Causal inference for Earth system sciences

Prof. Dr. Jakob Runge

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

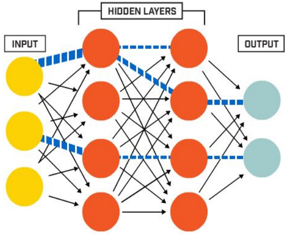
Chair of *Climate Informatics* at TUB

Prof. Dr. Imme Ebert-Uphoff

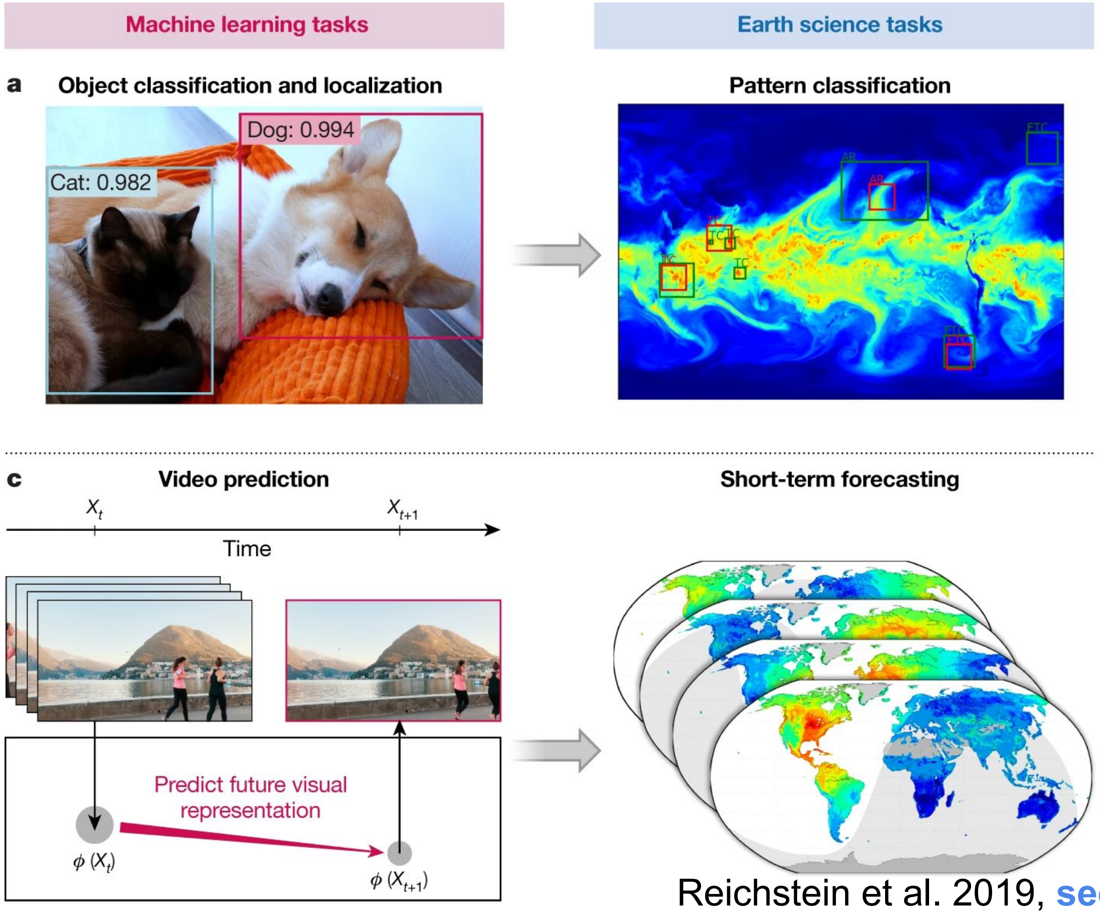
Machine Learning Lead - Cooperative Institute for Research in the Atmosphere (CIRA)

Research Prof. - Electrical & Computer Engineering

AI and machine learning in Earth system sciences



Turing-Award 2018
LeCun, Hinton, Bengio

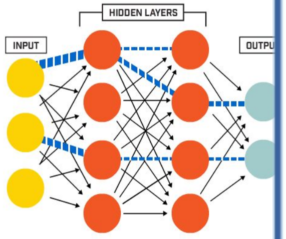


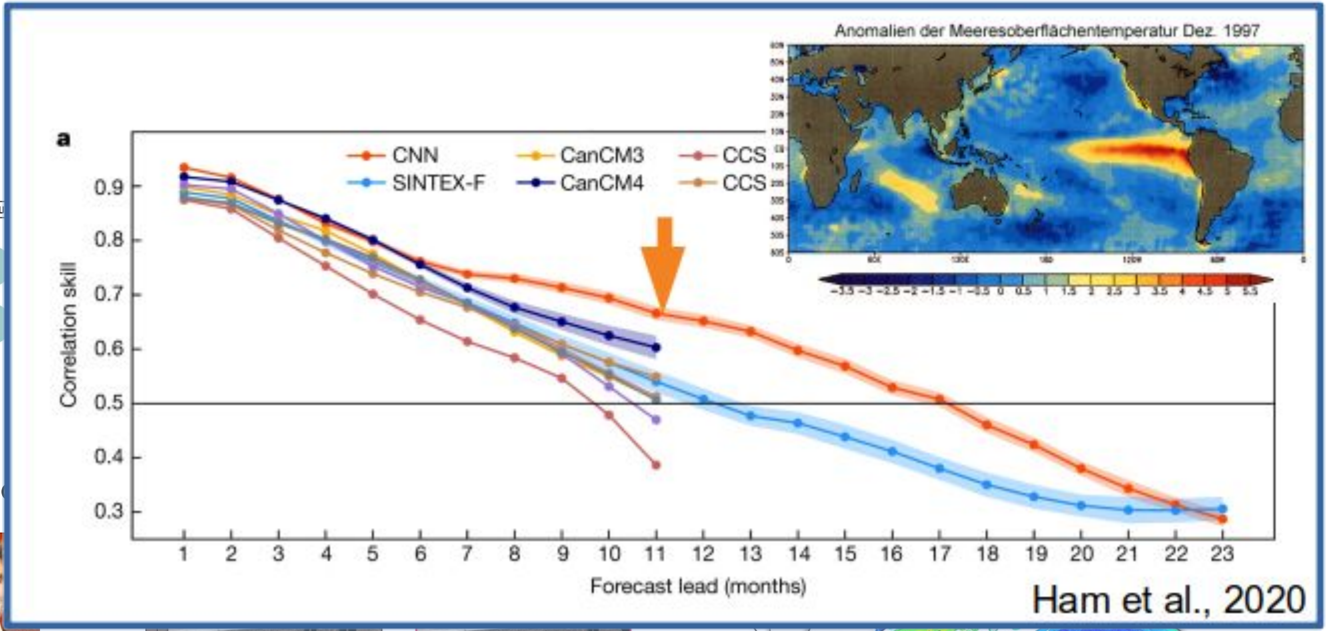
Reichstein et al. 2019, see AI4Good next week!

AI and machine learning in Earth system sciences

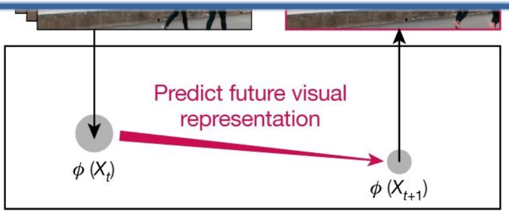
Machine learning tasks

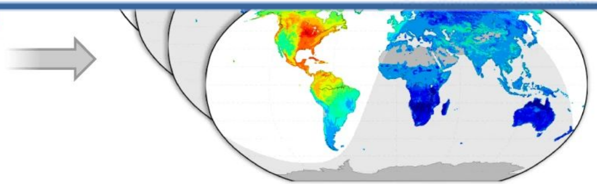
Earth science tasks





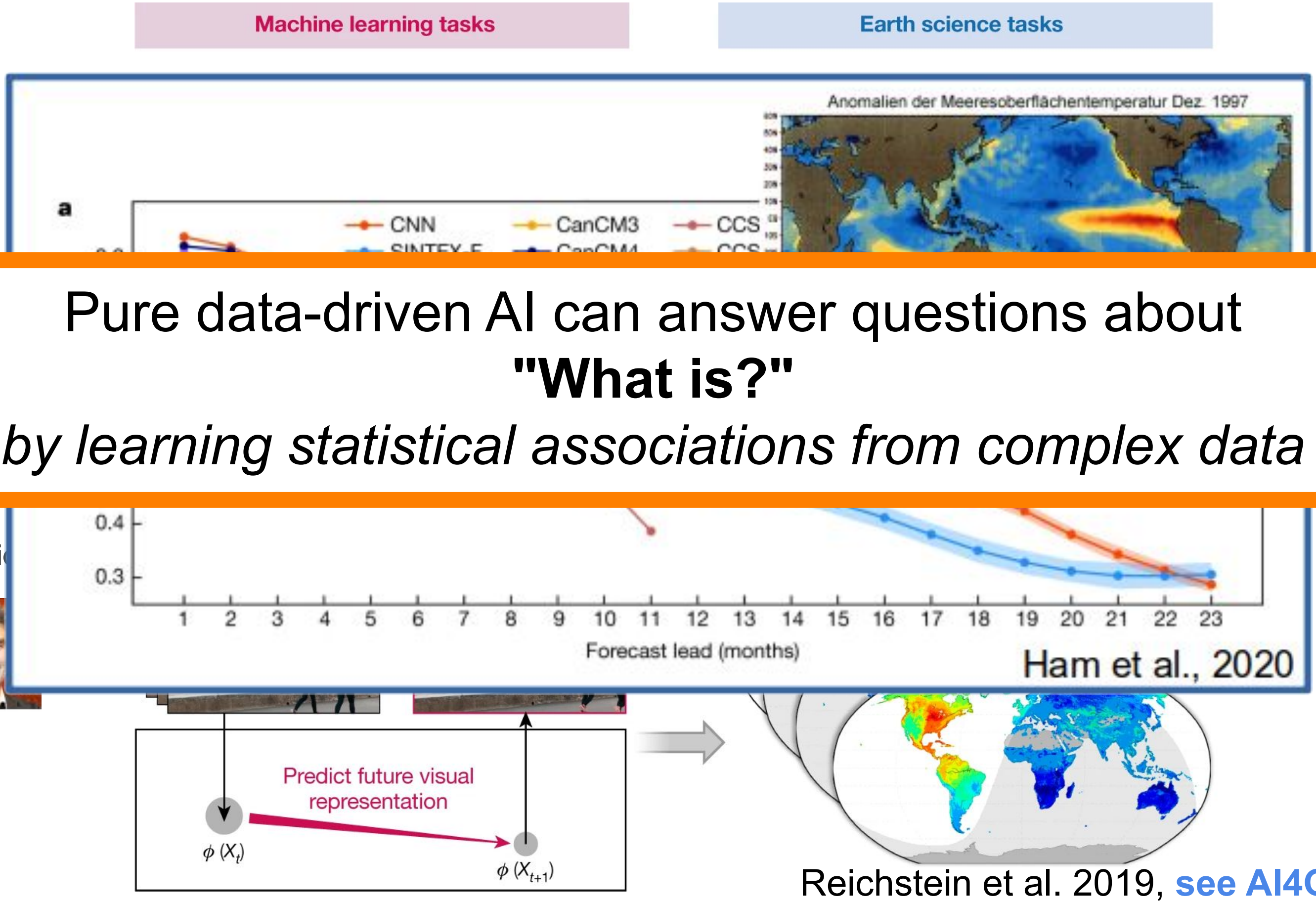
Turing-Award 2018
LeCun, Hinton, Bengi



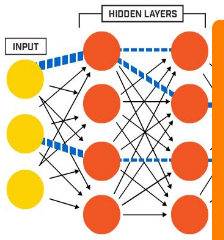


Reichstein et al. 2019, **see AI4Good next week!**

AI and machine learning in Earth system sciences



Pure data-driven AI can answer questions about
"What is?"
by learning statistical associations from complex data

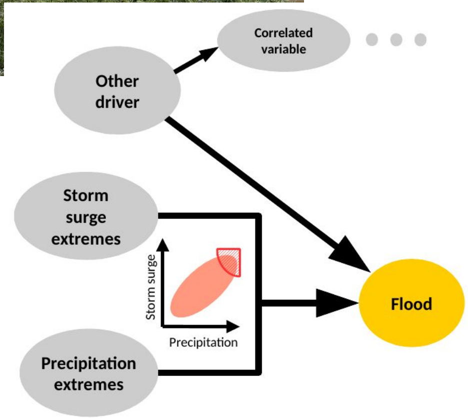


Turing-Award 2018
LeCun, Hinton, Bengi

Reichstein et al. 2019, **see AI4Good next week!**

Typical questions in Earth sciences

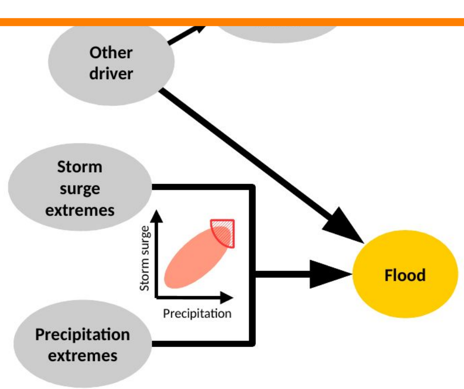
Typical questions in Earth sciences



Zscheischler et al. 2018

Typical questions in Earth sciences

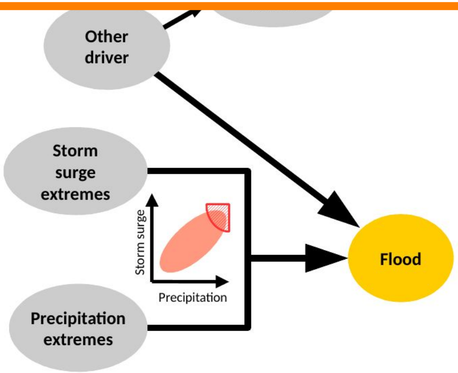
What causes extremes?

