This also applies in the opposite direction: we note that most work in causality starts from the premise that the causal variables are given. A central problem for AI and causality is, thus, causal representation learning, that is, the discovery of high-level causal variables from low-level observations. Finally, we delineate some implications of causality for machine learning and propose key research areas at the intersection of both communities.

I. INTRODUCTION

If we compare what machine learning can do to what animals accomplish, we observe that the former is rather limited at some crucial feats where natural intelligence excels. These include transfer to new problems and any form of generalization that is not from one data point to the next (sampled from the same distribution), but rather from one problem to the next—both have been termed *generalization*, but the latter is a much harder form thereof, sometimes referred to as *horizontal*, *strong*, or *out-of-distribution* generalization. This shortcoming is not too surprising, given that machine learning often disregards information that animals use heavily: interventions in the world, domain shifts, and temporal structure—by and large, we consider these factors a nuisance and try to engineer them away. In accordance with this, the majority of current successes of machine learning boil down to large-

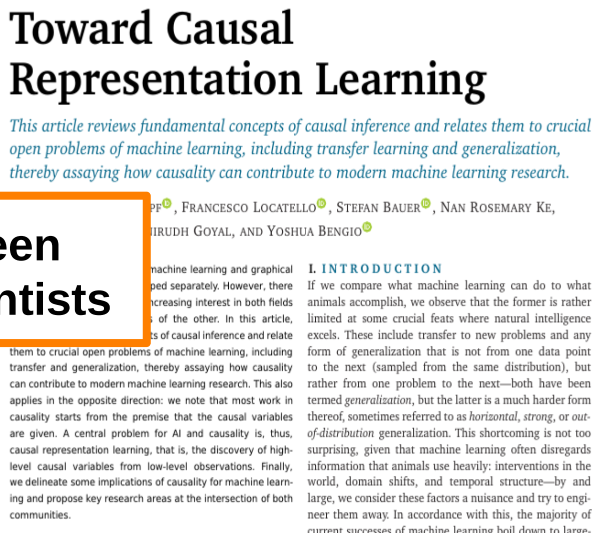
Challenges

Runge et al. (2019)



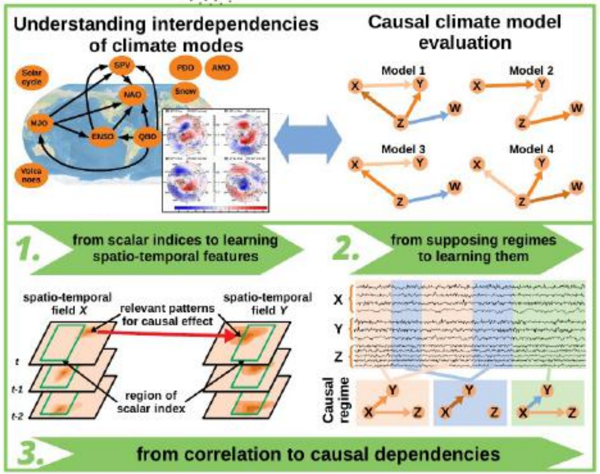
Need for close collaboration between method developers and domain scientists

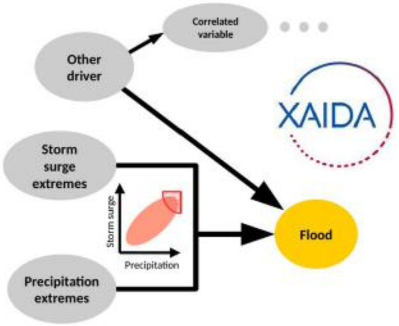
Schölkopf et al. (2020)



