Causal discovery in time series with unobserved confounders

Andreas Gerhardus*
March 17, 2021
Joint IS-ENES3/ESiWACE2 Virtual Workshop on New Opportunities for ML and AI in Weather and Climate Modelling

*Climate Informatics Group
DLR-Institute of Data Science
Jena, Germany
Introduction
Motivation: Complex dynamics of the climate system

System of interest:

\[
X^1 \\
X^2 \\
X^3
\]

Goal:
Contribute to a better understanding of Earth’s complex weather and climate system.
Climate Informatics in general:
Use modern tools of machine learning, statistics, and data science to aid climate and Earth system sciences.
Climate Informatics in general:
Use modern tools of machine learning, statistics, and data science to aid climate and Earth system sciences.

Focus of the Climate Informatics Group @DLR Jena*:
- Development of methods
- Provisioning of open-source software implementations† for application by domain scientists
- Methods based on the modern causal inference framework

*www.climateinformaticslab.com
†https://github.com/jakobrunge/tigramite
Causal inference
Causal inference:

- Defines notions of *cause* and *effect* in a mathematical framework.
- Casts causal questions within this framework.
- Specifies assumptions and develops methods for answering these questions.
Causal inference and causal discovery

Causal inference:
- Defines notions of *cause* and *effect* in a mathematical framework.
- Casts causal questions within this framework.
- Specifies assumptions and develops methods for answering these questions.

Sub-field: Causal discovery
- Specifies assumptions and develops methods for learning cause and effect relationships from observational data.
On the notion of causation

Correlation is not causation:
Statistical dependencies in observational data do not by themselves imply causal relationships.
⇒ Need assumptions to connect stat. dependencies and causation
On the notion of causation

Correlation is not causation:
Statistical dependencies in observational data do not by themselves imply causal relationships.
⇒ Need assumptions to connect stat. dependencies and causation

Working definition of causality:
Variable $X$ causes variable $Y$ if an experimental manipulation that changes $X$ (and only $X$) leads to a change of $Y$.
⇒ experimental mode of inferring causation
On the notion of causation

Correlation is not causation:
Statistical dependencies in observational data do not by themselves imply causal relationships.
⇒ Need assumptions to connect stat. dependencies and causation

Working definition of causality:
Variable X causes variable Y if an experimental manipulation that changes X (and only X) leads to a change of Y.
⇒ experimental mode of inferring causation

A theory of causality:
Framework of causal inference, largely developed and popularized by Judea Pearl, Peter Spirtes, Clark Glymour, Richard Scheines.

Textbooks: [Pearl, 2000, Spirtes et al., 2000, Peters et al., 2017].
Modelling causal relationships: Structural causal models

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

**Example** (scientifically oversimplified, for illustration only):

Structural causal model:

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

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

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

**Causal graph:**
Specifies the *direct causes* of each variable.
Modelling causal relationships: Structural causal models

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

**Example** (scientifically oversimplified, for illustration only):

Structural causal model:

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

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

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

**Causal graph:**
Specifies the *direct causes* of each variable.
Modelling causal relationships: Structural causal models

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

Example (scientifically oversimplified, for illustration only):
Structural causal model:

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

Causal graph:
Specifies the \textit{direct causes} of each variable.
Modelling causal relationships: Structural causal models

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

Example (scientifically oversimplified, for illustration only):

Structural causal model:

\[
\begin{align*}
X_{\text{clouds}} &= f_{\text{clouds}}(X_{\text{aerosols}}, X_{\text{env.facs.}}, \eta_{\text{clouds}}) \\
X_{\text{aerosols}} &= f_{\text{aerosols}}(X_{\text{env.facs.}}, \eta_{\text{aerosols}}) \\
X_{\text{env.facs.}} &= f_{\text{env.facs.}}(\eta_{\text{env.facs.}})
\end{align*}
\]

Causal graph:
Specifies the \textit{direct causes} of each variable.
Modelling causal relationships: Structural causal models

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

Example (scientifically oversimplified, for illustration only):
Structural causal model:

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

Causal graph:
Specifies the *direct causes* of each variable.
Modelling causal relationships: Structural causal models

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

Example (scientifically oversimplified, for illustration only):

Structural causal model:

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

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

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

Causal graph:
Specifies the direct causes of each variable.
Why is causal knowledge important?

**Scientific understanding:**
Knowledge of cause and effect relationships is an essential part of the physical understanding of natural processes.
Why is causal knowledge important?

**Scientific understanding:**
Knowledge of cause and effect relationships is an essential part of the physical understanding of natural processes.

**Robust prediction & forecasting:**
Predictive systems consistent with the underlying causal structures are thought to be more robust under changing environmental conditions.
Why is causal knowledge important?

