Causal Discovery for Climate Time Series in the Presence of Unobserved Variables

Andreas Gerhardus and Jakob Runge
German Aerospace Center, Institute of Data Science, Jena

EGU2020: Sharing Geoscience Online

DLR

Causal Relationships and Their Inference by Experimentation

- Bypassing a major philosophical debate, we adopt the following definition of causality:

  X is a cause Y if changing the value of X while keeping all other conditions the same leads to a different value of Y

- The classical method of empirically inferring causal relationships is by experimentation:

  Set up an experiment that changes the value of *X* without affecting other variables. If the value of *Y* changes when *X* changes, then *X* is a cause of *Y*

- Example:

  X: Ceiling light is on   *Y*: Room is illuminated

  X is a cause of *Y*: Turning the light on, the room is illuminated
  Y is not a cause of *X*: Illuminating the room by, say, a flashlight does not turn on the light

Inference of Causal Relationships from Observational Data?

- **Causal discovery aims to infer causal relationships from observational data** [1]
- Given the above definition of causality, this task comes with the following fundamental challenge:

  The data is already there, it has been generated without us controlling the experimental conditions. That is, we cannot intervene to change the value of some variables and then observe what happens to the other variables

- How about statistical measures such as correlation or non-linear generalizations thereof, e.g. mutual information?

  By themselves, statistical dependencies do not imply causation

- **Then, is causal discovery possible at all?**

  Yes… … when making some additional assumptions

[1]Selection of relevant textbooks:
Pearl, J. Causality: Models, Reasoning, and Inference
Spirtes, P., Glymour, C., and Scheines, R. Causation, Prediction, and Search.
Peters, J., Janzing, D., and Schölkopf, B. Elements of Causal Inference: Foundations and Learning Algorithms.

A Framework for Causal Discovery

- A common framework representing causal relationships is that of **causal graphs** and **structural causal models (SCMs)** [1]

- Causal graphs:

  Nodes represent variables, arrows represent direct causal relationships

  Example: $W \rightarrow X \rightarrow Z \leftarrow Y$

  → *X* and *Y* are direct causes of *Z*
  → *W* is a direct cause of *X* and an indirect cause of *Z*

- Structural causal models (which imply causal graphs):

  Specification of the functional relationships that determine the value of each variable from those of the other variables

  Example cont.: $X = f(W), \quad Z = g(X, Y)$

[1] See for example:
Pearl, J. Causality: Models, Reasoning, and Inference
Spirtes, P., Glymour, C., and Scheines, R. Causation, Prediction, and Search.
Bollen, K. Structural Equations with Latent Variables.

Causal Graphs and Statistical Independencies: Part 1

- **Assumption 1:**

  The observed data was generated by a process that is expressable as SCM

- Discussion:

  Equilibrium states of ordinary differential equations and random differential equations can be described by SCMs [1] [2]

- Consequence:

  The structure of the corresponding causal graph implies statistical independencies

  Example cont.: $W \rightarrow X \rightarrow Z \leftarrow Y$

  $\Rightarrow$ *W* conditionally independent of *Z* given *X* (causal influence is mediated by *X*)
  $\Rightarrow$ *X* and *W* are marginally independent of *Y* (colliding arrows at *Z* block influence)

  General rule: d-separation [3]

[1]Mooij, J. M., Janzing, D., and Schölkopf, B. From Ordinary Differential Equations to Structural Causal Models: The Deterministic Case.
[2]Bongers, S., Mooij, J. M.. From Random Differential Equations to Structural Causal Models: The Stochastic Case.
[3]Verma, T. S., Pearl, J. Causal Networks: Semantics and Expressiveness.

