Quantifying causal interactions from time series of complex systems

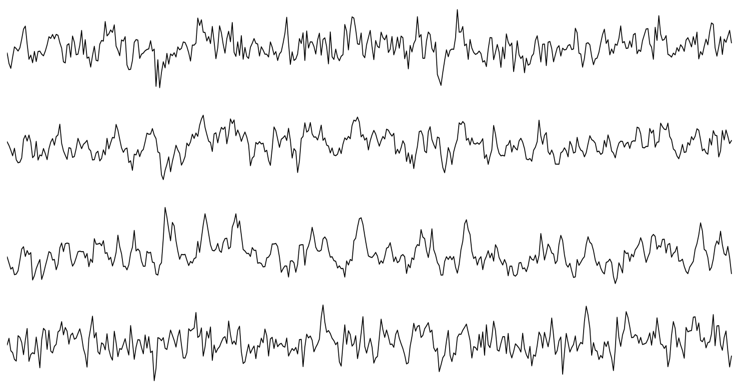
Jakob Runge

CIDNET Workshop June 16 – 20, 2014, MPI Dresden

Approach: Inferring interactions in processes from investigating their time series...

arbirator-robloxnews.blogspot.com

Earth from Apollo 13 (wikipedia.org)



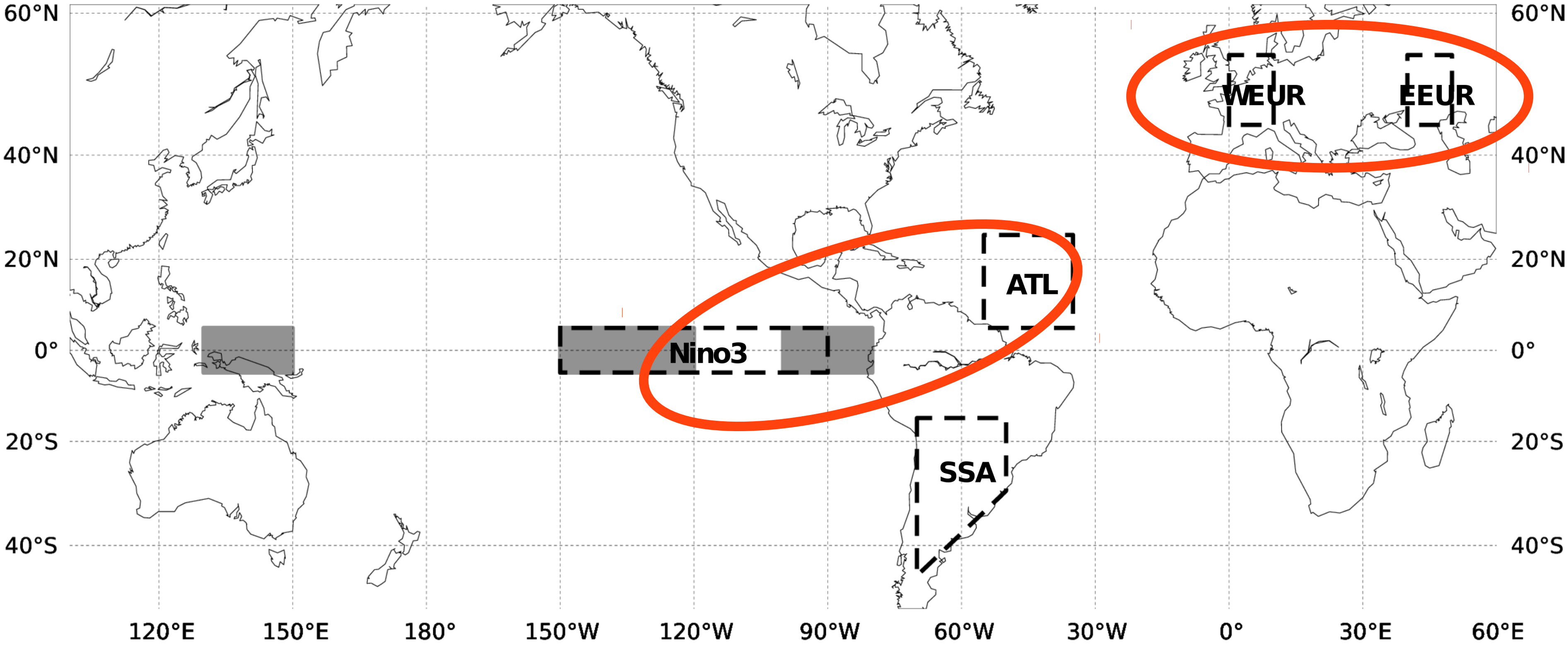
bigthink.com

Outline

1. Pitfalls in estimating **coupling delays** with lagged cross correlation / mutual information
2. Ambiguities in **interpreting mutual information and transfer entropy**
3. Representing causal interactions with **time series graphs**
4. Quantifying **causal strength** with momentary information transfer
5. Estimating conditional mutual information and coping with **autocorrelation in significance testing**

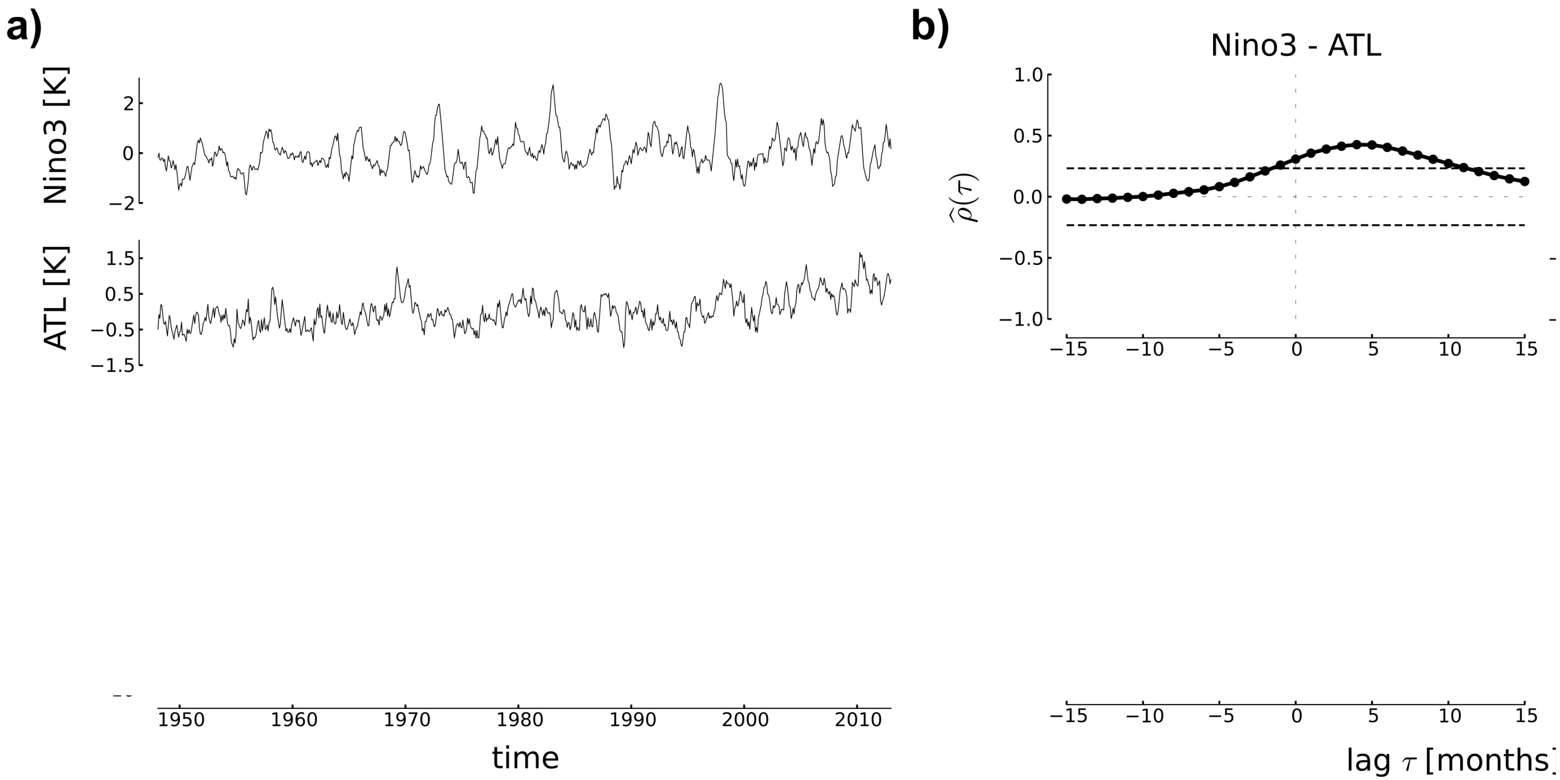
1. Inferring coupling delays

Example: ENSO teleconnections

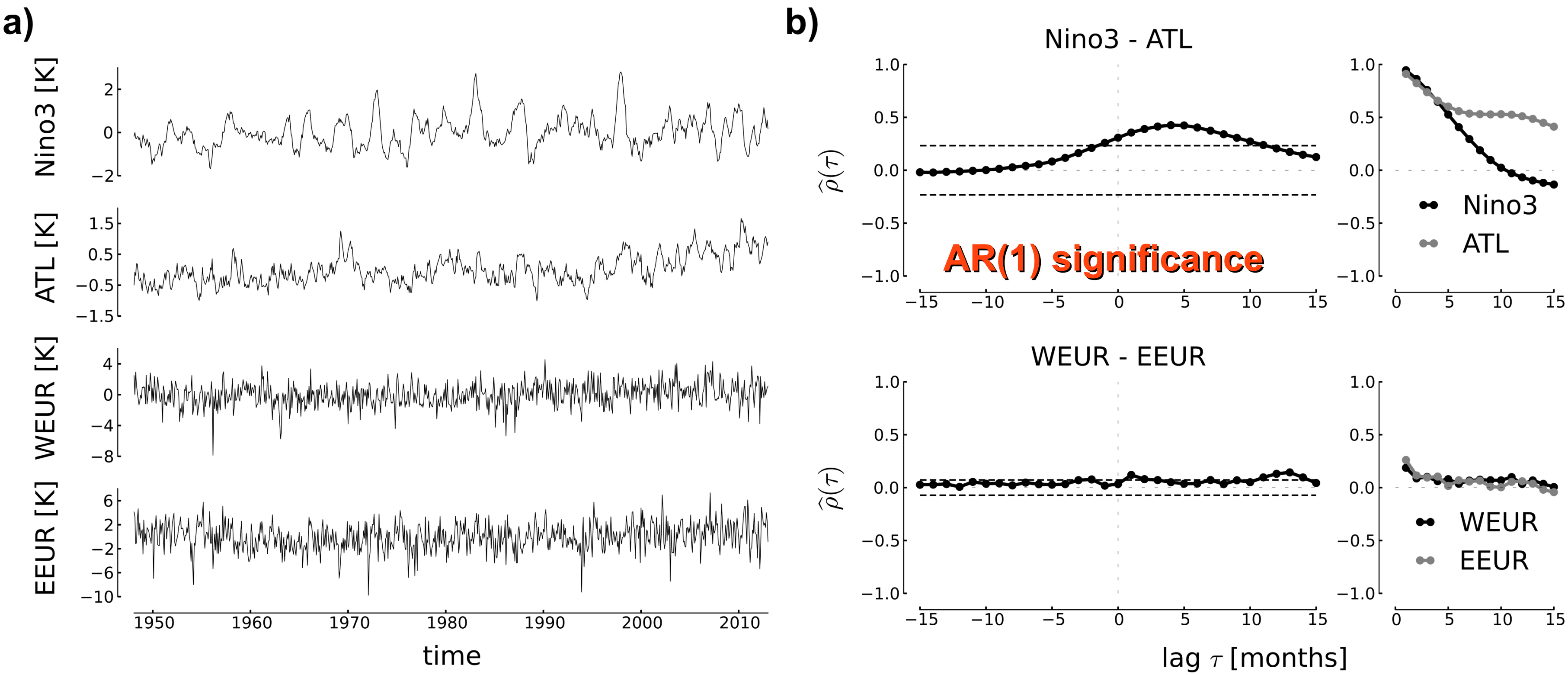


Reanalysis data: Monthly surface air temperature and pressure
Kalnay et al., 1996: The NCEP/NCAR 40-Year Reanalysis Project. Bulletin of the American Meteorological Society, 77(3), 437–471.

Lagged correlations $\rho(X_{t-\tau}; Y_t)$

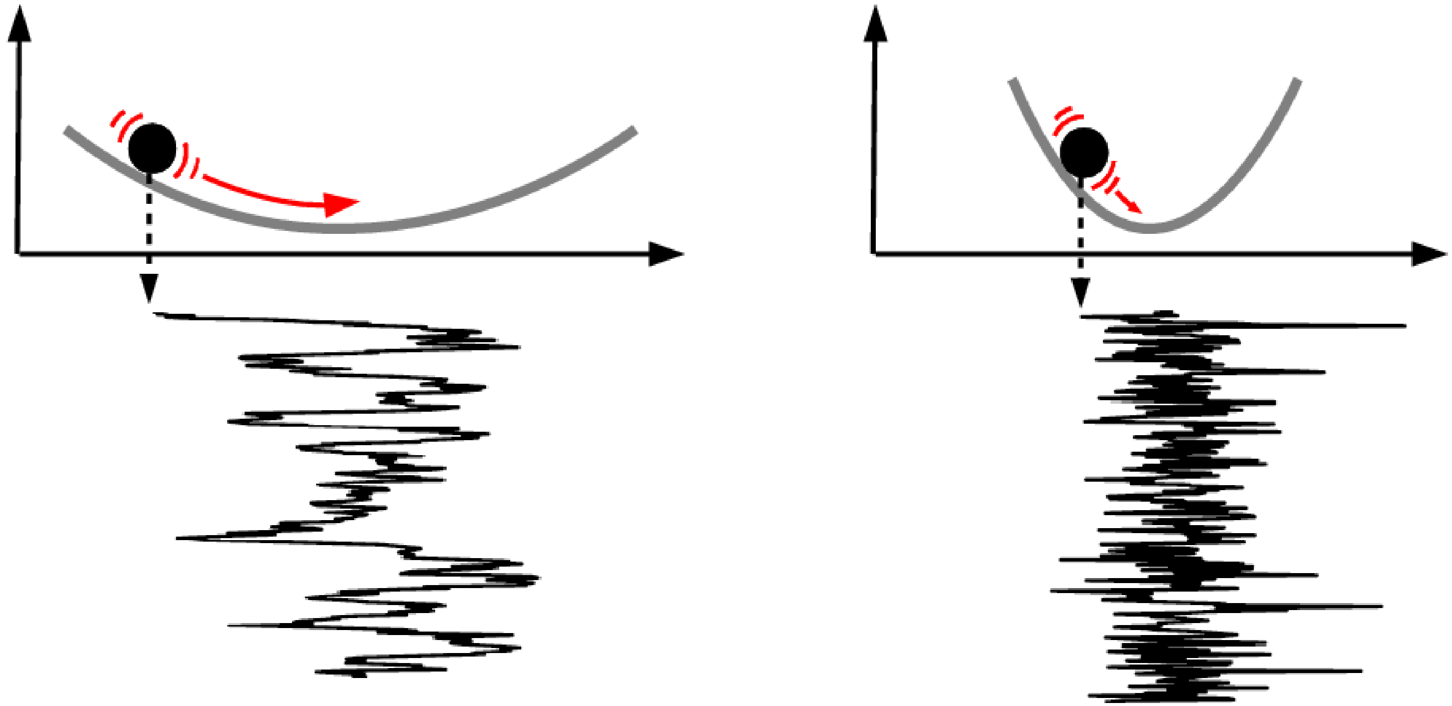


Lagged correlations



Toy model

$$X_t = aX_{t-1} + \varepsilon_t^X$$

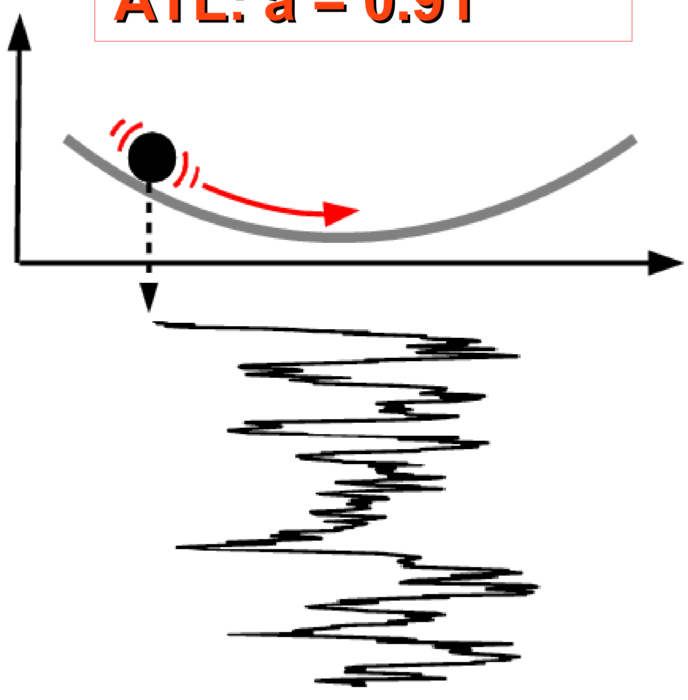


Toy model

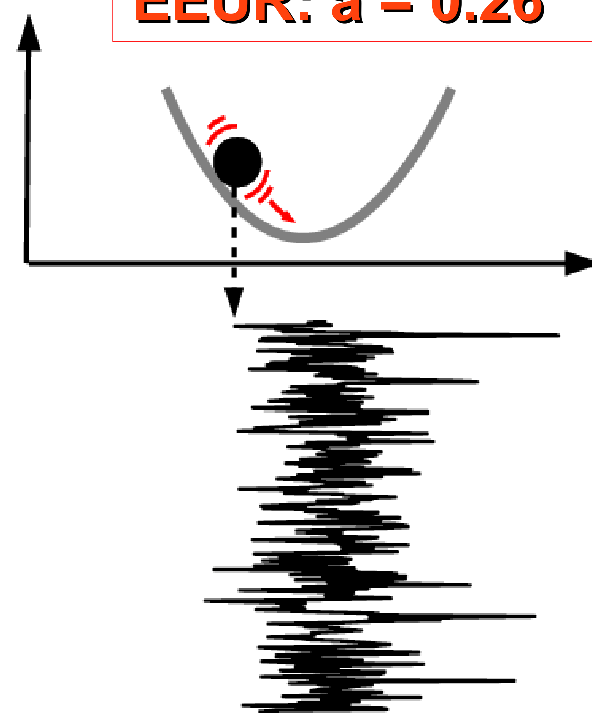
$$X_t = \underline{a} X_{t-1} + \varepsilon_t^X$$