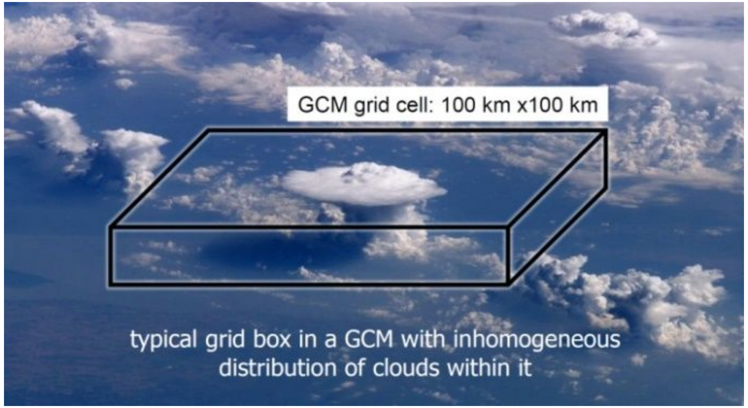


Zscheischler et al. 2018

Typical questions in Earth sciences

What causes extremes?



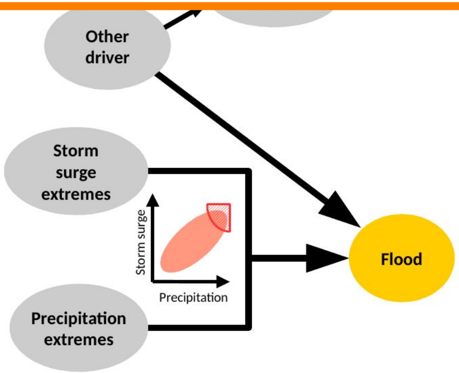


IPCC
P. Gentine

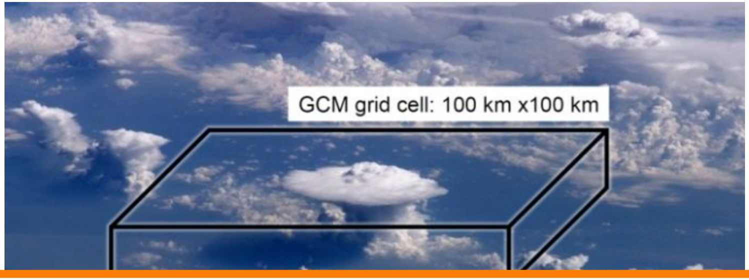
Zscheischler et al. 2018

Typical questions in Earth sciences

What causes extremes?

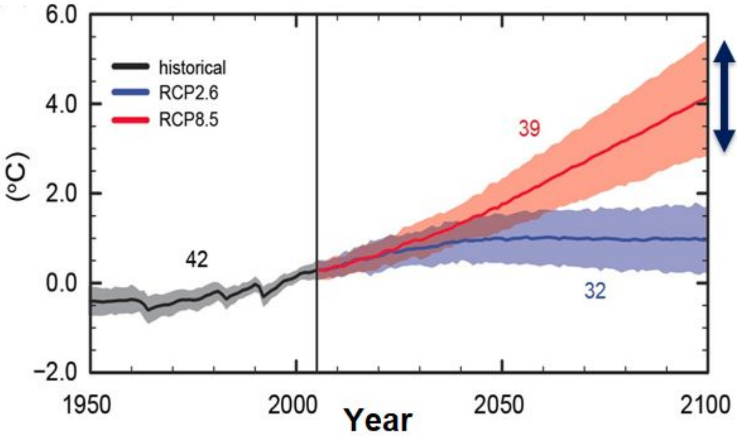


Zscheischler et al. 2018



Causal mechanism of aerosol-cloud interactions?

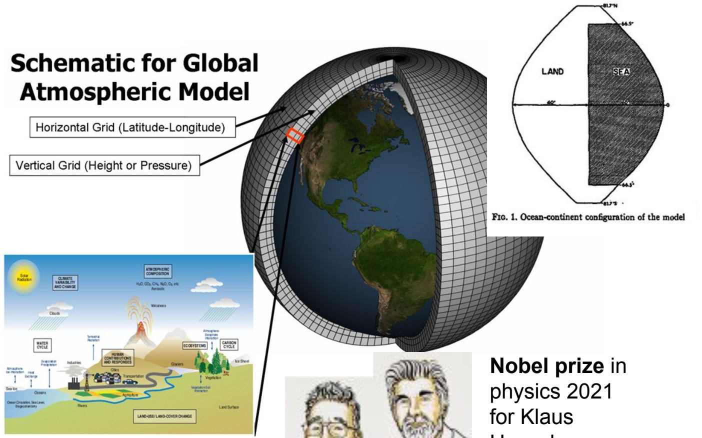
IPCC
P. Gentine



Two Types of Causality Studies

1) Experimental Study: when interventions are possible.

- Either in real system or in physical simulation models
- Supports necessary and sufficient conditions for causality.
- But: In climate science often infeasible or time-consuming!

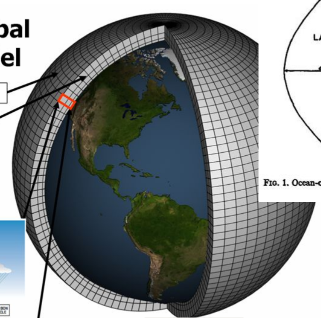


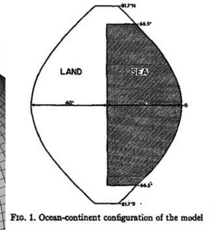
Nobel prize in physics 2021 for Klaus Hasselmann and Syukuro Manabe

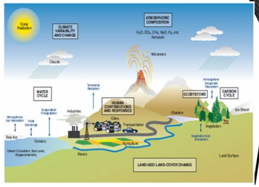
Two Types of Causality Studies

1) Experimental Study: when interventions are possible.

- Either in real system or in physical simulation models
- Supports necessary and sufficient conditions for causality.
- But: In climate science often infeasible or time-consuming!



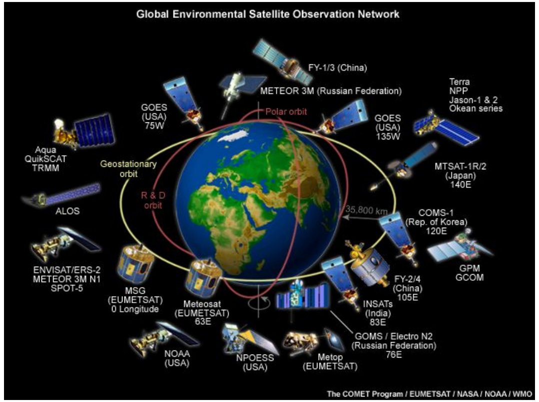




Nobel prize in physics 2021 for Klaus Hasselmann and Syukuro Manabe

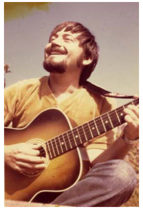
2) Observational Study: purely from observations / model output.

- Only supports necessary conditions for causality
- Weaker statements possible, but still powerful.
- Topic of this talk.



Causal inference

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



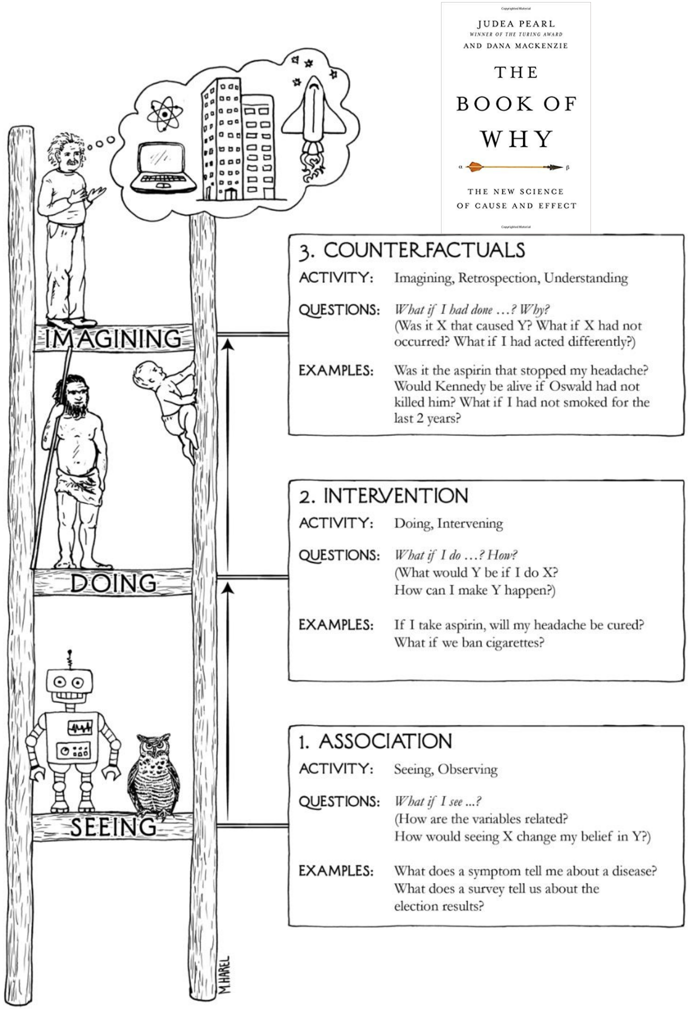
Judea Pearl
Turing-Award 2011
(theoretical framework, starting in 1980s)

Clark Glymour
(practical algorithms, starting in 1980s)

Spirtes, Glymour, Scheines

Ill. Niklas Elmehed © Nobel Prize Outreach

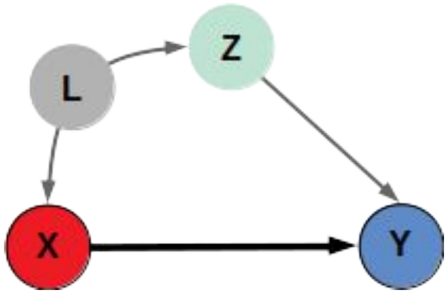
JD Angrist and GW Imbens
Nobel prize in economics 2021 (drawing conclusions from unintended/natural experiments)



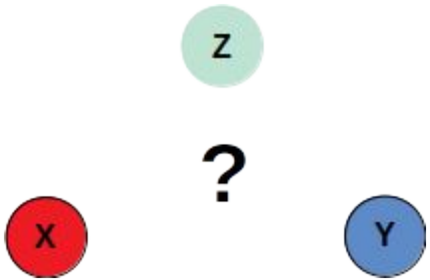
Causal inference and causal discovery

Two types of tasks:

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

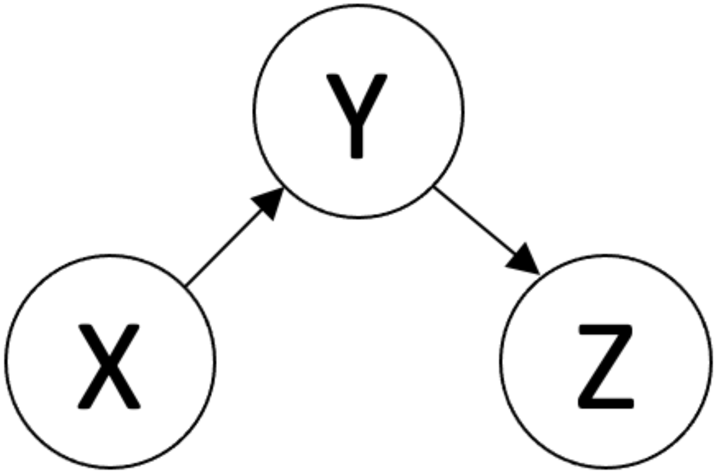


2. **Learn causal graphs** based on general assumptions (causal discovery) → start here in the following!