Causal Graphs and Statistical Independencies: Part 2

- **Assumption 2:**

  All statistical independencies are implied by d-separation on the causal graph [1]

- Discussion:

  Intuitively, this excludes „accidental“ independencies due to fine-tuned parameters

  Weaker forms of this assumption exist [2]

- Consequence:

  Statistical independencies constrain the structure of the causal graph

- **Constraint-based causal discovery:**

  Perform tests of statistical (in-)dependence in the observed data to constrain the causal graph as much as possible, thereby inferring causal relationships

[1]See notions of “minimality” in Pearl, J. Causality: Models, Reasoning, and Inference, and “faithfulness” in Spirtes, P., Glymour, C., and Scheines, R. Causation, Prediction, and Search.
[2]For example: Ramsey, J., Spirtes, P., and Zhang, J. Adjacency-Faithfulness and Conservative Causal Inference.

Unobserved Causally Relevant Variables

- In practice, we won‘t observe every single variable that is involved in the physical process under investigation
- However, some of the unobserved variables may be causally relevant:

  Z is causally relevant if it is a cause of two observed variables X and Y (and the causal influence of Z on Y is not entirely mediated through X, nor vice versa) [1]

  If Z is unobserved, it is called a hidden confounder or a hidden common cause
- This complicates the inference of causal relationships for the following reason:

  Say we observe a statistical dependence between X and Y, and this dependence cannot be blocked off by conditioning on some other observed variables

  If there are no hidden confounders, we can conclude that X causes Y or vice versa

  If there are hidden confounders, we cannot draw this conclusion

[1]See notion of “causally sufficient” in Spirtes, P., Glymour, C., and Scheines, R. Causation, Prediction, and Search.

Causal Discovery With and Without Causal Sufficiency

- Optional assumption: Causal Sufficiency

  There are no hidden confounders, i.e., all causally relevant variables are observed

- Two important examples of constraint-based causal discovery algorithms:
  - **PC-Algorithm:** **Assumes causal sufficiency** [1] [2]
  - **FCI-Algorithm:** **Does not assume causal sufficiency** [2] [3] [4]
- Comparison:
  - FCI makes fewer assumption as PC
  - FCI is computationally and statistically more involved, it tends to assert fewer causal relationsships

[1]Spirtes, P., Glymour, C. N. An algorithm for fast recovery of spares causal graphs.
[2]Spirtes, P., Glymour, C., and Scheines, R. Causation, Prediction, and Search.
[3]Spirtes, P., Meek, C., and Richardson, T. S. An algorithm for causal inference in the presence of latent variables and selection bias.
[4]Zhang, J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias.

Example of Discovering Causal Graphs with PC and FCI

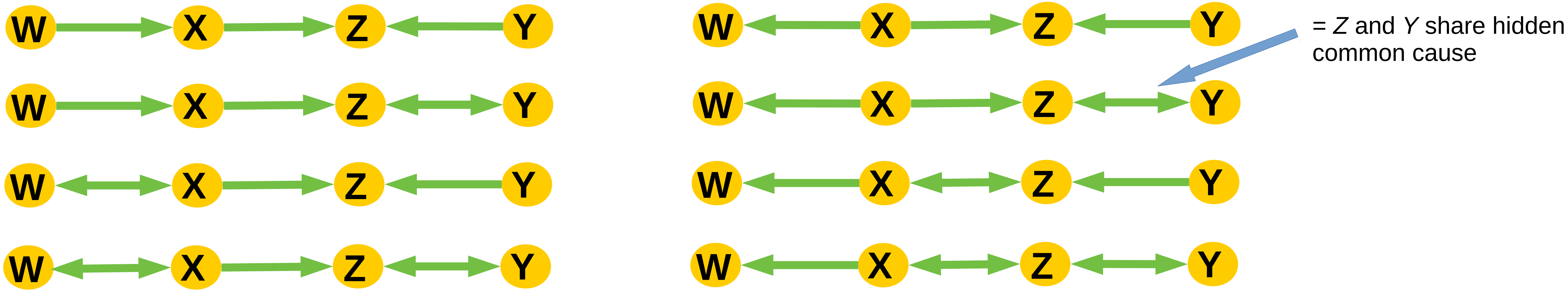
- Ground Truth: 

