erc
European Research Council
Established by the European Commission

DLR

DKRZ
DEUTSCHES
KLIMARECHENZENTRUM

berlin

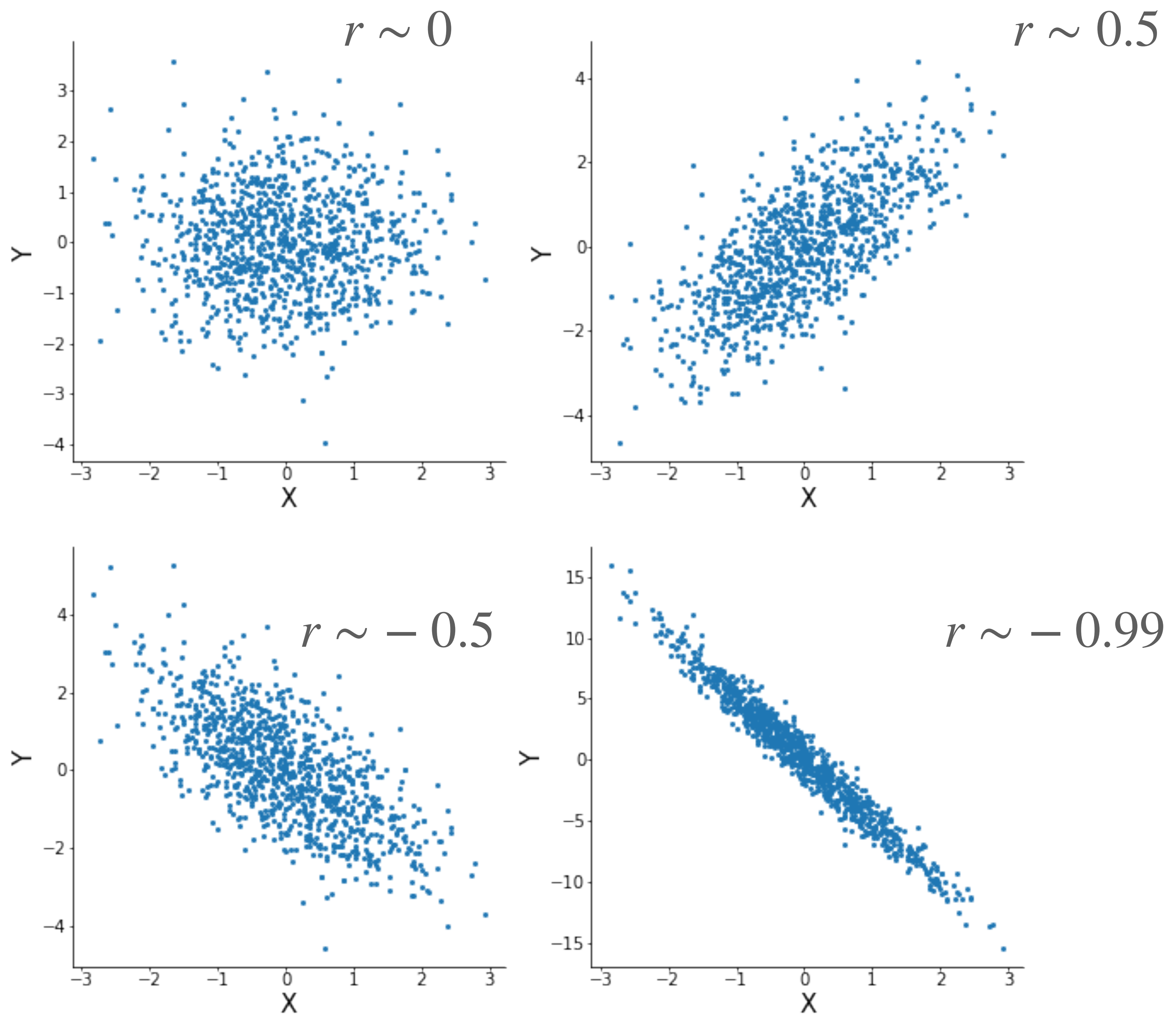
Causal Inference for Time Series and Applications to Climate

Urmi Ninad | TU Berlin | Berlin Applied Causal Graphs Workshop (23 April 2024)

Correlation in Statistical Inference

"A Causality-free Science"

- Pearson correlation coefficient r measures the correlation between two random variables X and Y (right)
- Karl Pearson said "Science in no case can demonstrate any inherent necessity in a sequence, nor prove with absolute certainty that it must be repeated"[1]
- ".. the idea of causation is extracted by conceptual processes from phenomena, it is neither a logical necessity, nor an actual experience. We can merely classify things as like; we cannot reproduce sameness, but *we can only measure how relatively like follows relatively like*. The wider view of the universe sees all phenomena as correlated, but not causally related."[1]



1: Pearson, "Grammar of Science", 1892

The Need for a Causal Framework

Formalising causal queries

- **Understanding**: *Why* do I observe what I observe?
  Eg.: Why can I, a human, reliably distinguish images of cats and dogs?