Next: Very quick (and incomplete) intro to Causality 101

- **For those new to causality.**
- **Everyone else: please take a 5 minute nap - or answer your email.**



Concept 1: Graphs as Language for causal models

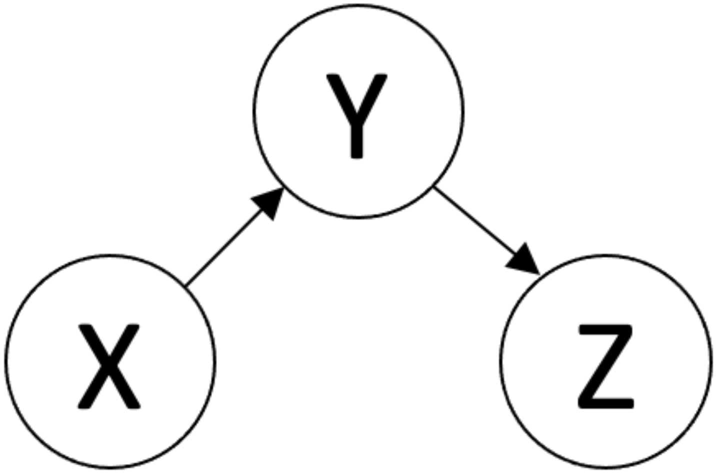
Express causal relationships as graph

- Variables are nodes of graph.
- Edges indicate causal connection between nodes.
- Arrows indicate direction: cause → effect.

In this example:

- Three variables: X,Y, Z.
- X is a cause of Y.
- Y is a cause of Z.

You should have a question here…



Concept 1: Graphs as Language for causal models

Express causal relationships as graph

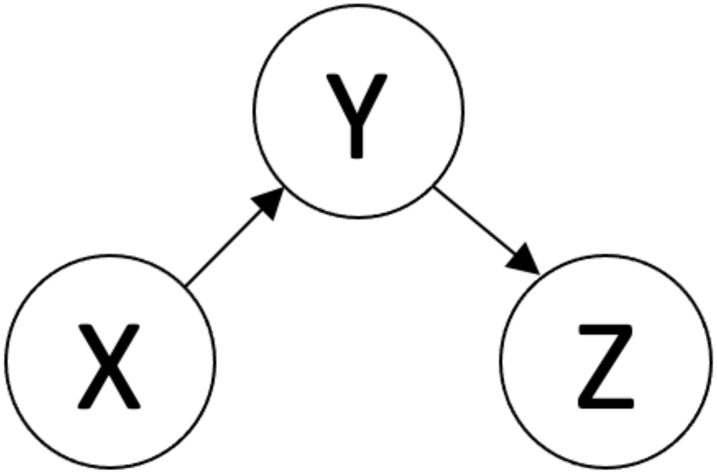
- Variables are nodes of graph.
- Edges indicate causal connection between nodes.
- Arrows indicate direction: cause → effect.

In this example:
- Three variables: X,Y, Z.
- X is a cause of Y.
- Y is a cause of Z.

You should have a question here…

If X causes Y and Y causes Z,
isn't X then also a cause of Z?

Should there be an arrow also from X → Z?

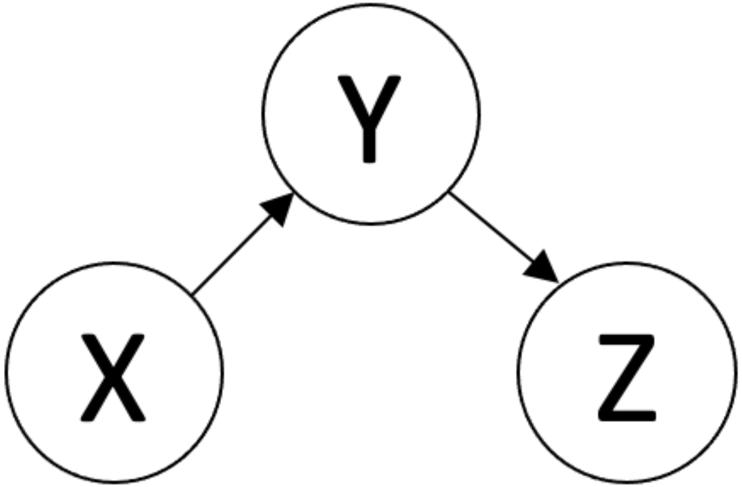


Concept 2: Direct vs. indirect connections

Arrows indicate direct causes only.

In this plot:
- X is a **direct** cause of Y.
- Y is a **direct** cause of Z.
- X is only an **indirect** cause of Z.

Goal of causal discovery: we want to identify only direct connections. Eliminate all others.



Why eliminate indirect connections?

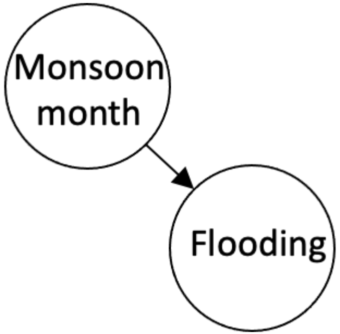
1) Sparsity, simplicity.

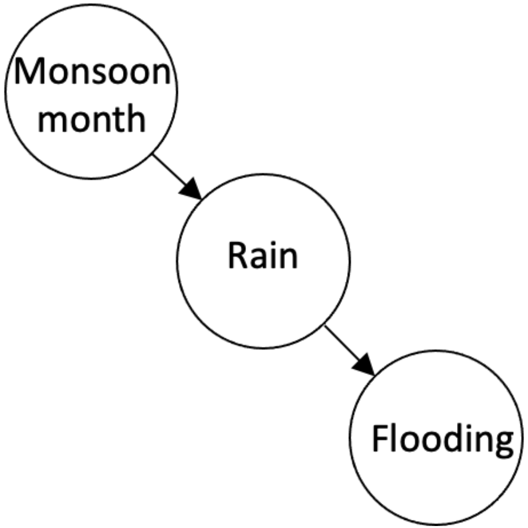
2) Only then can you understand effect of interventions!

Concept 3: Directness is relative property

One can always transform a direct connection into an indirect one by including an intermediate cause!

Toy example:





Monsoon month is **direct** cause of flooding in this model.

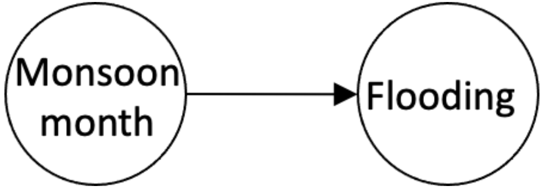
Monsoon month is only **indirect** cause of flooding in this model.

Both models are correct!

Directness is only defined *relative to variables included in model*.

Concept 4: Causality is probabilistic relationship

Example:



This graph implies:

1) Flooding is *more likely* in monsoon months, but *not* certain.
2) Flooding can also happen outside of monsoon months.

→ Supplement graph with probabilities.
→ **Probabilistic graphical model** (Bayesian network).

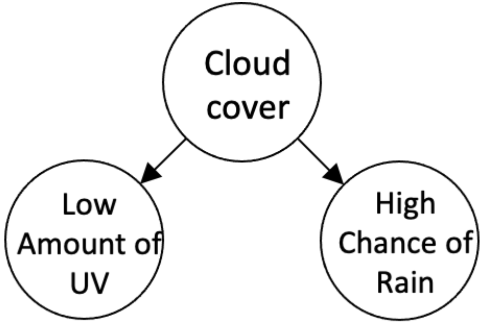
When learning these models from data:

Step 1: Identify **graph structure** from data – hard!
Step 2: Determine probabilities afterwards to estimate causal effects – easier!

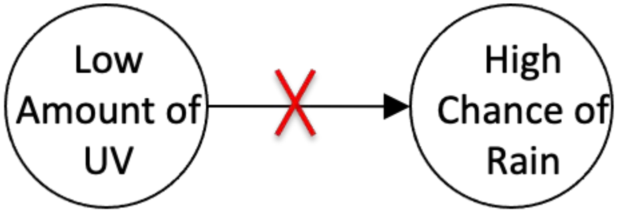
Often: Care only about graph structure.

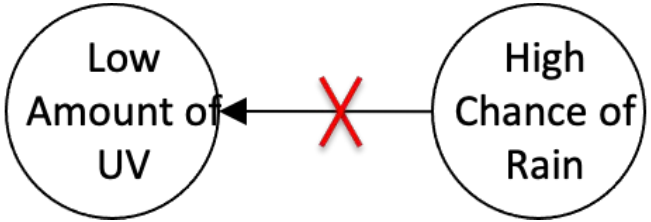
Concept 5: Hidden common causes (latent variables) makes things challenging!

Ex.: Cloud cover is common cause here
of “low UV” and “high chance of rain”.



If we remove the common cause (Cloud cover) in model:
→ Can no longer express a correct causal model with our standard arrow notation!





Three alternatives to dealing with latent variables:
1) Ensure to include all latent variables in model (usually impossible in earth science).
2) Consider arrows *only as a hypothesis* while *absence of arrows = absence of causality!*
3) Use latent causal discovery algorithms, but they are slow and tend to be statistically fragile.

How can we remove connections based on observed data?

The following 4 questions are equivalent:

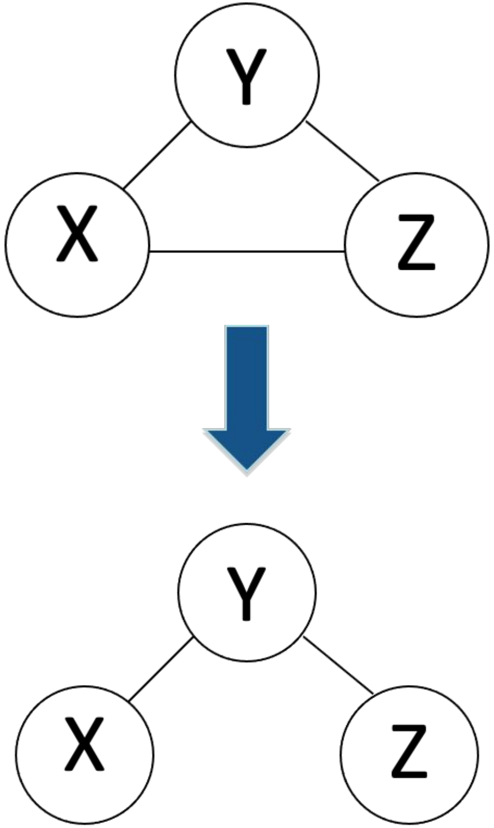
1) Can we eliminate edge between X and Z?

2) Is there direct connection between X and Z?

3) "Is X conditionally independent of Z *given* Y?"

4) Is P(X | Y , Z) ≈ P(X | Y) ?

If yes for any of the above: eliminate edge between X and Z.



→ Use conditional independence test:
Many statistical tests available to test for conditional independence.

A first simple algorithm for causal discovery - PC algorithm

Now we have:

Can use cond. independence test to detect and eliminate indirect connections (graph edges).

Basic algorithm for learning independence graph from data (PC algorithm):

1. Nodes of graph = observed variables.
2. **Start with fully connected graph** = assume that every variable is connected to every other variable.
3. **Eliminate as many edges as possible using conditional independence tests.**
4. Establish arrow directions (using constraints from independence tests or temporal constraints).

This is an elimination procedure.

Whatever edges are left at end: ***potential*** **causal connections** (causal hypotheses).

Why only *potential*?

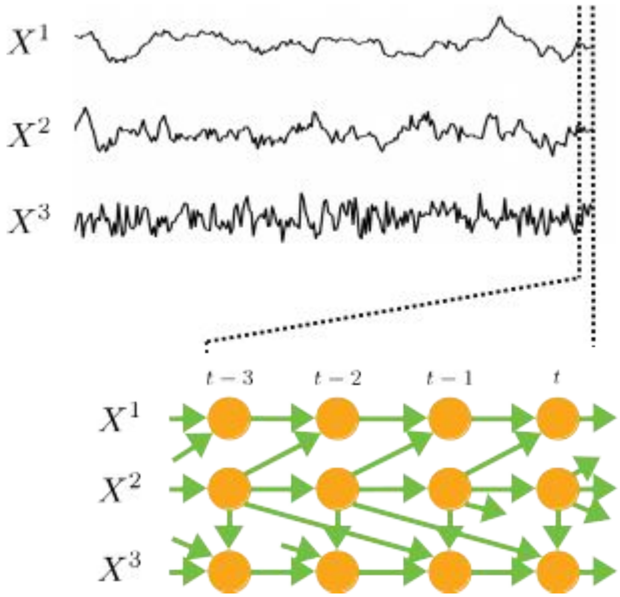
Because some of the connections might be due to latent variables (as discussed before) in this simple, but still powerful algorithm.

PC algorithm - named after **P**eter Spirtes and **C**lark Glymour who invented this algorithm.