  W conditionally independent of *Z* given *W*

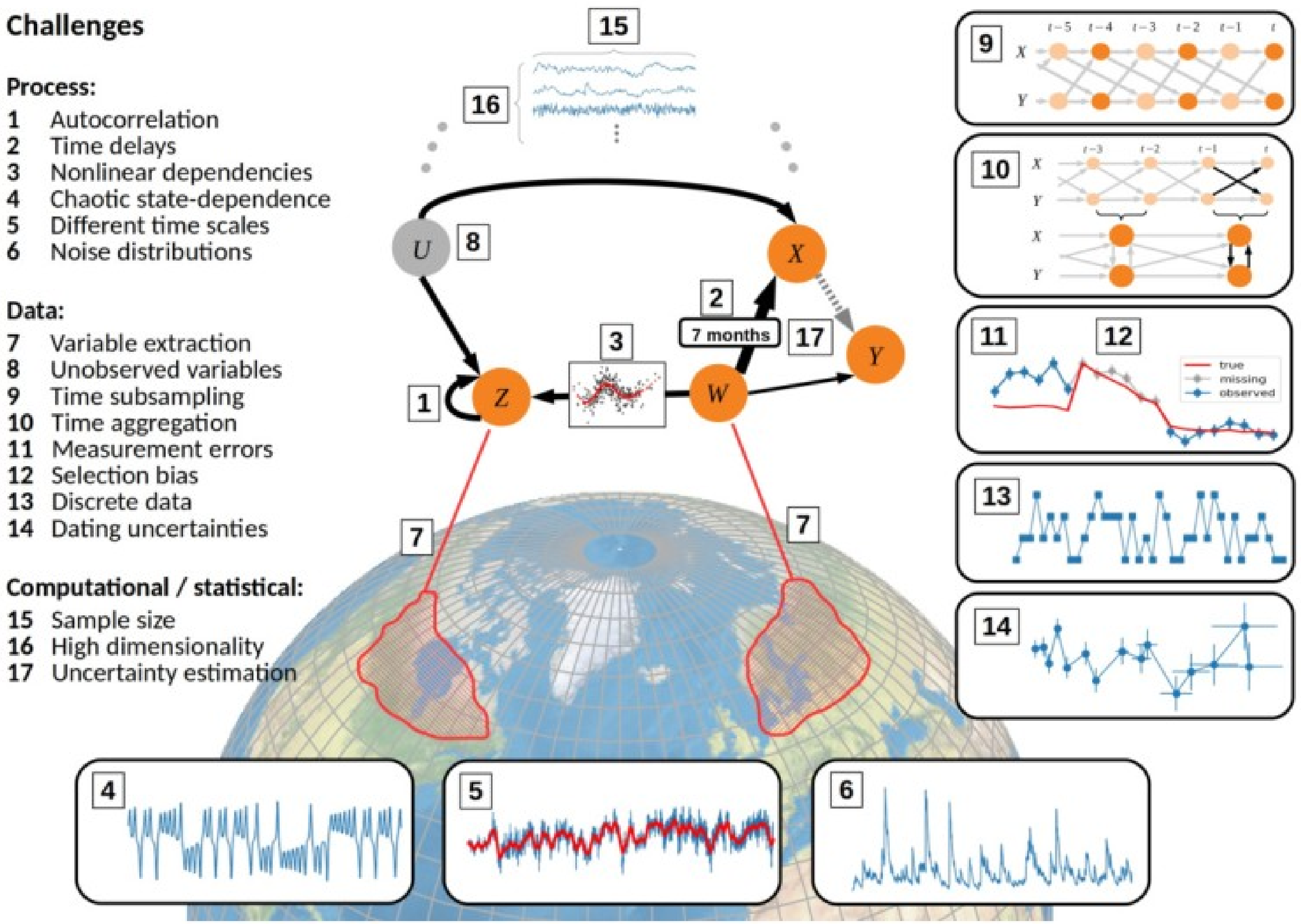
  X and *W* are marginally independent of *Y*

- Output of PC-Algorithm: 2 structures consistent with this exact set of independencies

- Output of FCI-Algorithm: 10 structures consistent with this exact set of independencies



Challenges for Causal Discovery in Climate Time Series



Picture from: Runge, J., Bathiany, S., Bollt, E. et al. Inferring causation from time series in Earth system sciences.