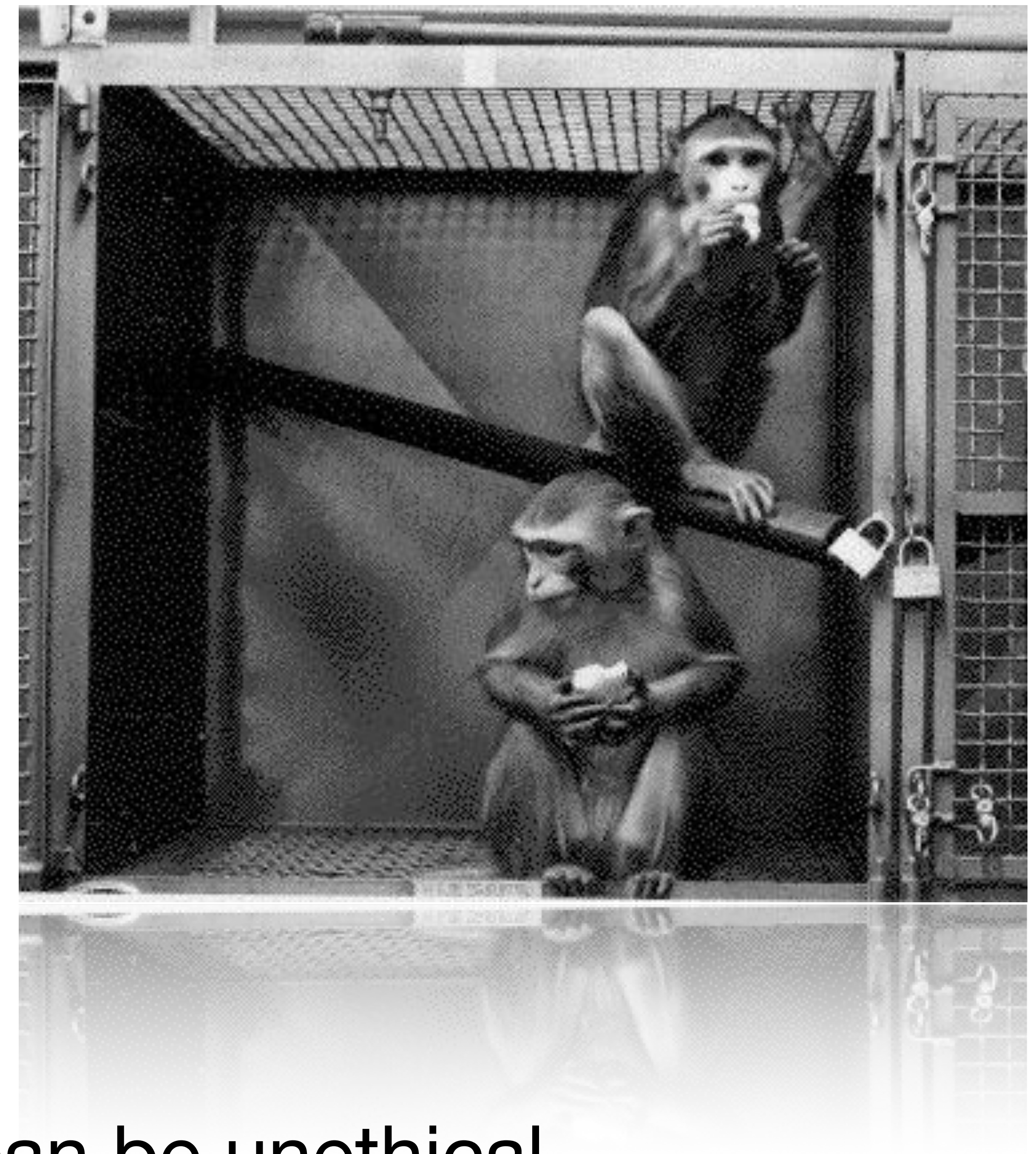
- **Attribution**: Did a certain event take place *due to* a certain action in the past? Would it have been different if a different action been chosen?
  Eg: Are extreme climatic events becoming more frequent because of anthropogenic contributions?

