Causal inference and complex network methods for the geosciences

Jakob Runge

September 28, 2017

https://jakobrunge.github.io/tigramite/

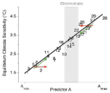
New research group

Climate Informatics

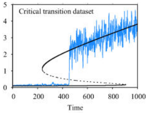
Open PhD and Postdoc positions
→ climateinformaticslab.com

Climate model evaluation

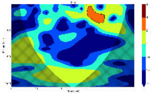
Climate sensitivity



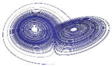
Extremes prediction



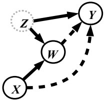
Time-scale causality



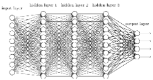
Data-driven dynamical models



Causal discovery theory



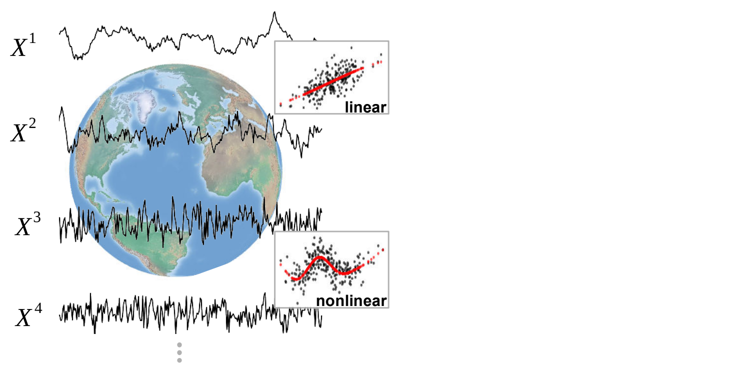
Deep learning



Problem setting

Goal

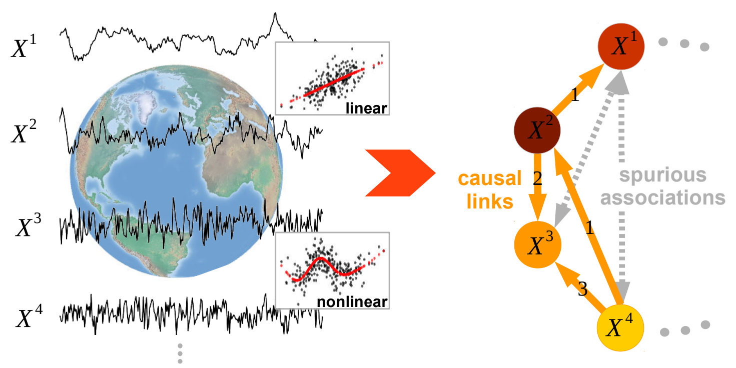
Learn causal interactions from time series of complex dynamical systems



Problem setting

Goal

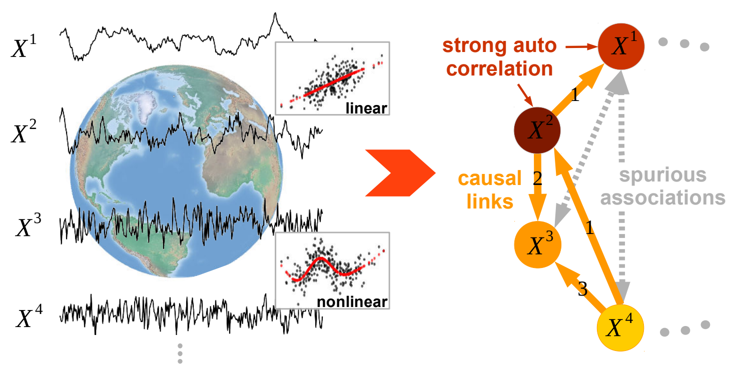
Learn causal interactions from time series of complex dynamical systems



Problem setting

Goal

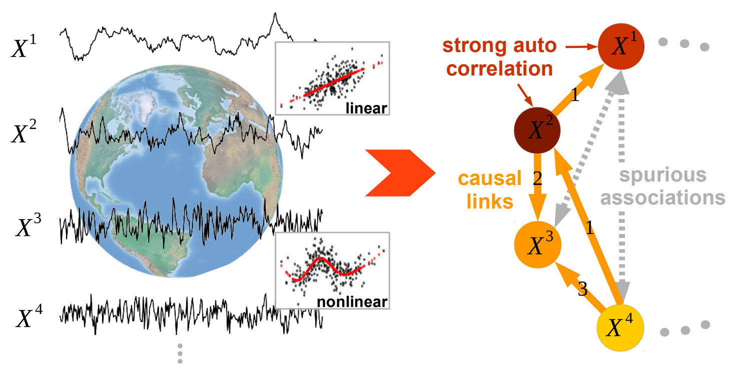
Learn causal interactions from time series of complex dynamical systems



Problem setting

Goal

Learn causal interactions from time series of complex dynamical systems

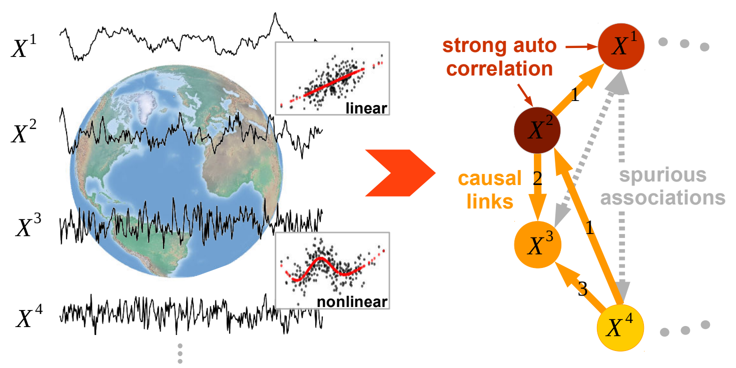


1. How to **formulate** causal inference for complex dynamical systems?

Problem setting

Goal

Learn causal interactions from time series of complex dynamical systems

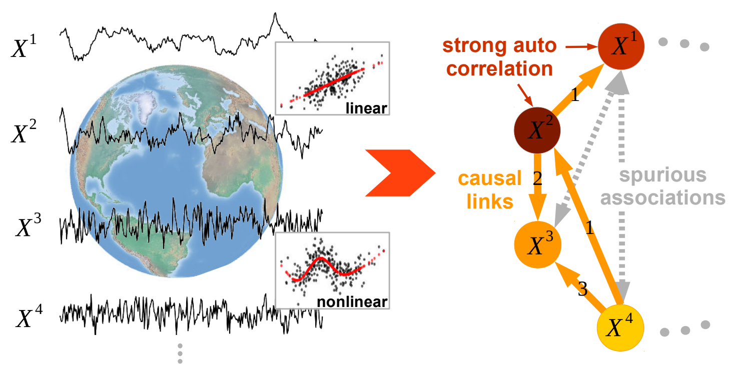


1. How to **formulate** causal inference for complex dynamical systems?
2. How to **detect** causal links?

Problem setting

Goal

Learn causal interactions from time series of complex dynamical systems



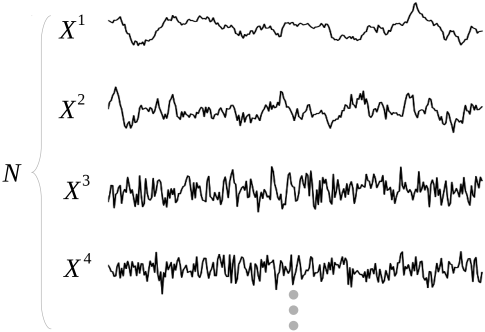
1. How to **formulate** causal inference for complex dynamical systems?
2. How to **detect** causal links?
3. How to **quantify** causal interactions?

How to formulate causal inference for complex dynamical systems?

Time series graphs

Time series graphs

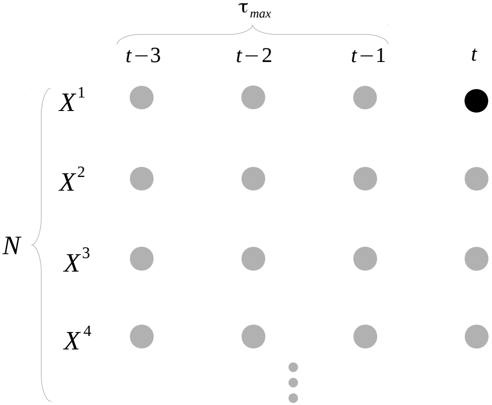
Definition



[Eichler, 2012]

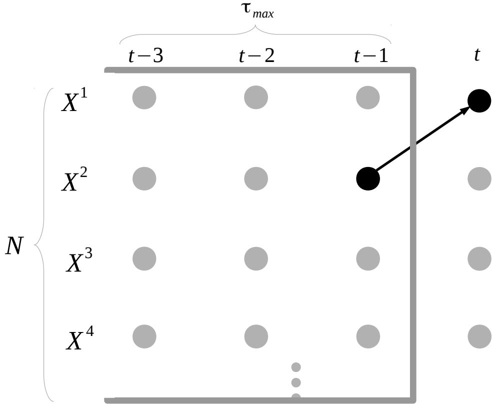
Time series graphs

Definition



[Eichler, 2012]

Definition



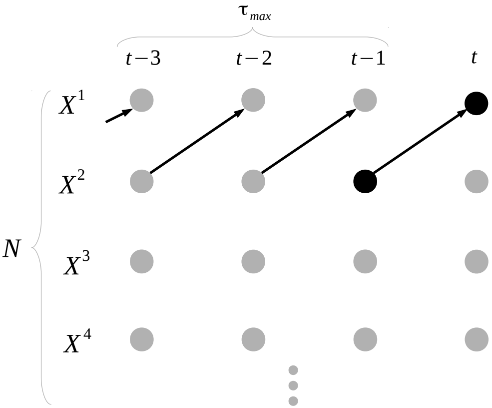
$$X^i_{t-\tau} \quad \not\perp\!\!\!\perp \quad X^j_t \mid \mathbf{X}^-_t$$

$X^i_{t-\tau}$ is *not* independent of X^j_t given $\mathbf{X}^-_t$

[Eichler, 2012]

Time series graphs

Definition



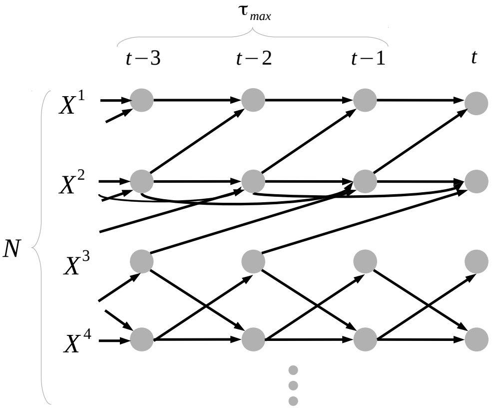
$X^i_{t-\tau} \quad \not\perp\!\!\!\perp \quad X^j_t \mid \mathbf{X}^-_t$

$X^i_{t-\tau}$ is *not* independent of X^j_t given $\mathbf{X}^-_t$

Assuming stationarity

[Eichler, 2012]

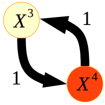
Definition



$X^i_{t-\tau} \not\perp\!\!\!\perp X^j_t \mid \mathbf{X}^-_t$

$X^i_{t-\tau}$ is *not* independent of X^j_t given $\mathbf{X}^-_t$

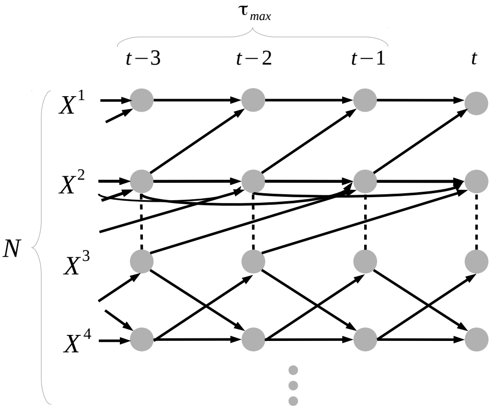
Assuming stationarity



[Eichler, 2012]

Time series graphs

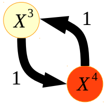
Definition



$X_{t-\tau}^i \not\perp\!\!\!\perp X_t^j \mid \mathbf{X}_t^-$

$X_{t-\tau}^i$ is *not* independent of X_t^j given $\mathbf{X}_t^-$

Assuming stationarity

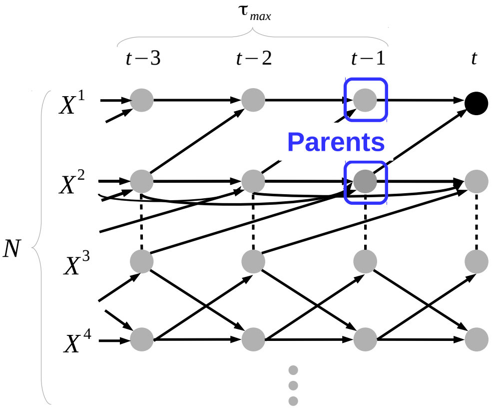


Contemporaneous links defined as $X_t^i \not\perp\!\!\!\perp X_t^j \mid \mathbf{X}_t^-$ left *undirected* here

[Eichler, 2012]

Time series graphs

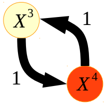
Definition



$X^i_{t-\tau} \quad \not\perp\!\!\!\perp \quad X^j_t \mid \mathbf{X}^-_t$

$X^i_{t-\tau}$ is *not* independent of X^j_t given $\mathbf{X}^-_t$

Assuming stationarity



Contemporaneous links defined as $X^i_t \not\perp\!\!\!\perp X^j_t \mid \mathbf{X}^-_t$ left *undirected* here

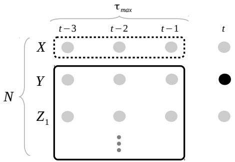
[Eichler, 2012]

How to detect causal links?