"internal dynamics"

Nino3: a = 0.95
ATL: a = 0.91



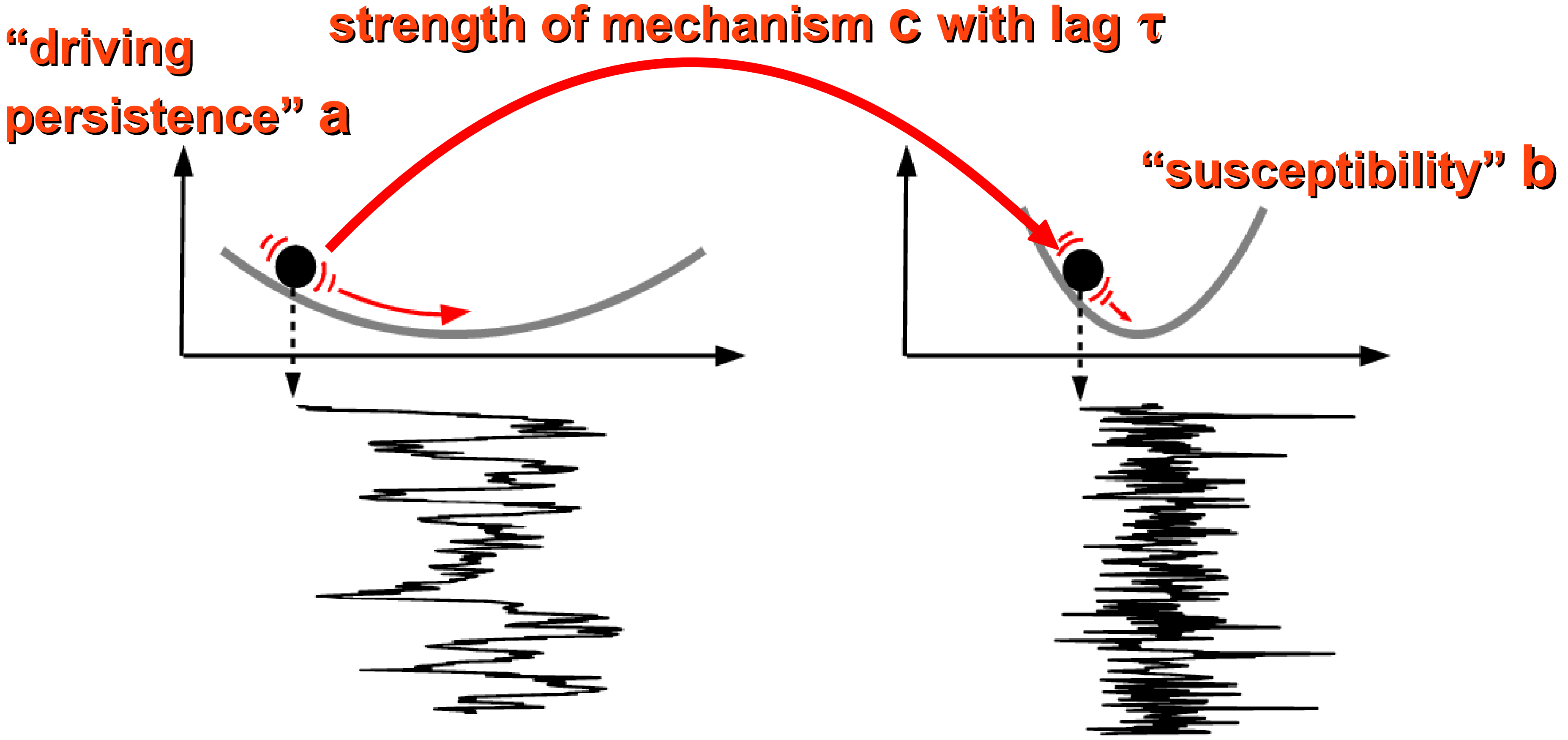
WEUR: a = 0.19
EEUR: a = 0.26



Toy model

$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + \underline{cX_{t-\tau}} + \varepsilon_t^Y$$



Analytical lagged correlations

c = 0.1,

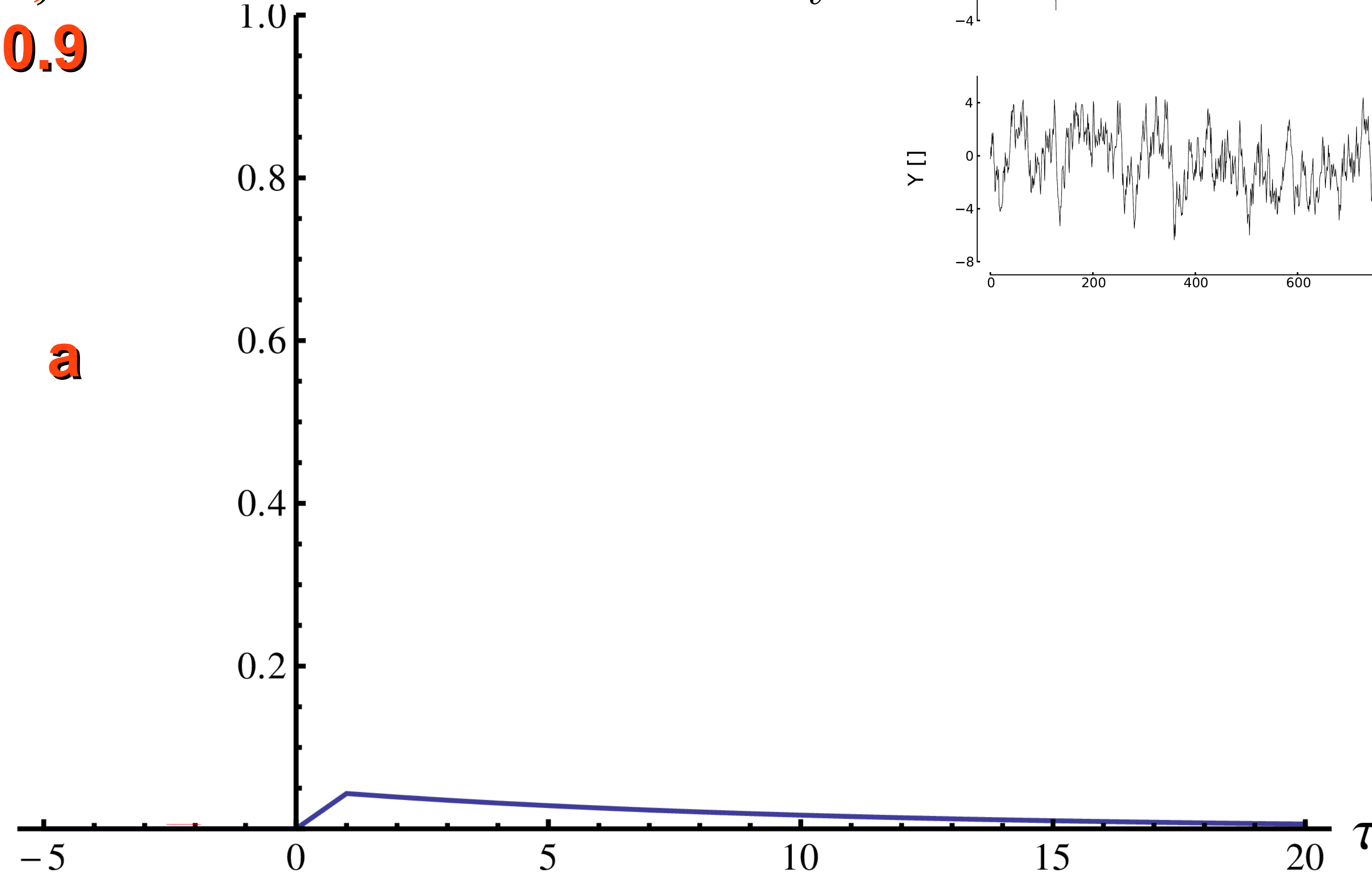
τ = 1,

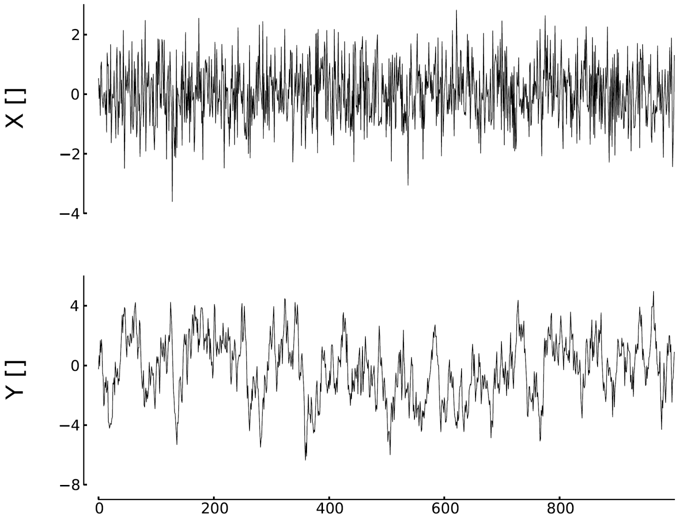
b = 0.9

$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + cX_{t-\tau} + \varepsilon_t^Y$$

a





Analytical lagged correlations

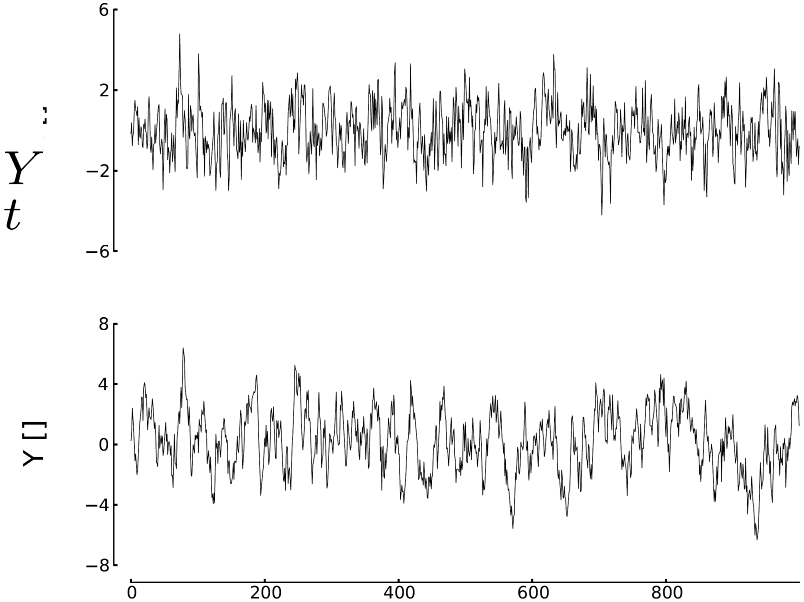
c = 0.1,

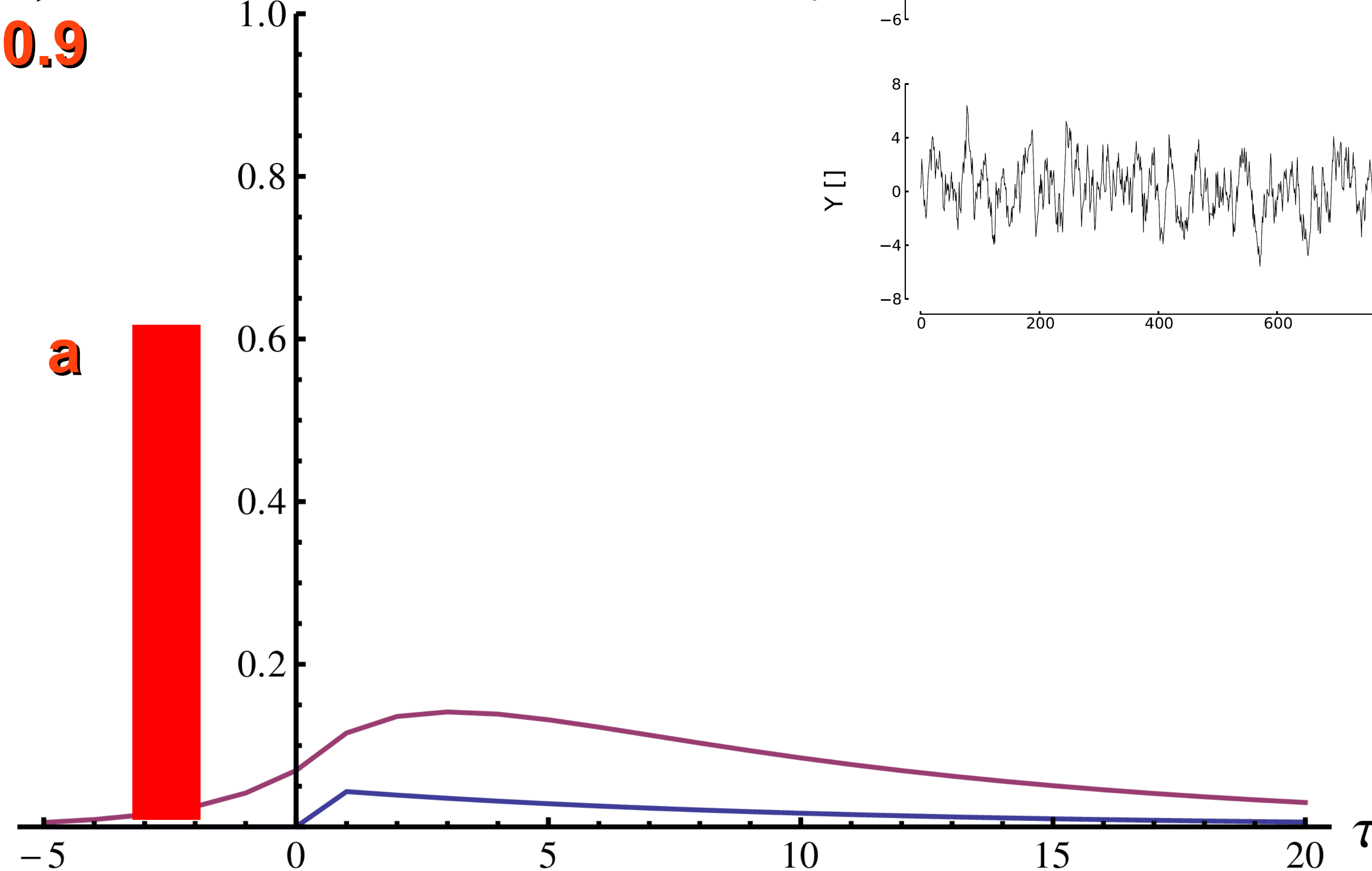
τ = 1,

b = 0.9

$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + cX_{t-\tau} + \varepsilon_t^Y$$





Analytical lagged correlations

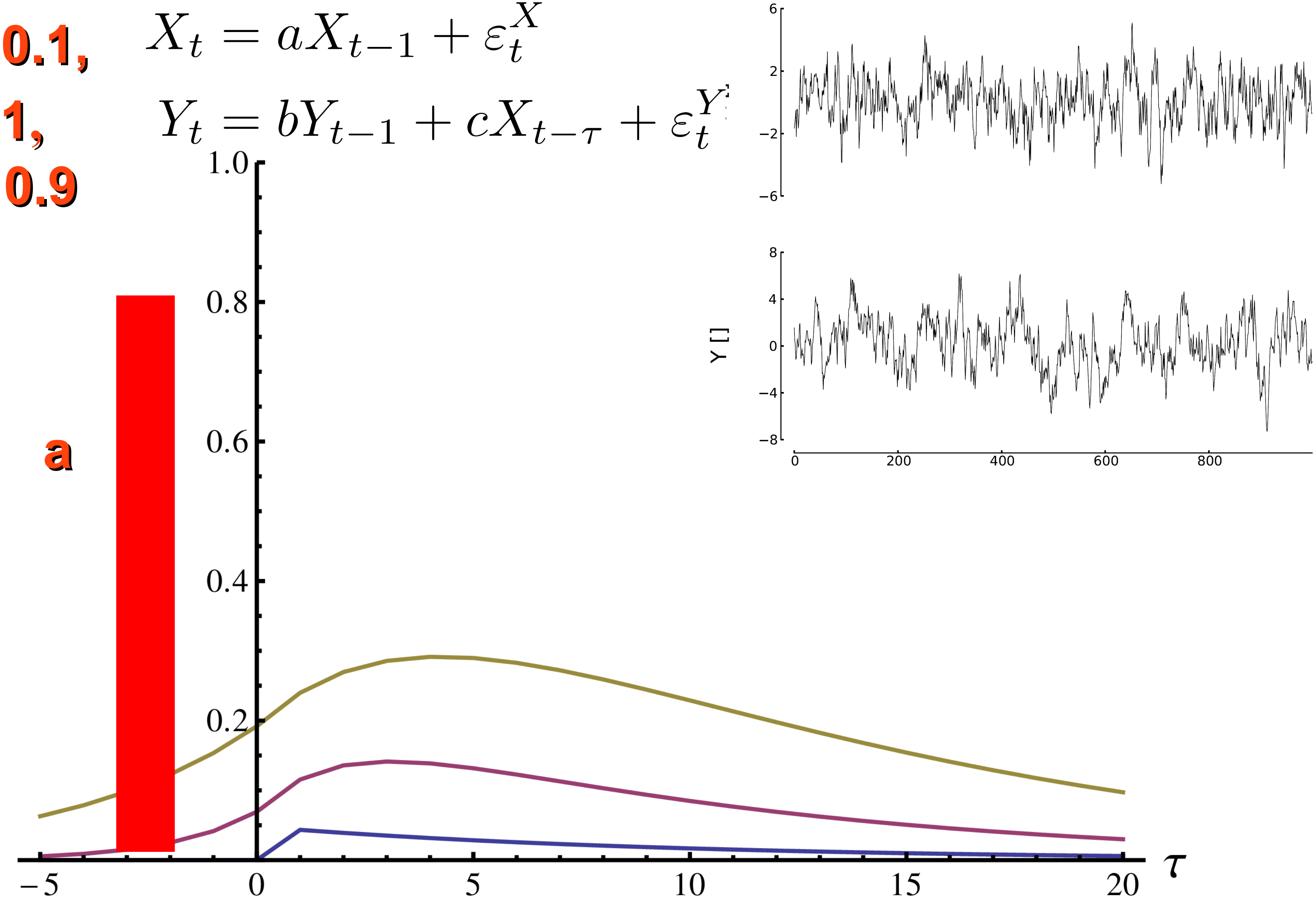
c = 0.1,

τ = 1,

b = 0.9

$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + cX_{t-\tau} + \varepsilon_t^Y$$