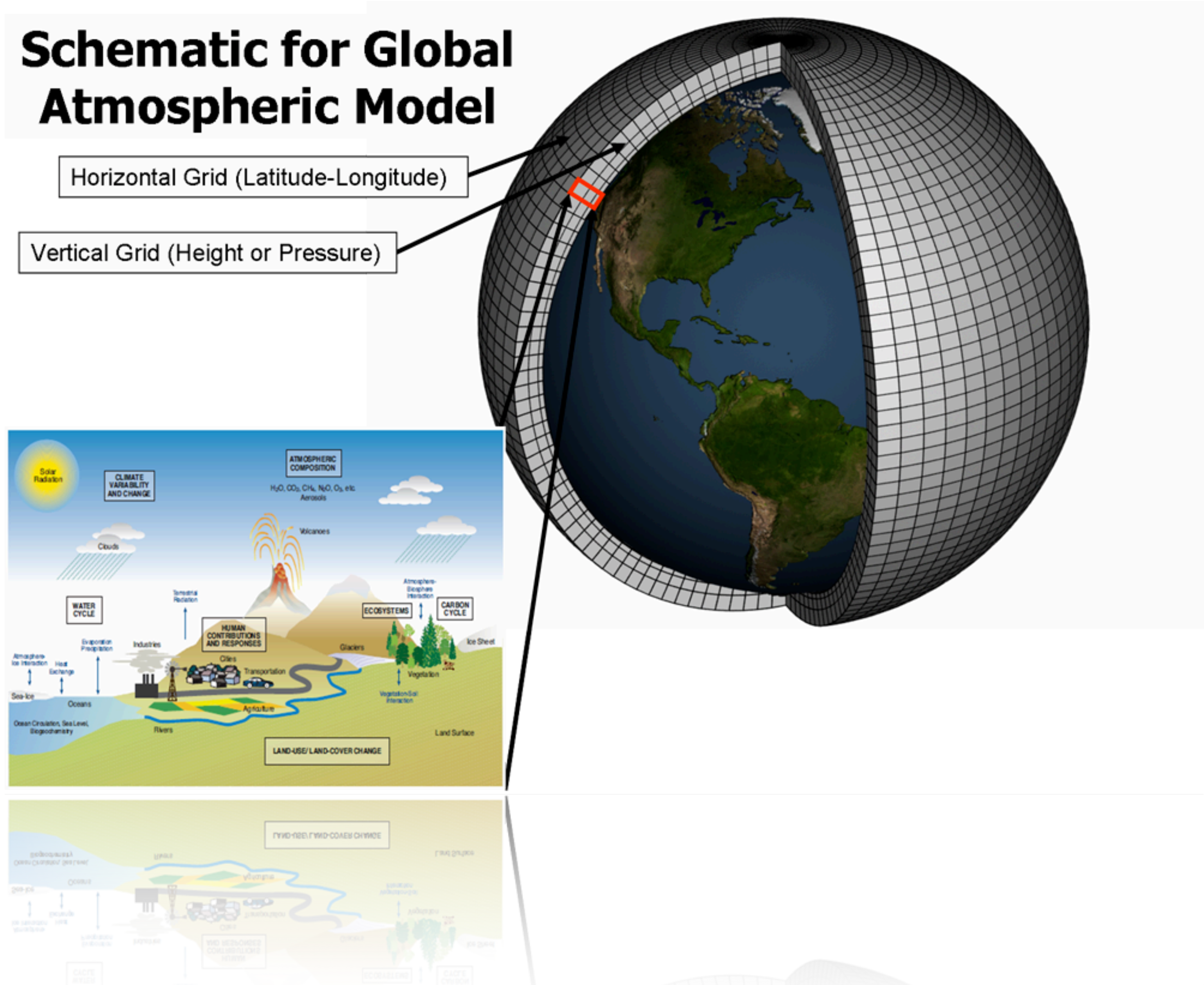
- **Decision making**: What should I do to achieve a *certain goal*?
  Eg: How can I enhance the cognitive function of a population?