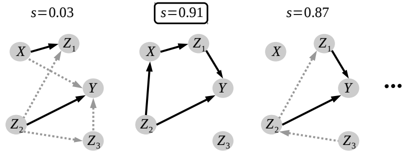
Causal discovery

Causal discovery overview

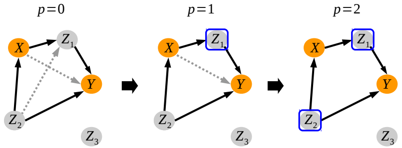
High-dimensional independence testing / Granger causality



Score-based / Bayesian network estimation

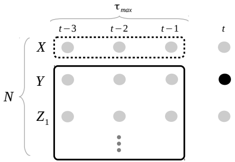


Constraint-based / PC algorithm



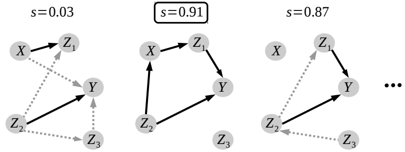
Causal discovery overview

High-dimensional independence testing / Granger causality

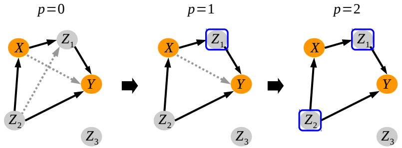


Hybrid-methods

Score-based / Bayesian network estimation

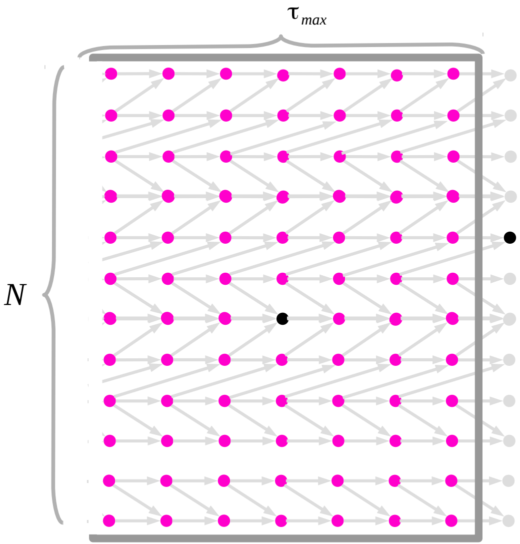


Constraint-based / PC algorithm



Causal discovery

Granger causality



Lag-specific Granger causality:

$$X^i_{t-\tau} \not\perp\!\!\!\perp X^j_t \mid \mathbf{X}^-_t$$

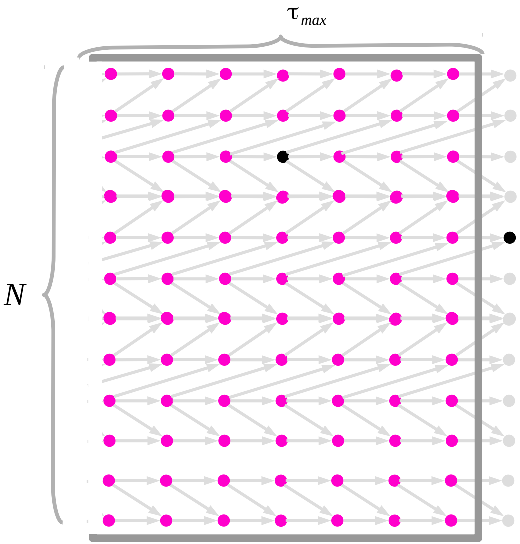
here implemented with ParCorr based on OLS / Ridge / Lasso,

non-parametric Gaussian

processes test → paper

Causal discovery

Granger causality



Lag-specific Granger causality:

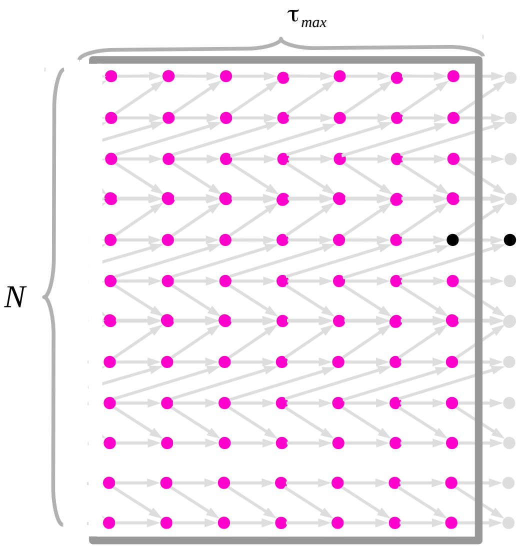
$$X^i_{t-\tau} \not\perp\!\!\!\perp X^j_t \mid \mathbf{X}^-_t$$

here implemented with ParCorr based on OLS / Ridge / Lasso,

non-parametric Gaussian

processes test → paper

Granger causality



Lag-specific Granger causality:

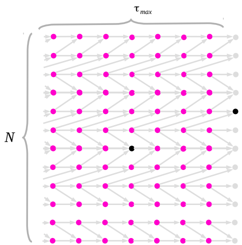
$$X^i_{t-\tau} \not\perp\!\!\!\perp X^j_t \mid \mathbf{X}^-_t$$

here implemented with ParCorr based on OLS / Ridge / Lasso,

non-parametric Gaussian

processes test → paper

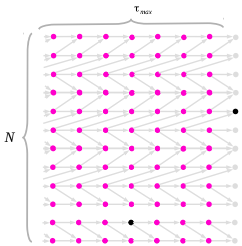
PCMCI = PC$_1$-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

PCMCI = PC_1-condition selection + MCI test



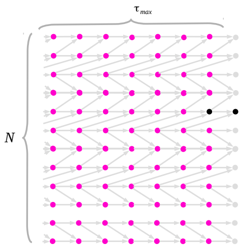
1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!

PCMCI = PC$_1$-condition selection + MCI test

1. Condition selection (Markov blanket)

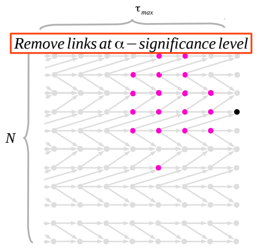
For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!



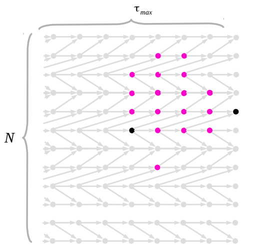
PCMCI = PC$_1$-condition selection + MCI test

1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!



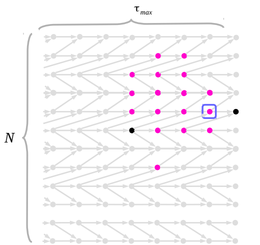
PCMCI = PC$_1$-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

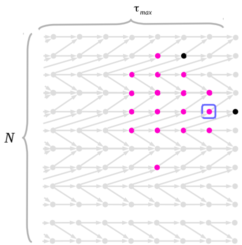
PCMCI = PC_1-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!

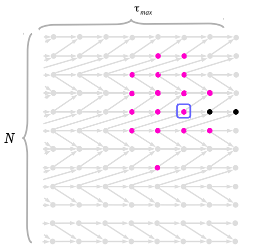
PCMCI = PC_1-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!

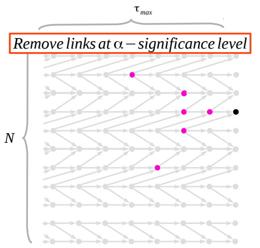
PCMCI = PC$_1$-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

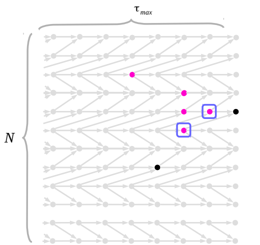
PCMCI = PC$_1$-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X^j_t)$ such that $X^j_t \perp\!\!\!\perp \mathbf{X}^-_t \setminus \tilde{\mathcal{P}}(X^j_t) \mid \tilde{\mathcal{P}}(X^j_t)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

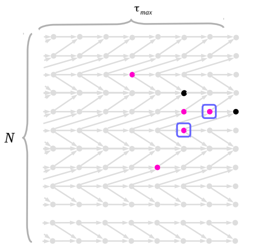
PCMCI = PC_1-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X^j_t)$ such that $X^j_t \perp\!\!\!\perp \mathbf{X}^-_t \setminus \tilde{\mathcal{P}}(X^j_t) \mid \tilde{\mathcal{P}}(X^j_t)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!

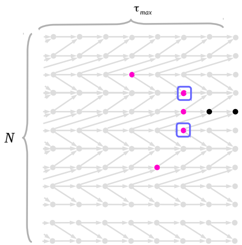
PCMCI = PC$_1$-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

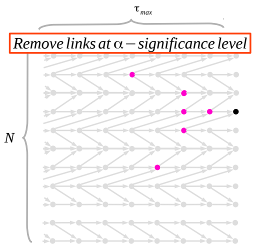
PCMCI = PC_1-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!

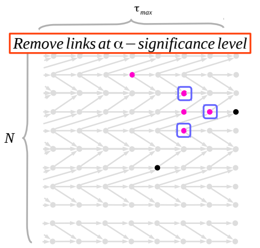
PCMCI = PC$_1$-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

PCMCI = PC$_1$-condition selection + MCI test



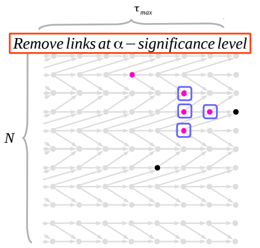
1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

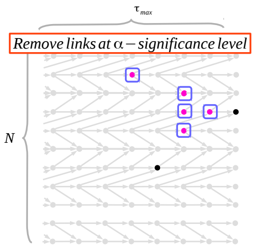
PCMCI = PC$_1$-condition selection + MCI test

1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!



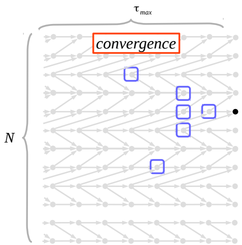
PCMCI = PC_1-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!

PCMCI = PC$_1$-condition selection + MCI test



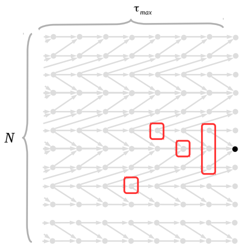
1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!