Applications

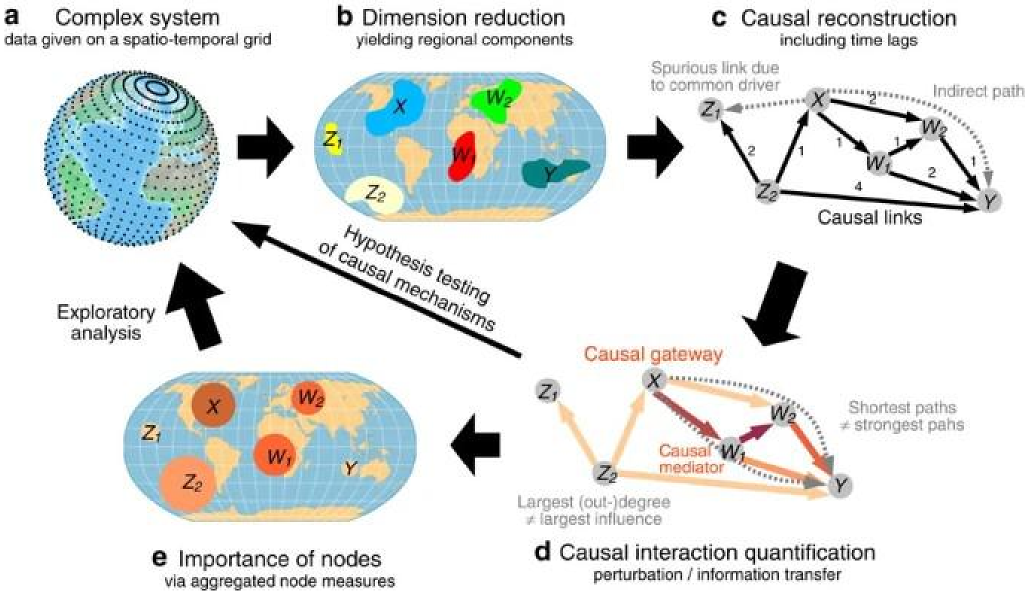
1. **Learning causal graphs** to understand mechanisms
2. **Quantifying causal mechanisms:** link strength and mediation analysis
3. Causally robust **forecasting**
4. Causally validating ML methods
5. **Evaluating climate models and constraining climate change projections**
6. Hybrid physical-ML modeling
7. Detection and attribution of extreme events
8. …





Causal mediation analysis

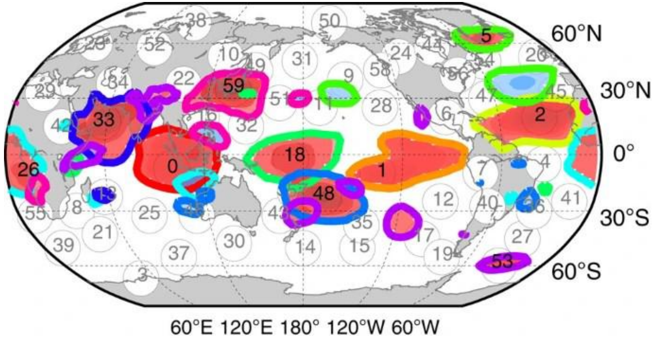
- Pathway mechanisms between **El Nino** and **Indian monsoon** through sea-level pressure system