PCMCI causal discovery framework for time series

Challenge:

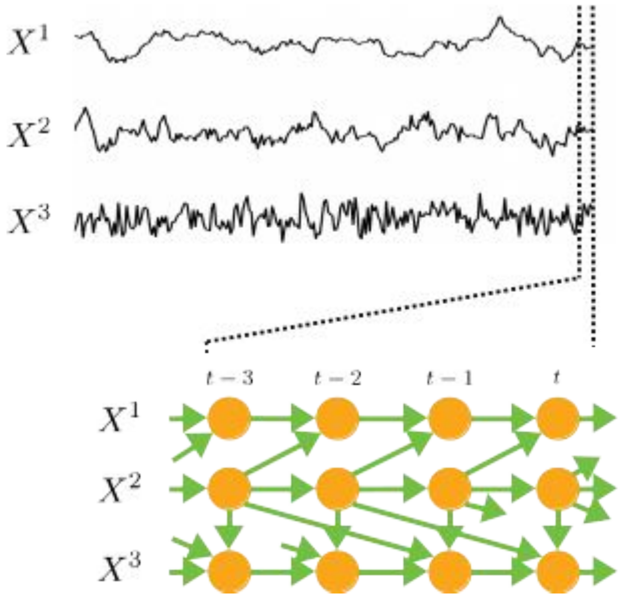
- Autocorrelation, time lags
- Contemporaneous links
- Latent variables



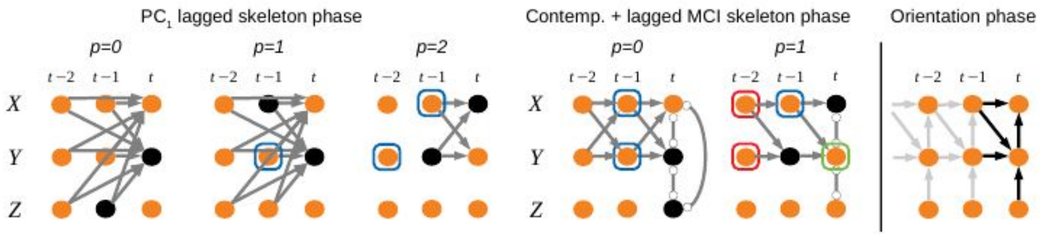
PCMCI causal discovery framework for time series

Challenge:

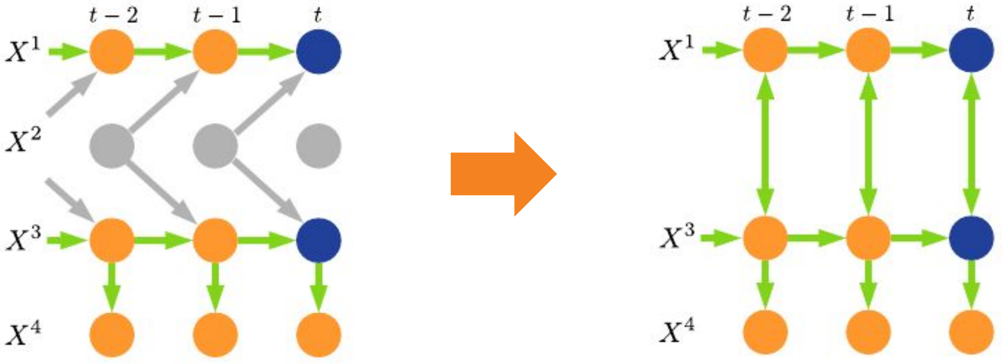
- Autocorrelation, time lags
- Contemporaneous links
- Latent variables



PCMCI, PCMCI+
(Runge et al. 2019, Runge 2020)



L(atent)PCMCI *(Gerhardus and Runge 2020)*



Causal Discovery 101 - Summary

- **Causal interpretation of reconstructed graphs requires caution:**
  - *depends on assumption that causal relations reliably leave imprint in statistical dependencies (Markov and Faithfulness assumption)*
  - *<u>some</u> methods assume no unobserved variables: links = potential cause-effect relationships*

Causal Discovery 101 - Summary

- **Causal interpretation of reconstructed graphs requires caution:**
  - *depends on assumption that causal relations reliably leave imprint in statistical dependencies (Makov and Faithfulness assumption)*
  - *<u>some</u> methods assume no unobserved variables: links = potential cause-effect relationships*

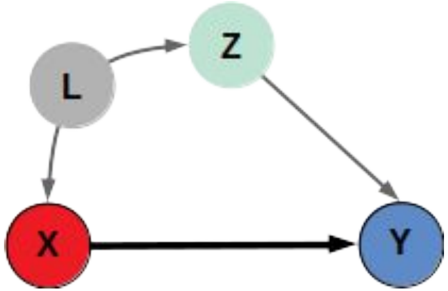
- But solid tool - **underutilized** in the geosciences where often just correlation and regression are used

- **Proposed Use:** Generate and test causal hypotheses

Causal Inference 101: Utilizing causal graphs

Task: Given causal graph and data, compute causal effect of intervention in terms of observational distribution *P(**V**)*

$$P\left(Y \middle| do(X{=}x)\right)$$

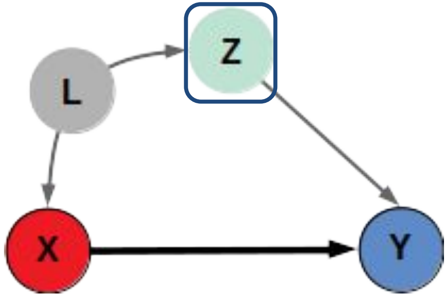


Causal Inference 101: Utilizing causal graphs

Task: Given causal graph and data, compute causal effect of intervention in terms of observational distribution *P(**V**)*

$$P\left(Y|do(X{=}x)\right)$$

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



Causal Inference 101: Utilizing causal graphs

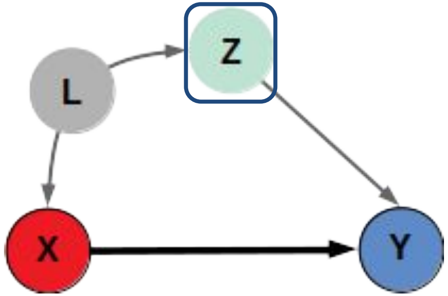
Task: Given causal graph and data, compute causal effect of intervention in terms of observational distribution *P(**V**)*

$$P\left(Y \middle| do(X{=}x)\right)$$

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

→ can be estimated with (deep) ML

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



Causal Inference 101: Utilizing causal graphs

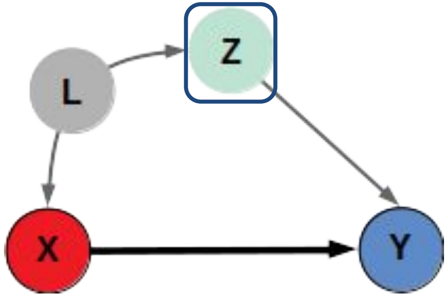
Task: Given causal graph and data, compute causal effect of intervention in terms of observational distribution *P(**V**)*

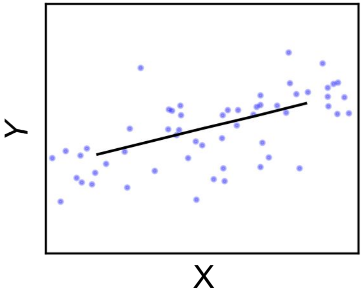
$$P\left(Y|do(X=x)\right)$$

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

→ can be estimated with (deep) ML

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





Causal Inference 101: Utilizing causal graphs

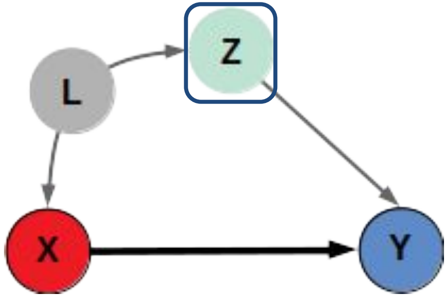
Task: Given causal graph and data, compute causal effect of intervention in terms of observational distribution *P(**V**)*

$$P\,(Y|do(X{=}x))$$

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

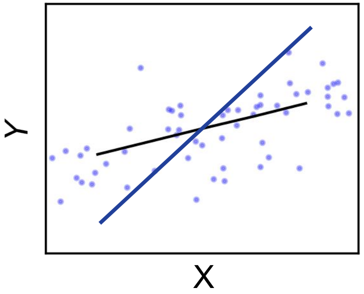
→ can be estimated with (deep) ML

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



Correlation
regression

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



Causal Inference 101: Utilizing causal graphs

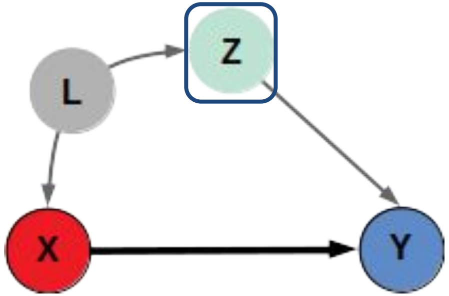
Task: Given causal graph and data, compute causal effect of intervention in terms of observational distribution *P(**V**)*

$$P\left(Y|do(X{=}x)\right)$$

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

→ can be estimated with (deep) ML

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

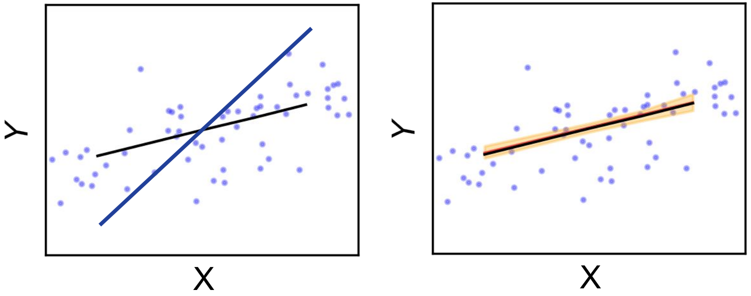


Correlation regression

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

Causal regression

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



Causal Inference 101: Utilizing causal graphs

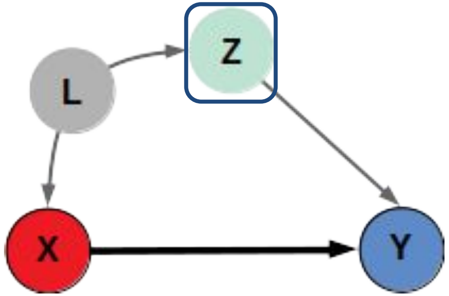
Task: Given causal graph and data, compute causal effect of intervention in terms of observational distribution *P(**V**)*

$$P\left(Y \middle| do(X=x)\right)$$

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

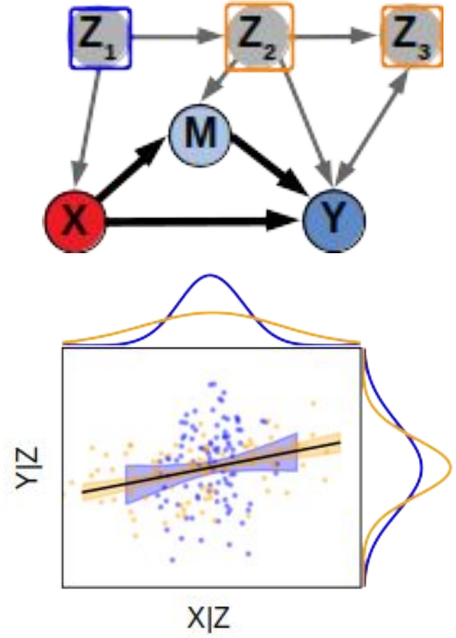
→ can be estimated with (deep) ML

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

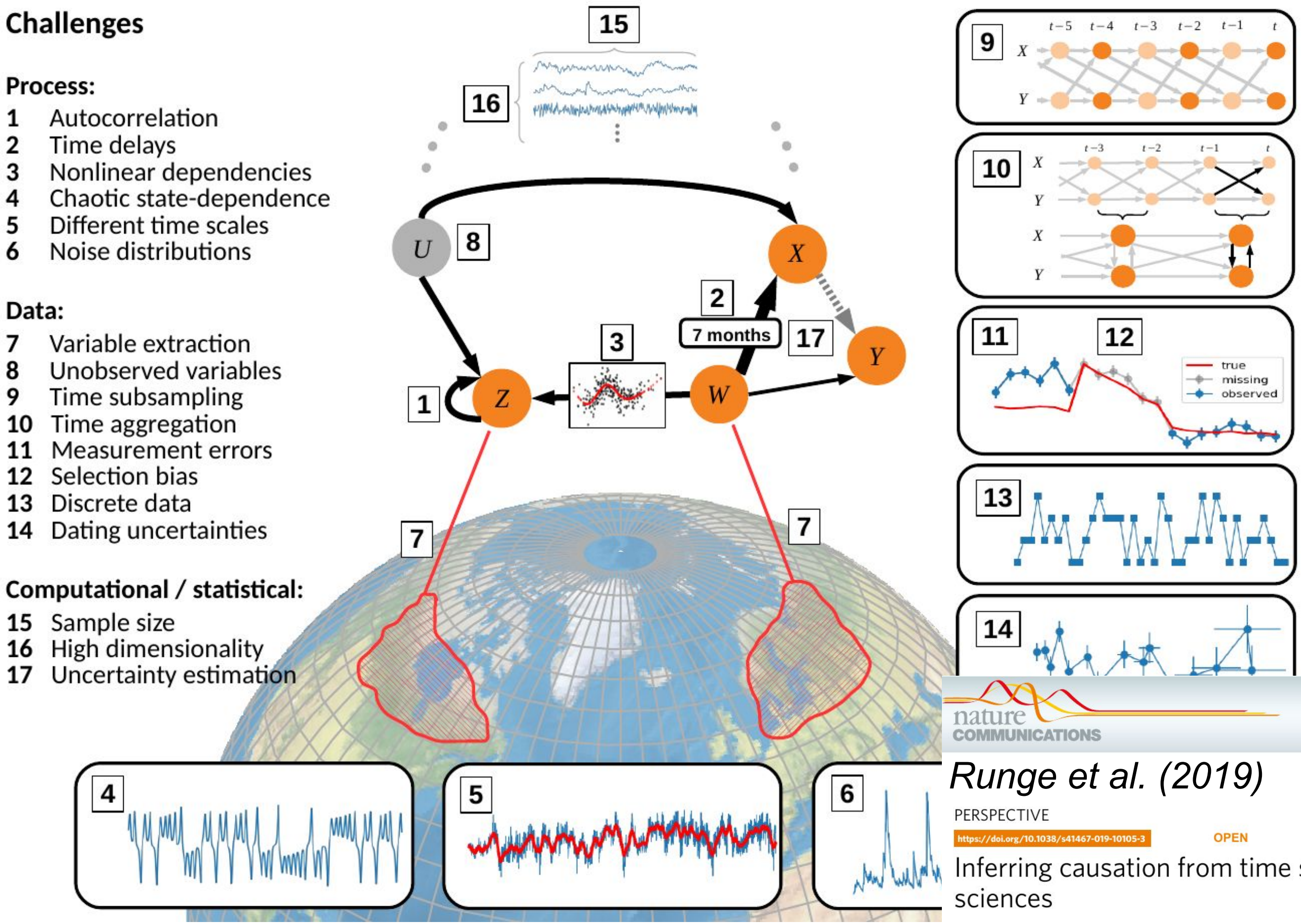


Optimal causal effect estimators

(Runge NeurIPS 2021)