PCMCI = PC_1-condition selection + MCI test

1. Condition selection (Markov blanket)

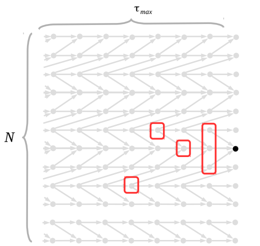
For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!



PCMCI = PC_1-condition selection + MCI test

1. Condition selection (Markov blanket)

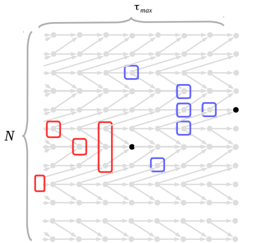
For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!



2. Momentary conditional independence (MCI) test

For $i, j \in \{1, \ldots, N\}$ and $0 \leq \tau \leq \tau_{\text{max}}$:
Test

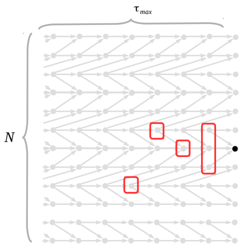
MCI: $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \tilde{\mathcal{P}}(X_t^j), \tilde{\mathcal{P}}(X_{t-\tau}^i)$



PCMCI = PC$_1$-condition selection + MCI test

1. Condition selection (Markov blanket)

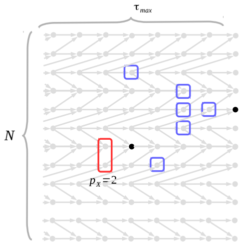
For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC$_1$ algorithm: tuned to high power with liberal α, false pos. control in next step!



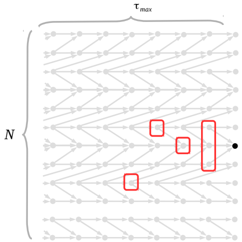
2. Momentary conditional independence (MCI) test

For $i, j \in \{1, \ldots, N\}$ and $0 \leq \tau \leq \tau_{\max}$:
Test

MCI: $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \tilde{\mathcal{P}}(X_t^j), \tilde{\mathcal{P}}(X_{t-\tau}^i)$

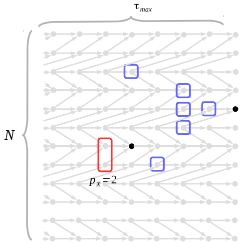


PCMCI = PC_1-condition selection + MCI test



1. Condition selection (Markov blanket)

For $j \in \{1, \ldots, N\}$: Estimate superset of parents $\tilde{\mathcal{P}}(X_t^j)$ such that $X_t^j \perp\!\!\!\perp \mathbf{X}_t^- \setminus \tilde{\mathcal{P}}(X_t^j) \mid \tilde{\mathcal{P}}(X_t^j)$ with iterative PC_1 algorithm: tuned to high power with liberal α, false pos. control in next step!



2. Momentary conditional independence (MCI) test

For $i, j \in \{1, \ldots, N\}$ and $0 \leq \tau \leq \tau_{\max}$:
Test

MCI: $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \tilde{\mathcal{P}}(X_t^j), \tilde{\mathcal{P}}(X_{t-\tau}^i)$

Flexible regarding conditional independence tests

here ParCorr (OLS), Gaussian processes $\rightarrow$ paper

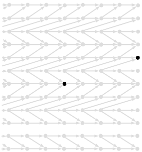
PCMCI = PC_1-condition selection + MCI test

0 → ***1*** *Condition-selection significance level* α

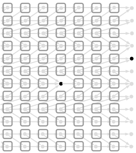
PCMCI = PC_1-condition selection + MCI test

0 → *Condition-selection significance level α* **1**

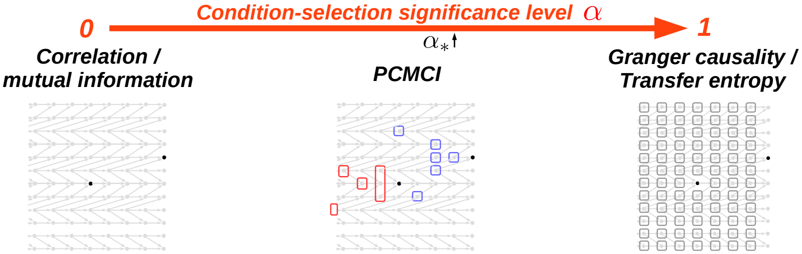
Correlation / mutual information



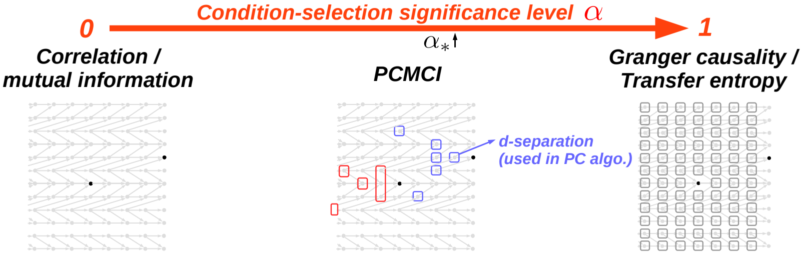
Granger causality / Transfer entropy



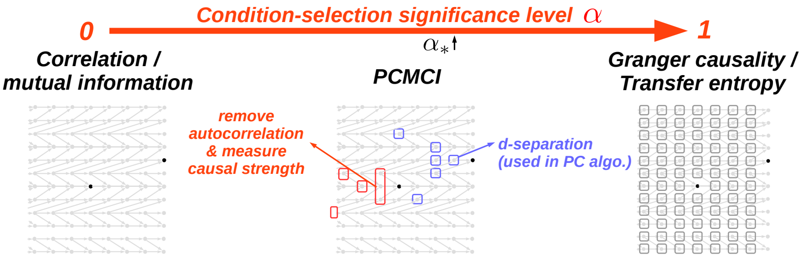
PCMCI = PC_1-condition selection + MCI test



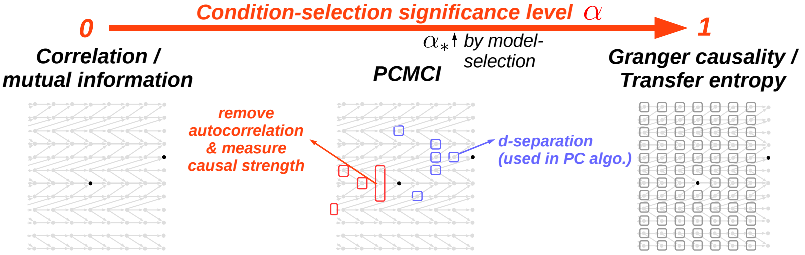
PCMCI = PC_1-condition selection + MCI test



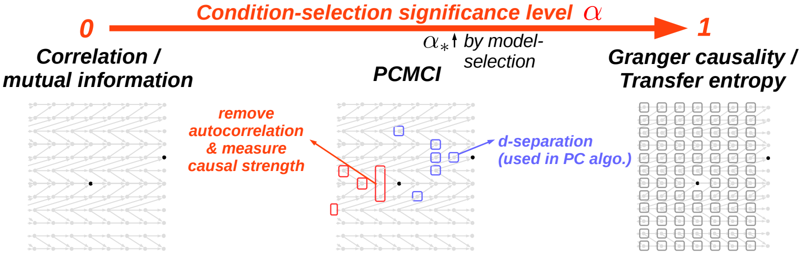
PCMCI = PC_1-condition selection + MCI test



PCMCI = PC_1-condition selection + MCI test



PCMCI = PC_1-condition selection + MCI test



More theory in paper:

If PC_1 identifies parents, then

- MCI has unbiased detection power for linear links in additive models

$$I^{\mathrm{MCI}}_{X\to Y} = I\left(\eta^X_{t-\tau}; c\eta^X_{t-\tau} + \eta^Y_t\right)$$

- MCI is well-calibrated also for autocorrelated data
- effect size of MCI is larger than GC $\rightarrow$ more power also for low dimensions

Conditional independence tests

Conditional independence tests $X \perp\!\!\!\perp Y \mid Z$

Assuming linear model: Partial correlation (ParCorr)

1. Regress out influence of Z with OLS

$$X = Z\beta_X + \epsilon_X$$

$$Y = Z\beta_Y + \epsilon_Y$$

   Ridge and Lasso implemented with `scikit-learn` on standardized time series

   - Ridge regularization: LOO-cross-validated regularization parameter $\alpha \in \{0.1, 1, 2, \ldots, 500\}$
   - Lasso regularization: multi-task lasso, $\alpha \in \{0.0001, 0.001, 0.01, 0.1, 1\}$ using 5-fold cross-validation, max. iterations $= 100$

2. Test independence of residuals with t-test

   - OLS: $T - D_Z - 2$ degrees of freedom

Conditional independence tests $X \perp\!\!\!\perp Y \mid Z$

Assuming nonlinear additive Gaussian: GPDC

1. Regress out influence of Z with Gaussian process assuming
$$X = f_X(Z) + \epsilon_X$$
$$Y = f_Y(Z) + \epsilon_Y$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$
GP regression implemented using `sklearn`
    - Radial Basis Function (RBF) + White Noise kernel
    - bandwidth estimated with MLE
2. Test independence of uniformized residuals with *distance correlation coefficient* [Székely et al., 2007]
$$\mathcal{R}\left(r_X, r_Y\right)$$
using pre-computed null distribution (for every T)

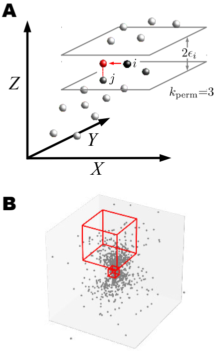
Conditional independence tests $X \perp\!\!\!\perp Y \mid Z$

General: Conditional mutual information (CMI)

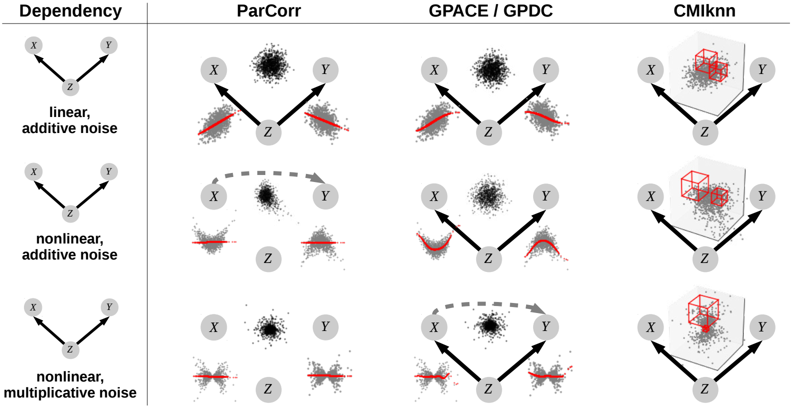
$$I(X;Y|Z) = \int dz\ p(z) \int\int dxdy\ p(x,y|z) \log \frac{p(x,y|z)}{p(x|z)\cdot p(y|z)}$$

Estimated with kNN-estimator [Kraskov et al., 2004, Frenzel and Pompe, 2007]

Free parameter: number of nearest neighbors $k \sim$ locally adaptive bandwidth



Conditional independence tests $X \perp\!\!\!\perp Y \mid Z$



Causal assumptions

Causal interpretation assumes [Spirtes et al., 2000]:

- **Causal Markov Condition:** "All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"
- **Causal Sufficiency:** Measured variables include all of the common causes
- **Faithfulness / Stableness:** "Independencies in data arise not from incredible coincidence, but rather from causal structure"; violated by purely deterministic dependencies

Causal assumptions

Causal interpretation assumes [Spirtes et al., 2000]:

- **Causal Markov Condition:** "All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"
- **Causal Sufficiency:** Measured variables include all of the common causes
- **Faithfulness / Stableness:** "Independencies in data arise not from incredible coincidence, but rather from causal structure"; violated by purely deterministic dependencies
- **No contemporaneous effects**: (but can be extended)
- **Stationarity:** time series case
- **Parametric assumptions** of independence tests