- **Robust prediction and forecasting**: Given I observe X, *what* is Y?
  Predictive systems consistent with the underlying causal structures may show a better out-of-distribution generalisation.
  Eg: Will it rain tomorrow?

Figure sources:
1. Getty
2. CNN, Ahr Valley 2021 floods
3. Tiny-Giant.net
4. Apple weather app

do-experiments are Hard to Do





- Interventions can be unethical
- Interventions can be impossible or highly impractical
- Interventions can be expensive

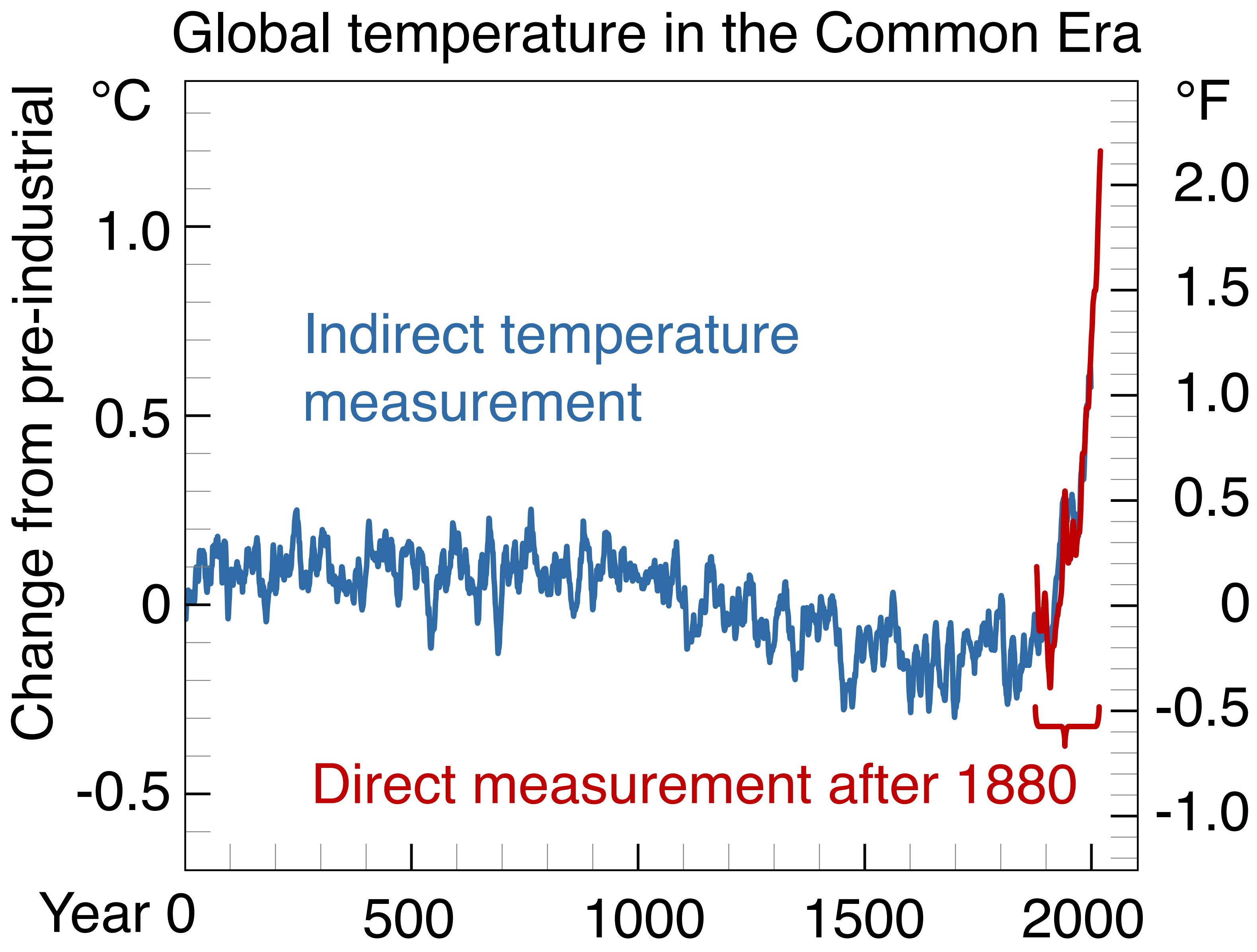


Image source: Wikipedia

(The Obligatory xkcd Slide)