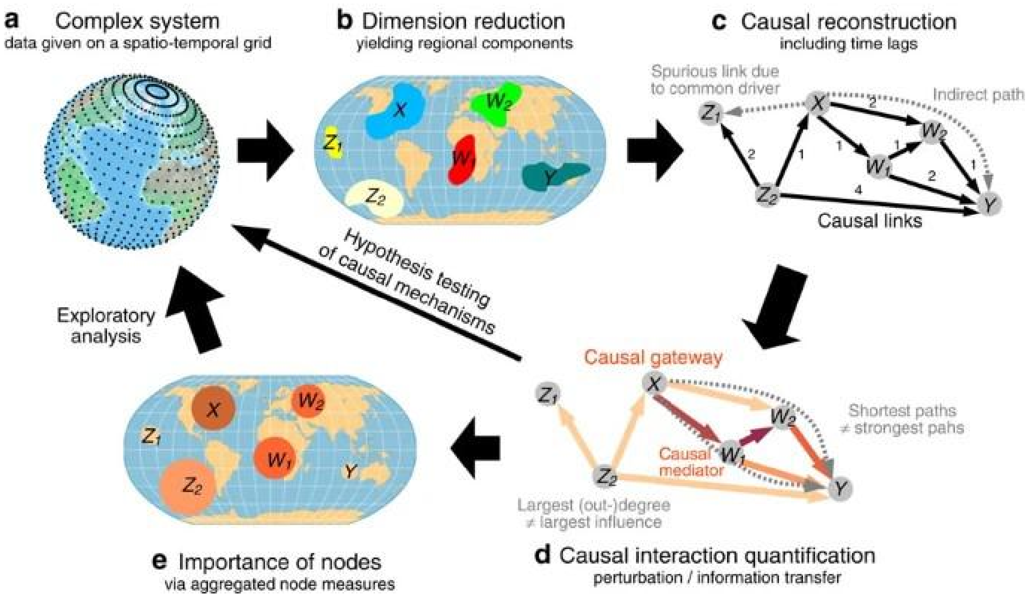


Runge et al., NatComm 2015

Causal mediation analysis

- Pathway mechanisms between **El Nino** and **Indian monsoon** through sea-level pressure system