Causal assumptions

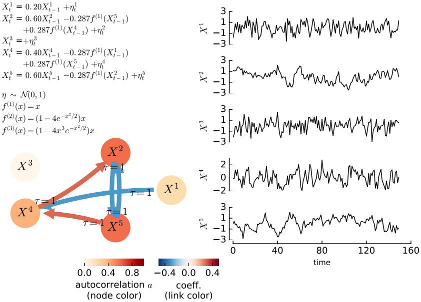
Causal interpretation assumes [Spirtes et al., 2000]:

- **Causal Markov Condition:** "All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"
- **Causal Sufficiency:** Measured variables include all of the common causes
- **Faithfulness / Stableness:** "Independencies in data arise not from incredible coincidence, but rather from causal structure"; violated by purely deterministic dependencies
- **No contemporaneous effects**: (but can be extended)
- **Stationarity:** time series case
- **Parametric assumptions** of independence tests

More discussion $\rightarrow$ appendix

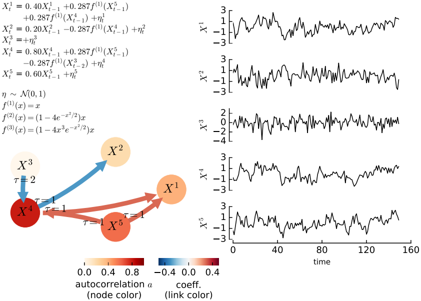
Numerical experiments

Model setup



- random coupling topologies, time lags, linear
- fixed link strength within each network
- different autocorrelations for variables
- $\tau_{\max} = 5,\ T = 150$, varying $N = 5..60 \to N \cdot \tau_{\max} > T$

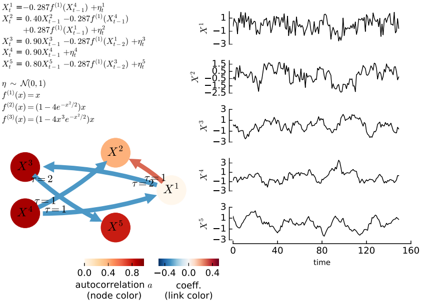
Model setup



- random coupling topologies, time lags, linear
- fixed link strength within each network
- different autocorrelations for variables
- $\tau_{\max} = 5,\ T = 150$, varying $N = 5..60 \rightarrow N \cdot \tau_{\max} > T$

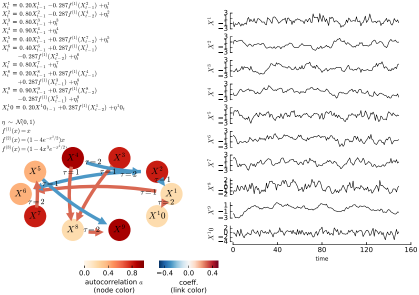
Numerical experiments

Model setup



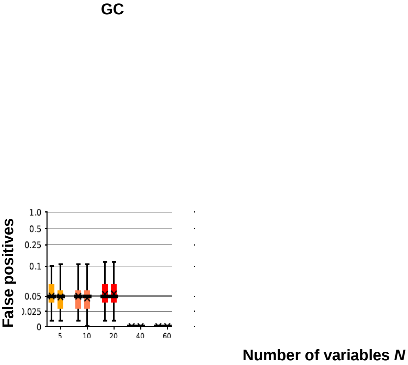
- random coupling topologies, time lags, linear
- fixed link strength within each network
- different autocorrelations for variables
- $\tau_{\max} = 5,\ T = 150$, varying $N = 5..60 \rightarrow N \cdot \tau_{\max} > T$

Model setup



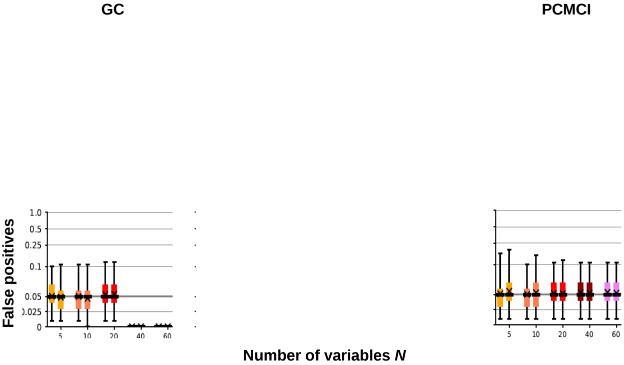
- random coupling topologies, time lags, linear
- fixed link strength within each network
- different autocorrelations for variables
- $\tau_{\max} = 5$, $T = 150$, varying $N = 5..60 \rightarrow N \cdot \tau_{\max} > T$

Similarly well-calibrated tests



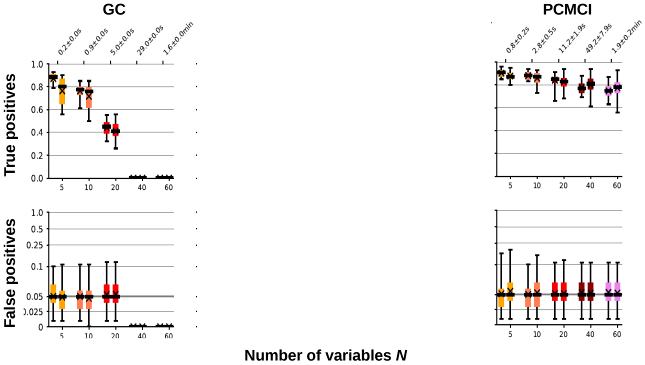
Numerical experiments

Similarly well-calibrated tests

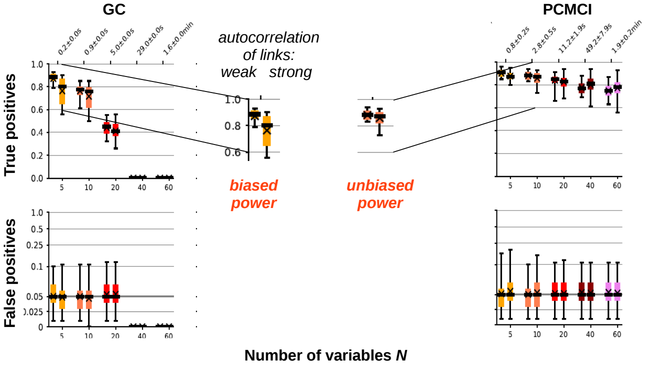


Numerical experiments

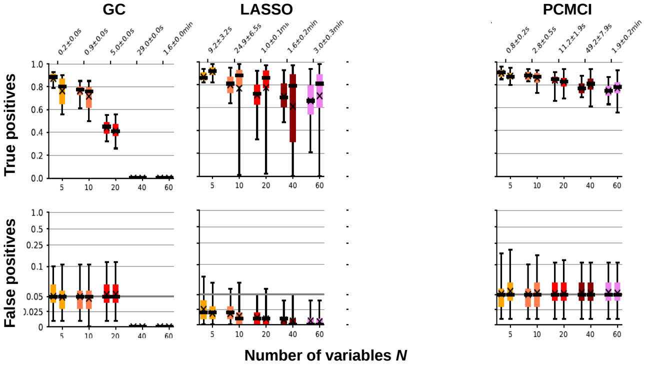
GC suffers from curse of dimensionality and power bias



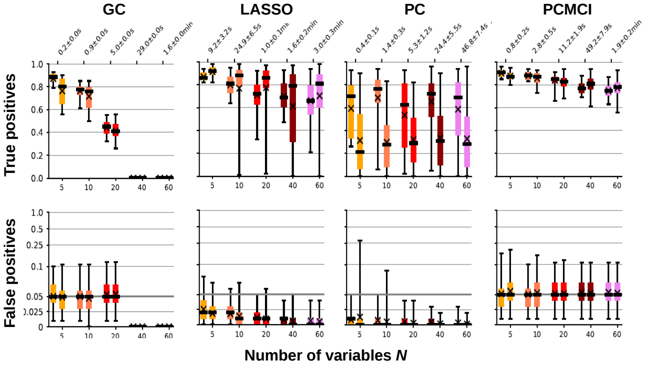
GC suffers from curse of dimensionality and power bias



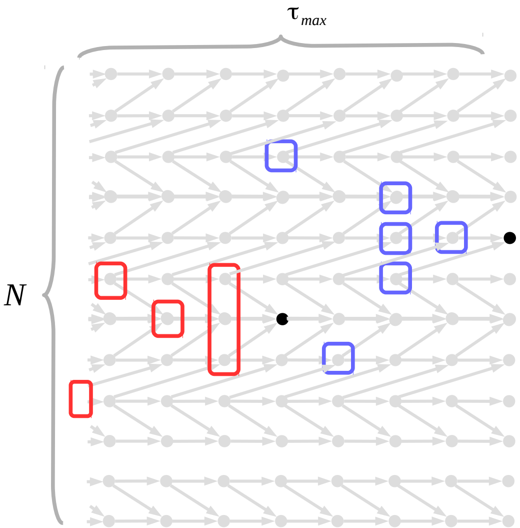
Lasso not well-calibrated and power bias



PC algorithm also low and biased power

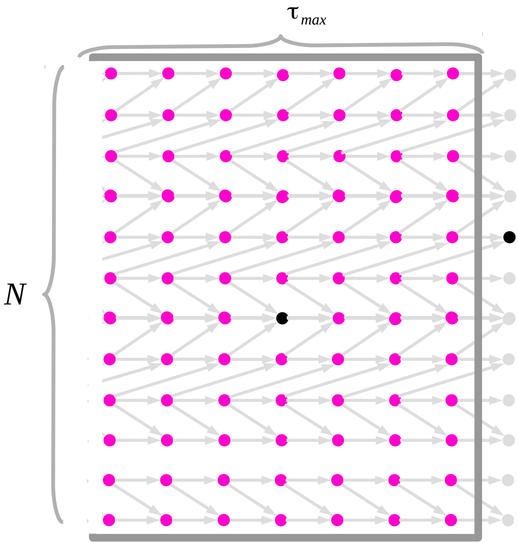


Key idea again



Numerical experiments

Key idea again



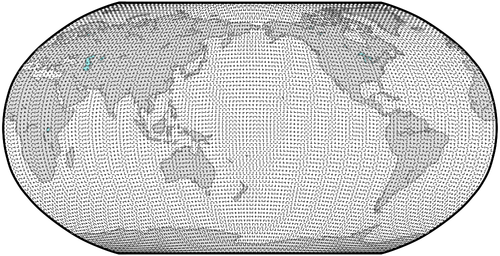
Applications

Applications

- **Causal hypothesis testing**
  [Runge et al., 2014, Runge et al., 2015c, Kretschmer et al., 2016]
- **Variable selection** for model building
- ...or **prediction schemes**
  [Runge et al., 2015a, Kretschmer et al., 2017]
- **Pathway analysis** [Runge et al., 2015b, Runge, 2015]

Global sea-level pressure interactions

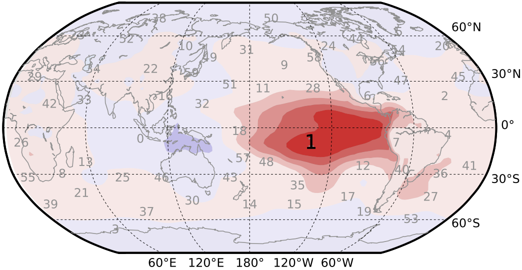
Sea-level pressure system [Kalnay et al., 1996]



- detrended, anomalized, winter-only (DJF) of 1981–2012
- dimension reduction using Varimax-rotated PCA [Vejmelka et al., 2014]
- time resolution: 3-days, $\tau_{\mathsf{max}} = 3$ weeks

Global sea-level pressure interactions

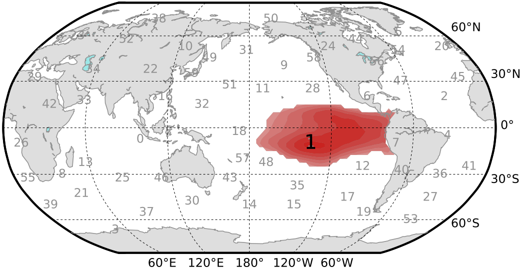
Sea-level pressure system [Kalnay et al., 1996]