Causal inference for earth science - special challenges



nature COMMUNICATIONS

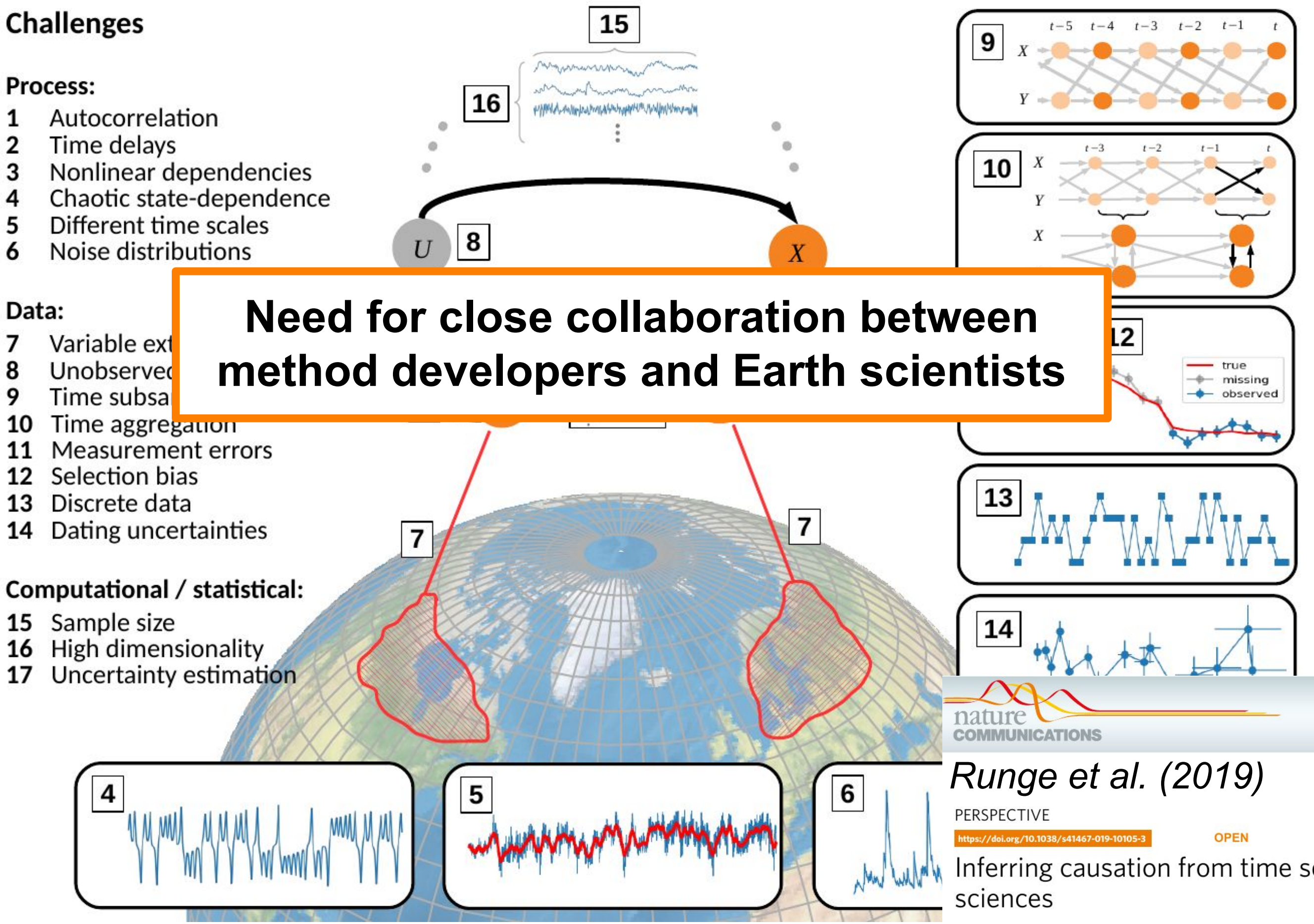
Runge et al. (2019)

PERSPECTIVE

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

Inferring causation from time series in Earth system sciences

Causal inference for earth science - special challenges



Need for close collaboration between method developers and Earth scientists

nature COMMUNICATIONS

Runge et al. (2019)

PERSPECTIVE

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

OPEN

Inferring causation from time series in Earth system sciences

Application use cases

- Learning causal graphs to understand mechanisms
- Quantifying causal mechanisms: link strength and mediation analysis

- Causally robust forecasting (e.g. Kretschmer et al. 2017, DiCapua 2019)

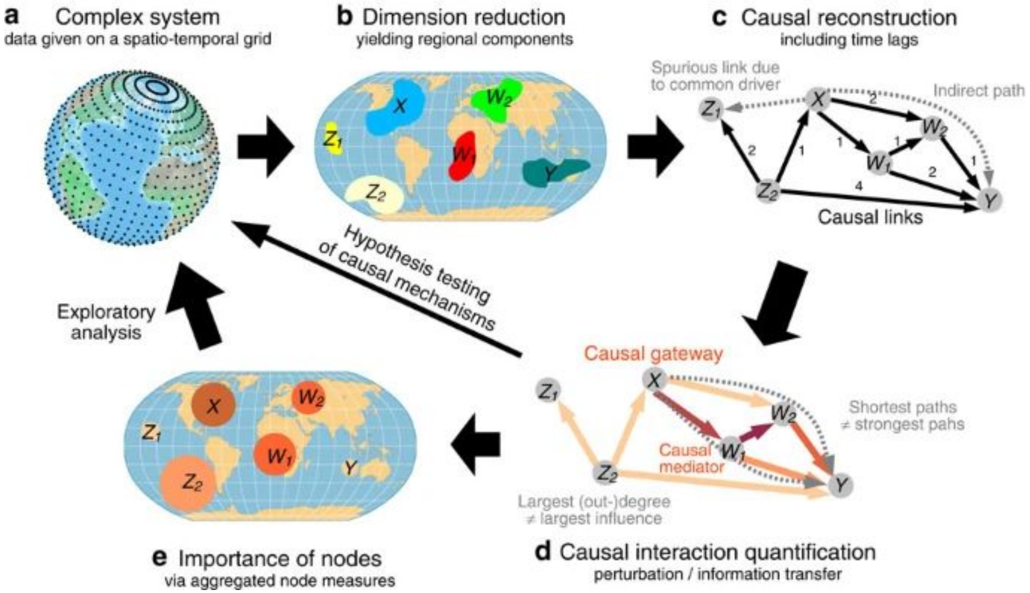
- Detection and attribution of extreme events (e.g. Hannart et al. 2016)

- Evaluating climate models and constraining climate change projections
- Hybrid physical-ML modeling

- ...

Causal inference and dimension-reduction

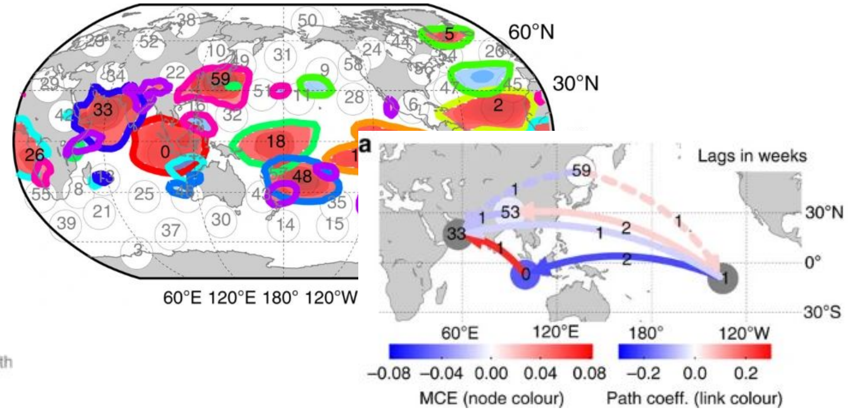
- **Motivation:** High-dimensionality and redundancy of spatio-temporal data
- **Idea**: First extract ‘modes of variability’

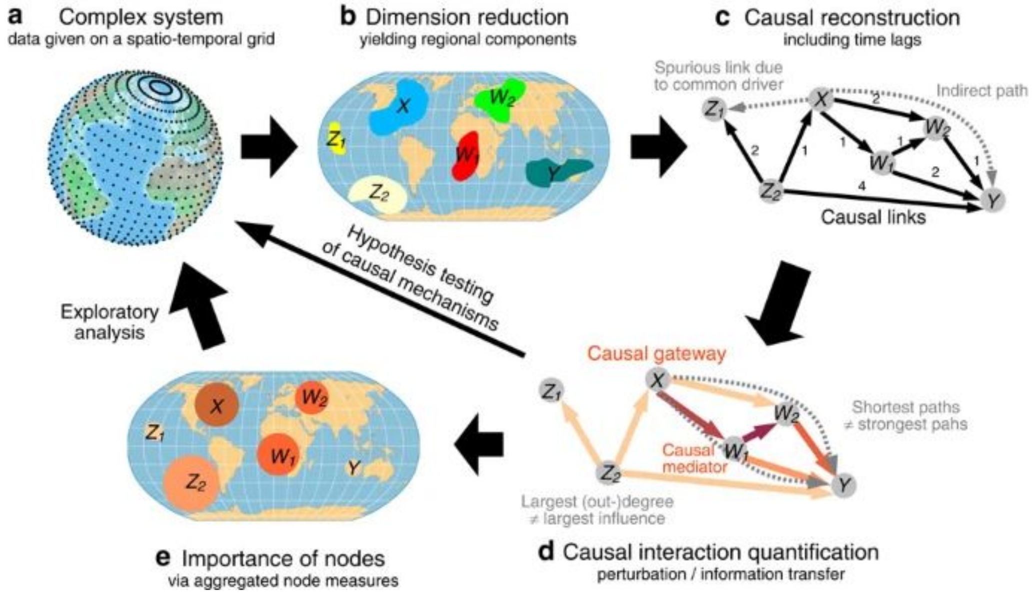


Runge et al. NatComm. (2015)

Causal inference and dimension-reduction

- **Motivation:** High-dimensionality and redundancy of spatio-temporal data
- **Idea**: First extract ‘modes of variability’