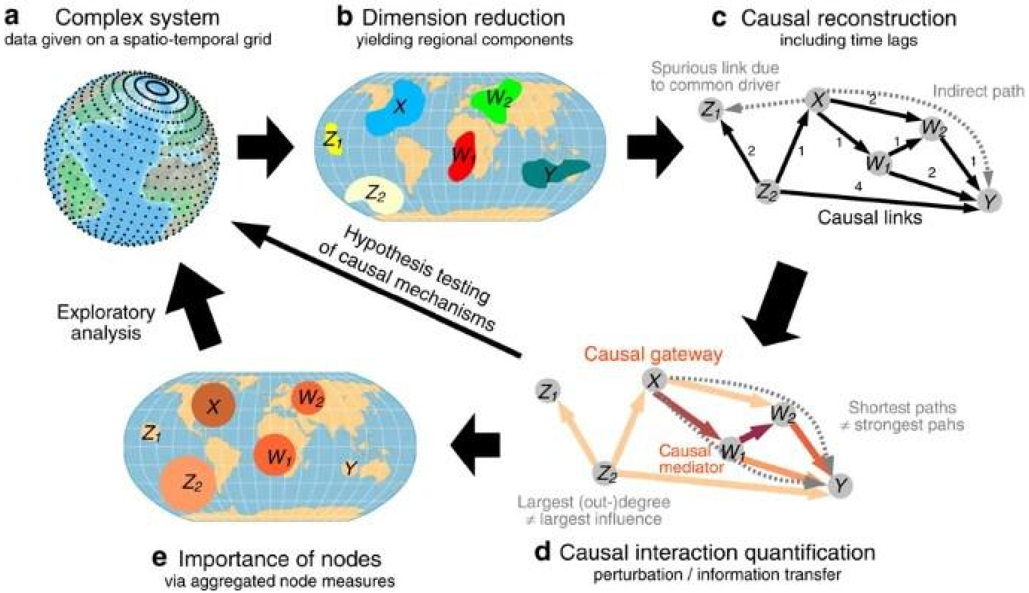


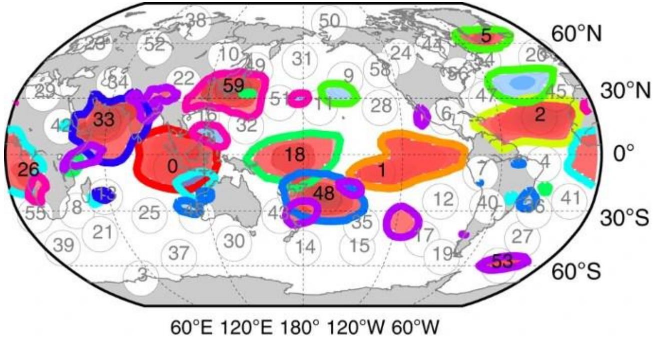


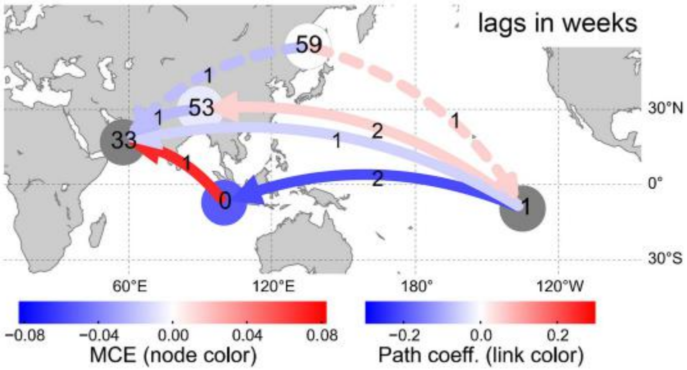
Runge et al., NatComm 2015

Causal mediation analysis

- Pathway mechanisms between **El Nino** and **Indian monsoon** through sea-level pressure system



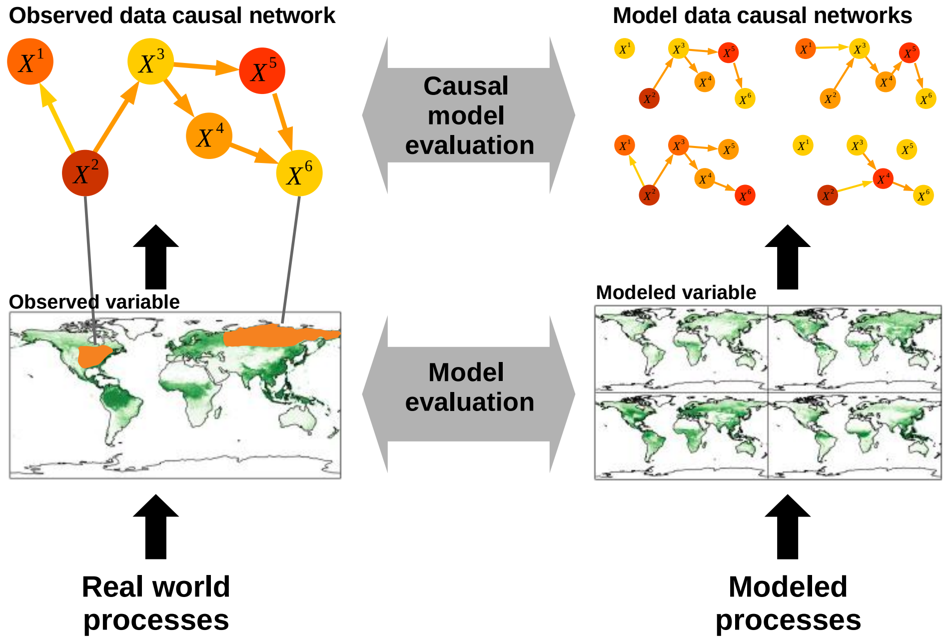




Runge et al., NatComm 2015

Simulation model evaluation

- Simple statistics can be right for the wrong reasons
- **Idea**: Compare climate models and observations in terms of causal relationships




nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-15195-y OPEN

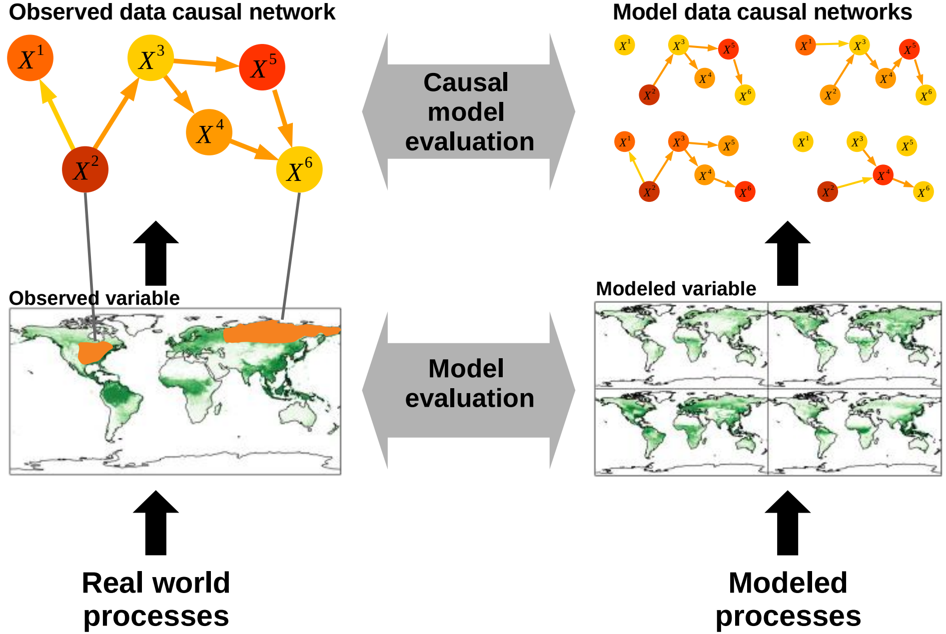
Causal networks for climate model evaluation and constrained projections

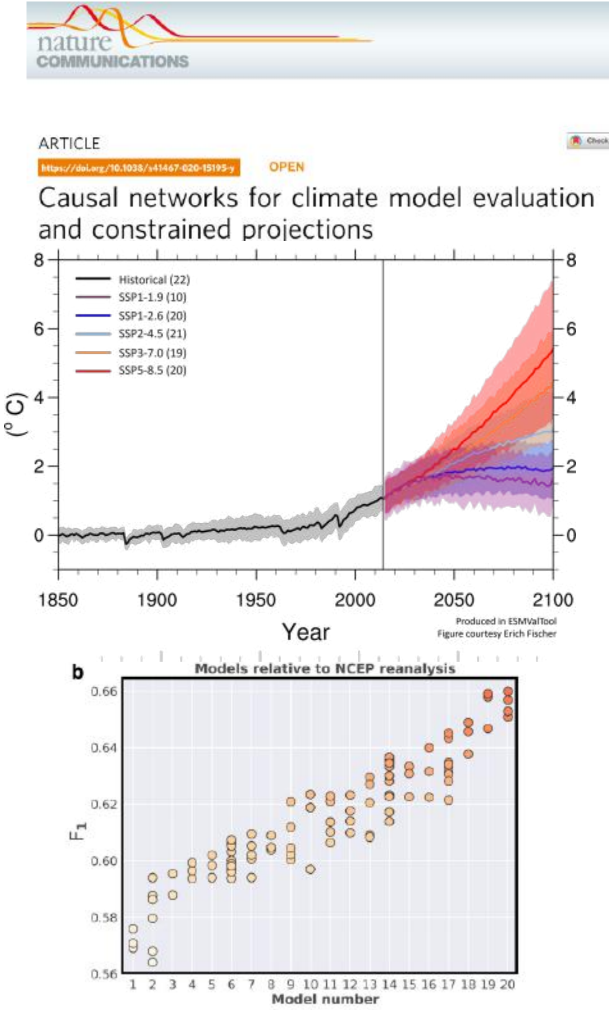
Peer Nowack[1,2,3,4], Jakob Runge[5,1], Veronika Eyring[6,7] & Joanna D. Haigh[1,2]

Nowack et al. NatComm. (2020)

Simulation model evaluation

- Simple statistics can be right for the wrong reasons
- **Idea**: Compare climate models and observations in terms of causal relationships