- detrended, anomalized, winter-only (DJF) of 1981–2012
- dimension reduction using Varimax-rotated PCA [Vejmelka et al., 2014]
- time resolution: 3-days, $\tau_{\max} = 3$ weeks

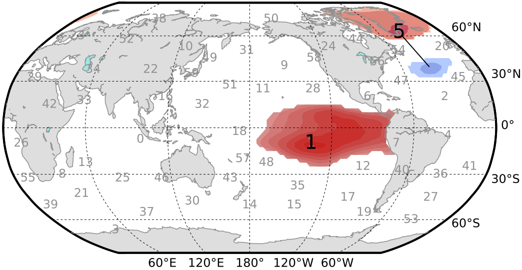
Global sea-level pressure interactions

Sea-level pressure system [Kalnay et al., 1996]



- detrended, anomalized, winter-only (DJF) of 1981–2012
- dimension reduction using Varimax-rotated PCA [Vejmelka et al., 2014]
- time resolution: 3-days, $\tau_{\max} = 3$ weeks

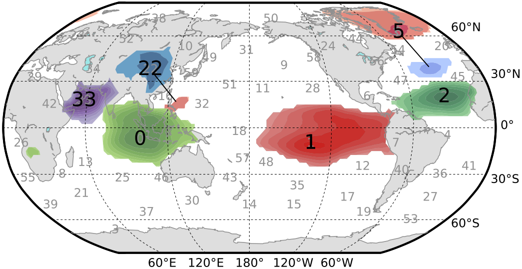
Sea-level pressure system [Kalnay et al., 1996]



- detrended, anomalized, winter-only (DJF) of 1981–2012
- dimension reduction using Varimax-rotated PCA [Vejmelka et al., 2014]
- time resolution: 3-days, $\tau_{\max} = 3$ weeks

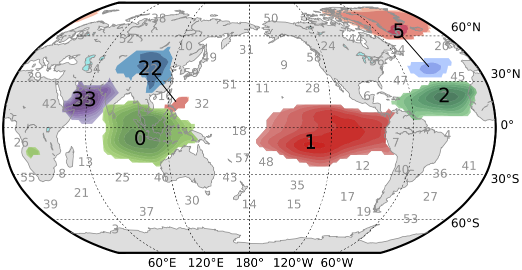
Global sea-level pressure interactions

Sea-level pressure system [Kalnay et al., 1996]



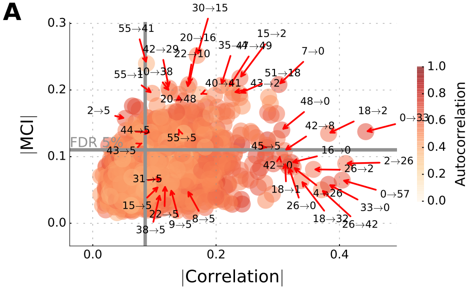
- detrended, anomalized, winter-only (DJF) of 1981–2012
- dimension reduction using Varimax-rotated PCA [Vejmelka et al., 2014]
- time resolution: 3-days, $\tau_{\max} = 3$ weeks

Sea-level pressure system [Kalnay et al., 1996]



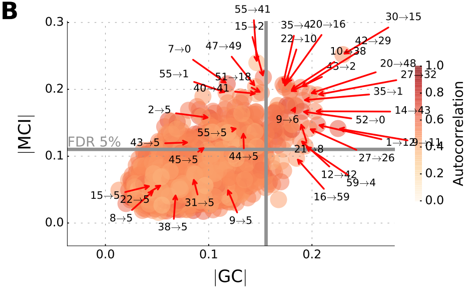
- detrended, anomalized, winter-only (DJF) of 1981–2012
- dimension reduction using Varimax-rotated PCA [Vejmelka et al., 2014]
- time resolution: 3-days, $\tau_{\max} = 3$ weeks
- $N\tau_{\max} = 60 \cdot 7 = 420$ with a comparably small sample size of about 950 samples and partially strong autocorrelations

Spurious correlation vs MCI



- even strong correlations are spurious

Granger causality vs MCI



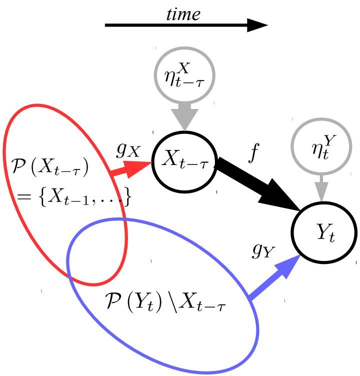
- many even strong causal links overlooked with GC

How to quantify causal interactions?

Causal strength

MCI and causal strength

Defining causal strength



$$X_{t-\tau} = g_X\left(\mathcal{P}\left(X_{t-\tau}\right)\right) + \eta^X_{t-\tau}$$

$$Y_t = g_Y\left(\mathcal{P}\left(Y_t\right) \setminus \{X_{t-\tau}\}\right) + \tilde{\eta}^Y_t$$

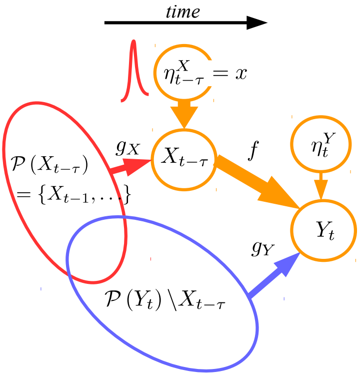
with link $X_{t-\tau} \to Y_t$ represented as $\tilde{\eta}^Y_t = f(X_{t-\tau}) + \eta^Y_t$

Causal strength

$$I\left(\eta^X_{t-\tau}; \tilde{\eta}^Y_t | \mathcal{P}\left(X_{t-\tau}\right)\right)$$

measures "momentary" dependence in $\tilde{\eta}^Y_t$ on $X_{t-\tau}$ that does not come through the parents of $X_{t-\tau}$

Defining causal strength



$$X_{t-\tau} = g_X(\mathcal{P}(X_{t-\tau})) + \eta^X_{t-\tau}$$

$$Y_t = g_Y(\mathcal{P}(Y_t) \setminus \{X_{t-\tau}\}) + \tilde{\eta}^Y_t$$

with link $X_{t-\tau} \to Y_t$ represented as $\tilde{\eta}^Y_t = f(X_{t-\tau}) + \eta^Y_t$

Causal strength

$$I\left(\eta^X_{t-\tau}; \tilde{\eta}^Y_t | \mathcal{P}(X_{t-\tau})\right)$$

measures "momentary" dependence in $\tilde{\eta}^Y_t$ on $X_{t-\tau}$ that does not come through the parents of $X_{t-\tau}$

MCI and causal strength

Definition

$$\text{MCI:}\;\; X_{t-\tau} \;\perp\!\!\!\perp\; Y_t \mid \mathcal{P}(Y_t) \setminus \{X_{t-\tau}\}, \mathcal{P}(X_{t-\tau}) \tag{1}$$

MCI and causal strength

Definition

$$I^{\mathrm{MCI}}_{X\to Y}(\tau) = I\left(X_{t-\tau}\ ;\ Y_t \mid \mathcal{P}(Y_t)\setminus\{X_{t-\tau}\}, \mathcal{P}(X_{t-\tau})\right) \tag{1}$$

MCI and causal strength

Definition

$$I^{\mathrm{MCI}}_{X\to Y}(\tau) = I\left(X_{t-\tau}\ ;\ Y_t \mid \mathcal{P}(Y_t)\setminus\{X_{t-\tau}\}, \mathcal{P}(X_{t-\tau})\right) \tag{1}$$

1. MCI measures causal strength

$$\begin{aligned} I^{\mathrm{MCI}}_{X\to Y} &= I\left(g_X\left(\mathcal{P}_{X_{t-\tau}}\right) + \eta^X_{t-\tau}\ ;\ \ g_Y\left(\mathcal{P}_{Y_t}\setminus\{X_{t-\tau}\}\right) + \tilde{\eta}^Y_t\ \ \mid\ \ \ldots\right) \\ &= I\left(\eta^X_{t-\tau}; \tilde{\eta}^Y_t \mid \mathcal{P}_{X_{t-\tau}}\right) \quad \square \end{aligned}$$

MCI and causal strength

Definition

$$I^{\mathrm{MCI}}_{X\to Y}(\tau) = I\left(X_{t-\tau}\ ;\ Y_t \mid \mathcal{P}(Y_t)\setminus\{X_{t-\tau}\}, \mathcal{P}(X_{t-\tau})\right) \tag{1}$$

1. MCI measures causal strength

$$\begin{aligned} I^{\mathrm{MCI}}_{X\to Y} &= I\left(g_X\left(\mathcal{P}_{X_{t-\tau}}\right) + \eta^X_{t-\tau}\ ;\ \ g_Y\left(\mathcal{P}_{Y_t}\setminus\{X_{t-\tau}\}\right) + \tilde{\eta}^Y_t\ \ \mid\ \ \ldots\right) \\ &= I\left(\eta^X_{t-\tau}; \tilde{\eta}^Y_t \mid \mathcal{P}_{X_{t-\tau}}\right) \quad \square \end{aligned}$$

2. MCI has unbiased detection power for linear links

$$\begin{aligned} \tilde{\eta}^Y_t &= cX_{t-\tau} + \eta^Y_t = c\left(g_X\left(\mathcal{P}_{X_{t-\tau}}\right) + \eta^X_{t-\tau}\right) + \eta^Y_t \\ \implies I^{\mathrm{MCI}}_{X\to Y} &= I\left(\eta^X_{t-\tau}; c\eta^X_{t-\tau} + \eta^Y_t\right) \end{aligned}$$

MCI and causal strength

Definition

$$I^{\mathrm{MCI}}_{X\to Y}(\tau) = I\left(X_{t-\tau}\ ;\ Y_t \mid \mathcal{P}(Y_t)\setminus\{X_{t-\tau}\}, \mathcal{P}(X_{t-\tau})\right) \tag{1}$$

1. MCI measures causal strength

$$\begin{aligned} I^{\mathrm{MCI}}_{X\to Y} &= I\left(g_X\left(\mathcal{P}_{X_{t-\tau}}\right) + \eta^X_{t-\tau}\ ;\ \ g_Y\left(\mathcal{P}_{Y_t}\setminus\{X_{t-\tau}\}\right) + \tilde{\eta}^Y_t\ \ \mid\ \ \ldots\right) \\ &= I\left(\eta^X_{t-\tau}; \tilde{\eta}^Y_t \mid \mathcal{P}_{X_{t-\tau}}\right) \quad \square \end{aligned}$$

2. MCI has unbiased detection power for linear links

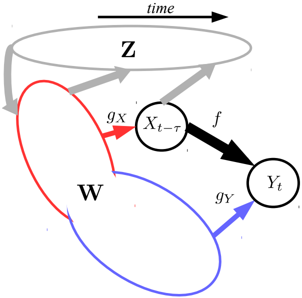
$$\begin{aligned} \tilde{\eta}^Y_t &= cX_{t-\tau} + \eta^Y_t = c\left(g_X\left(\mathcal{P}_{X_{t-\tau}}\right) + \eta^X_{t-\tau}\right) + \eta^Y_t \\ \Longrightarrow\ I^{\mathrm{MCI}}_{X\to Y} &= I\left(\eta^X_{t-\tau}; c\eta^X_{t-\tau} + \eta^Y_t\right) \end{aligned}$$

3. MCI leads to well-calibrated test

$$\tilde{\eta}^Y_t = \eta^Y_t \quad \Longrightarrow \quad I^{\mathrm{MCI}}_{X\to Y} = I\left(\eta^X_{t-\tau}; \eta^Y_t\right) = 0$$

1. **Generally: GC $\leq$ MCI ($\rightarrow$ GC has lower power)**

$$I^{\mathrm{GC}}_{X\to Y}(\tau) = I\left(X_{t-\tau}; Y_t | \mathbf{X}^-_t \setminus \{X_{t-\tau}\}\right)$$



$$\begin{aligned} I\left((X,Z);Y|W\right) &= \underbrace{I\left(X;Y|W\right)}_{\mathrm{MCI}} + \underbrace{I\left(Z;Y|W,X\right)}_{=0\ \text{(Markov)}} \\ &= \underbrace{I\left(Z;Y|W\right)}_{\geq 0} + \underbrace{I\left(X;Y|WZ\right)}_{\mathrm{GC}} \end{aligned}$$

$$\Longrightarrow I^{\mathrm{MCI}}_{X\to Y}(\tau) \geq I^{\mathrm{GC}}_{X\to Y}(\tau) \quad \square$$