Scientific understanding:
Knowledge of cause and effect relationships is an essential part of the physical understanding of natural processes.

Robust prediction & forecasting:
Predictive systems consistent with the underlying causal structures are thought to be more robust under changing environmental conditions.

Attribution:
Questions of the type *Why did this event happen?* or *Is this due to climate change?* are of causal nature.
How to obtain causal knowledge?

1. **Experimentation:**
   Deliberately manipulate the system and observe the consequences.
How to obtain causal knowledge?

2. Simulation:
Experimentation inside a simulated version of the system.
3. Causal discovery:
Learn from observational data, given certain assumptions.
Causal discovery
Today’s approach to causal discovery:
Learn causal graph from statistical tests of (conditional) independencies*
in observational data

⇒ *Cl-based* causal discovery

*Conditional independence:
For random variables $X$, $Y$, and $Z$ with distribution $p$: $X$ and $Y$ are conditionally independent $Z$, denoted as $X \perp \!\!\!\!\!\!\perp Y \mid Z$, if $p(x|y,z) = p(x|z)$ for all $x, y, z$. 
Learning causal relationships from statistical independencies

Today’s approach to causal discovery:
Learn causal graph from statistical tests of (conditional) independencies* in observational data

⇒ CI-based causal discovery

Enabling assumptions:

1. Observational data is generated by a structural causal model (this true SCM is unknown)

2. No accidental independencies ⇒ more on this later

3. Optional: No unobserved confounders ⇒ more on this later

*Conditional independence:
For random variables $X$, $Y$, and $Z$ with distribution $p$: $X$ and $Y$ are conditionally independent $Z$, denoted as $X \perp \perp Y \mid Z$, if $p(x|y, z) = p(x|z)$ for all $x, y, z$. 
Causal graphs and (conditional) independencies

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

Intuition:
- Statistical dependencies derive from causal relationships
- Conditioning can block and open the *flow of information*
Causal graphs and (conditional) independencies

Example:

- $X$ influences $Y$: $X \perp \perp Y$
- $Y$ influences $Z$: $Y \perp \perp Z$
- $X$ influences $Z$ through $Y$: $X \perp \perp Z$
Causal graphs and (conditional) independencies

Example:

- $X$ influences $Y$: $X \not\perp\!\!\!\!\perp Y$
- $Y$ influences $Z$: $Y \not\perp\!\!\!\!\perp Z$
- $X$ influences $Z$ through $Y$: $X \not\perp\!\!\!\!\perp Z$
- Knowing $Y$, $X$ does not say more about $Z$: $X \perp\!\!\!\!\perp Z \mid Y$
Causal graphs and (conditional) independencies

Example:

- X influences Y: \( X \perp \perp Y \)
- Y influences Z: \( Y \perp \perp Z \)
- X influences Z through Y: \( X \perp \perp Z \)
- Knowing Y, X does not say more about Z: \( X \perp \perp Z \mid Y \)

General rule: d-separation
Graphical criterion to read off all (conditional) independencies implied by the structure of a given causal graph [Pearl, 1985, Pearl, 1988].

No accidental independencies:
There are no independencies beyond those implied by the causal graph.
CI-based causal discovery without unobserved confounders

Idea:

- Perform statistical tests of (conditional) independence in observational data
- Use test results to constrain the structure of the causal graph
CI-based causal discovery without unobserved confounders

Idea:

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

Example 1:

Test decisions:

- $X \not\perp\!\!\!\!\perp Y$
- $Y \not\perp\!\!\!\!\perp Z$
- $X \not\perp\!\!\!\!\perp Z$

Possible causal graphs:
CI-based causal discovery without unobserved confounders

Idea:

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

Example 2:

Test decisions:

- \(X \perp \!\!\!\!\!\!\!\!\perp Y\)
- \(Y \perp \!\!\!\!\!\!\!\!\perp Z\)
- \(X \perp \!\!\!\!\!\!\!\!\perp Z\)
- \(X \perp \!\!\!\!\!\!\!\!\!\perp Z | Y\)

Possible causal graphs:

observationally equivalent graphs
Unobserved confounders make causal discovery more difficult

Without unobserved confounders:

\[ X \perp\!\!\!\!\!\!\perp Y \implies \]

\[ X \perp\!\!\!\!\!\!\perp Y \]

or

\[ X \perp\!\!\!\!\!\!\perp Y \]

\[ X \leftrightarrow Y \]
Unobserved confounders make causal discovery more difficult

Without unobserved confounders:

\[ X \perp \!
\perp \! Y \Rightarrow \]

With unobserved confounders:

\[ X \perp \!
\perp \! Y \Rightarrow \]

\[ X \perp \!
\perp \! L \perp \!
\perp \! Y \]
Unobserved confounders make causal discovery more difficult