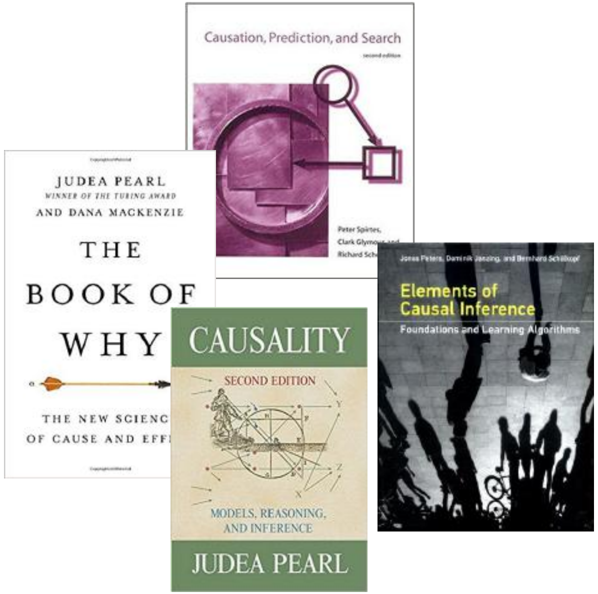


Nowack et al. NatComm. (2020)

Take-home message

Take-home message

- **Causal inference:** Framework to answer causal questions from empirical data



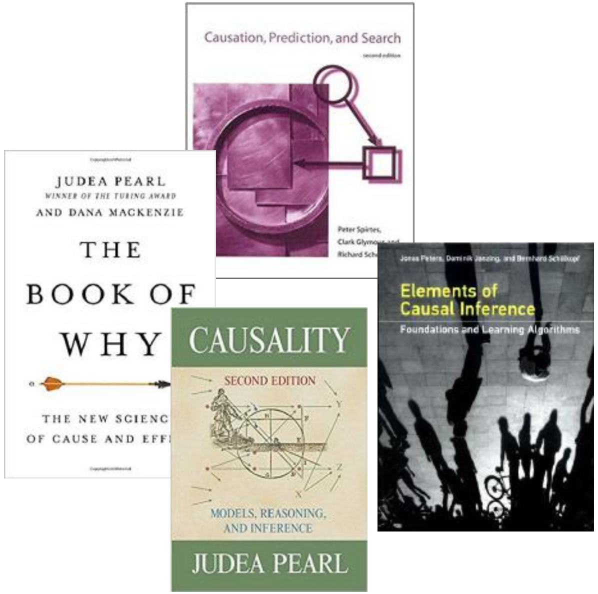
Take-home message

- **Causal inference:** Framework to answer causal questions from empirical data

- Two settings:

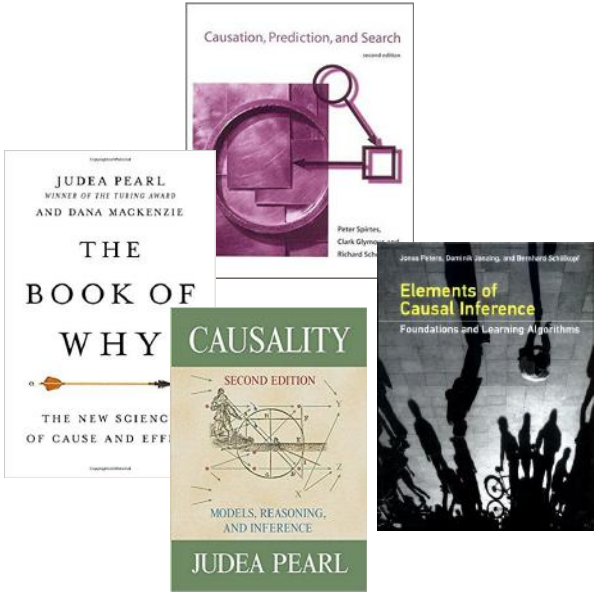
  1) **Utilize qualitative causal knowledge (graphs)**

  2) **Learn causal graphs** (then utilize them)