MCI, GC, and PC

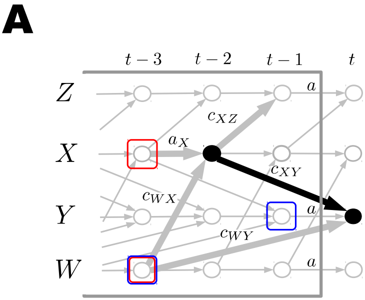
2. **Single PC test has more power, but is non-iid**

$$
\begin{aligned}
I^{\mathrm{PC}}_{X\to Y}(\tau) &= I\left(X_{t-\tau}; Y_t | \mathcal{P}(Y_t) \setminus \{X_{t-\tau}\}\right) \\
&= I\left(\eta^X_{t-\tau}, \mathcal{P}(X_{t-\tau}); Y_t | \mathcal{P}(Y_t) \setminus \{X_{t-\tau}\}\right) \\
&= \underbrace{I\left(\mathcal{P}(X_{t-\tau}); Y_t | \mathcal{P}(Y_t) \setminus \{X_{t-\tau}\}\right)}_{\text{typically } \textit{non-iid}} \\
&\quad + \underbrace{I\left(\eta^X_{t-\tau}; Y_t | \mathcal{P}(Y_t) \setminus \{X_{t-\tau}\}, \mathcal{P}(X_{t-\tau})\right)}_{\mathrm{MCI}}
\end{aligned}
$$

$$
\implies I^{\mathrm{MCI}}_{X\to Y}(\tau) \leq I^{\mathrm{PC}}_{X\to Y}(\tau)
$$

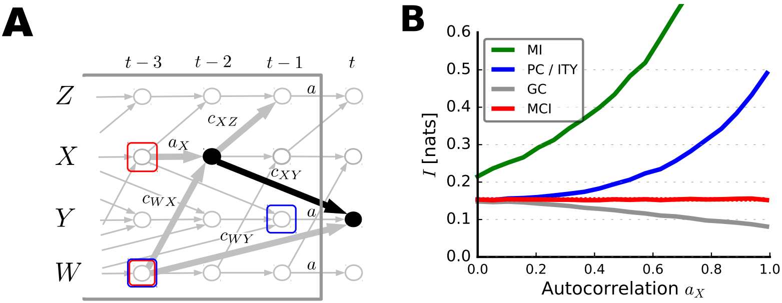
MCI and causal strength

Effect size analysis for simple model

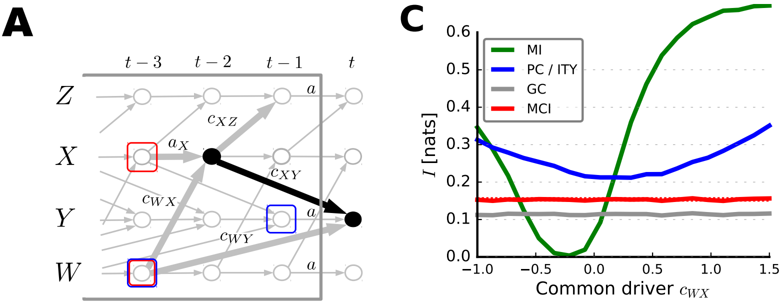


MCI and causal strength

Effect size analysis for simple model

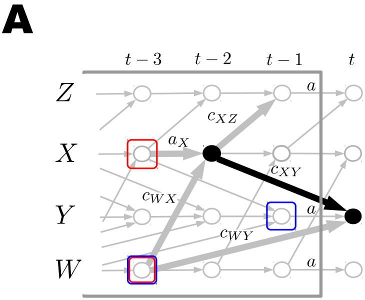


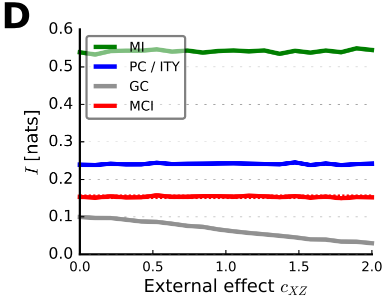
Effect size analysis for simple model



MCI and causal strength

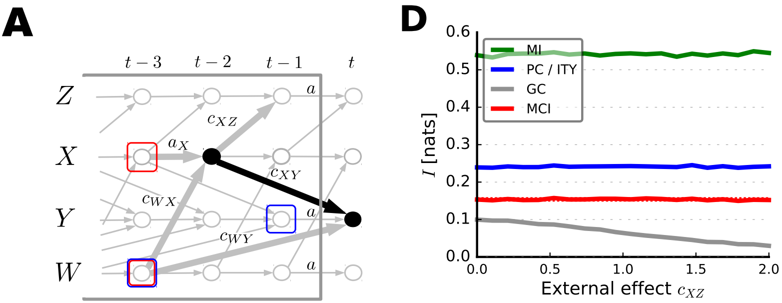
Effect size analysis for simple model





MCI and causal strength

Effect size analysis for simple model

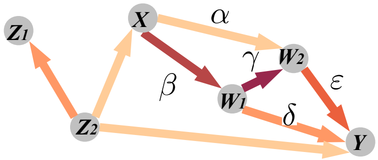


General proof for 'unbiased' detection power $\rightarrow$ paper

Quantifying causal pathways

Linear approach: Mediated Causal Effect (MCE)

$$Y_t = f(\vec{\mathcal{P}_Y}) + error = \vec{\mathcal{P}_Y} \cdot \vec{B} + error$$

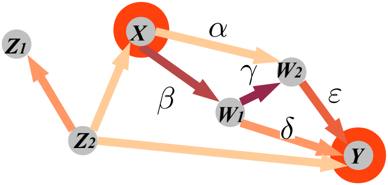


- Direct links: path coefficients α, β, γ, δ, ε

Quantifying causal pathways

Linear approach: Mediated Causal Effect (MCE)

$$Y_t = f(\vec{\mathcal{P}_Y}) + error = \vec{\mathcal{P}_Y} \cdot \vec{B} + error$$



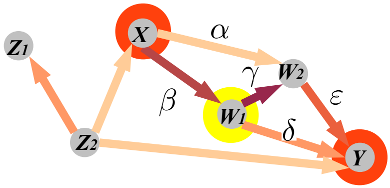
- Direct links: path coefficients α, β, γ, δ, ε
- Indirect causal effect:

$$\mathrm{CE}_{X \to Y} = \alpha\varepsilon + \beta\delta + \beta\gamma\varepsilon$$

Quantifying causal pathways

Linear approach: Mediated Causal Effect (MCE)

$$Y_t = f(\vec{\mathcal{P}_Y}) + error = \vec{\mathcal{P}_Y} \cdot \vec{B} + error$$

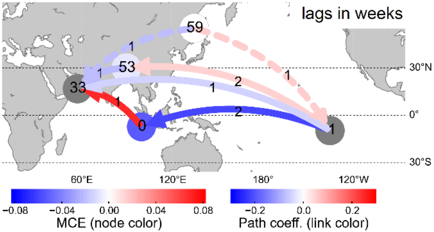


- Direct links: path coefficients α, β, γ, δ, ε
- Indirect causal effect:
  $$\mathrm{CE}_{X\to Y} = \alpha\varepsilon + \beta\delta + \beta\gamma\varepsilon$$
- Mediated causal effect:
  $$\mathrm{MCE}_{X\to Y|W_1} = \beta\delta + \beta\gamma\varepsilon$$

Quantifying causal pathways

Climate application: East Pacific → Monsoon pathway

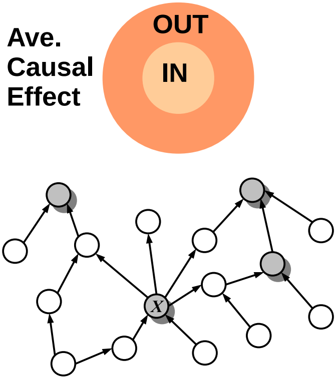
Here whole year analysis [Runge et al., 2015b]



- causal approach to atmospheric teleconnections

Climate application

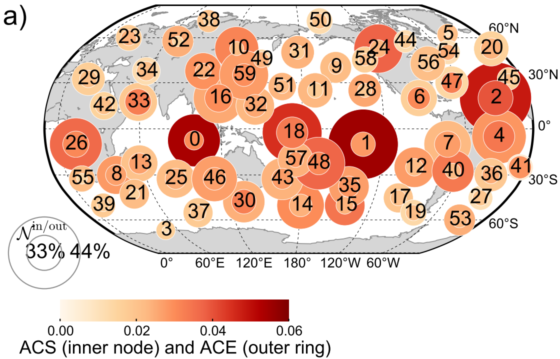
Here whole year analysis [Runge et al., 2015b]



- Average Causal Effect (ACE)
$$\mathrm{ACE}(i) = \frac{1}{N-1}\sum_{j\neq i}\mathrm{CE}^{\max}_{i\to j}$$
- Average Causal Susceptibility
$$\mathrm{ACS}(j) = \frac{1}{N-1}\sum_{i\neq j}\mathrm{CE}^{\max}_{i\to j}$$

Climate application

Here whole year analysis [Runge et al., 2015b]

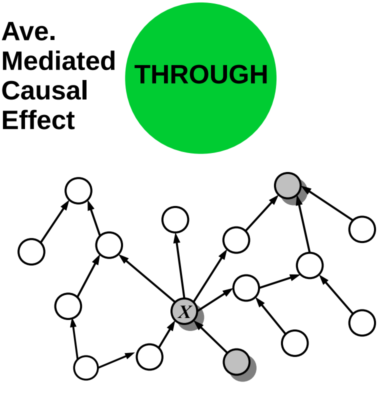


- uplifts over tropical oceans are major drivers

Network analysis

Climate application

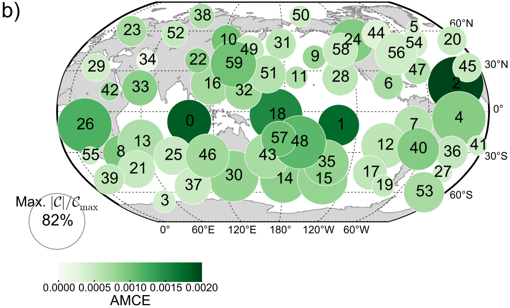
Here whole year analysis [Runge et al., 2015b]



- Average Mediated Causal Effect (AMCE)
$$\mathrm{AMCE}(i) = \frac{1}{|\mathcal{C}_k|} \sum_{(i,j)\in\mathcal{C}_k} \max_{\tau} |\mathrm{MCE}_{i\to j|k}(\tau)|$$

Climate application

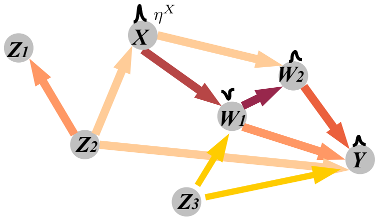
Here whole year analysis [Runge et al., 2015b]



- uplifts over tropical oceans also major mediators

Quantifying causal pathways

Information-theoretic approach [Runge, 2015]



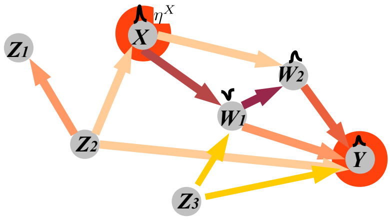
$I(X; Y|Z) =$

$\int dz\, p(z) \iint dxdy\;\; p(x, y|z) \log \frac{p(x,y|z)}{p(x|z) \cdot p(y|z)}$