Analytical lagged correlations

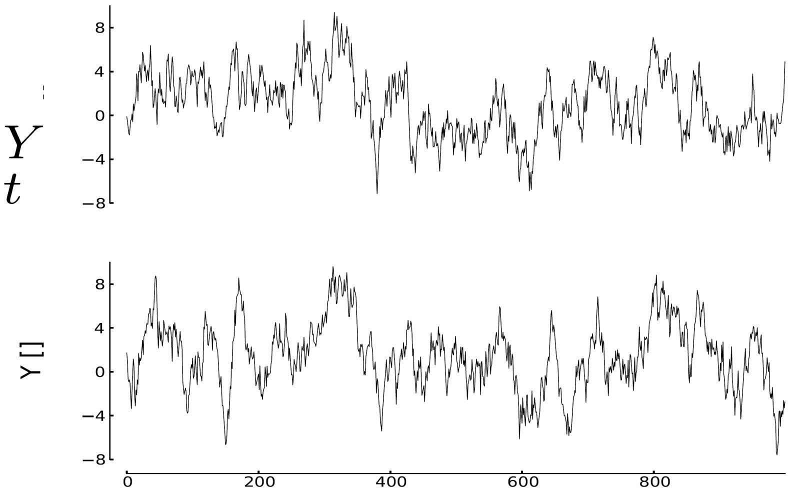
c = 0.1,

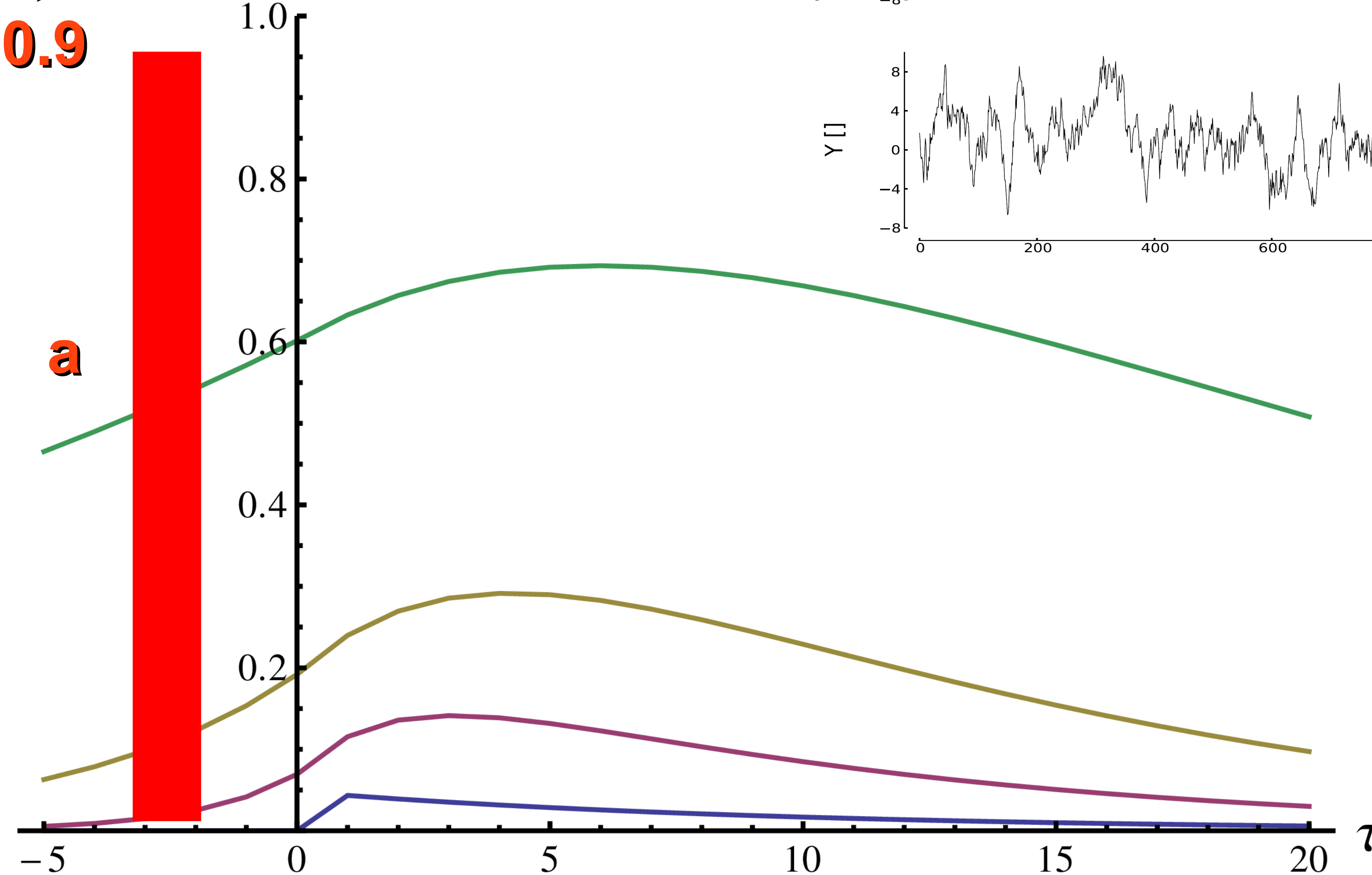
τ = 1,

b = 0.9

$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + cX_{t-\tau} + \varepsilon_t^Y$$





Analytical lagged correlations

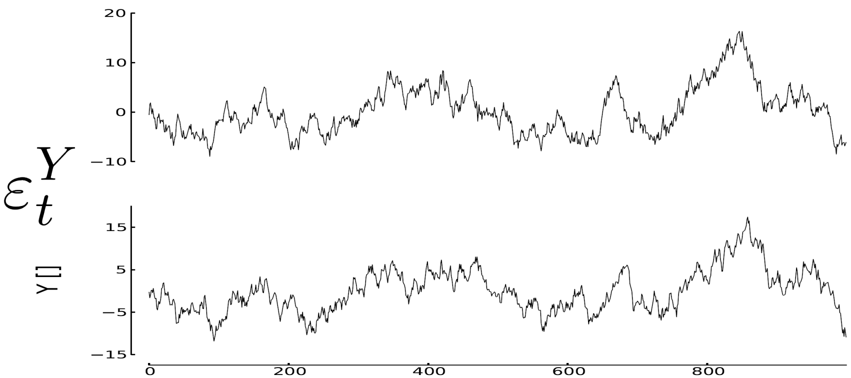
c = 0.1,

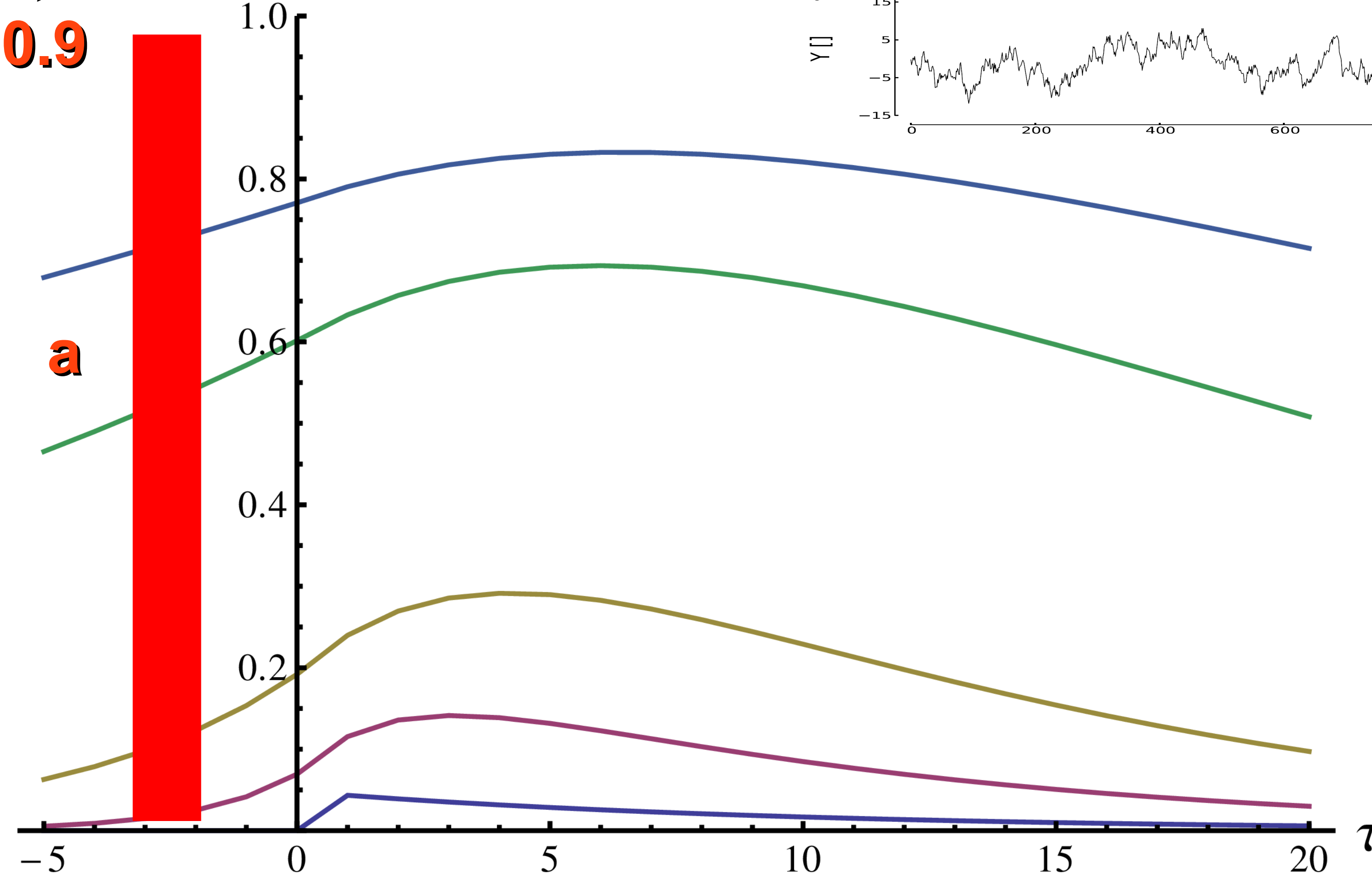
τ = 1,

b = 0.9

$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + cX_{t-\tau} + \varepsilon_t^Y$$





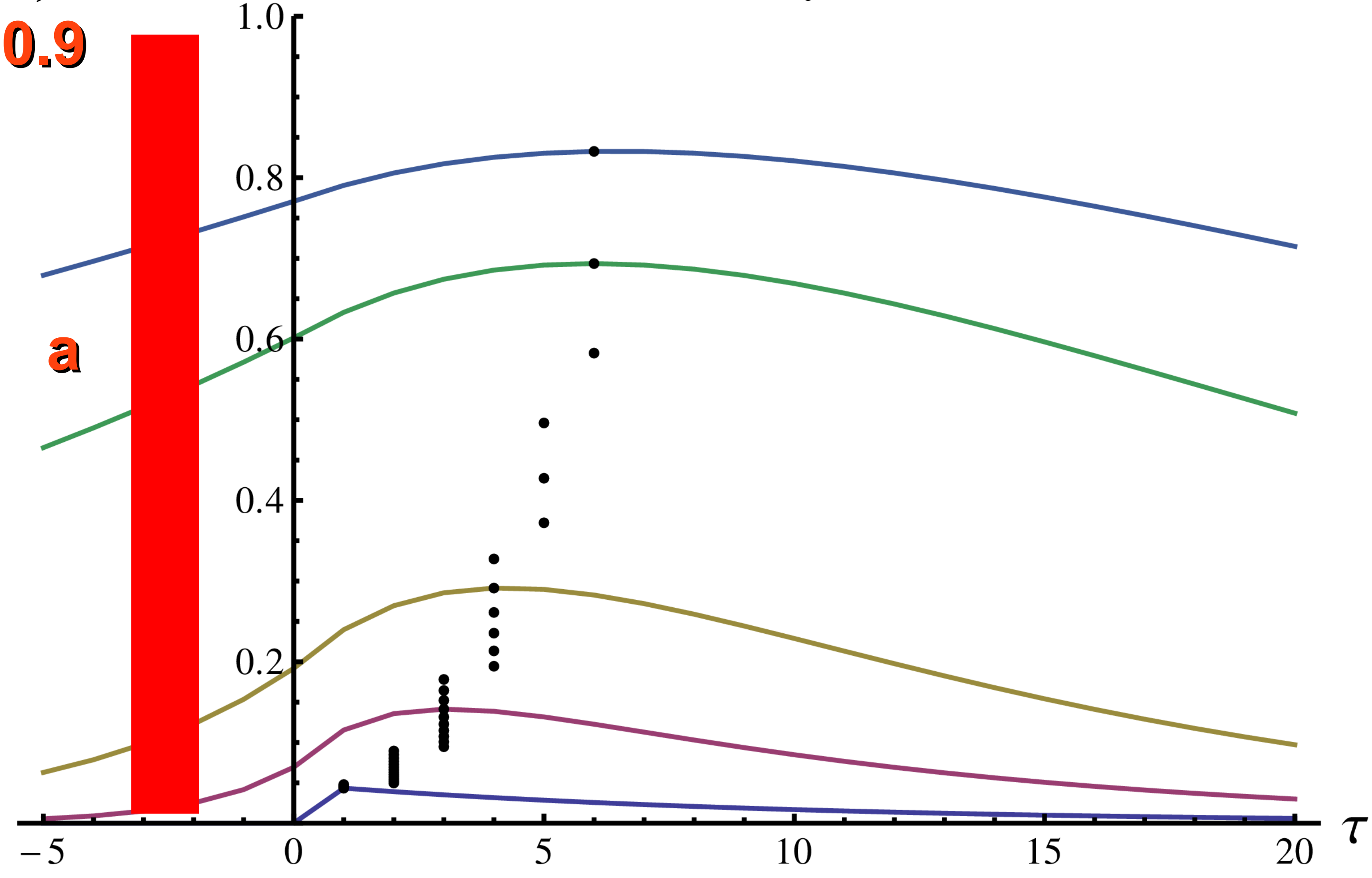
Analytical lagged correlations

c = 0.1,
τ = 1,
b = 0.9

$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + cX_{t-\tau} + \varepsilon_t^Y$$