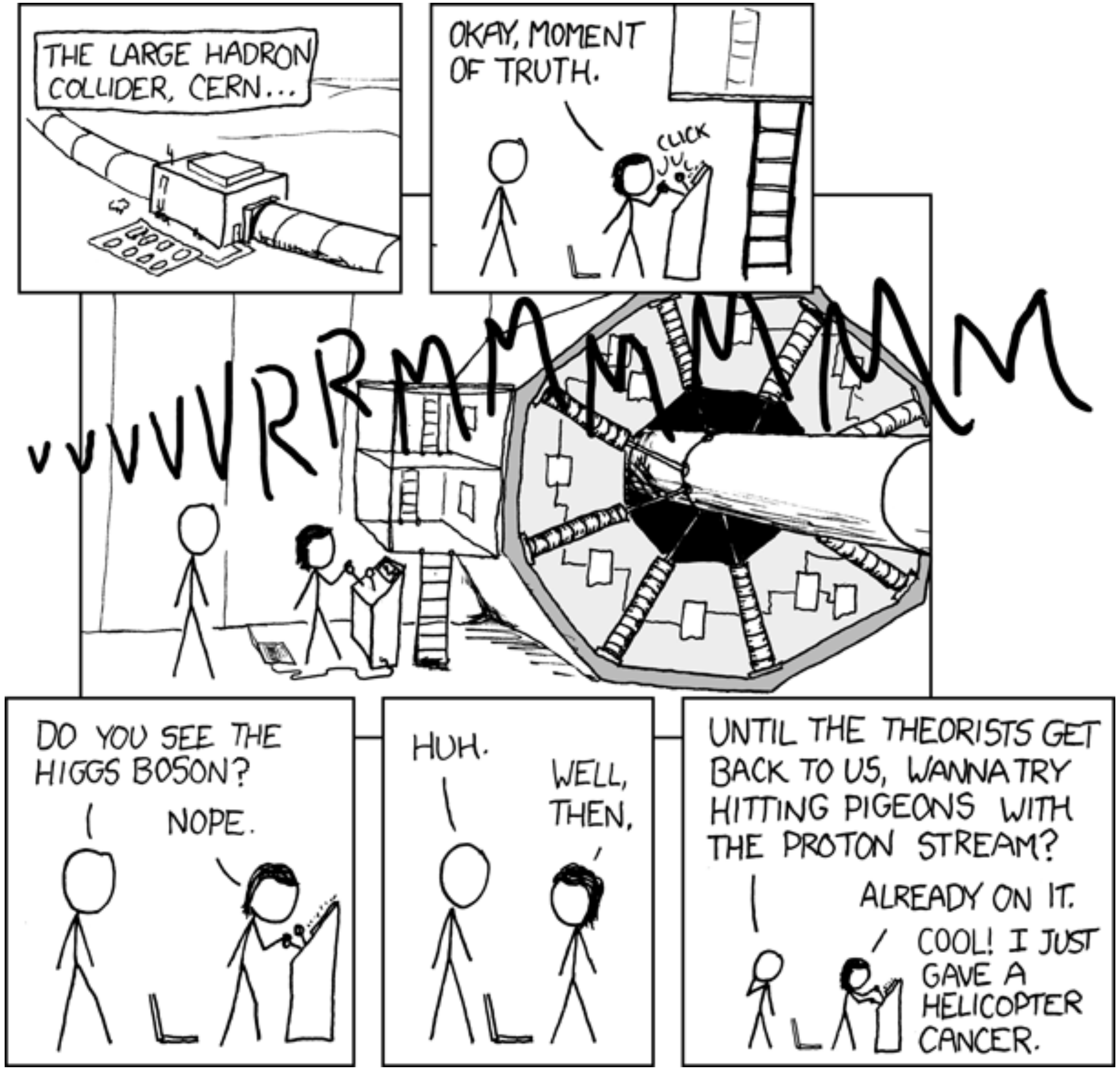
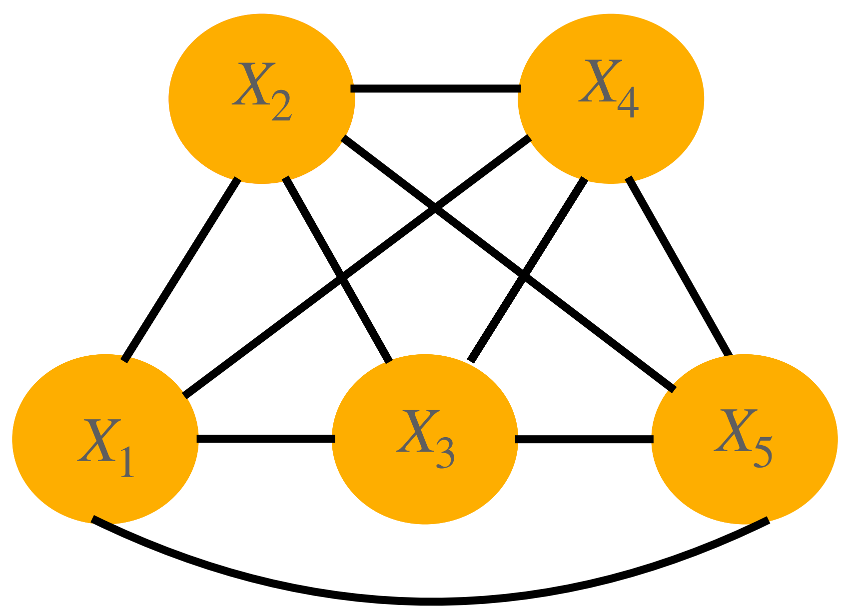