- Direct links: Momentary information transfer (MIT)

$$I^{\mathrm{MIT}}_{X \to Y} = I(X; Y|\mathcal{P}_Y, \mathcal{P}_X)$$

Information-theoretic approach [Runge, 2015]



$$X_t = f(\mathcal{P}_X) + \eta^X_t$$

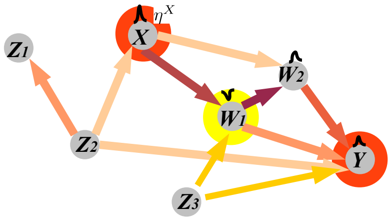
$$I^{\mathrm{MITP}}_{X\to Y} = I(\eta^X_{t-3}\ ;\ Y_t \mid \mathcal{P}_{paths})$$

$$I\,(X;Y|Z) = \int dz\, p(z) \iint dxdy\ p(x,y|z) \log \frac{p(x,y|z)}{p(x|z)\cdot p(y|z)}$$

- Direct links: Momentary information transfer (MIT)
$$I^{\mathrm{MIT}}_{X\to Y} = I(X;Y|\mathcal{P}_Y, \mathcal{P}_X)$$
- Indirect paths: Momentary information transfer along paths (MITP)
$$I^{\mathrm{MITP}}_{X\to Y}(\tau) = I(X;Y|\mathcal{P}_{paths})$$

Quantifying causal pathways

Information-theoretic approach [Runge, 2015]



$$X_t = f(\mathcal{P}_X) + \eta_t^X$$

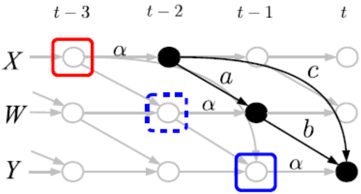
$$I^{\mathrm{MITP}}_{X\to Y} = I(\eta^X_{t-3}\ ;\ Y_t \mid \mathcal{P}_{paths})$$

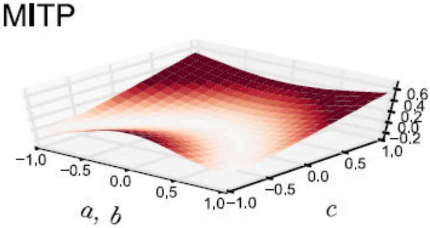
$$\mathcal{I}^{\mathrm{MII}}_{X\to Y|W_1} = \mathcal{I}(\eta^X_{t-3}; W_{1,t-2}; Y_t|\mathcal{P}_{paths})$$

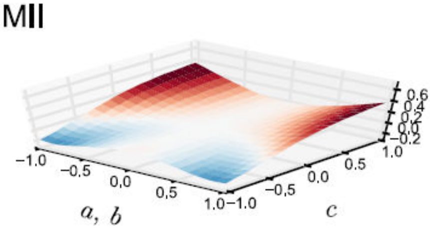
$$I\,(X;Y|{\color{red}Z}) = {\color{red}\int dz\, p(z)} \iint dxdy\;\; p(x,y|{\color{red}z}) \log \frac{p(x,y|{\color{red}z})}{p(x|{\color{red}z})\cdot p(y|{\color{red}z})}$$

- Direct links: Momentary information transfer (MIT)
$$I^{\mathrm{MIT}}_{X\to Y} = I(X;Y|\mathcal{P}_Y,\mathcal{P}_X)$$
- Indirect paths: Momentary information transfer along paths (MITP)
$$I^{\mathrm{MITP}}_{X\to Y}(\tau) = I(X;Y|\mathcal{P}_{paths})$$
- Mediation: Momentary interaction information (MII)
$$\mathcal{I}^{\mathrm{MII}}_{X\to Y|\mathbf{W}}(\tau) = I^{\mathrm{MITP}}_{X\to Y}(\tau) - \underbrace{I(X;Y \mid \mathcal{P}_{paths},\mathbf{W})}_{\text{MITP conditioned on } \mathbf{W}}$$

Information-theoretic approach [Runge, 2015]







Discussion and conclusion

Discussion and conclusion

- framework for **reliable** large-scale time-lagged causal discovery

Python code on `https://jakobrunge.github.io/tigramite/`

Paper: J. Runge, D. Sejdinovic, S. Flaxman (2017): https://arxiv.org/abs/1702.07007

Discussion and conclusion

- framework for **reliable** large-scale time-lagged causal discovery
- **flexible** regarding (non-parametric) conditional independence tests
  $\rightarrow$ nodes can be multivariate, variables discrete, ...

Python code on `https://jakobrunge.github.io/tigramite/`

Paper: J. Runge, D. Sejdinovic, S. Flaxman (2017):
https://arxiv.org/abs/1702.07007

Discussion and conclusion

- framework for **reliable** large-scale time-lagged causal discovery
- **flexible** regarding (non-parametric) conditional independence tests
  $\rightarrow$ nodes can be multivariate, variables discrete, ...
- **interpretable** MCI statistic measures causal strength
  $\rightarrow$ ranking causal links in large-scale analyses

Python code on `https://jakobrunge.github.io/tigramite/`

Paper: J. Runge, D. Sejdinovic, S. Flaxman (2017):
https://arxiv.org/abs/1702.07007

Discussion and conclusion

- framework for **reliable** large-scale time-lagged causal discovery
- **flexible** regarding (non-parametric) conditional independence tests
  → nodes can be multivariate, variables discrete, ...
- **interpretable** MCI statistic measures causal strength
  → ranking causal links in large-scale analyses
- Causal **pathway** analysis [PRE Dec 2015]

Python code on `https://jakobrunge.github.io/tigramite/`

Paper: J. Runge, D. Sejdinovic, S. Flaxman (2017):
https://arxiv.org/abs/1702.07007

Discussion and conclusion

- framework for **reliable** large-scale time-lagged causal discovery
- **flexible** regarding (non-parametric) conditional independence tests
  $\rightarrow$ nodes can be multivariate, variables discrete, ...
- **interpretable** MCI statistic measures causal strength
  $\rightarrow$ ranking causal links in large-scale analyses
- Causal **pathway** analysis [PRE Dec 2015]
- Causal complex **network measures** [Nature Comm. 2015]

Python code on `https://jakobrunge.github.io/tigramite/`

Paper: J. Runge, D. Sejdinovic, S. Flaxman (2017):
https://arxiv.org/abs/1702.07007

Discussion and conclusion

- framework for **reliable** large-scale time-lagged causal discovery
- **flexible** regarding (non-parametric) conditional independence tests
  $\rightarrow$ nodes can be multivariate, variables discrete, ...
- **interpretable** MCI statistic measures causal strength
  $\rightarrow$ ranking causal links in large-scale analyses
- Causal **pathway** analysis [PRE Dec 2015]
- Causal complex **network measures** [Nature Comm. 2015]
- Optimal **prediction** [PRE May 2015]

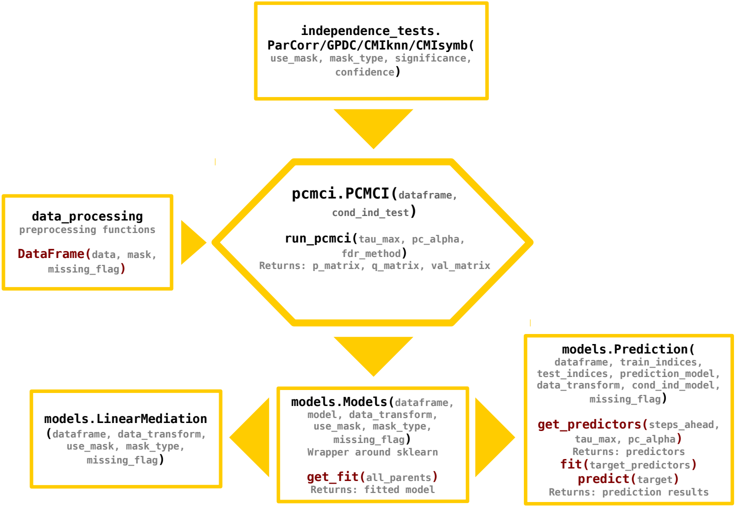
Python code on `https://jakobrunge.github.io/tigramite/`

Paper: J. Runge, D. Sejdinovic, S. Flaxman (2017):
https://arxiv.org/abs/1702.07007

Tigramite

Tigramite 3.0 → tigramite_tutorial.ipynb

https://jakobrunge.github.io/tigramite/



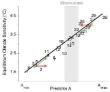
New research group

Climate Informatics

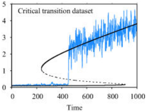
Open PhD and Postdoc positions
→ climateinformaticslab.com

Climate model evaluation

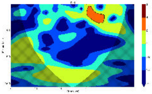
Climate sensitivity



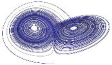
Extremes prediction



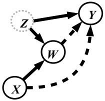
Time-scale causality



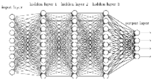
Data-driven dynamical models



Causal discovery theory



Deep learning



Thank you !!!

References i

Eichler, M. (2012).
Graphical modelling of multivariate time series.
Probability Theory and Related Fields, 153(1):233–268.

Frenzel, S. and Pompe, B. (2007).
Partial mutual information for coupling analysis of multivariate time series.
Physical Review Letters, 99(20):204101.

References ii

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996).
The NCEP/NCAR 40-year reanalysis project.
Bulletin of the American Meteorological Society, 77(3):437–471.

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004).
Estimating mutual information.
Physical Review E, 69(6):066138.

References iii

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J. (2016).

Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation.
Journal of Climate, 29(11):4069–4081.

Kretschmer, M., Coumou, D., and Runge, J. (2017).
Early prediction of weak stratospheric polar vortex states using causal precursors.
under review.

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

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

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

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

Runge, J., Riedl, M., Müller, A., Stepan, H., Wessel, N., and Kurths, J. (2015c).
Quantifying the causal strength of multivariate cardiovascular couplings with momentary information transfer.
Physiological Measurement, 36(4):813–825.

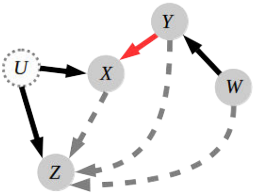
Spirtes, P., Glymour, C., and Scheines, R. (2000).
Causation, Prediction, and Search, volume 81.
The MIT Press, Boston.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007).
MEASURING AND TESTING DEPENDENCE BY CORRELATION OF DISTANCES.
The Annals of Statistics, 35(6):2769–2794.

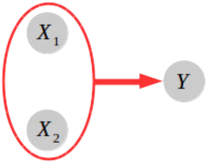
Vejmelka, M., Pokorná, L., Hlinka, J., Hartman, D., Jajcay, N., and Paluš, M. (2014).
Non-random correlation structures and dimensionality reduction in multivariate climate data.
Climate Dynamics, 44(9-10):2663–2682.