a



Analytical lagged correlations

c = 0.1,

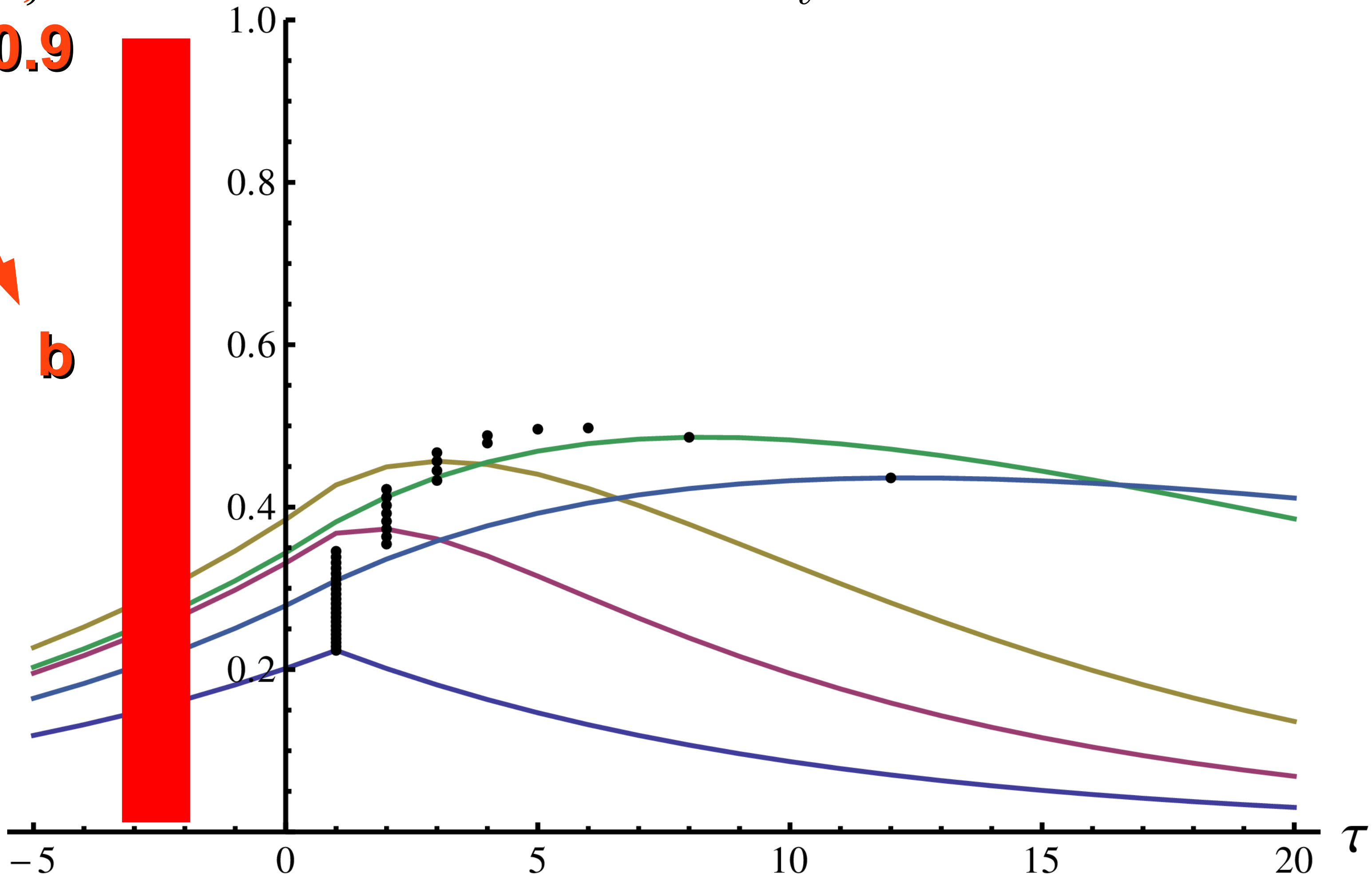
τ = 1,

a = 0.9

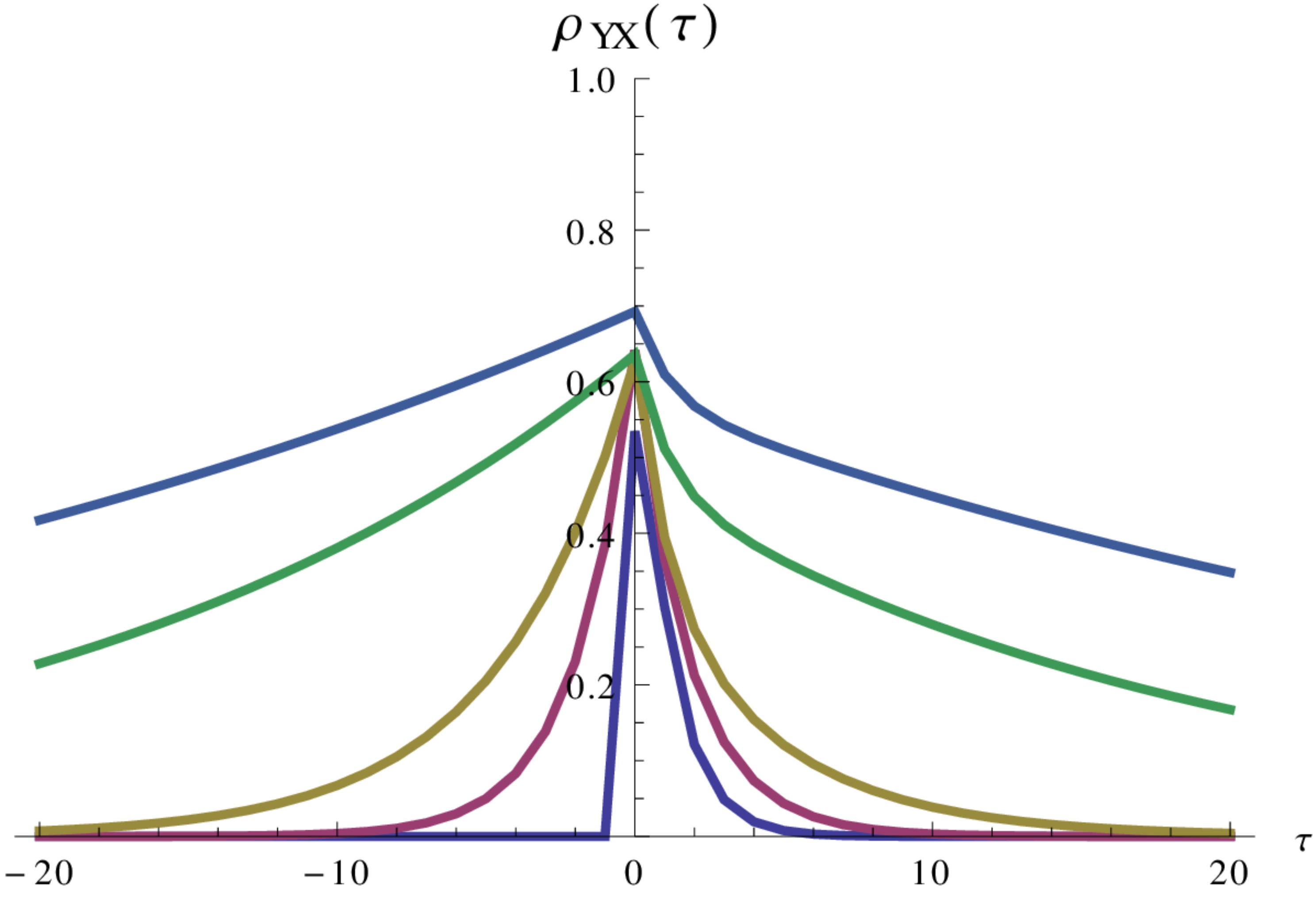
$$X_t = aX_{t-1} + \varepsilon_t^X$$

$$Y_t = bY_{t-1} + cX_{t-\tau} + \varepsilon_t^Y$$

b



Who drives whom?



Analysis summary

➡ height of peak strongly varies with “driving persistence” *a*

➡ lag is strongly shifted for large “susceptibility” *b*

Analysis summary

➡ height of peak strongly varies with "driving persistence" *a*

➡ lag is strongly shifted for large "susceptibility" *b*

for the same small

mechanism strength $c = 0.1$

and mechanism delay $\tau = 1$

Analysis summary

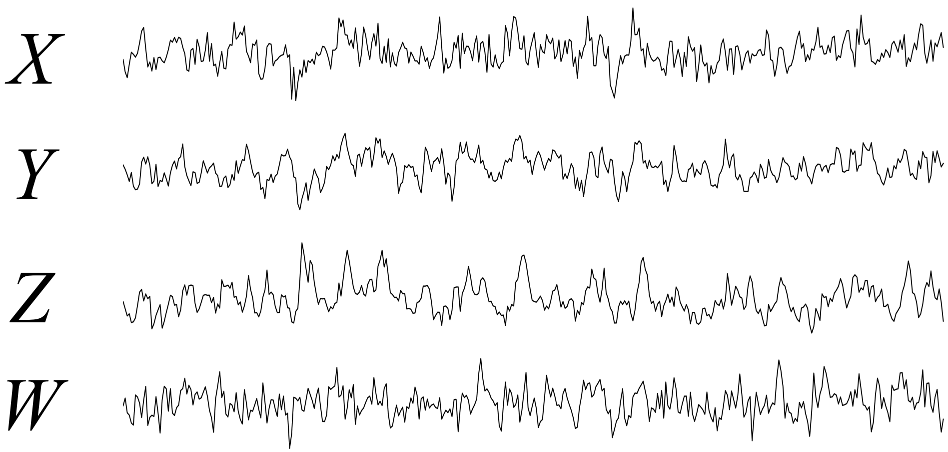
➡ **height of peak strongly varies with “driving persistence” *a***

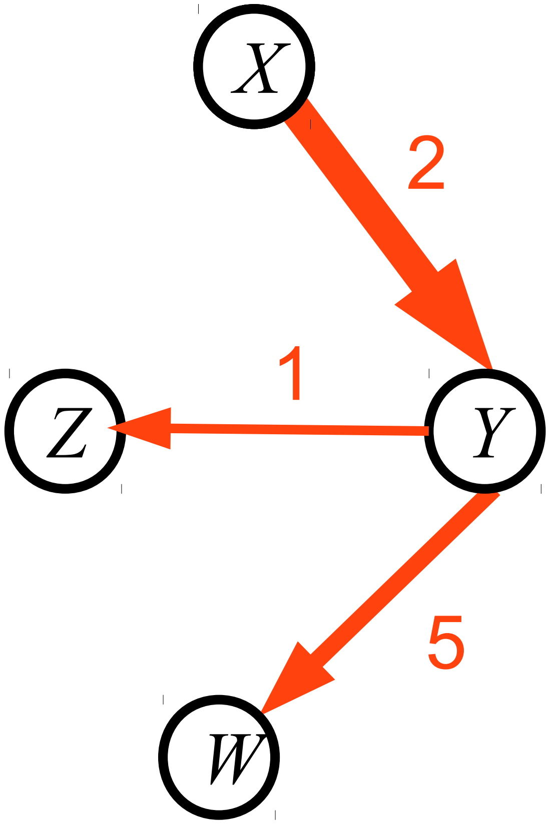
➡ **lag is strongly shifted for large “susceptibility” *b***

Same thing happens for lagged mutual information!

2. Measuring causal strength

How to interprete measures of coupling strength





Measures based on (conditional) mutual information

$$\mathrm{I}(\mathrm{X};\mathrm{Y}\ |Z) = \int p(z) \iint p(x,y|z) \log \frac{p(x,y|z)}{p(x|z)\cdot p(y|z)} \mathrm{d}x\mathrm{d}y\ \mathrm{d}z$$
$$= H(X|Z) + H(Y|Z) - H(X,Y|Z)$$

Estimation via *k*-nearest neighbor statistics

Commonly used approach: (Multivariate) Transfer Entropy (TE)

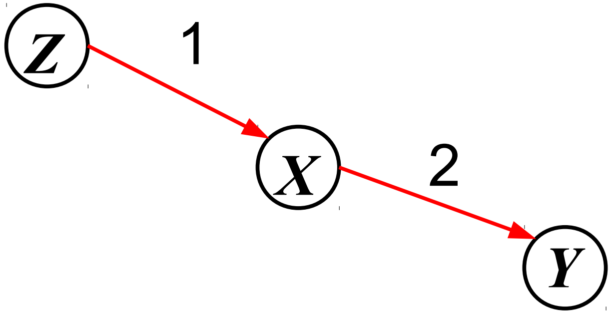
= Generalized Granger Causality

$$I^{\mathrm{TE}}_{X\to Y} = I(X_t^-; Y_t \,|\, Y_t^-, \ldots)$$

$$X_t^- = (X_{t-1}, X_{t-2}, \ldots)$$

T. Schreiber, Phys.Rev.Lett. 85, 461 (2000)
Barnett et al., Physical Review Letters, 103, 238701 (2009)

But what does TE measure?

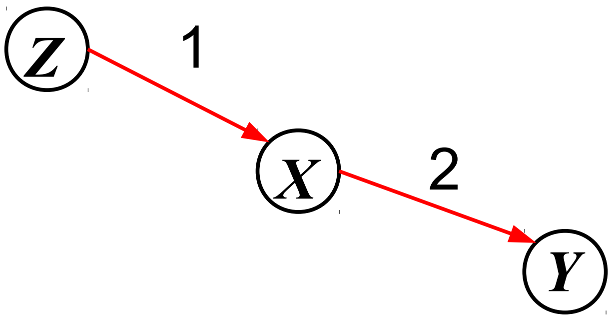


$$Z_t = \eta_t^Z$$

$$X_t = a_X Z_{t-1} + \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + \eta_t^Y$$

But what does TE measure?



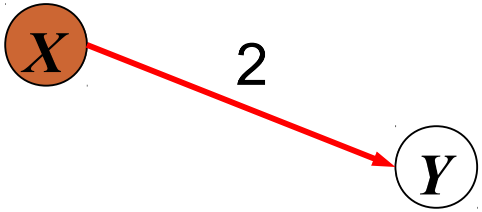
$$Z_t = \eta_t^Z$$

$$X_t = a_X Z_{t-1} + \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + \eta_t^Y$$

$$I_{X \to Y}^{\mathrm{TE}} = \frac{1}{2} \ln \left(1 + \frac{c_{XY}^2 (\sigma_X^2 + a_X^2 \sigma_Z^2)}{\sigma_Y^2} \right)$$

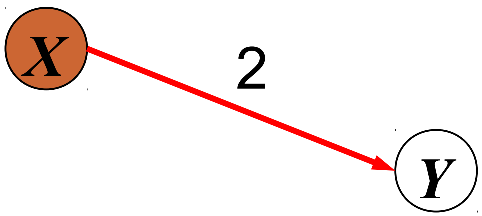
But what does TE measure?



$$X_t = a_X X_{t-1} + \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + \eta_t^Y$$

But what does TE measure?

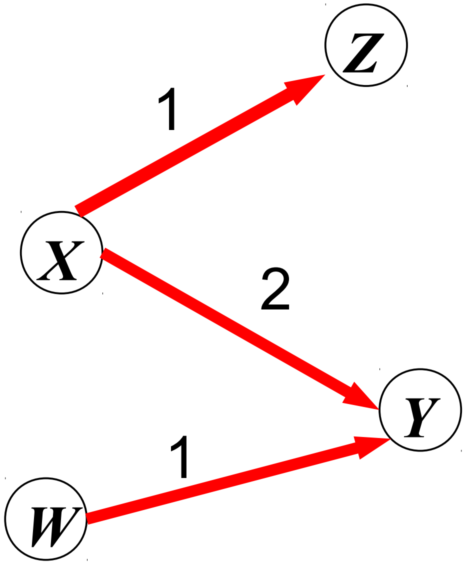


$$X_t = a_X X_{t-1} + \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + \eta_t^Y$$

$$I_{X\to Y}^{\mathrm{TE}} = \frac{1}{2}\ln\left(1 + \frac{(c_{XY}^2\sigma_X^2)/(1-a_X^2)}{\sigma_Y^2}\right)$$

But what does conditional TE measure?



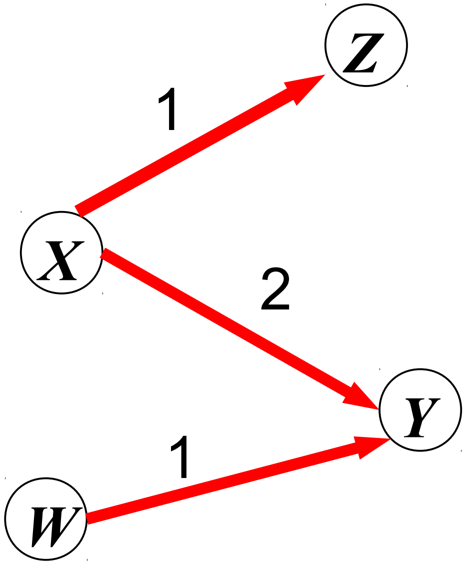
$$Z_t = c_{XZ} X_{t-1} + \eta_t^Z$$

$$X_t = \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + c_{WY} W_{t-1} + \eta_t^Y$$

$$W_t = \eta_t^W$$

But what does conditional TE measure?



$$Z_t = c_{XZ} X_{t-1} + \eta_t^Z$$

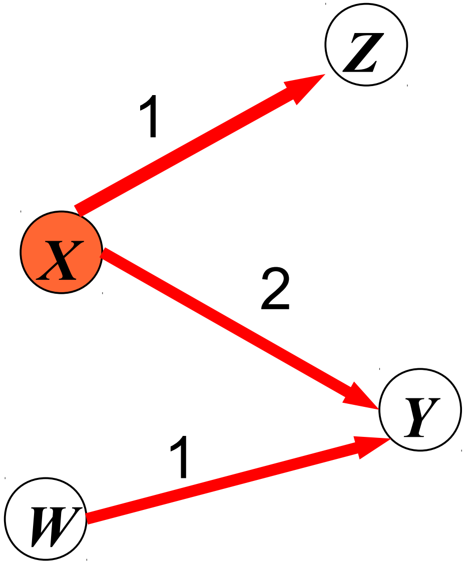
$$X_t = \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + c_{WY} W_{t-1} + \eta_t^Y$$

$$W_t = \eta_t^W$$

$$I^{\mathrm{TE}}_{X \to Y} = \frac{1}{2} \ln \left(1 + \frac{c_{XY}^2 \sigma_X^2 \sigma_Z^2}{\sigma_Y^2 (c_{XZ}^2 \sigma_X^2 + \sigma_Z^2)} \right)$$

What does the Mutual Information measure?



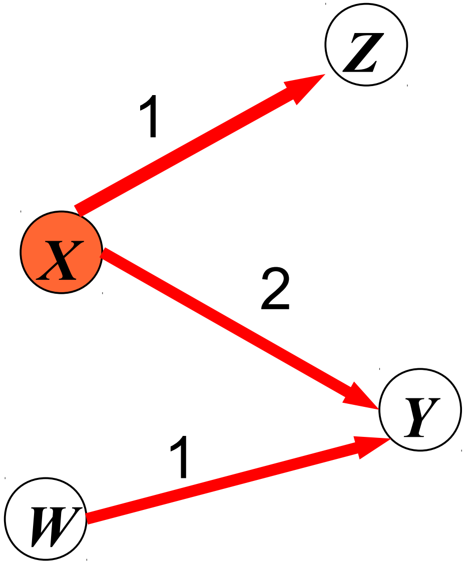
$$Z_t = c_{XZ} X_{t-1} + \eta_t^Z$$

$$X_t = a_X X_{t-1} + \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + c_{WY} W_{t-1} + \eta_t^Y$$

$$W_t = \eta_t^W$$

What does the Mutual Information measure?



$$Z_t = c_{XZ} X_{t-1} + \eta_t^Z$$

$$X_t = a_X X_{t-1} + \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + c_{WY} W_{t-1} + \eta_t^Y$$

$$W_t = \eta_t^W$$

$$I_{X\to Y}^{\mathrm{MI}} = \frac{1}{2} \ln \left(1 + \frac{(c_{XY}^2 \sigma_X^2)/(1-a_X^2)}{c_{WY}^2 \sigma_W^2 + \sigma_Y^2} \right)$$

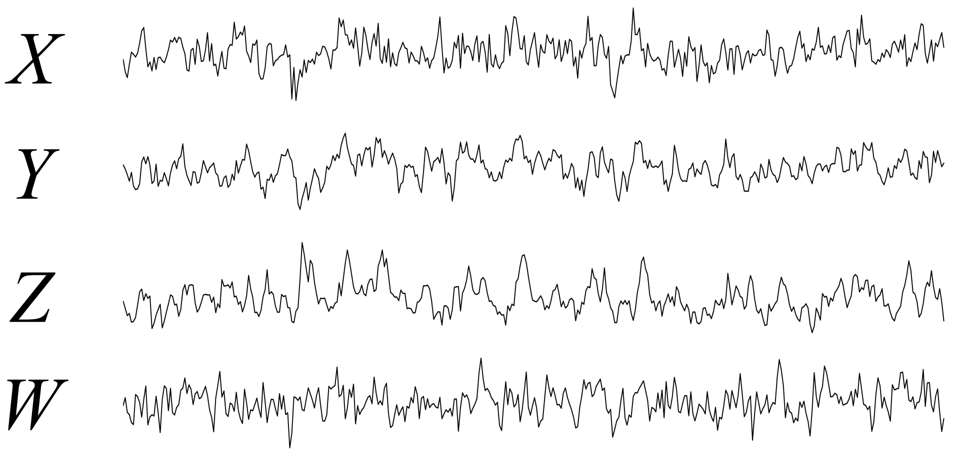
Analysis summary

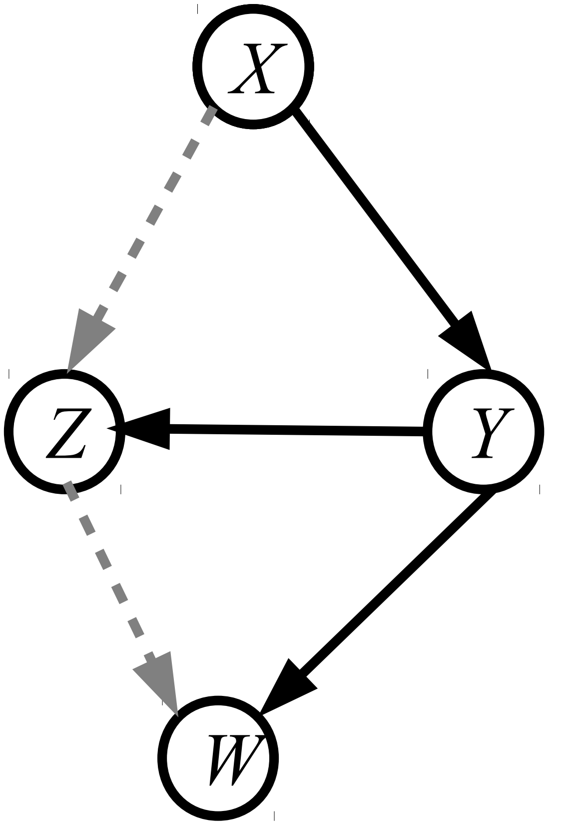
- TE(X → Y) depends on external or internal driving of X and even on processes driven by X
- MI depends on external or internal driving of X and other drivers of Y

Are these well-interpretable/precise measures of the coupling strength between X and Y?