Image source: xkcd comics

Causal Structure Learning

Query: What Can we Do without Prior Assumptions?



X_1	X_2	X_3	X_4	X_5
0.7	0.6	1.33	2.4	0.01
0.5	0.56	0.98	2.2	…
0.75	0.5	1.56	…	…
0.6	…	…	…	…
0.56	…	…	…	…
0.53	…	…	…	…
0.69	…	…	…	…
0.5	…	…	…	…
0.7	…	…	…	…
…	…	…	…	…

- Given a dataset of samples for five random variables, what can we say about their causal relationships?
- We are supplied with conditional independence tests:

$$X_1 \perp X_3 \mid X_2 \ , \ X_1 \perp X_4 \ , \ X_1 \perp X_5$$

$$X_2 \perp X_4 \ , \ X_2 \perp X_5$$

$$X_3 \perp X_5 \mid X_4$$

- What can we say about the causal graph between these variables?

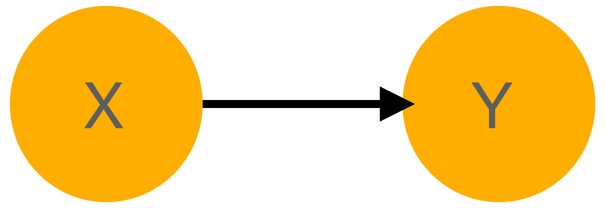
Answer: Nothing :/

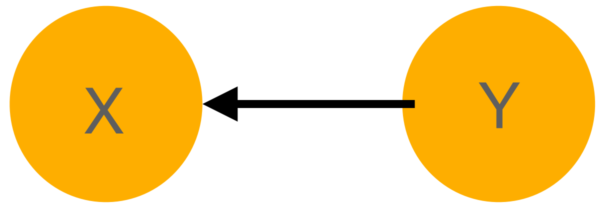
Interlude: Reichenbach's Common Cause Principle[1]

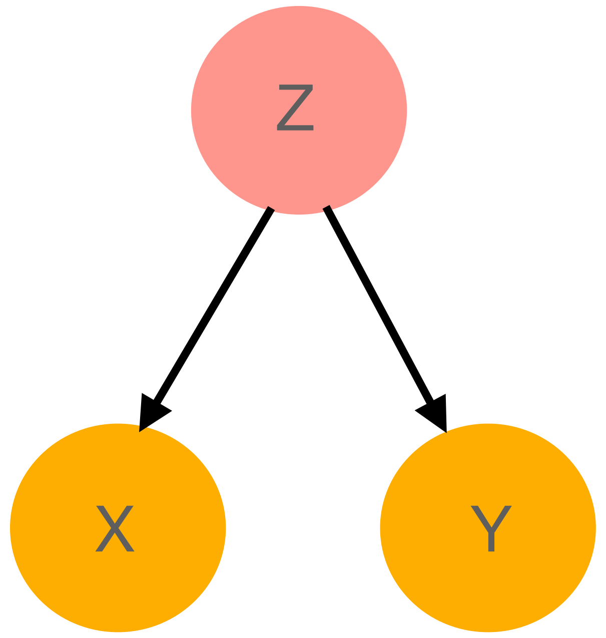
An intuition to formalise the connection between causality and statistical dependence

If two random variables X and Y are statistically dependent $(X \not\perp\!\!\!\perp Y)$, then :

1. X is (possibly indirectly) causing Y , or
2. Y is (possibly indirectly) causing X, or
3. there is a (possibly unobserved) common cause Z that (possibly indirectly) causes both X and Y

1.