- Focus here:

  Autocorrelation

- Challenge posed:

  High rate of wrong statistical decisions

- Goal:

  Modify and adapt causal discovery algorithms to become statistically more reliable and informative on autocorrelated climate time series data

Previous Work: Causal Discovery in Climate Time Series With Causal Sufficiency

- In previous work we introduced the **PCMCI-Algorithm**, a modification of the PC-Algorithm to **better handle autocorrelated time series** [1]
- Central ideas of PCMCI:
  - MCI conditional independence tests for well calibrated tests with improved detection power [1]
  - Make fewer tests of statistical (in-)dependencies in total
- Result:

  Significantly improved recall and well controlled false positives
- **Limitation** of PCMCI:

  PCMCI makes the **assumption of causal sufficiency**, i.e., it assumes that there are no hidden confounders

[1]Runge, J., Nowack, P., Kretschmer, M. et al. Detecting and quantifying causal associations in large nonlinear time series datasets.

Current Work: Causal Discovery in Climate Time Series Without Causal Sufficiency

- Currently, we are working on an algorithm that **generalizes the key PCMCI ideas to the FCI-Algorithm, i.e., the case when there may be hidden confounders**
- This requires significant changes and new conceptual ideas
- Numerical experiments are under way

- Outlook:
  - Allow for selection bias
  - Approach other challenges for causal discovery in climate time series [1]

[1]See Runge, J., Bathiany, S., Bollt, E. et al. Inferring causation from time series in Earth system sciences for a discussion

References

- *Pearl, J. Causality: Models, Reasoning, and Inference. Cambridge University Press, New York, USA, 2nd edition (2009)*
- *Spirtes, P., Glymour, C., and Scheines, R. Causation, Prediction, and Search. The MIT Press, Cambridge, Massachusetts, 2nd edition (2000).*
- *Peters, J., Janzing, D., and Schölkopf, B. Elements of Causal Inference: Foundations and Learning Algorithms. The MIT Press, Cambridge, Massachusetts (2017).*
- *Bollen, K. Structural Equations with Latent Variables. John Wiley & Sons, New York, USA (1989).*
- *Mooij, J. M., Janzing, D., and Schölkopf, B. From Ordinary Differential Equations to Structural Causal Models: The Deterministic Case. UAI'13: Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (2013)*
- *Bongers, S., Mooij, J. M.. From Random Differential Equations to Structural Causal Models: The Stochastic Case. arXiv preprint arXiv:1803.08784 (2018).*
- *Verma, T. S., Pearl, J. Causal Networks: Semantics and Expressiveness. Machine Intelligence and Pattern Recognition vol. 9 (1990).*
- *Ramsey, J., Spirtes, P., and Zhang, J. Adjacency-Faithfulness and Conservative Causal Inference. UAI'06: Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (2006)*
- *Spirtes, P., Glymour, C. N. An algorithm for fast recovery of spares causal graphs. Social Science Computer Review vol. 9, issue 1 (1991)*
- *Spirtes, P., Meek, C., and Richardson, T. S. An algorithm for causal inference in the presence of latent variables and selection bias. UAI'95: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence (1995)*
- *Zhang, J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. Artificial Intelligence vol. 172, issues 16–17 (2008)*
- *Runge, J., Bathiany, S., Bollt, E. et al. Inferring causation from time series in Earth system sciences. Nat Commun 10, 2553 (2019).*
- *Runge, J., Nowack, P., Kretschmer, M. et al. Detecting and quantifying causal associations in large nonlinear time series datasets. Science Advances vol. 5, no. 11 (2019)*