Time series graphs
+
momentary information transfer

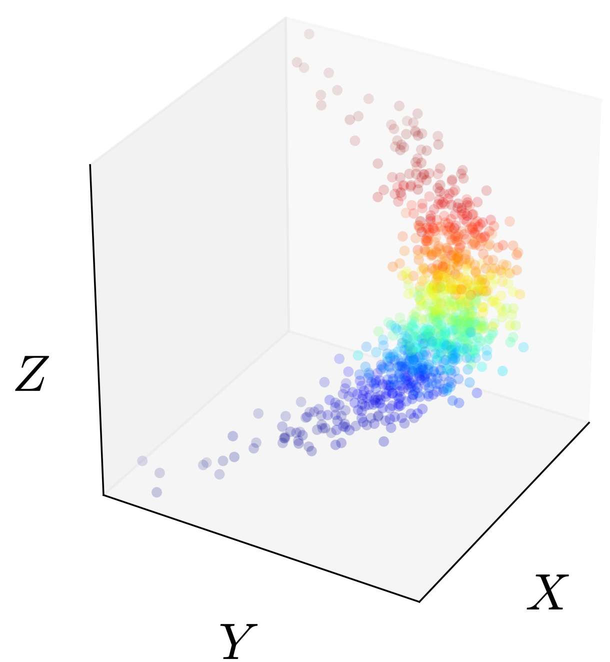
Conditional independence



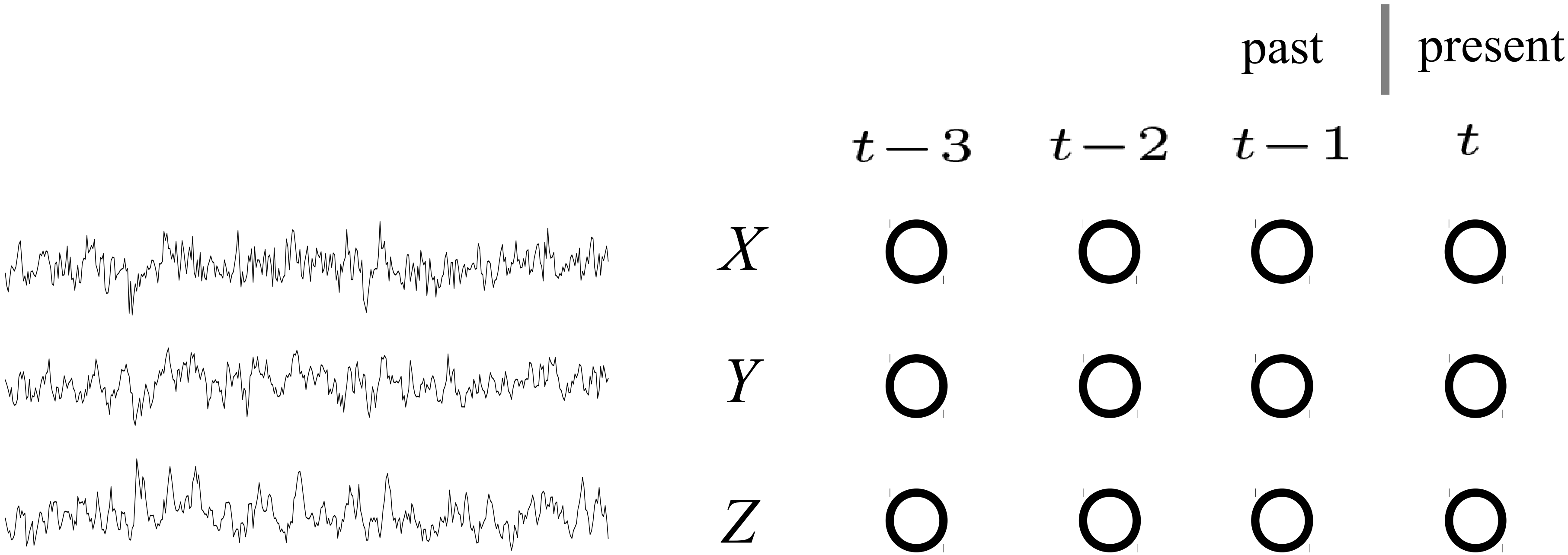


Conditional independence of X and Y given Z:

$$X \perp\!\!\!\perp Y | Z \iff I(X;Y|Z) = 0$$



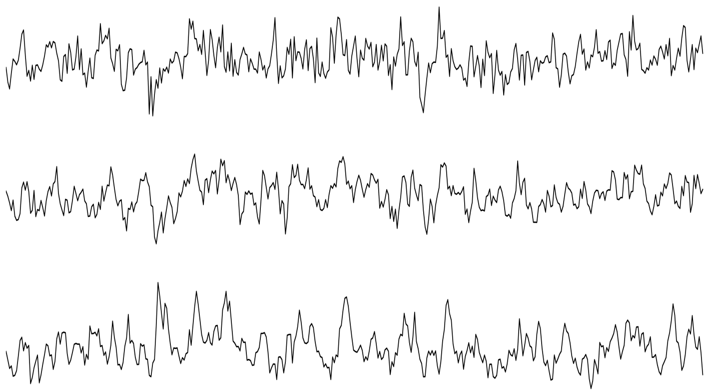
Conditional independence for time series

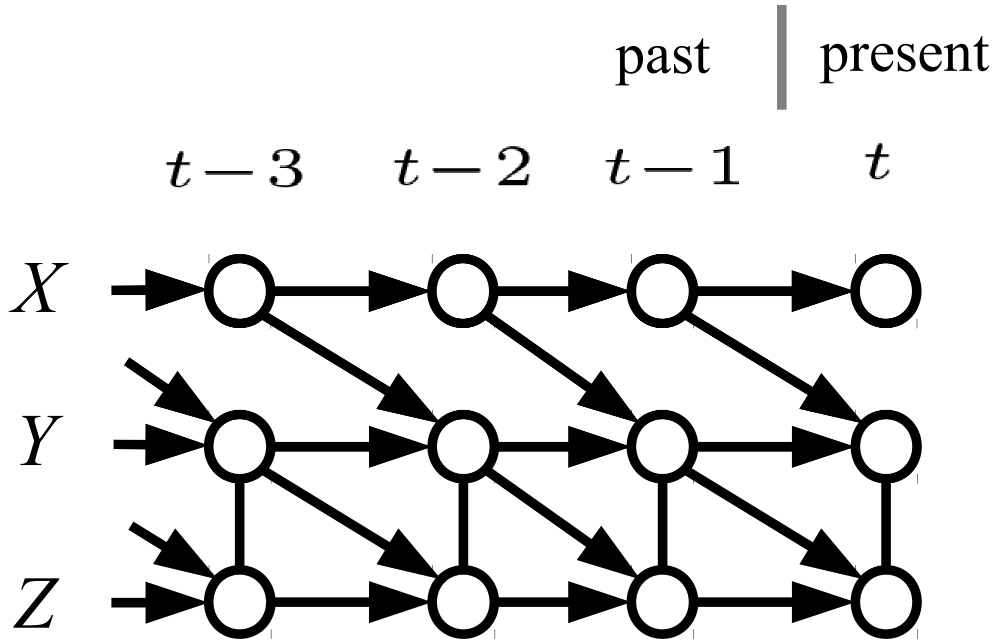


Time series graphs/graphical models

S. L. Lauritzen, Graphical Models, Oxford, 1996
R. Dahlhaus, Metrika 51, 157 (2000)
M. Eichler, Probability Theory and Related Fields 1 (2012)

Conditional independence for time series

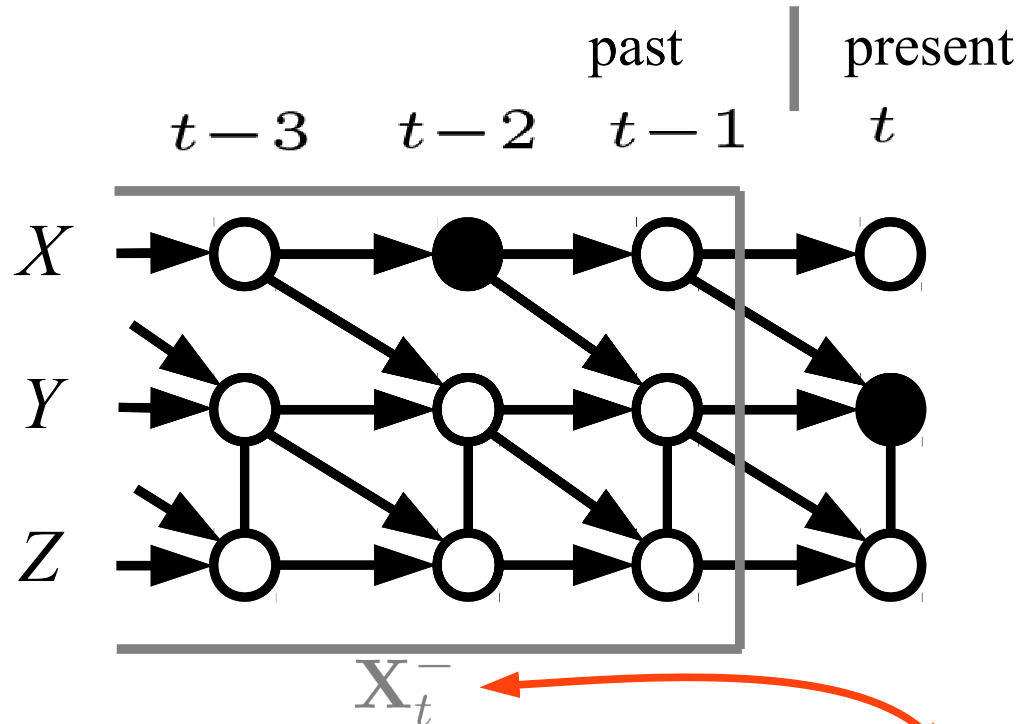




Time series graphs/graphical models

S. L. Lauritzen, Graphical Models, Oxford, 1996
R. Dahlhaus, Metrika 51, 157 (2000)
M. Eichler, Probability Theory and Related Fields 1 (2012)

Time series graphs



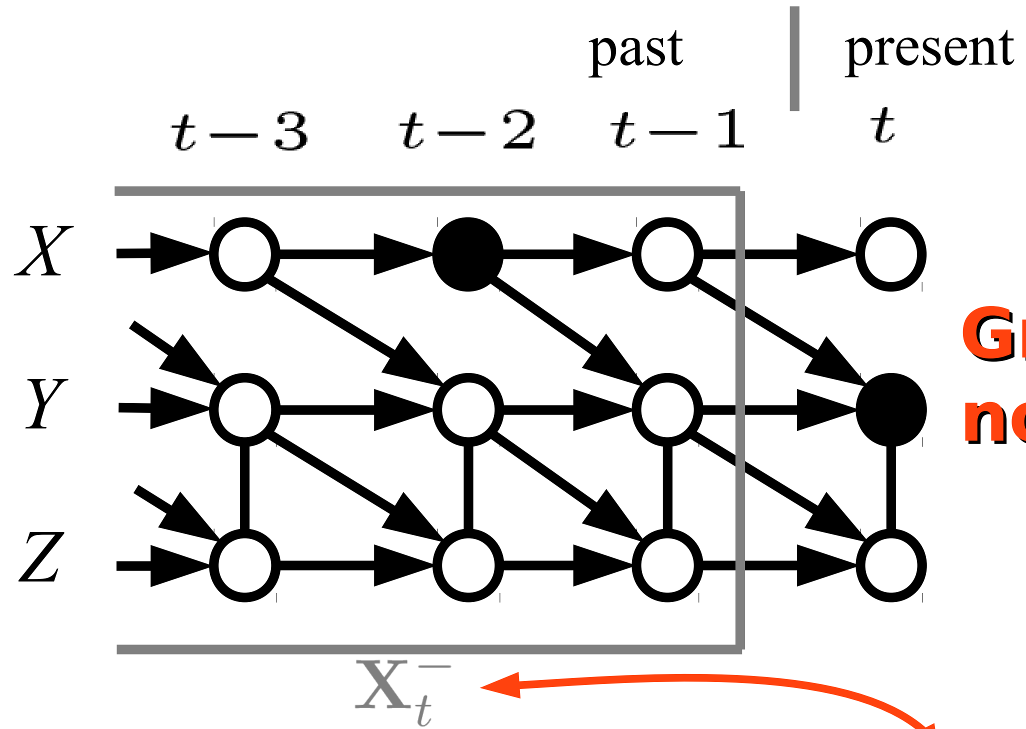
Directed link: $X_{t-\tau} \not\perp\!\!\!\perp Y_t \mid \mathbf{X}_t^- \setminus \{X_{t-\tau}\}$

Contemp. link: $X_t \not\perp\!\!\!\perp Y_t \mid \mathbf{X}_{t+1}^- \setminus \{X_t, Y_t\}$

Not so important for now

M. Eichler, Probability Theory and Related Fields 1 (2012)

Time series graphs



Granger-non-causality

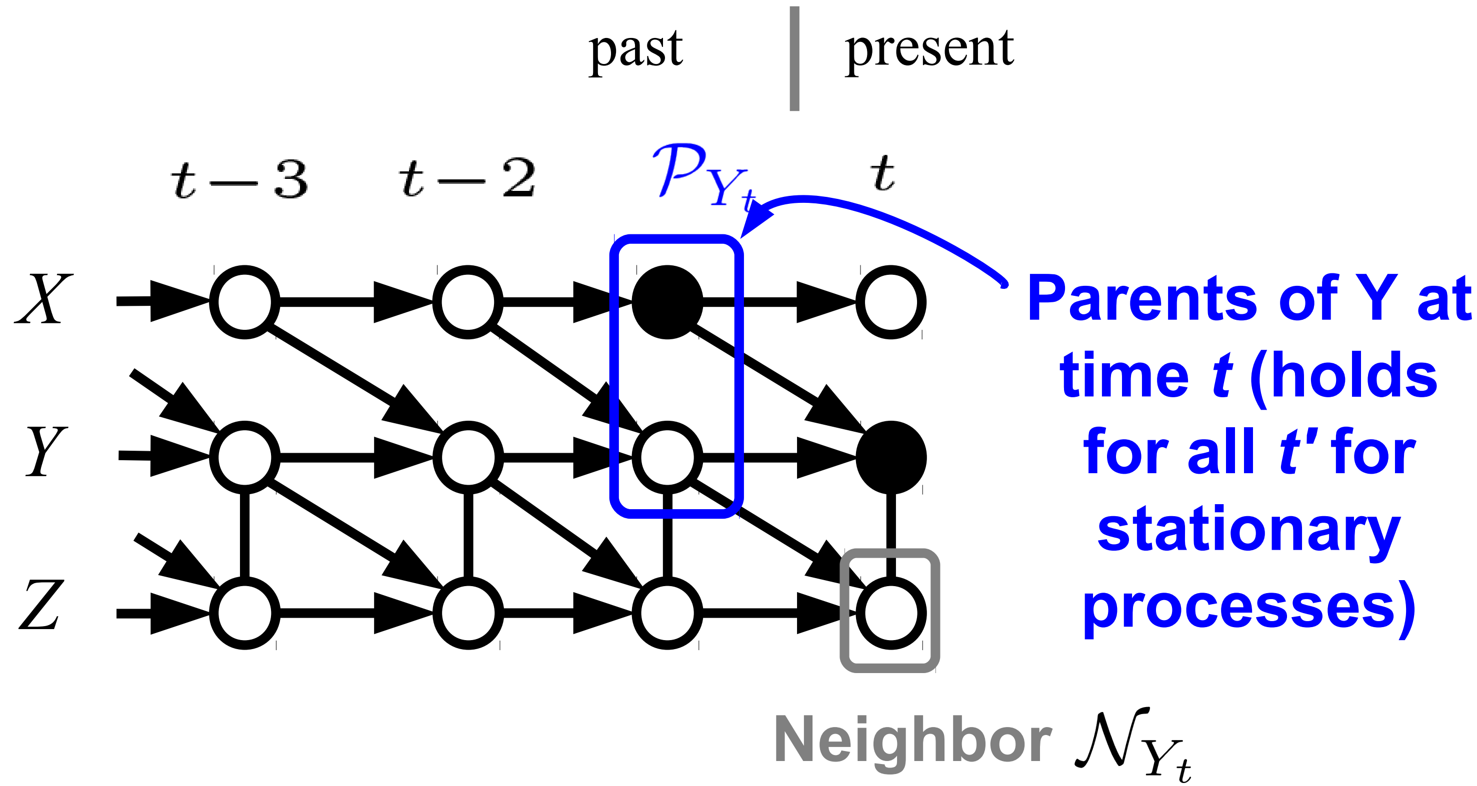
Directed link: $X_{t-\tau} \not\perp\!\!\!\perp Y_t \mid \mathbf{X}_t^- \setminus \{X_{t-\tau}\}$

Contemp. link: $X_t \not\perp\!\!\!\perp Y_t \mid \mathbf{X}_{t+1}^- \setminus \{X_t, Y_t\}$