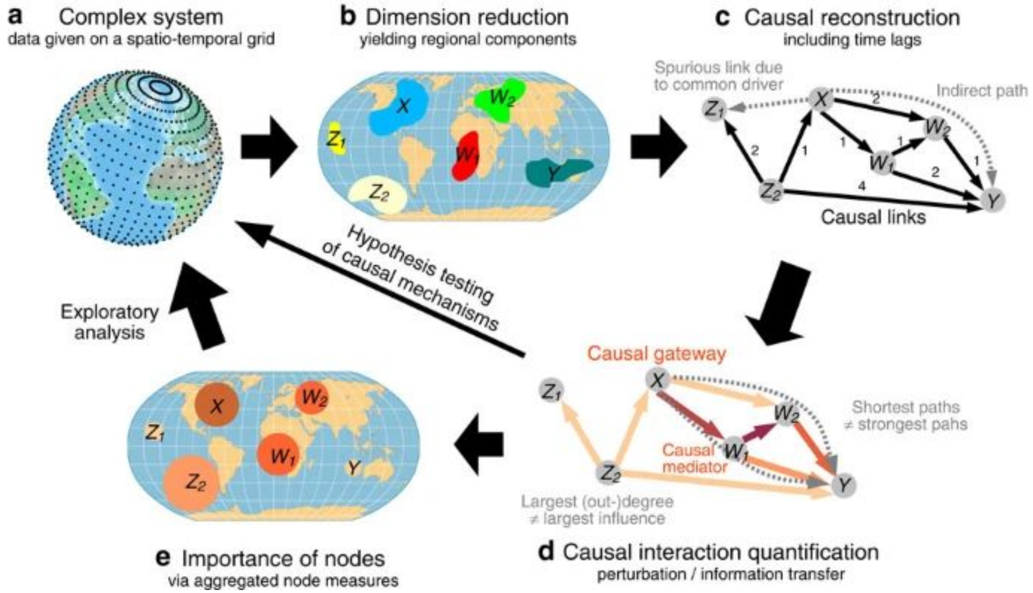


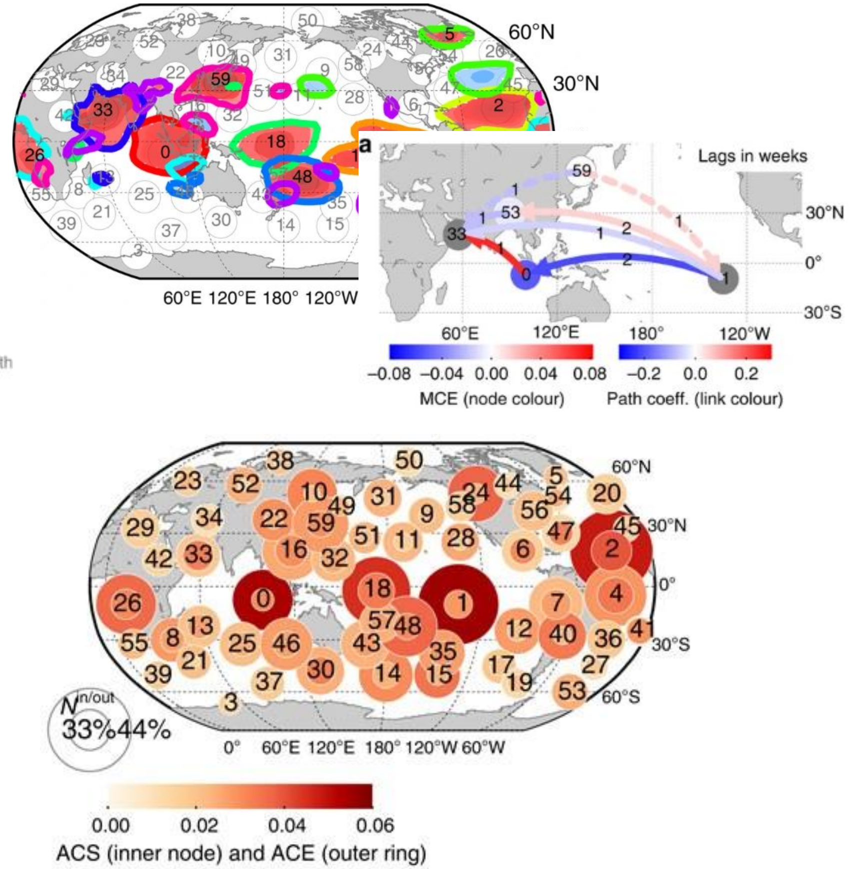


Runge et al. NatComm. (2015)

Causal inference and dimension-reduction

- **Motivation:** High-dimensionality and redundancy of spatio-temporal data
- **Idea**: First extract ‘modes of variability’

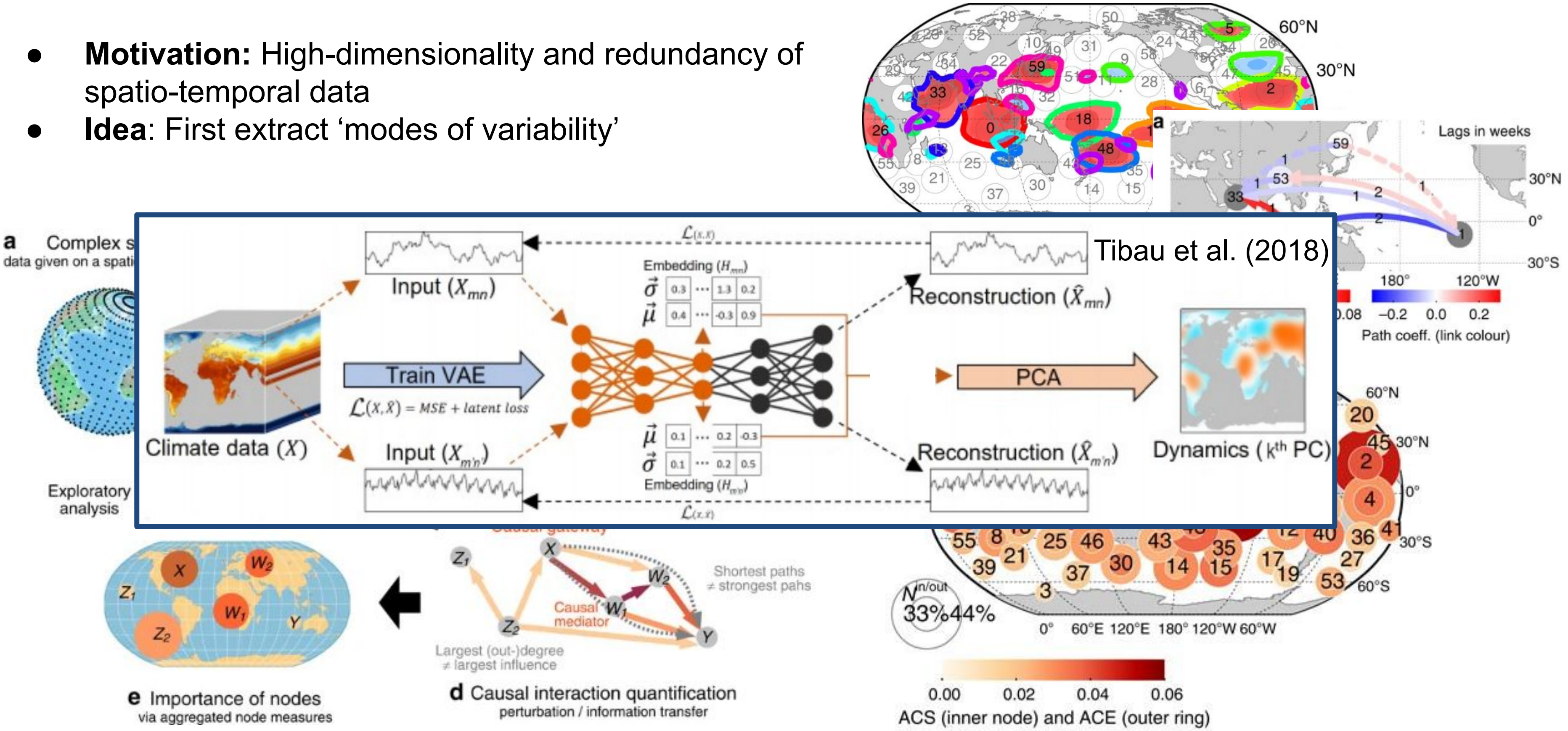




Runge et al. NatComm. (2015)

Causal inference and dimension-reduction

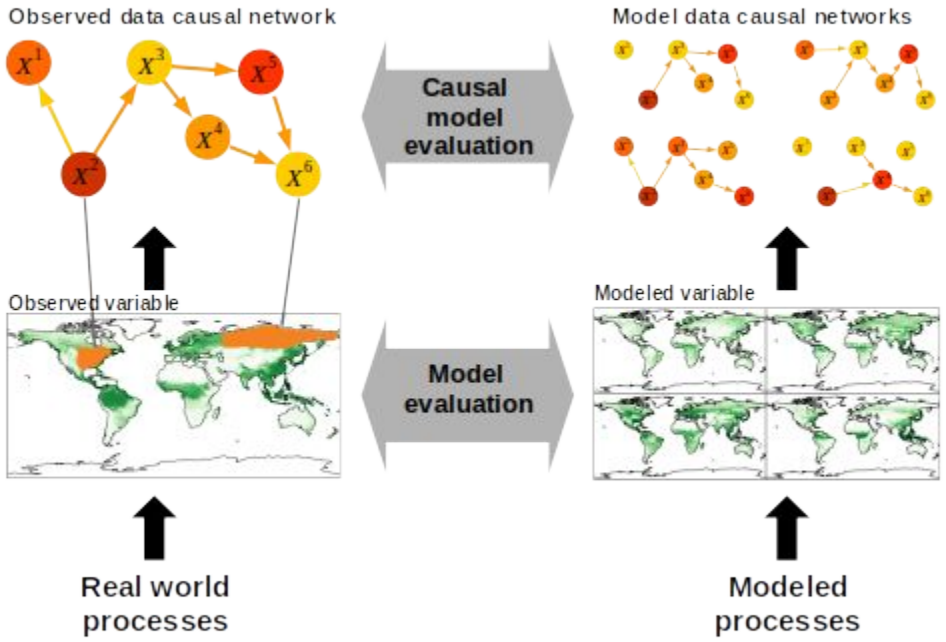
- **Motivation:** High-dimensionality and redundancy of spatio-temporal data
- **Idea**: First extract ‘modes of variability’

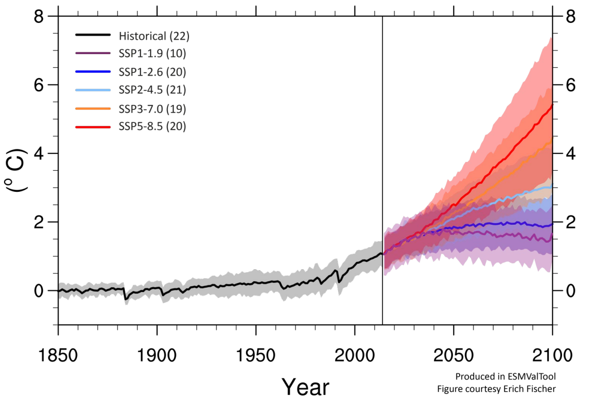


Runge et al. NatComm. (2015)

Causal physical model evaluation

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

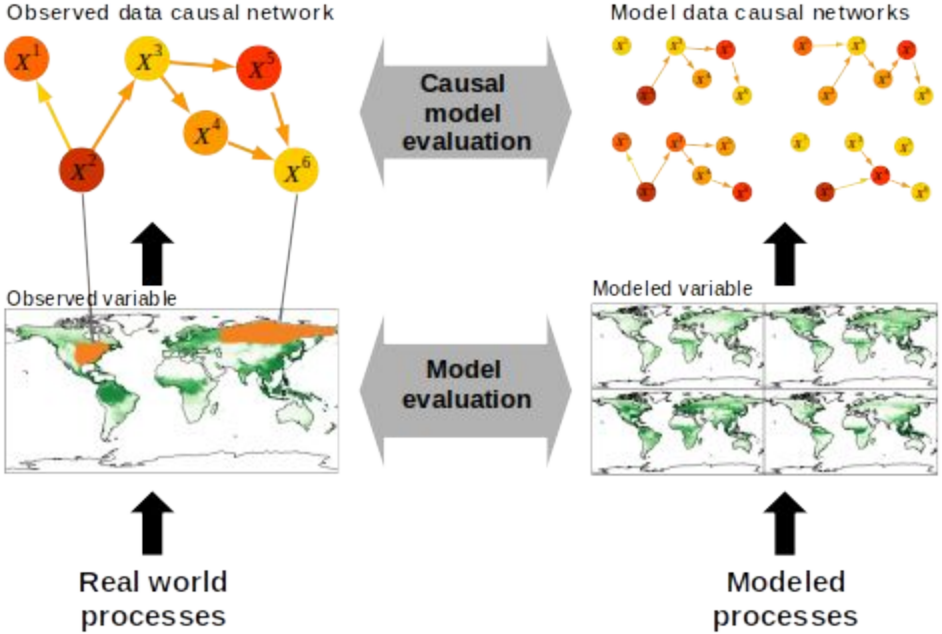


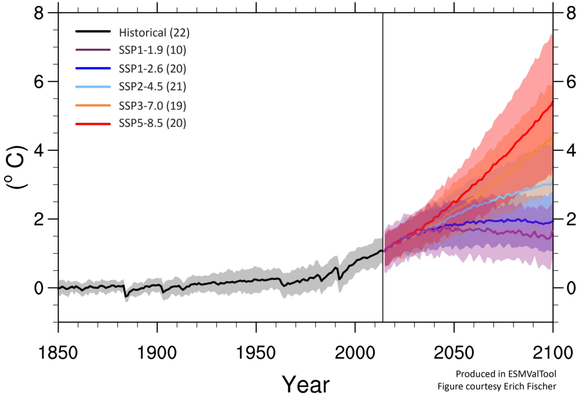


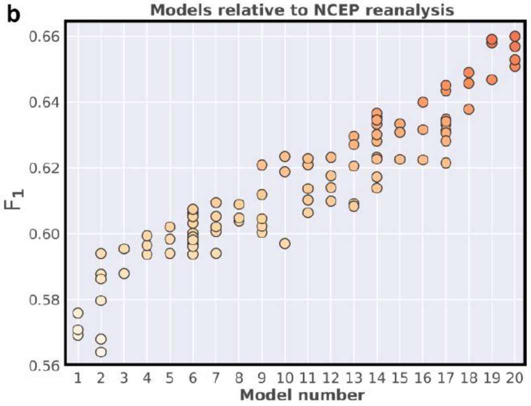
Nowack et al. NatComm. (2020)