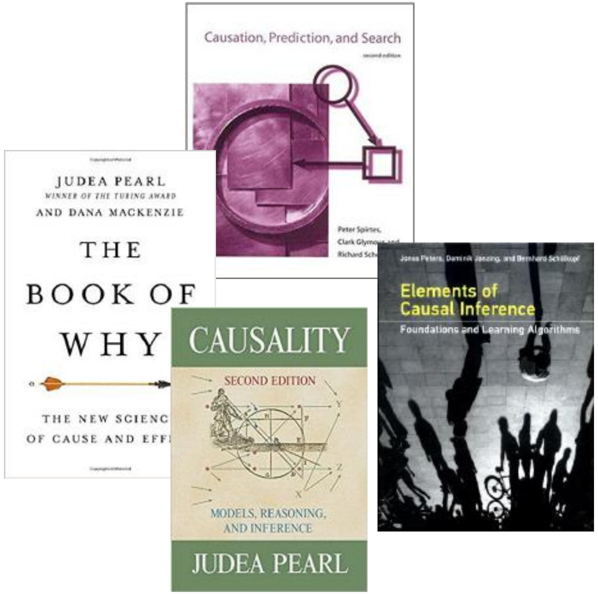
Take-home message

- **Causal inference:** Framework to answer causal questions from empirical data
- Two settings:
  1) **Utilize qualitative causal knowledge (graphs)**
  2) **Learn causal graphs** (then utilize them)
- Causal reasoning **requires assumptions** about underlying system and data



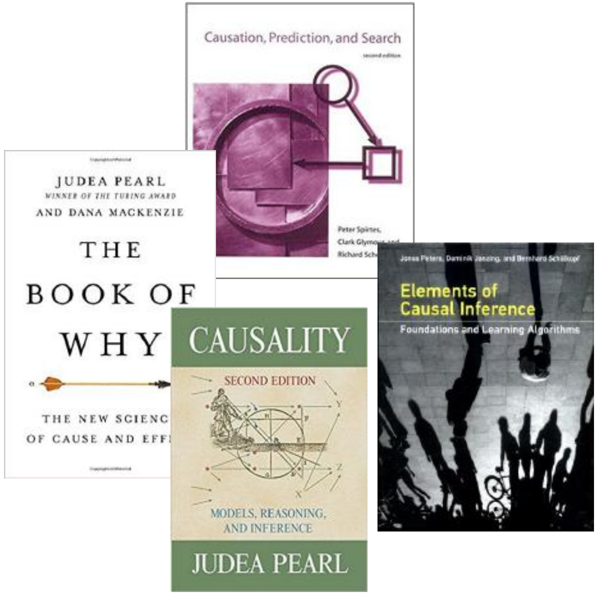
Take-home message

- **Causal inference:** Framework to answer causal questions from empirical data
- Two settings:
  1) **Utilize qualitative causal knowledge (graphs)**
  2) **Learn causal graphs** (then utilize them)
- Causal reasoning **requires assumptions** about underlying system and data
- Causal inference well **complements AI and machine learning** on complex datasets



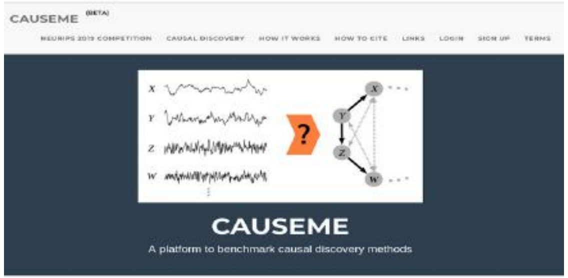
Take-home message

- **Causal inference:** Framework to answer causal questions from empirical data
- Two settings:
  1) **Utilize qualitative causal knowledge (graphs)**
  2) **Learn causal graphs** (then utilize them)
- Causal reasoning **requires assumptions** about underlying system and data
- Causal inference well **complements AI and machine learning** on complex datasets



Software:

- Tigramite, pcalg, TETRAD, causalfusion, CauseMe, ...



Thank you! Questions?