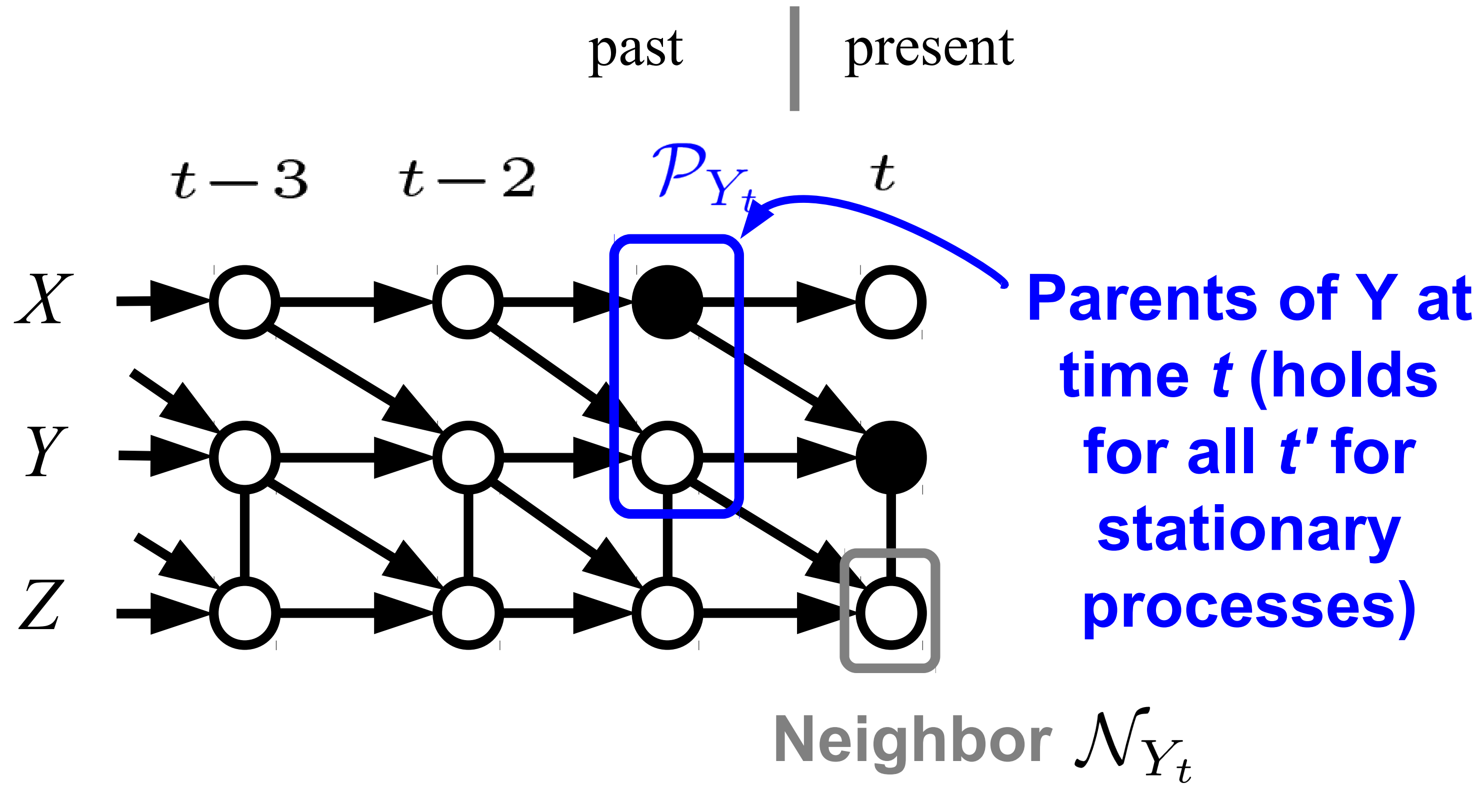
Not so important for now

M. Eichler, Probability Theory and Related Fields 1 (2012)

Markov property



Markov property



Markov Property:

Separation in graph ==> independence

$$\mathbf{X}_t^- \setminus \mathcal{P}_{Y_t} \perp\!\!\!\perp Y_t \mid \mathcal{P}_{Y_t}$$

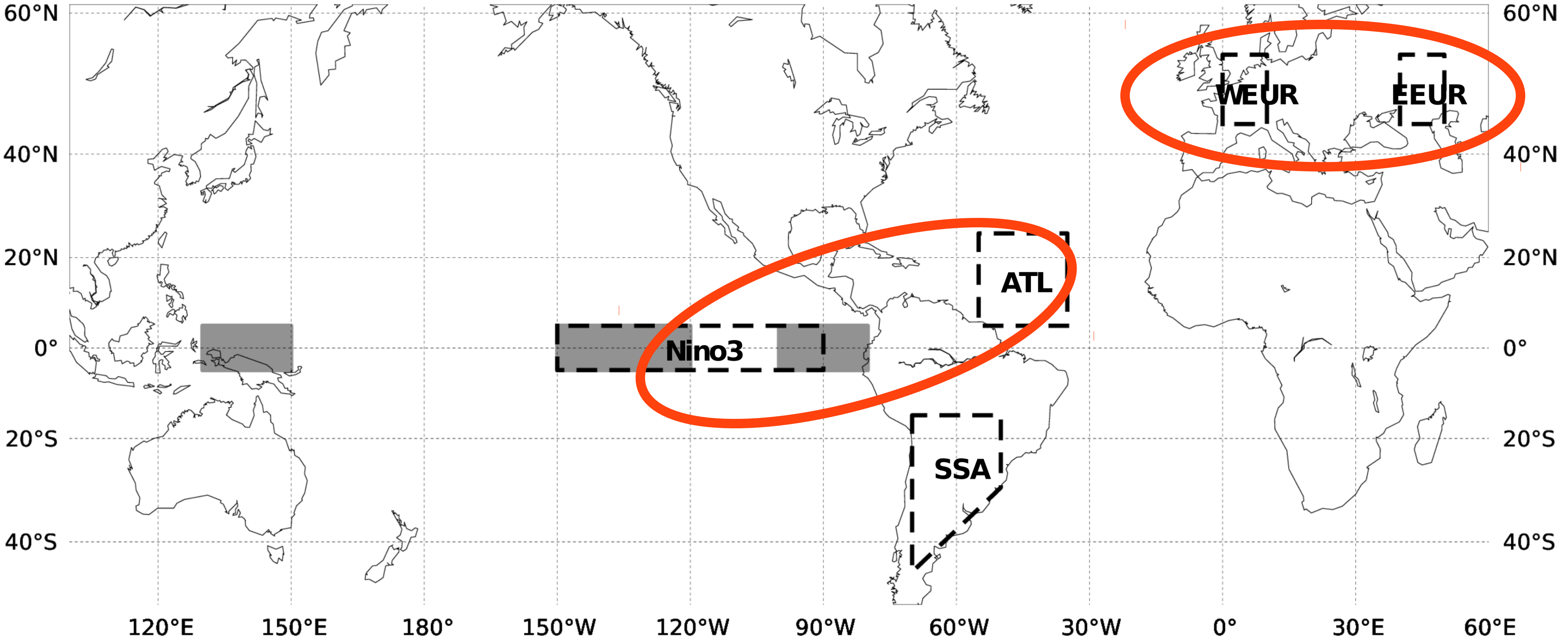
Spirtes (2000), Pearl (2000), Eichler (2012)

Estimation? Iterative PC-algorithm / Jie Sun's algorithm

P. Spirtes, C. Glymour, and R. Scheines, Causation, Prediction, and Search (MIT, Cambridge, MA, 2000).

J. Runge et al., Phys. Rev. Lett. 108, 258701 (2012)

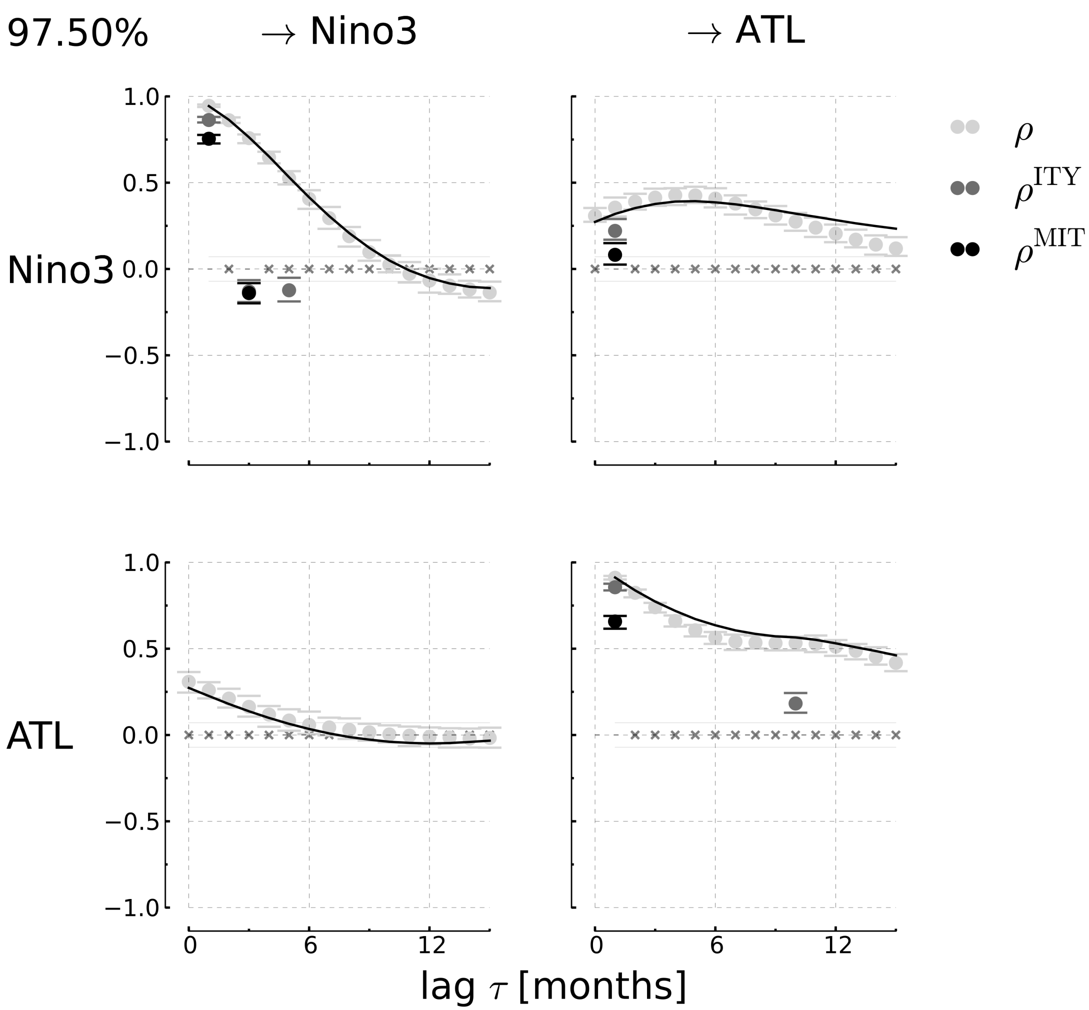
Re-examined: ENSO teleconnections



Reanalysis data: Monthly surface air temperature and pressure
Kalnay et al., 1996: The NCEP/NCAR 40-Year Reanalysis Project. Bulletin of the American Meteorological Society, 77(3), 437–471.

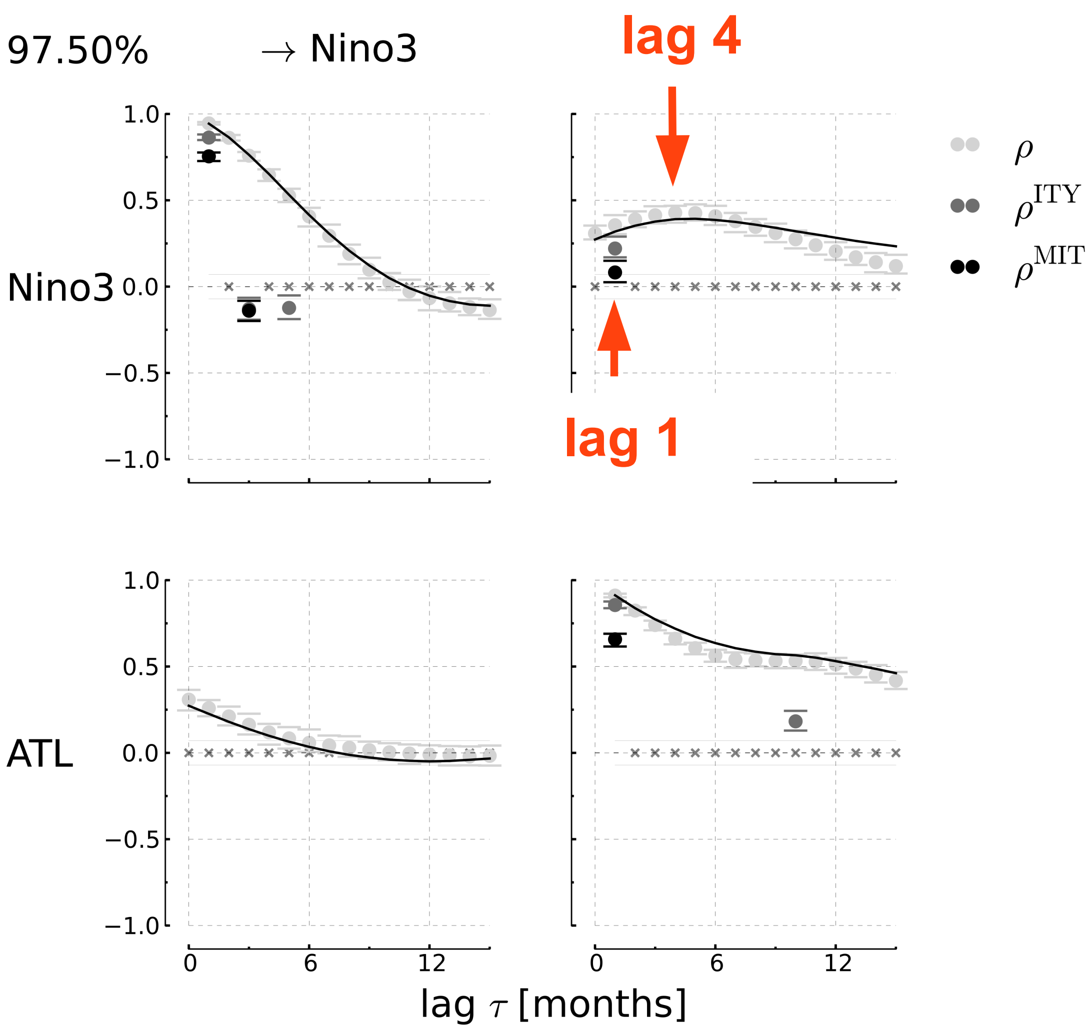
Pacific - Atlantic teleconnection

Correlation / ITY / MIT



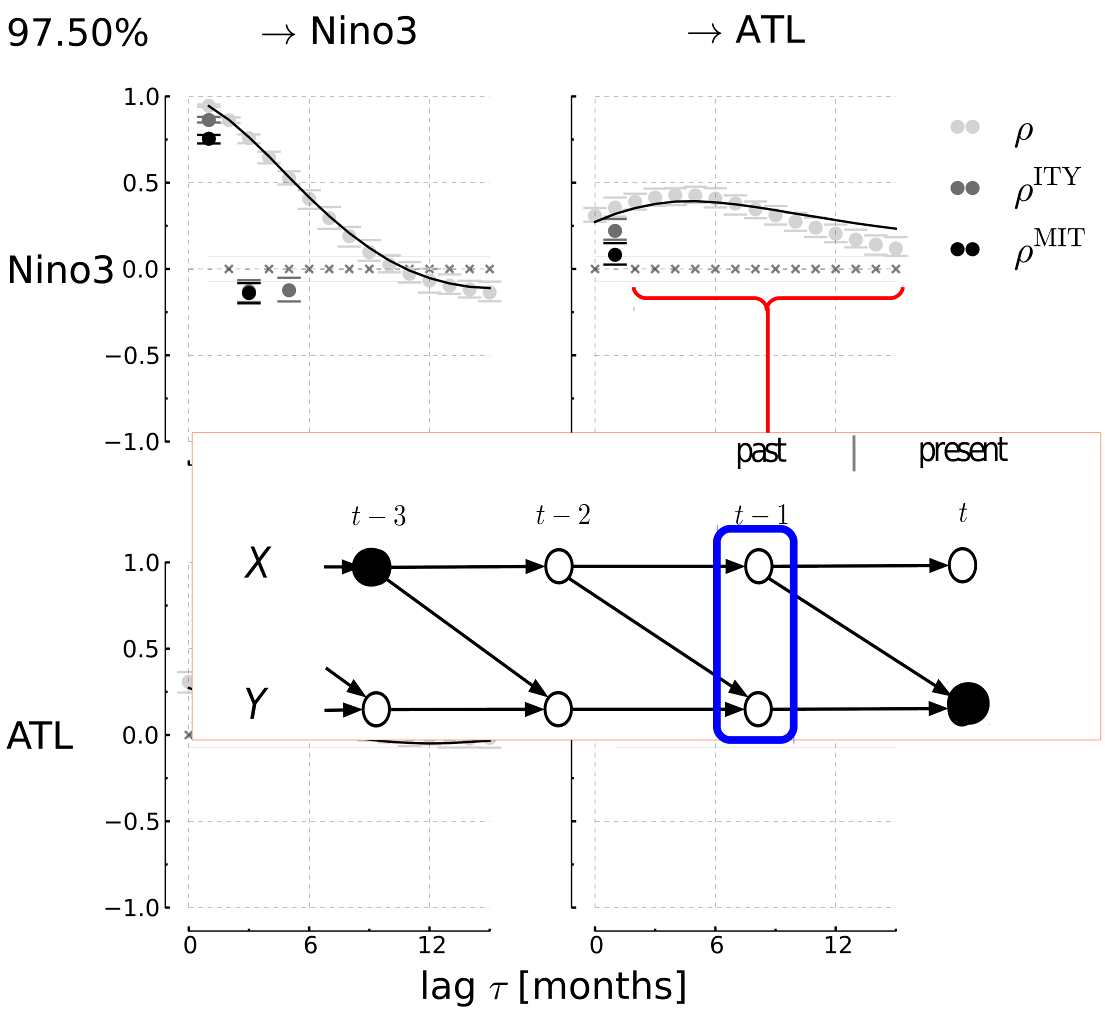
Pacific - Atlantic teleconnection

Correlation / ITY / MIT

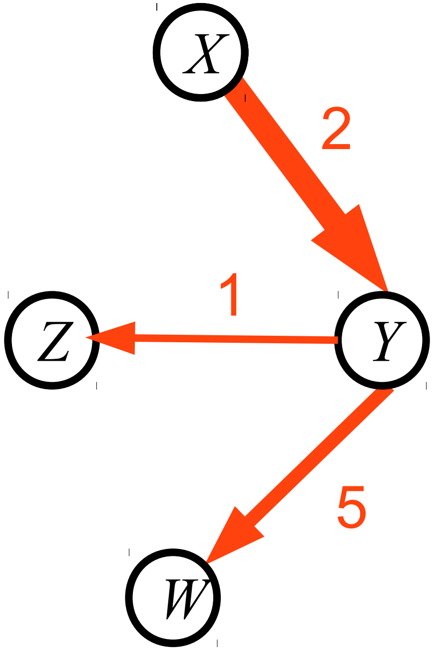


Pacific - Atlantic teleconnection

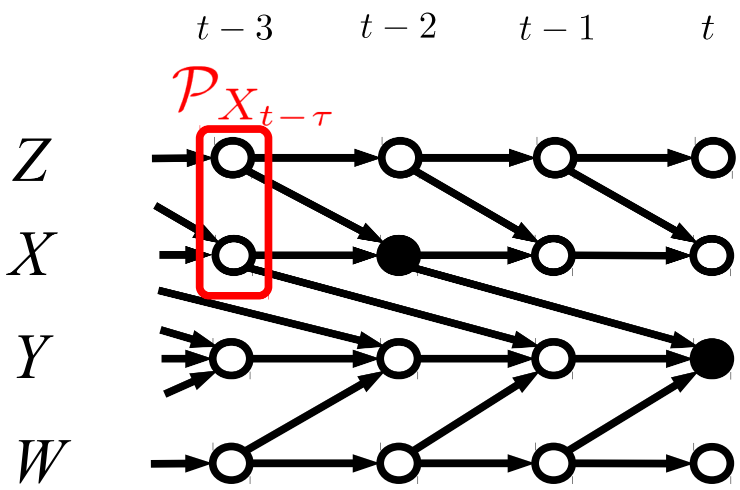
Correlation / ITY / MIT



What is a well interpretable coupling strength?