Causal physical model evaluation

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



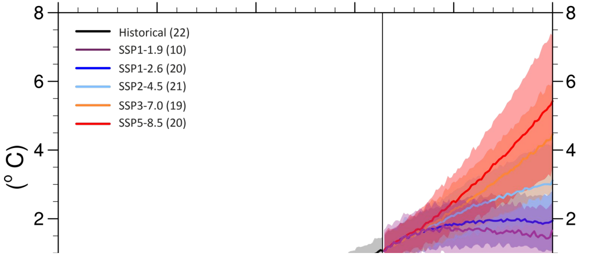




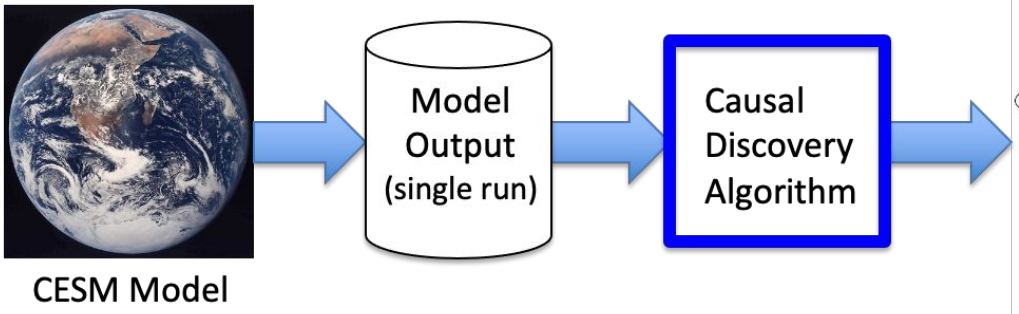
Nowack et al. NatComm. (2020)

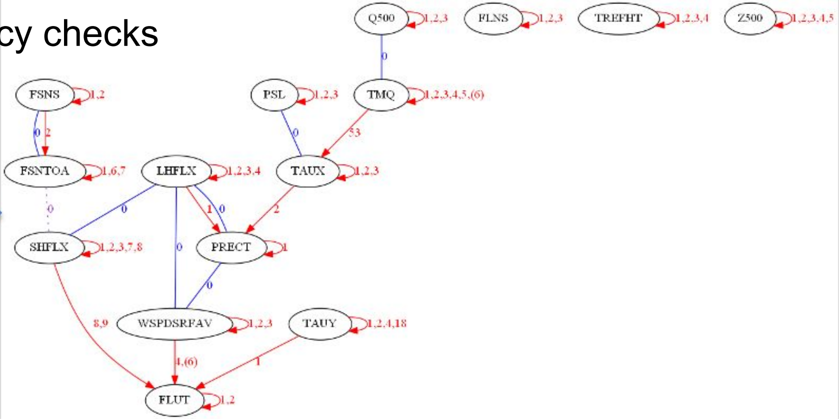
Causal physical model evaluation

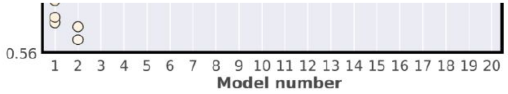
- **Motivation:** Simple statistics can be right for the wrong reasons
- **Idea**: Compare climate models and observations in terms of causal relationships



Baker et al. (2016), Hammerling et al. (2018): error/consistency checks





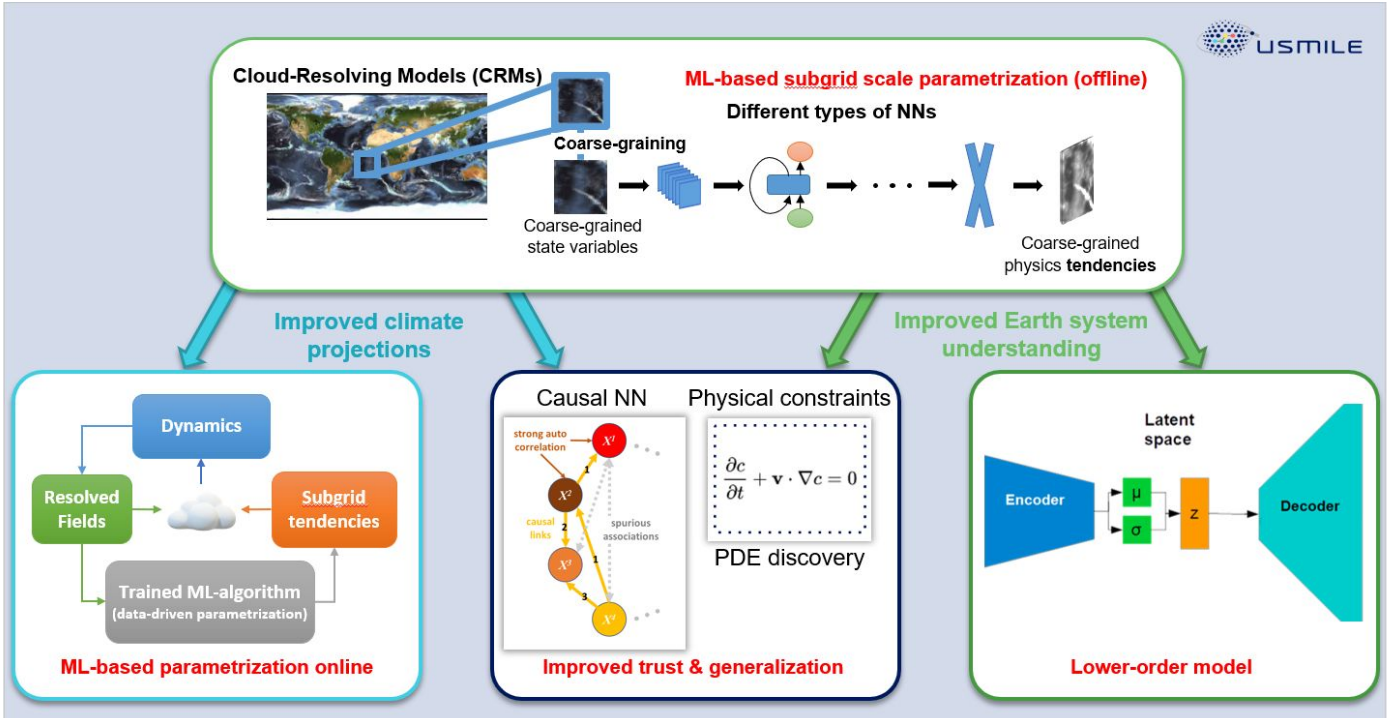


Nowack et al. NatComm. (2020)

Causal ML-hybrid modeling

Also see AI4Good talk by Reichstein next week!

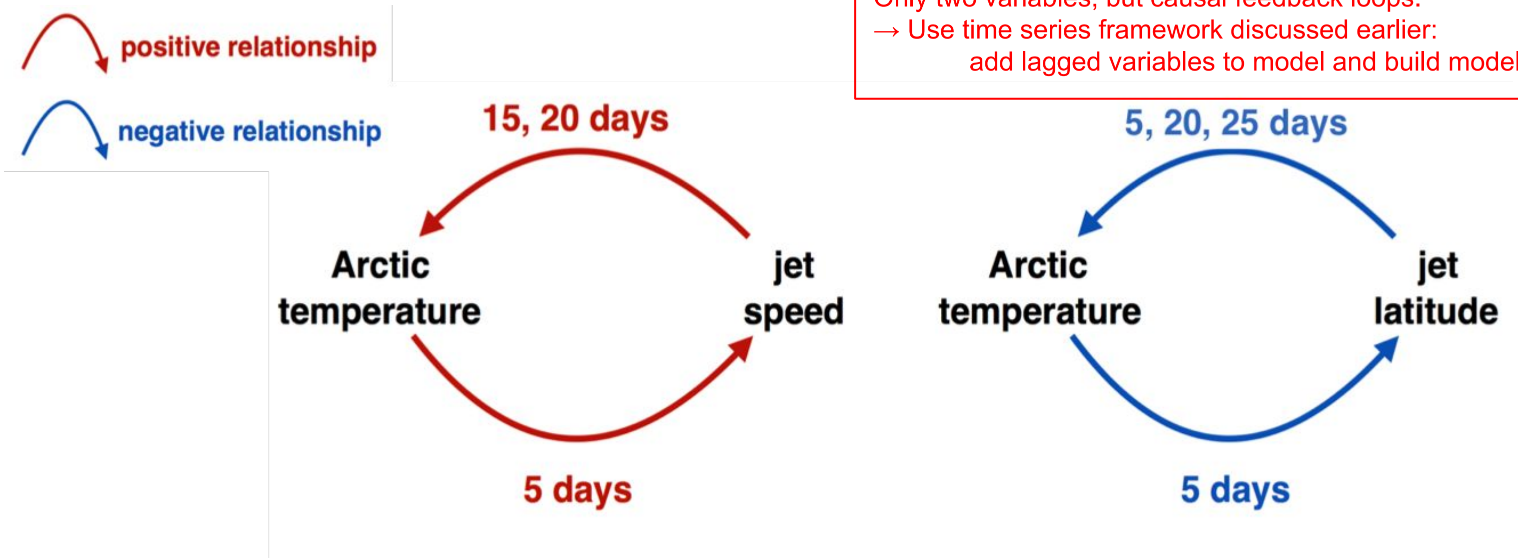
- Several previous AI4Good talks by Tapio Schneider, Bjorn Stevens, Chris Bretherton, Eyring and Gentine
- **Idea:** Restrict input for neural nets to causal drivers (see AI4Good by Eyring and Gentine)



Application: effect of arctic temp on jet stream

Science question: What is the effect of arctic temperature on speed / latitude of jet stream, and vice versa?

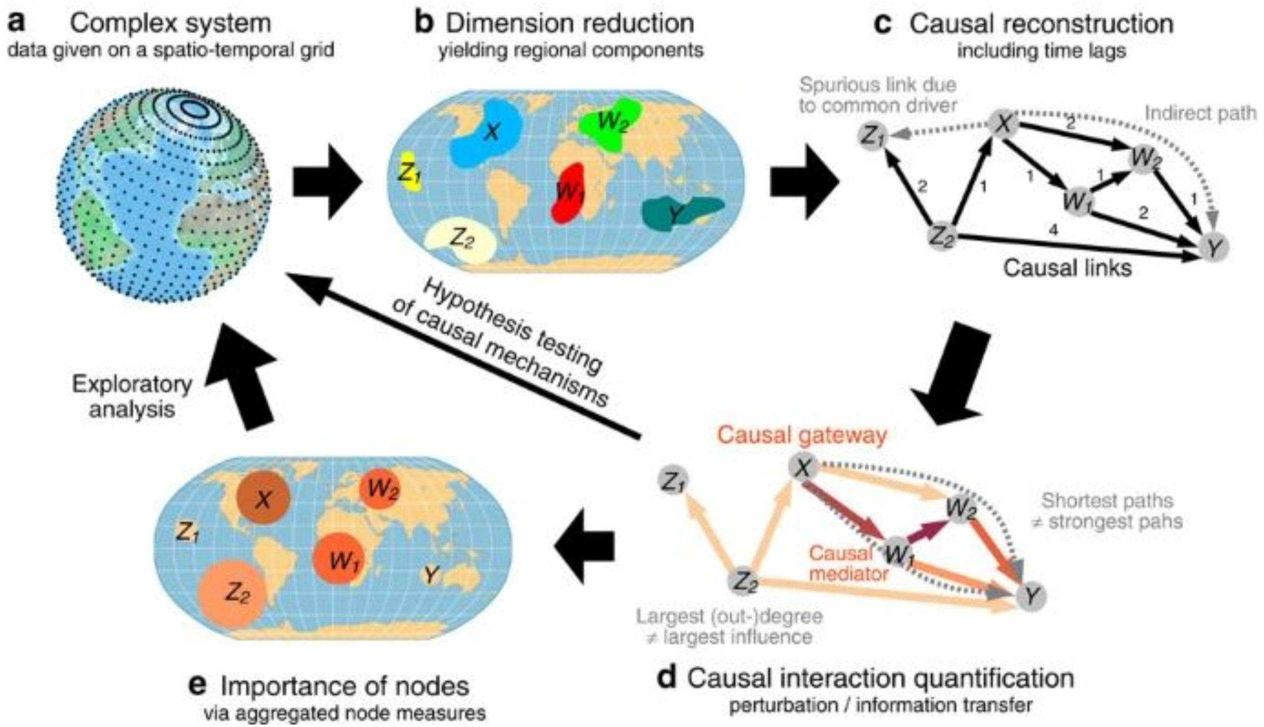
Only two variables, but causal feedback loops.
→ Use time series framework discussed earlier:
add lagged variables to model and build model including those.



- **Dominant relationships: Positive for jet speed, negative for jet latitude.**
- **Both are thus positive (reinforcing) feedback loops.**
- **Get time lags from analysis, too.**

Samarasinghe, McGraw, Barnes, Ebert-Uphoff, *A study of links between the Arctic and the midlatitude jet stream using Granger and Pearl causality*, Environmetrics, 2018.

Application: Spatially-distributed systems (Approach 1)



Approach 1:
Start with data on grid
→ dimensionality reduction
→ smaller number of nodes
→ build causal graph in reduced node space

Runge, J., Petoukhov, V., Donges, J.F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M. and Kurths, J., 2015. Identifying causal gateways and mediators in complex spatio-temporal systems. Nature communications, 6(1), pp.1-10.