Without unobserved confounders:

\[ X \perp\!\!\!\!\!\!\perp Y \quad \Rightarrow \quad X \rightarrow Y \text{ or } X \leftarrow Y \]

With unobserved confounders:

\[ X \perp\!\!\!\!\!\!\perp Y \quad \Rightarrow \quad X \rightarrow Y \text{ or } X \leftarrow Y \text{ or } \boxed{X \leftrightarrow Y} \]
Our research:
Causal discovery for time series
CI-based causal discovery for time series

Particularities:

- Variables are resolved in time
Particularities:

- Variables are resolved in time
- Autocorrelation
CI-based causal discovery for time series

Particularities:
- Variables are resolved in time
- Autocorrelation

Additional assumption:
- Stationary causal structure
CI-based causal discovery for time series

Particularities:
- Variables are resolved in time
- Autocorrelation

Additional assumption:
- Stationary causal structure
Additional statistical challenges:

- High dimensionality (resolving in time)
- Ill-calibrated statistical tests of independence (autocorrelation)
- Low detection power (autocorrelation)

⇒ standard algorithms often yield bad statistical performance
CI-based causal discovery for time series

Additional statistical challenges:

- High dimensionality (resolving in time)
- Ill-calibrated statistical tests of independence (autocorrelation)
- Low detection power (autocorrelation)

\[ \Rightarrow \text{standard algorithms often yield bad statistical performance} \]

Our contribution:
Statistical problems alleviated by specialized algorithms\(^\dagger\) developed by the Climate Informatics Group @DLR Jena:

- PCMCI time-lagged links only & no unobserved confounders [Runge et al., 2019]
- PCMCI\(^\dagger\) no unobserved confounders [Runge, 2020]
- LPCMCI (Latent-PCMCI) [Gerhardus and Runge, 2020]

\(^\dagger\)available at: https://github.com/jakobrunge/tigramite
CI-based causal discovery for time series

Additional statistical challenges:

- High dimensionality (resolving in time)
- Ill-calibrated statistical tests of independence (autocorrelation)
- Low detection power (autocorrelation)

⇒ standard algorithms often yield bad statistical performance

Our contribution:

Statistical problems alleviated by specialized algorithms† developed by the Climate Informatics Group @DLR Jena:

- PCMCI time-lagged links only & no unobserved confounders [Runge et al., 2019]
- PCMCI† no unobserved confounders [Runge, 2020]
- LPCMCI (Latent-PCMCI) [Gerhardus and Runge, 2020]

† available at: https://github.com/jakobrunge/tigramite
LPCMCI: Latent-PCMCI

LPCMCI allows for:

- Contemporaneous links

(Also PCMCI does)
LPCMCI: Latent-PCMCI

LPCMCI allows for:

- Contemporaneous links
- Unobserved confounders

(also PCMCI\(^+\) does)
LPCMCI: Latent-PCMCI

LPCMCI allows for:

- Contemporaneous links
- Unobserved confounders

(also PCMCI+ does)
LPCMCI: Latent-PCMCI

LPCMCI allows for:

- Contemporaneous links
- Unobserved confounders

Basic idea:
More powerful CI tests by iterative learning of and subsequent conditioning on direct causes.

(also PCMCI\(^+\) does)
LPCMCI achieves strong gains in recall

Results of numerical experiments:
For autocorrelated continuous data LPCMCI shows strong gains in recall as compared to the current state of the art algorithm*

*the SVAR-FCI algorithm by [Malinsky and Spirtes, 2018]

$N = 5, T = 500, \tau_{\text{max}} = 5, \lambda = 0.3$
ParC corr, $\alpha = 0.01$
**High-recall causal discovery for autocorrelated time series with latent confounders.**  

**Causal Structure Learning from Multivariate Time Series in Settings with Unmeasured Confounding.**  
**A Constraint - Propagation Approach to Probabilistic Reasoning.**

**Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.**
Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

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

References


Backup
Conditioning sets are extended with known causal parents
Conditioning sets are extended with known causal parents
Conditioning sets are extended with known causal parents
Conditioning sets are extended with known causal parents
Conditioning sets are extended with known causal parents

\[ X \rightarrow t - 3 \rightarrow t - 2 \rightarrow t - 1 \rightarrow t \]

\[ Y \rightarrow t - 3 \rightarrow t - 2 \rightarrow t - 1 \rightarrow t \]

\[ Z \rightarrow t - 3 \rightarrow t - 2 \rightarrow t - 1 \rightarrow t \]
Conditioning sets are extended with known causal parents

- For $X$: $t - 3$, $t - 2$, $t - 1$, $t$
- For $Y$: $t - 3$, $t - 2$, $t - 1$, $t$
- For $Z$: $t - 3$, $t - 2$, $t - 1$, $t$