Ansatz for a well interpretable measure of coupling strength



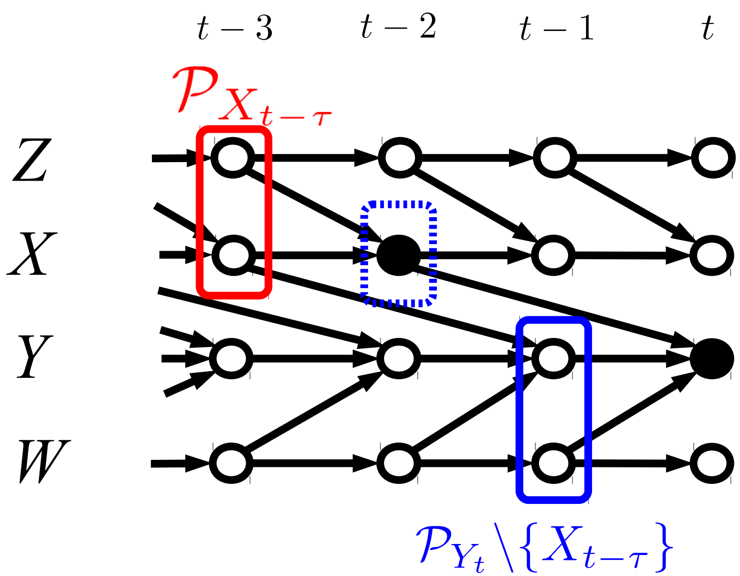
Source entropy of X:

$$H(X_{t-\tau}|\mathcal{P}_{X_{t-\tau}})$$

→ dynamical noise in a stochastic system
(→ uncertainty in a chaotic deterministic system)
→ input from unobserved variables

B. Pompe and J. Runge, Phys. Rev. E 83, 051122 (2011)

Ansatz for a well interpretable measure of coupling strength



Source entropy of X:

$$H(X_{t-\tau}|\mathcal{P}_{X_{t-\tau}})$$

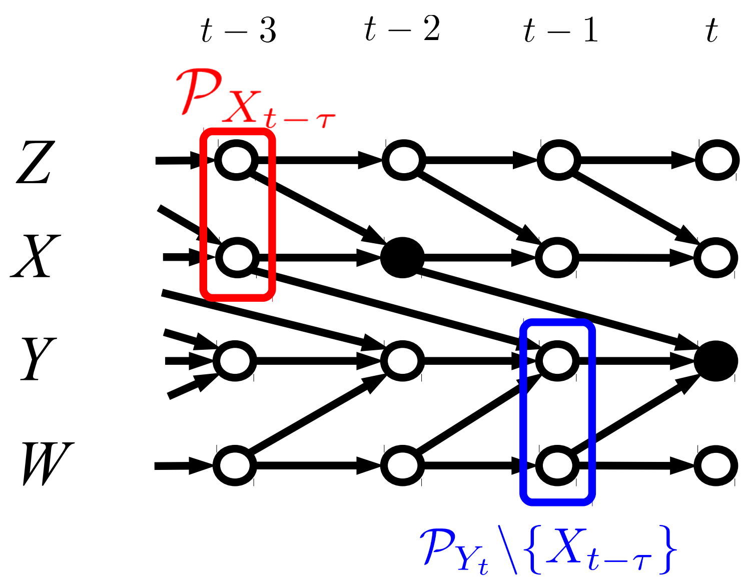
Source entropy of Y:

$$H(Y_t|\mathcal{P}_{Y_t})$$

→ dynamical noise in a stochastic system
(→ uncertainty in a chaotic deterministic system)
→ input from unobserved variables

B. Pompe and J. Runge, Phys. Rev. E 83, 051122 (2011)

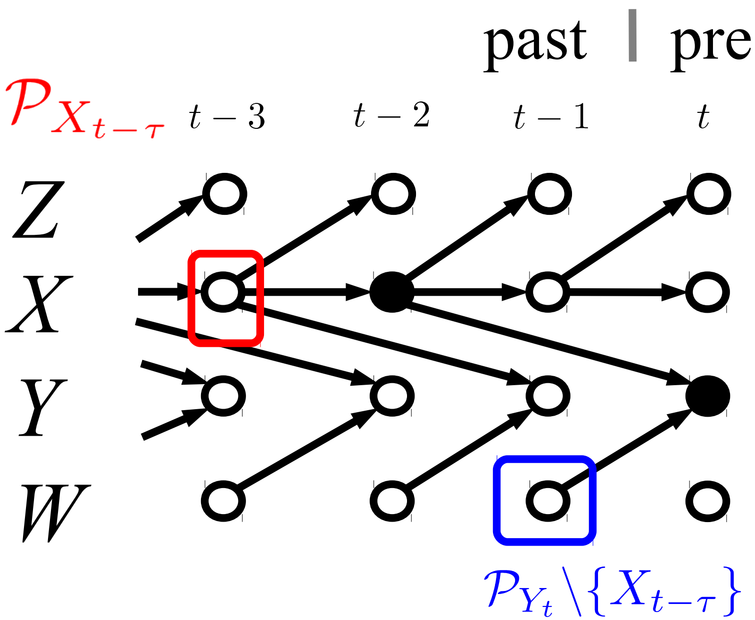
Momentary Information Transfer (MIT)



$$
\begin{aligned}
I^{\mathrm{MIT}}_{X\to Y}(\tau) &\equiv I(X_{t-\tau}; Y_t | \mathcal{P}_{Y_t} \setminus \{X_{t-\tau}\}, \mathcal{P}_{X_{t-\tau}}) \\
&= H(Y_t | \mathcal{P}_{Y_t} \setminus \{X_{t-\tau}\}, \mathcal{P}_{X_{t-\tau}}) - H(Y_t | \mathcal{P}_{Y_t})
\end{aligned}
$$

J. Runge, J. Heitzig, M. Marwan, and J. Kurths,
Quantifying causal coupling strength: ...
Phys. Rev. E 86, 061121 (2012)

What does MIT measure?



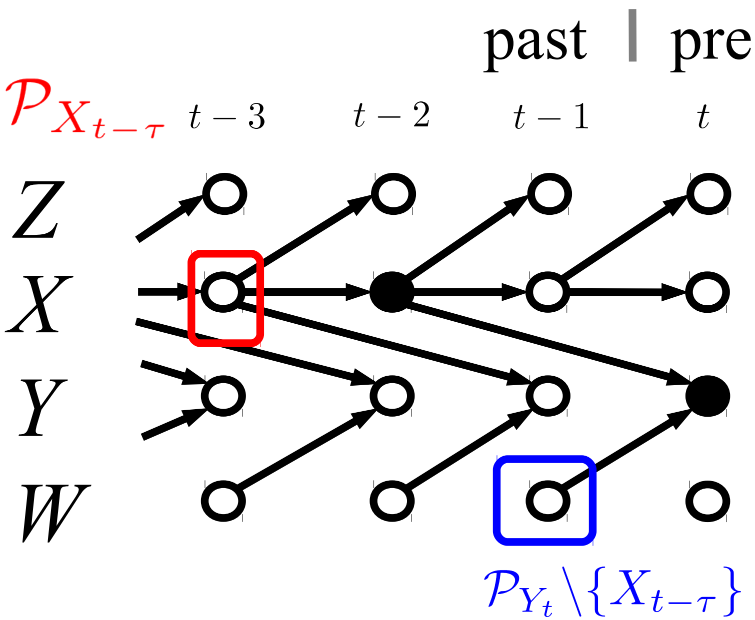
$$Z_t = c_{XZ} X_{t-1} + \eta_t^Z$$

$$X_t = a_X X_{t-1} + \eta_t^X$$

$$Y_t = c_{XY} X_{t-2} + c_{WY} W_{t-1} + \eta_t^Y$$

$$W_t = \eta_t^W$$

What does MIT measure?



$$Z_t = c_{XZ} X_{t-1} + \eta_t^Z$$

$$X_t = a_X X_{t-1} + \eta_t^X$$

$$Y_t = \mathbf{c}_{XY} X_{t-2} + c_{WY} W_{t-1} + \eta_t^Y$$

$$W_t = \eta_t^W$$

$$I_{X \to Y}^{\mathrm{MIT}} = \frac{1}{2} \ln \left(1 + \frac{c_{XY}^2 \sigma_X^2}{\sigma_Y^2} \right)$$

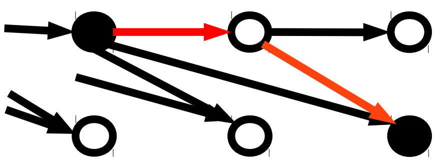
Coupling Strength Autonomy Theorem

Additive Models:

$$X_t = g_X(\mathcal{P}_{X_t}) + \eta_t^X$$

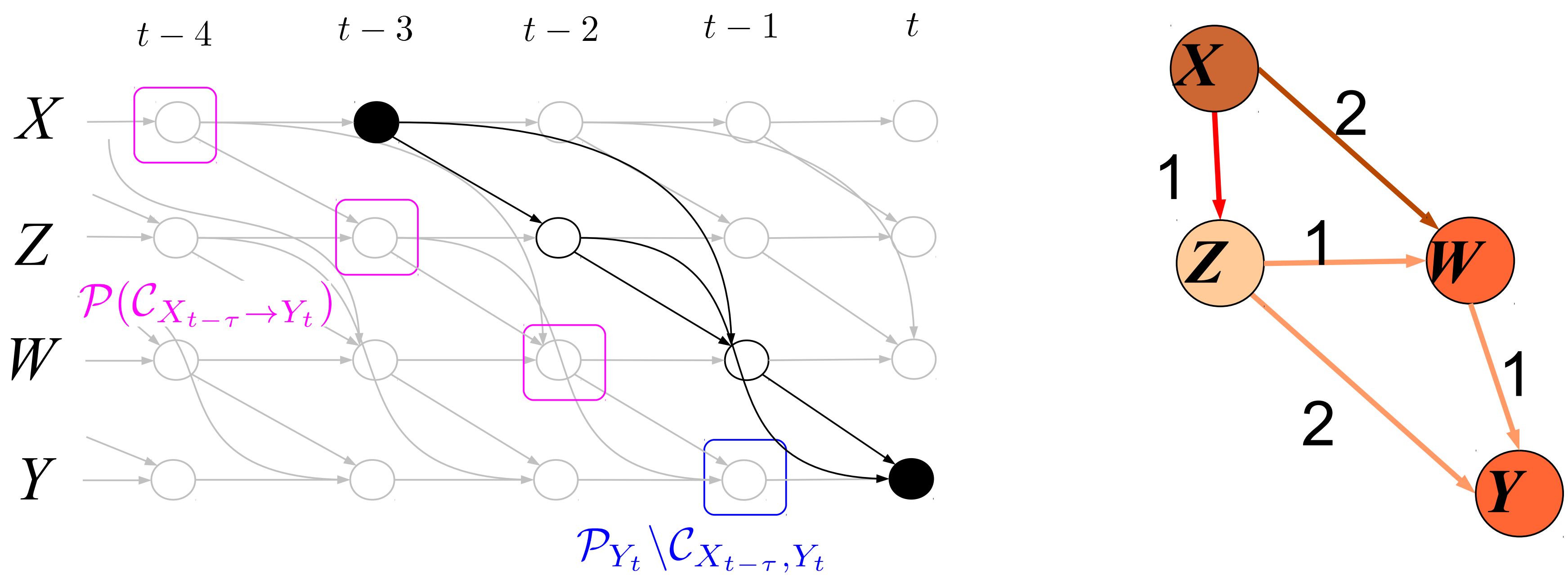
$$Y_t = cX_{t-\tau} + g_Y(\mathcal{P}_{Y_t} \setminus \{X_{t-\tau}\}) + \eta_t^Y$$

Under "no sidepath"-constraint:



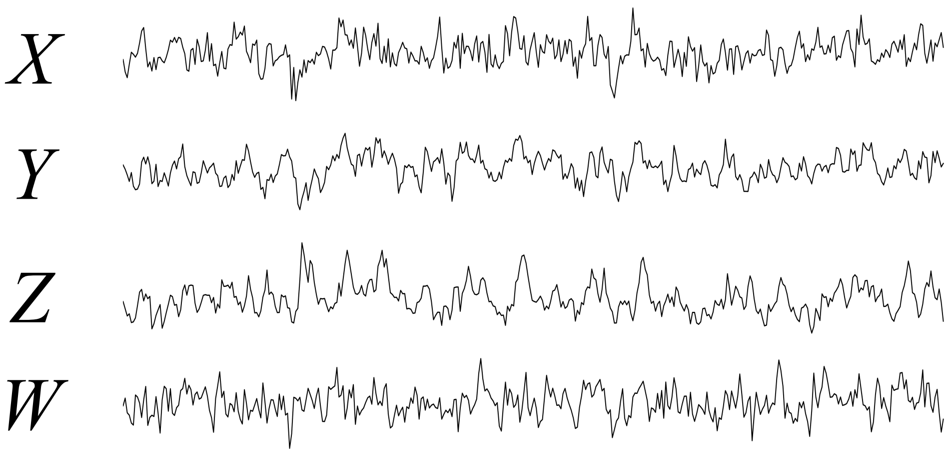
$$I^{\text{MIT}}_{X \to Y}(\tau) = I(\eta^X_{t-\tau}; c\eta^X_{t-\tau} + \eta^Y_t)$$

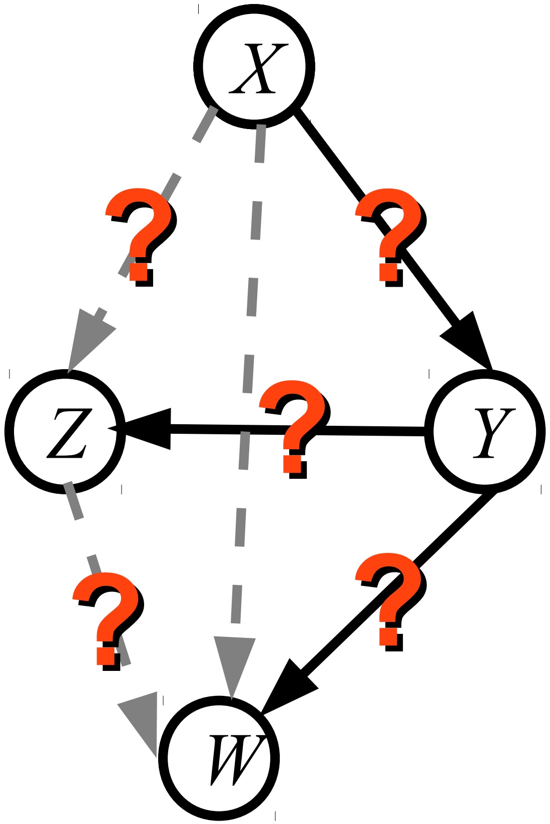
Path-based measures



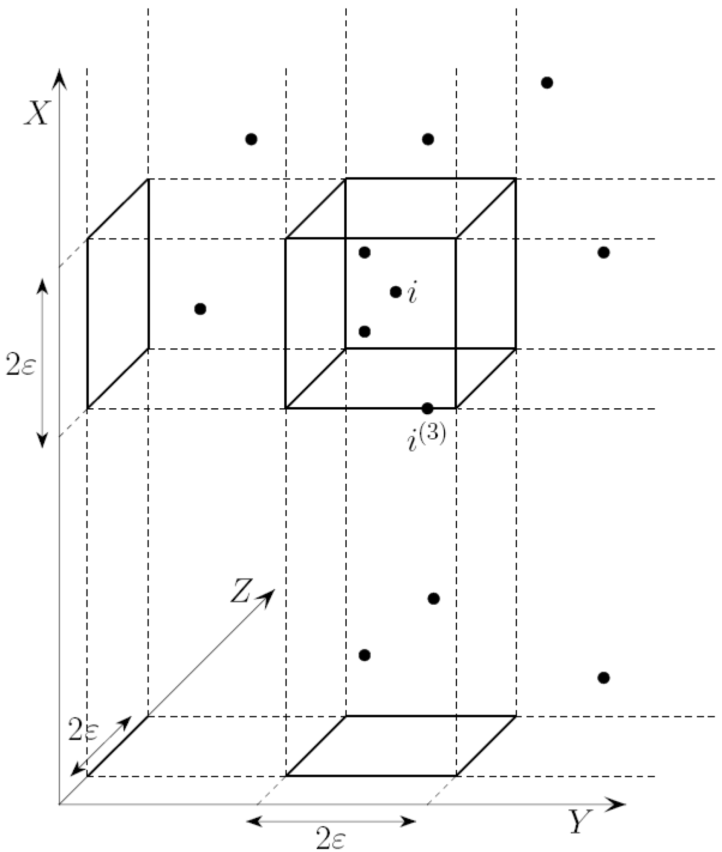
a) Time series graph **b)** Process graph

Significance testing under strong autocorrelations





Estimation of CMI via k-nearest-neighbor estimator



$$\widehat{I}_{XY|Z} = \psi(k) + \frac{1}{T}\sum_{t=1}^{T}\left[\psi(k_{Z,t}) - \psi(k_{XZ,t}) - \psi(k_{YZ,t})\right]$$

Frenzel & Pompe, Phys. Rev. Lett., 99(20), 204101. (2007)

Kraskov et al., Phys. Rev. E 69, 066138 (2004)