Application: Spatially-distributed systems (Approach 2)

Approach 2: Use all nodes from original global grid
→ Challenge: Grid spacing can create artifacts!
→ Need to map data to special grid (here: Fekete points)

Goal: track information flow in the atmosphere.
Nodes: grid points (each with associated time series)
Input: Atmospheric field on global grid
Output: Causal interactions between grid points

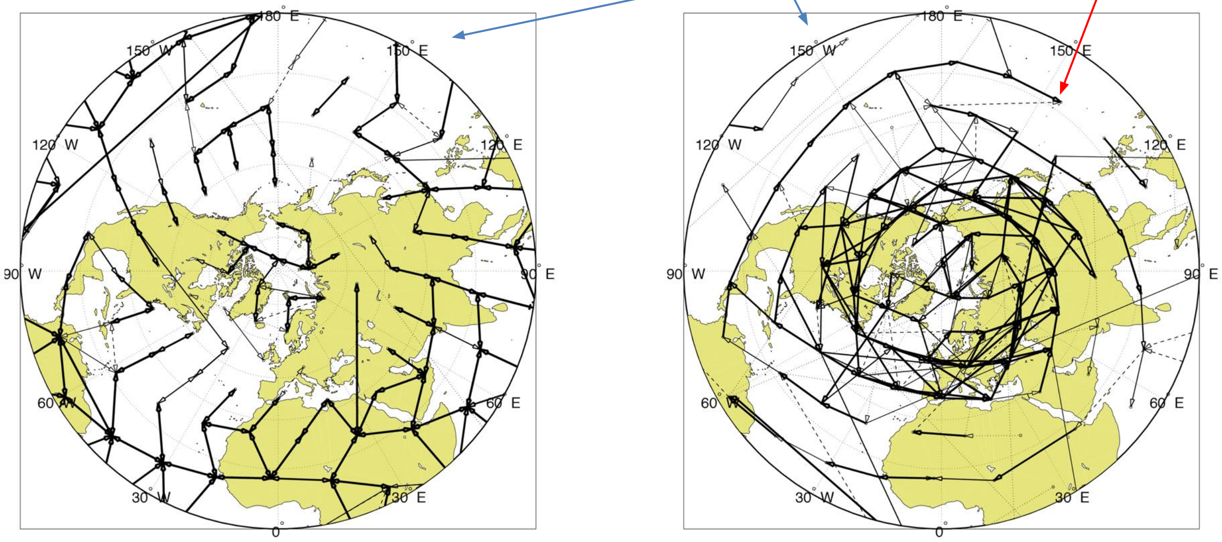
Information flow in the atmosphere

Sample input: **500 mb geopotential height**

Source:
NCEP/NCAR Reanalysis, 1948-2011

Results for boreal winter (Dec-Feb)

Algorithm: PC stable
(variation of PC algorithm)
with lagged variables



(a) 0-day-delay

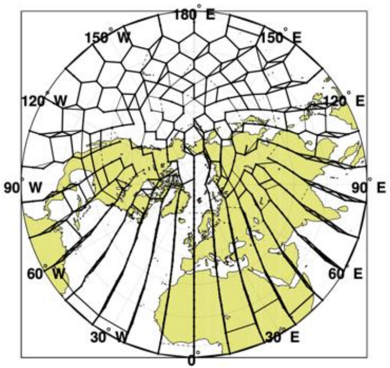
(b) 1-day-delay

Ebert-Uphoff and Deng, A New Type of Climate Network based on Probabilistic Graphical Models: Results of Boreal Winter versus Summer, Geophysical Research Letters, vol. 39, L19701, 2012.

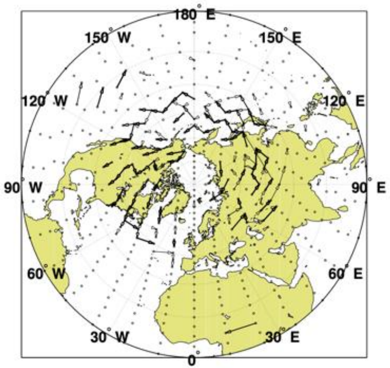
Application: Spatially-distributed systems (Approach 2)

Example of grid **artifacts**

Using equal area grid with 918 points
→ results distorted by uneven distance between neighboring points!

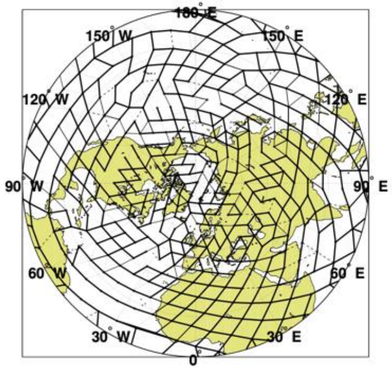


(a) Travel < 1 day

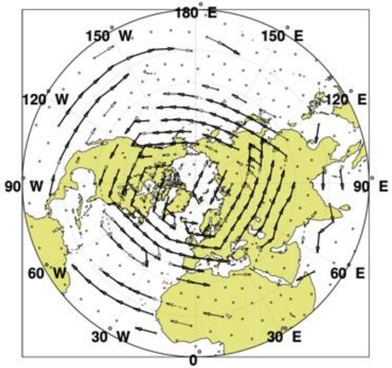


(b) Travel ≈ 1 day

Using Fekete grid with 800 points
→ points equally spaced
→ only small grid artifacts



(a) Travel < 1 day



(b) Travel ≈ 1 day

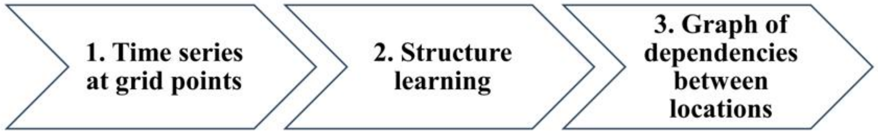
Ebert-Uphoff and Deng. "Causal discovery from spatio-temporal data with applications to climate science." In 2014 13th International Conference on Machine Learning and Applications, pp. 606-613. IEEE, 2014.

Application: Spatially-distributed system (Approach 3)

Approach 3: Start with data in global grid
→ Transform into spectral space
→ Causal discovery in spectral space

Causal discovery in Spectral Space
Goal: track interactions *between* processes occurring at different spatial scales

1) Causal discovery in grid space:



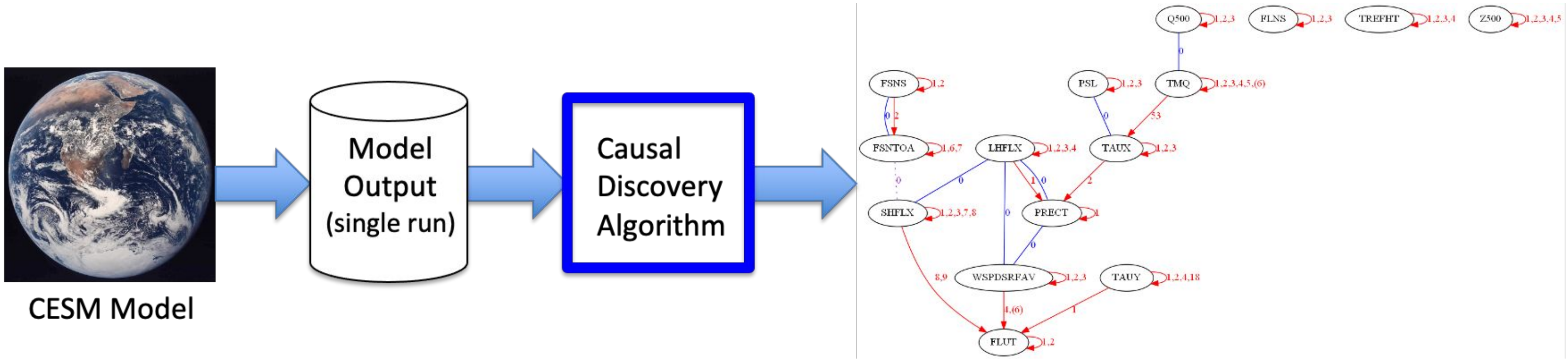
2) Causal discovery in spectral space using spherical harmonics (SH) for decomposition:

Use spherical harmonics (SH) decomposition for functions on a sphere

Nodes: real and imaginary coefficients of SH decomposition

Samarasinghe, Deng, Ebert-Uphoff, A Causality-Based View of the Interaction between Synoptic- and Planetary-Scale Atmospheric Disturbance, Journal of the Atmospheric Sciences, 77 (3): 925–941, 2020.

Determine “causal signatures” of climate model runs.



- Calculate “causal signature” for individual model outputs

  initial conditions), then compare their “signature”.

- First experiments: use only 15 variables, use **global avera**

- Applications: effect of compression, error check, understanding of differences between ensemble members or models.

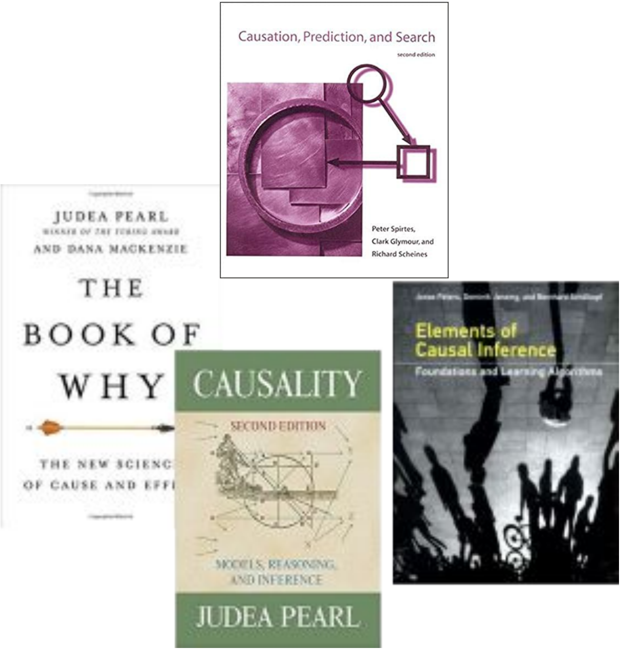
Baker et al., *[LINK]* Geoscentific Model Development, 2016

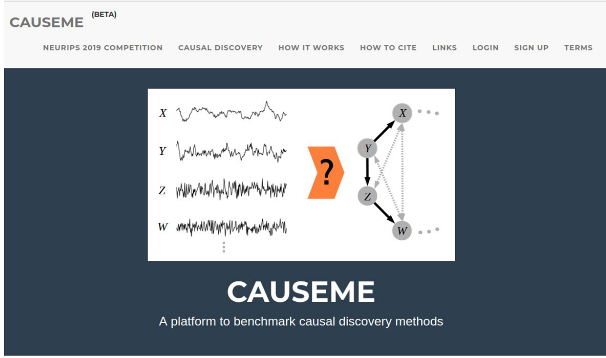
Take-home message

- **Causal inference**: Framework to answer causal questions from empirical data
- Two settings:
  - **Utilize qualitative causal knowledge** (graphs)
  - **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
- Primary goal of causal discovery is to generate ***hypotheses*** (for further investigation)
- **LOTS of opportunity for causality in earth science!**

Software:

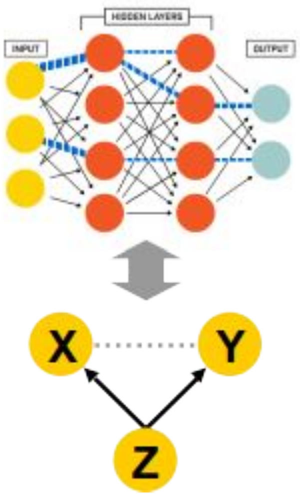
- Tigramite, pcalg, TETRAD, daggity, causalfusion, causeme platform, ...





Join the causal inference group at DLR / TU Berlin

Open postdoc positions – www.climateinformaticslab.com



erc Causal Earth

iMIRACLI

XAIDA

HELMHOLTZ AI

Join CIRA and AI2ES

Open position:
Data Visualization Researcher
(Research Associate II)

Apply by Feb 7, 2022.
See posting at
https://www.ai2es.org/opportunities/hiring/

More examples of AI research topics for weather/climate applications - from Imme's work

Cooperative Institute for Research in the Atmosphere (CIRA)

https://www.cira.colostate.edu/

Use AI to detect in satellite imagery:

- Convection initiation (severe weather)
- Cloud properties, vertical profiles (aviation),
- Gravity waves (understand energy transfer),
- Rapid intensification of tropical cyclones.

Use AI (image-to-image translation) to generate:

- Synthetic radar imagery (severe weather)
- Synthetic passive microwave imagery (tropical cyclones).

Use AI to emulate radiative transfer equations
→ speed up numerical weather prediction models.

Funded by NOAA/NASA.

NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)

https://www.ai2es.org/

Make AI trustworthy for earth science:

- Simplify AI methods to make them more robust and interpretable;
- Develop explainable AI (XAI) methods for weather/climate/coastal application;
- Work with risk communication scientists and forecasters to identify AI and XAI needs for operational weather forecasting settings.

Funded by National Science Foundation (NSF)

Thank you! Questions?