Causal discovery challenges

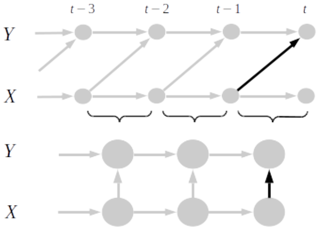
Latent variables



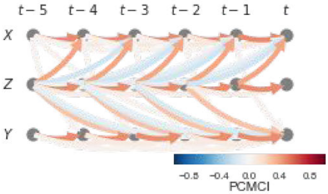
Synergistic dependencies



Contemporaneous effects



Long-range memory



Causal assumptions

Causal assumptions

Causal interpretation assumes [Spirtes et al., 2000]:

- **Causal Markov Condition:** "All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"
- **Causal Sufficiency:** Measured variables include all of the common causes
- **Faithfulness / Stableness:** "Independencies in data arise not from incredible coincidence, but rather from causal structure"; violated by purely deterministic dependencies

Causal assumptions

Causal interpretation assumes [Spirtes et al., 2000]:

- **Causal Markov Condition:** "All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"
- **Causal Sufficiency:** Measured variables include all of the common causes
- **Faithfulness / Stableness:** "Independencies in data arise not from incredible coincidence, but rather from causal structure"; violated by purely deterministic dependencies
- **No contemporaneous effects**: (but can be extended)
- **Stationarity:** time series case
- **Parametric assumptions** of independence tests

Causal assumptions

Causal interpretation assumes [Spirtes et al., 2000]:

- **Causal Markov Condition:** "All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"
- **Causal Sufficiency:** Measured variables include all of the common causes
- **Faithfulness / Stableness:** "Independencies in data arise not from incredible coincidence, but rather from causal structure"; violated by purely deterministic dependencies
- **No contemporaneous effects**: (but can be extended)
- **Stationarity:** time series case
- **Parametric assumptions** of independence tests

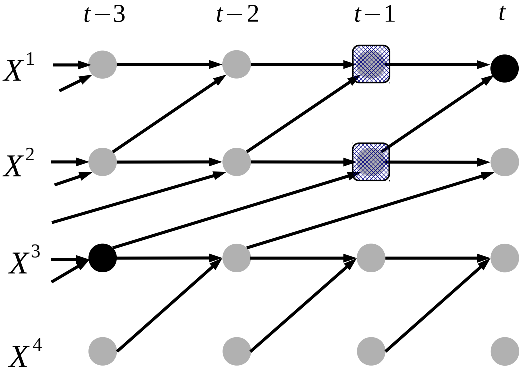
More discussion $\rightarrow$ appendix

Causal assumptions

Causal Markov Condition

"All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"

Formally: Upon specifying a complete graph that contains all common causes: separation in graph entails *at least* implied conditional independencies in process

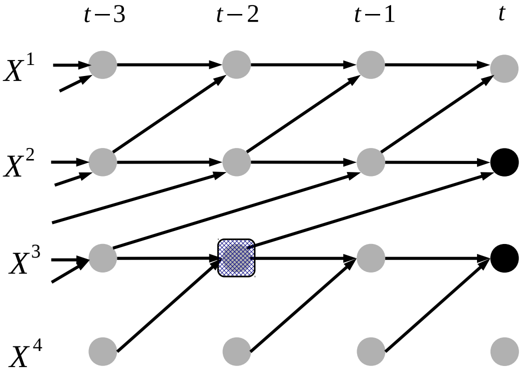


Causal assumptions

Causal Markov Condition

"All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"

Formally: Upon specifying a complete graph that contains all common causes: separation in graph entails *at least* implied conditional independencies in process

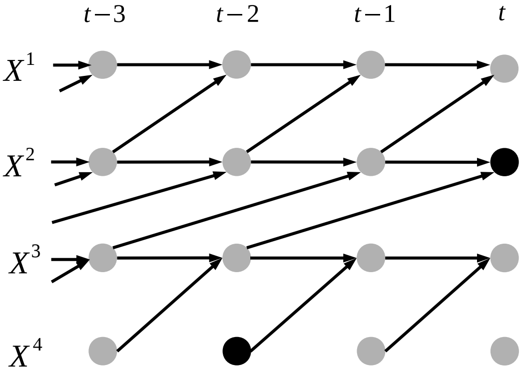


Causal assumptions

Causal Markov Condition

"All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"

Formally: Upon specifying a complete graph that contains all common causes: separation in graph entails *at least* implied conditional independencies in process

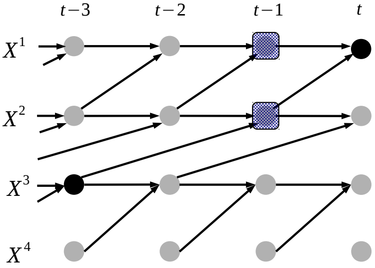


Causal assumptions

Causal Markov Condition

"All the relevant probabilistic information that can be obtained from the system is contained in its direct causes"

Formally: Upon specifying a complete graph that contains all common causes: separation in graph entails *at least* implied conditional independencies in process



Structural equation modeling framework:

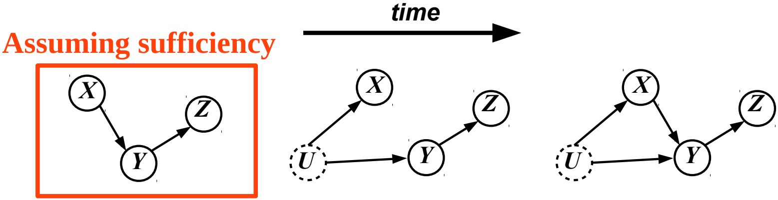
$X_t^j = f(\mathbf{X}_t^-, \eta_t^j) \qquad \eta_t^j \perp\!\!\!\perp \mathbf{X}_{t+1}^- \setminus X_t^j$

Causal assumptions

Causal Sufficiency

R. Scheines: "Theory of causal inference is about the inferential effect of a variety of assumptions far more than it is an endorsement of particular assumptions"

Given estimates $X \perp\!\!\!\perp Z \mid Y$ and no other independencies. Assuming only Markov condition and faithfulness allows for several different graphs:

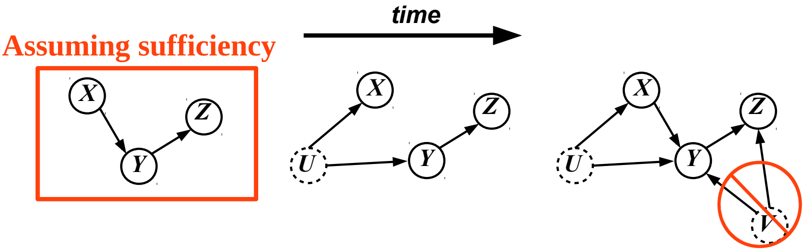


Causal assumptions

Causal Sufficiency

R. Scheines: "Theory of causal inference is about the inferential effect of a variety of assumptions far more than it is an endorsement of particular assumptions"

Given estimates $X \perp\!\!\!\perp Z \mid Y$ and no other independencies. Assuming only Markov condition and faithfulness allows for several different graphs:



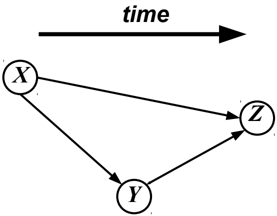
Causal assumptions

Faithfulness

If there are any independence relations in the population that are not a consequence of the Causal Markov condition (or d-separation), then the population is unfaithful.

For example, given three variables and assuming the Causal Markov and Sufficiency Conditions, suppose we measure these (in-)dependencies:

$$
\begin{array}{rl}
{\color{red} X \perp\!\!\!\perp Z} & X \not\!\perp\!\!\!\perp Z \mid Y \\
X \not\!\perp\!\!\!\perp Y & (X \not\!\perp\!\!\!\perp Y \mid Z) \\
Y \not\!\perp\!\!\!\perp Z & Y \not\!\perp\!\!\!\perp Z \mid X
\end{array}
$$



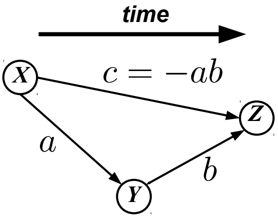
Causal assumptions

Faithfulness

If there are any independence relations in the population that are not a consequence of the Causal Markov condition (or d-separation), then the population is unfaithful.

For example, given three variables and assuming the Causal Markov and Sufficiency Conditions, suppose we measure these (in-)dependencies:

$$X \perp\!\!\!\perp Z \qquad X \not\!\perp\!\!\!\perp Z \mid Y$$
$$X \not\!\perp\!\!\!\perp Y \qquad (X \not\!\perp\!\!\!\perp Y \mid Z)$$
$$Y \not\!\perp\!\!\!\perp Z \qquad Y \not\!\perp\!\!\!\perp Z \mid X$$



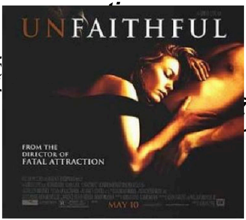
Causal assumptions

Faithfulness

If there are any independence relations in the population that are not a consequence of the Causal Markov condition (or d-separation), then the population is unfaithful.

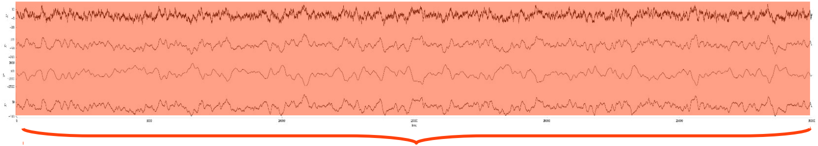
For example, given three variables and assuming the Causal Markov and Sufficiency Conditions, suppose we measure these (in-)dependencies:

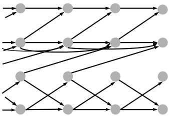
$$X \perp\!\!\!\perp Z \qquad X \not\!\perp\!\!\!\perp Z \mid Y$$
$$X \not\!\perp\!\!\!\perp Y \qquad (X \not\!\perp\!\!\!\perp Y \mid Z)$$
$$Y \not\!\perp\!\!\!\perp Z \qquad Y \not\!\perp\!\!\!\perp Z \mid X$$



Causal assumptions

Stationarity of causal structure

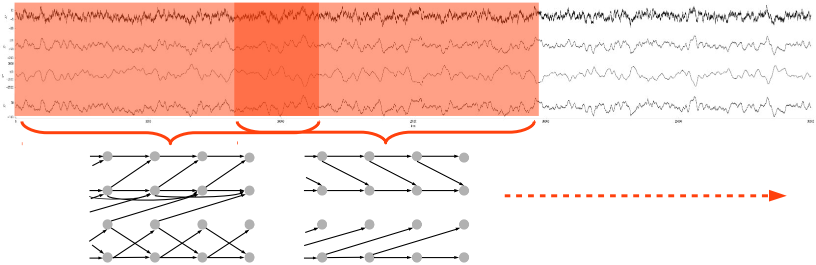




structurally stationary for all samples

Causal assumptions

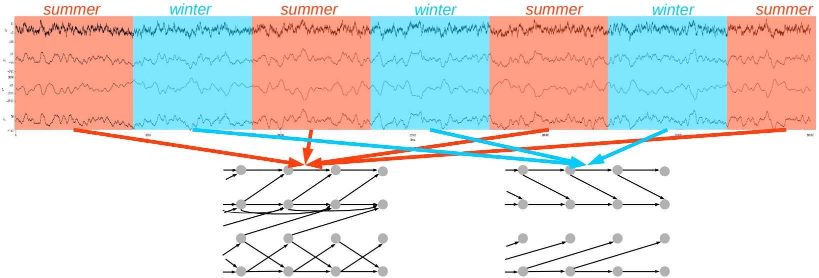
Stationarity of causal structure



structurally stationary within sliding windows

Causal assumptions

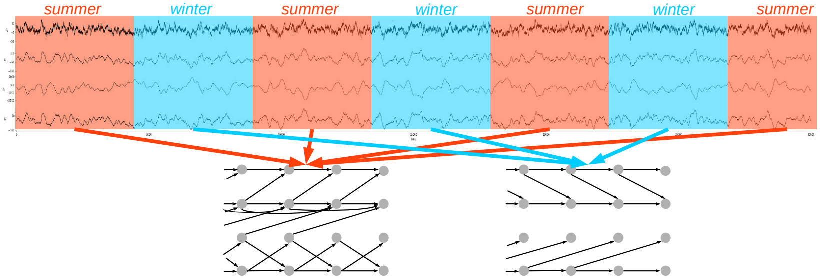
Stationarity of causal structure



periodically structurally stationary

Causal assumptions

Stationarity of causal structure

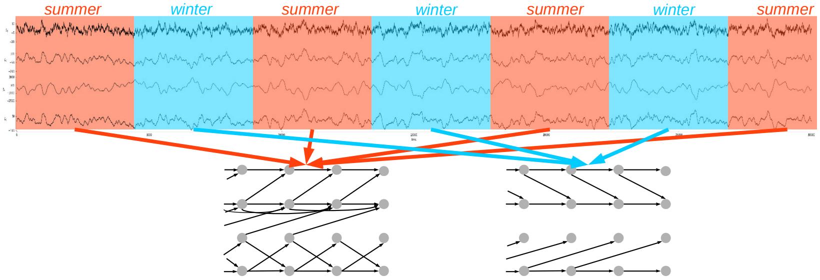


periodically structurally stationary

- structurally, *not* necessarily same strengh / parameters

Causal assumptions

Stationarity of causal structure



periodically structurally stationary

- structurally, *not* necessarily same strengh / parameters
- *masking* implemented in `Tigramite`