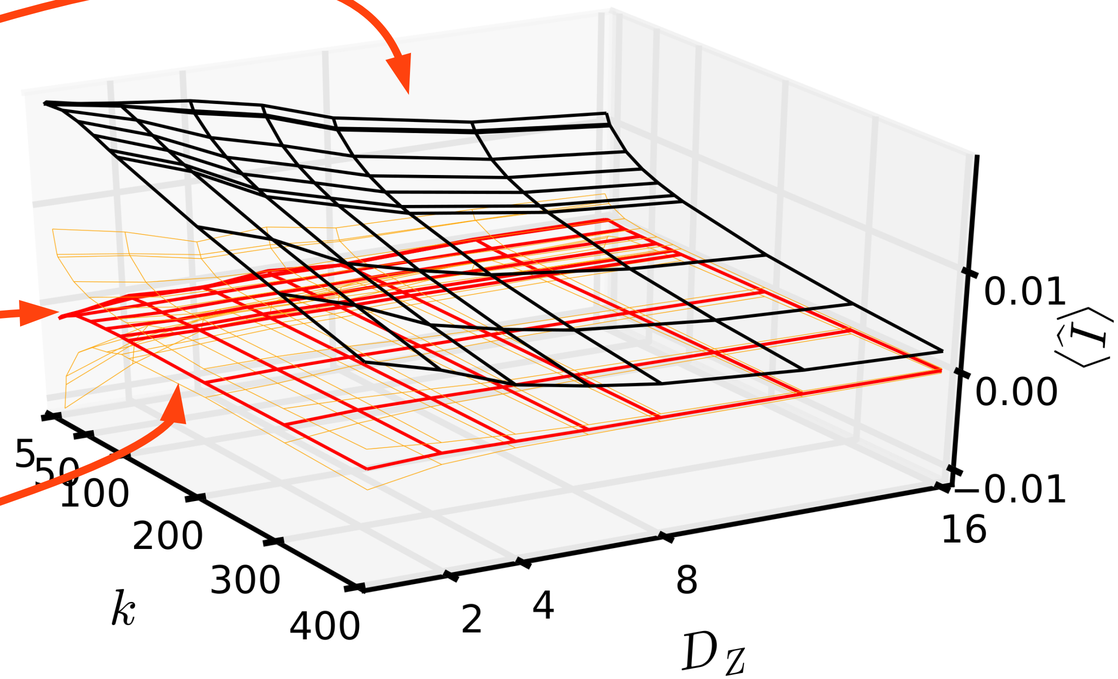
parameter *k* ~ bandwidth in KDE

(here k in joint space defines epsilon in all dimensions)

Much better than binning, still: Bias for short samples and large dimension

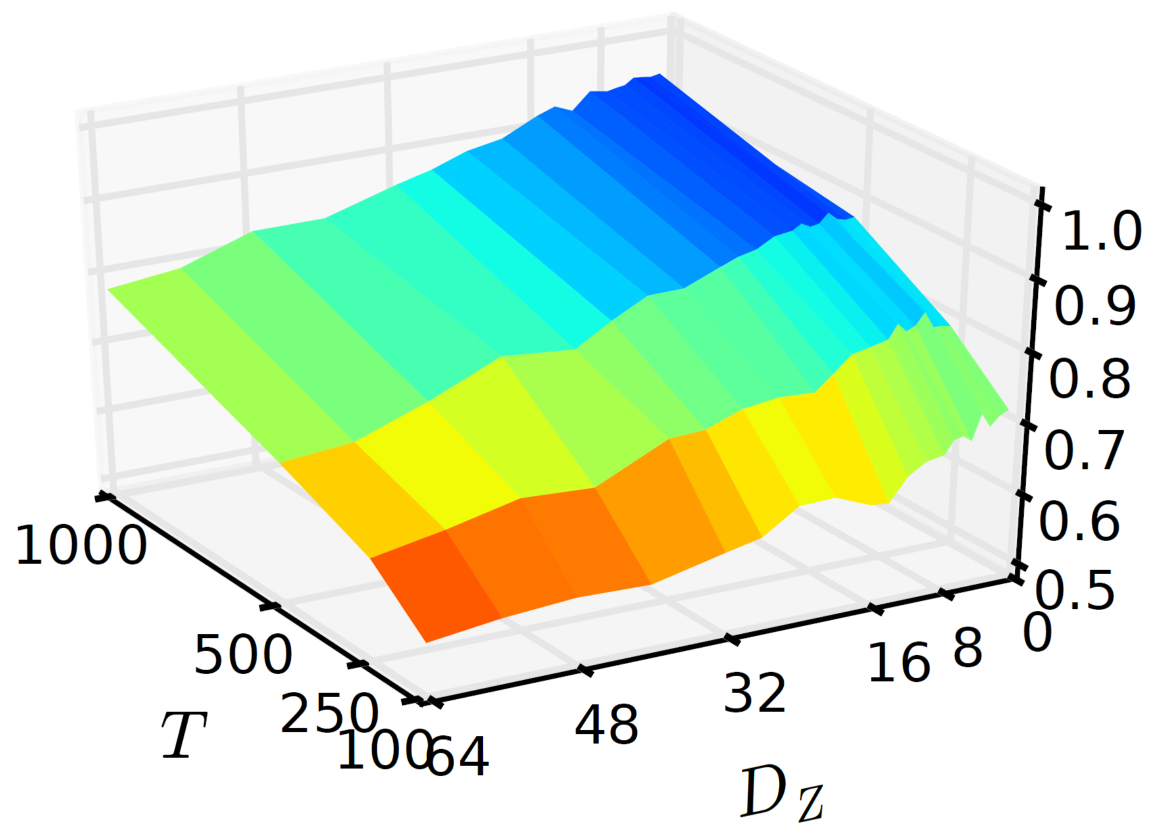
biased for large Dz and dependent X and Y

But: unbiased for independent X and Y



optimal *k* for best statistical power as conditional independence test

Power as independence test: AUC for multivariate Gaussian



Significance testing

- Need to know sample distribution of estimator for independent processes
- **Partial correlation**: analytical distribution known for Multivariate Gaussian (Student's *t),*
  But: assuming i.i.d. samples
- **Conditional mutual information (kNN)***:* Nothing known
  → shuffle test...

Significance testing
Partial correlation

What happens for autocorrelated time series?

$$X_t = aX_{t-1} + \eta_t^X$$

$$Y_t = aY_{t-1} + cX_{t-1} + \eta_t^Y$$

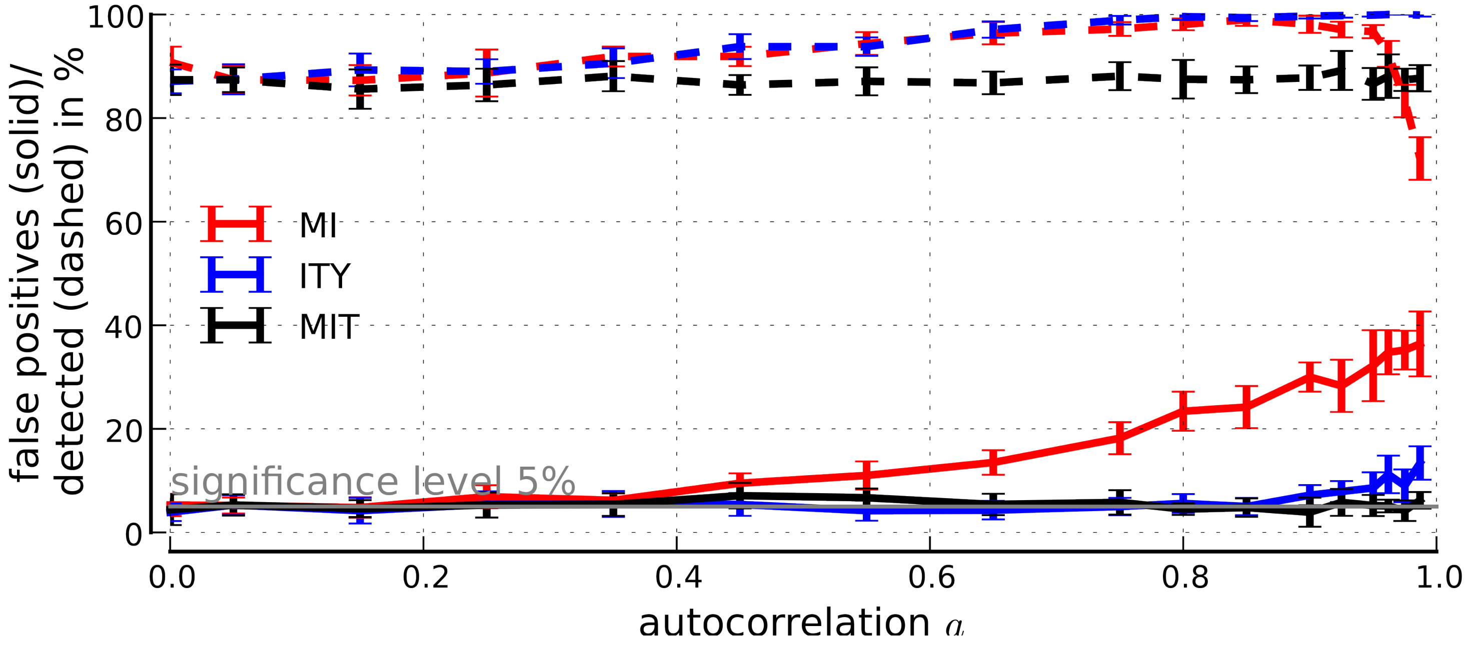
Significance testing
Partial correlation

What happens for autocorrelated time series?

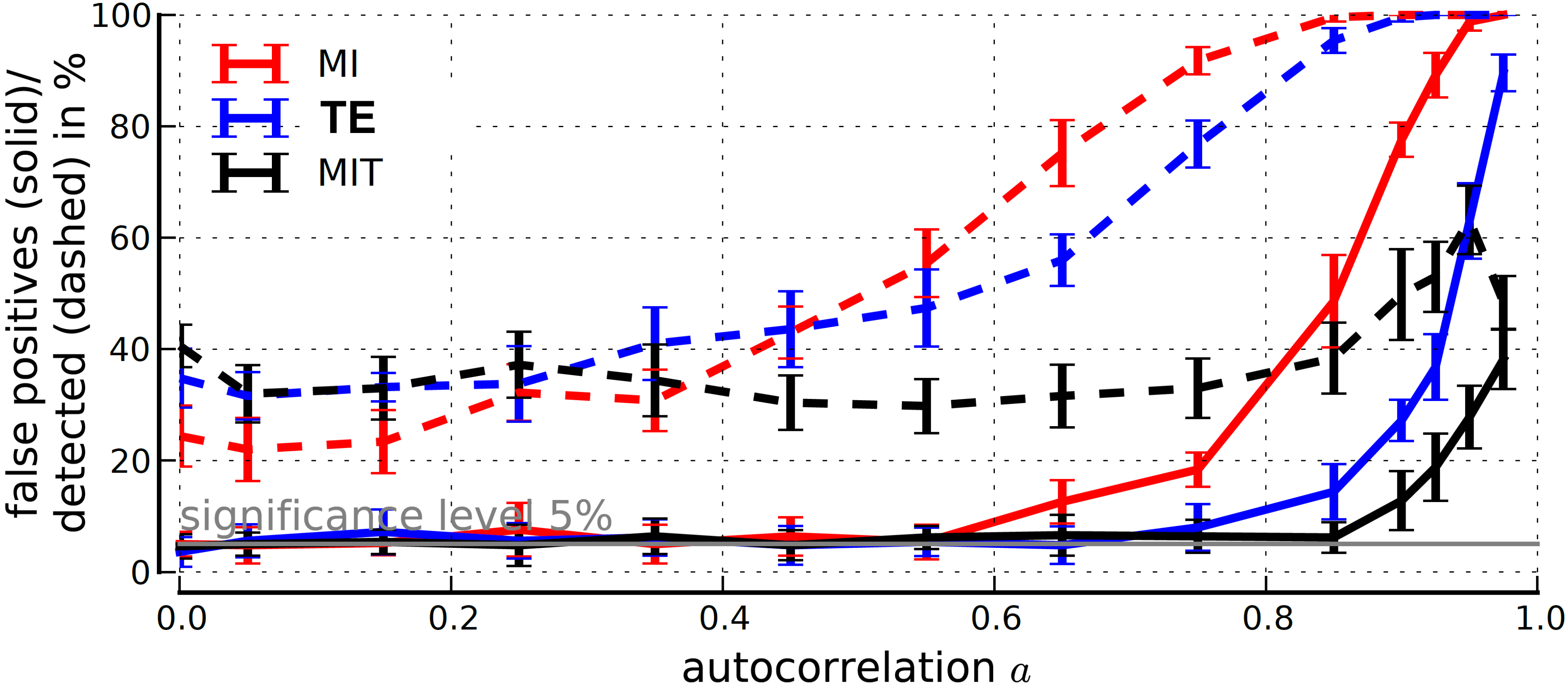
$$X_t = aX_{t-1} + \eta_t^X$$

$$Y_t = aY_{t-1} + cX_{t-1} + \eta_t^Y$$

b)



Significance testing
Conditional mutual information



Conclusions

Unconditional (Correlation, Mutual Information) lag functions or Transfer Entropy

- … are not suitable to infer coupling delays (not goal of TE)
- … are counterintuitive/ambiguous as measures of strength of mechanism
- … have large false positive rate in significance tests under high autocorrelations

Time series graph + Momentary information transfer

- ➔ … yield precise coupling delays
- ➔ … provide at least a more precisely defined measure of causal strength (also partial correlation MIT)
- ➔ … reduce the effect of autocorrelation in significance testing

Challenges: Eichler's list ...plus:

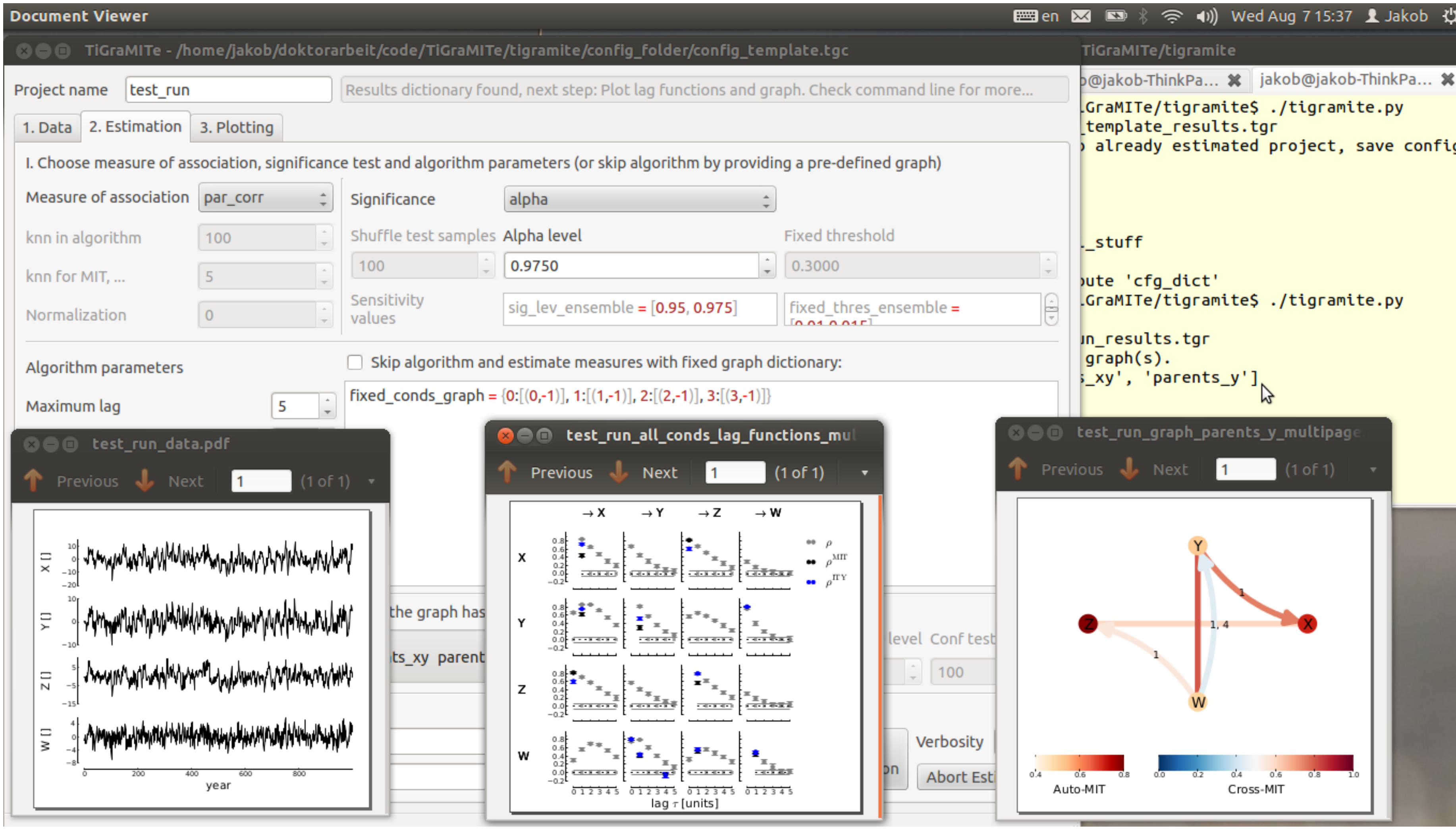
- PC – Algorithm: Iterative testing → multiple testing problem → significance/posterior prob. of links difficult to estimate...**but: only way without model!**
- Faithfulness assumption
- CMI: shuffle tests computationally expensive
- Estimation of CMIs bias for higher dimensions → difficult to compare causal strength! **→ desperate search for information-theoretic characterization of causal strength**
- ...

Need to improve CMI estimators
→ smartly include assumptions

TiGraMITe

Python script for **Ti**me series **Gra**ph and **M**omentary **I**nformation **T**ransfer **e**stimation

www.pik-potsdam.de/members/jakrunge



References

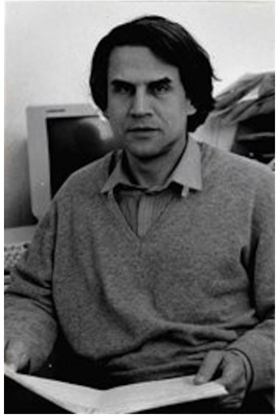
J. **Runge,** V. Petoukhov, and J. Kurths,
Quantifying the strength and delay of climatic interactions: ...
Journal of Climate 27(2), 720-739 (2014)

J. Runge, J. Heitzig, M. Marwan, and J. Kurths,
Quantifying causal coupling strength: ...
Phys. Rev. E 86, 061121 (2012)

J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths,
Escaping the curse of dimensionality in estimating multivariate transfer entropy
Phys. Rev. Lett. 108, 258701 (2012)

B. Pompe, **J. Runge**,
Momentary information transfer as a coupling measure of time series
Phys. Rev. E, 83, 051122 (2011)

V. Petoukhov



J. Heitzig

B. Pompe

J. Kurths

www.pik-potsdam.de/members/jakrunge

Sponsors: DFG

PIK
Potsdam Institute for Climate Impact Research

Humboldt-Universität zu Berlin

Studienstiftung des deutschen Volkes