2.


3.


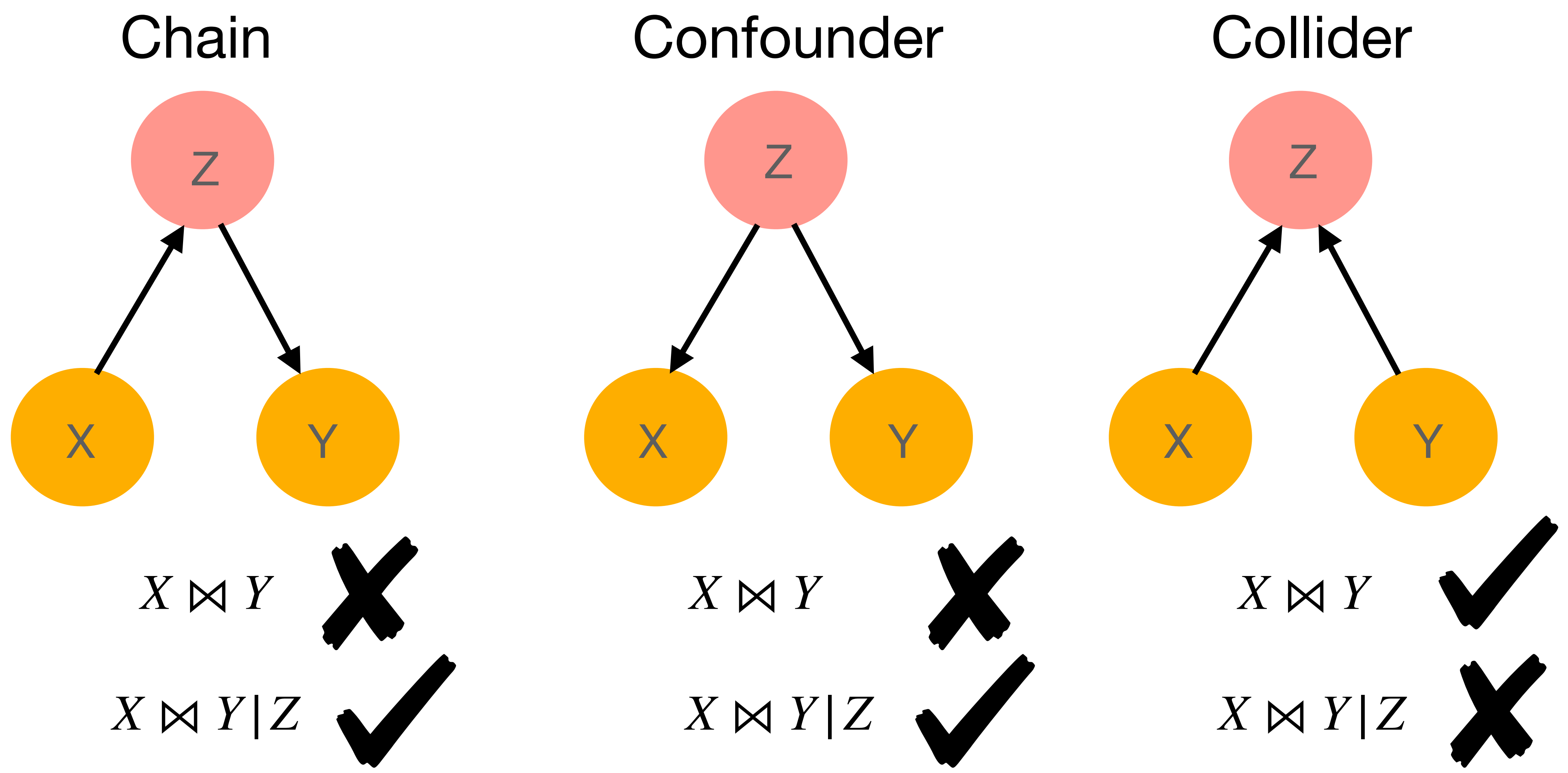
Generally: Statistical dependence $\Longrightarrow$ Causal 'connectedness': **Causal Markov Condition**

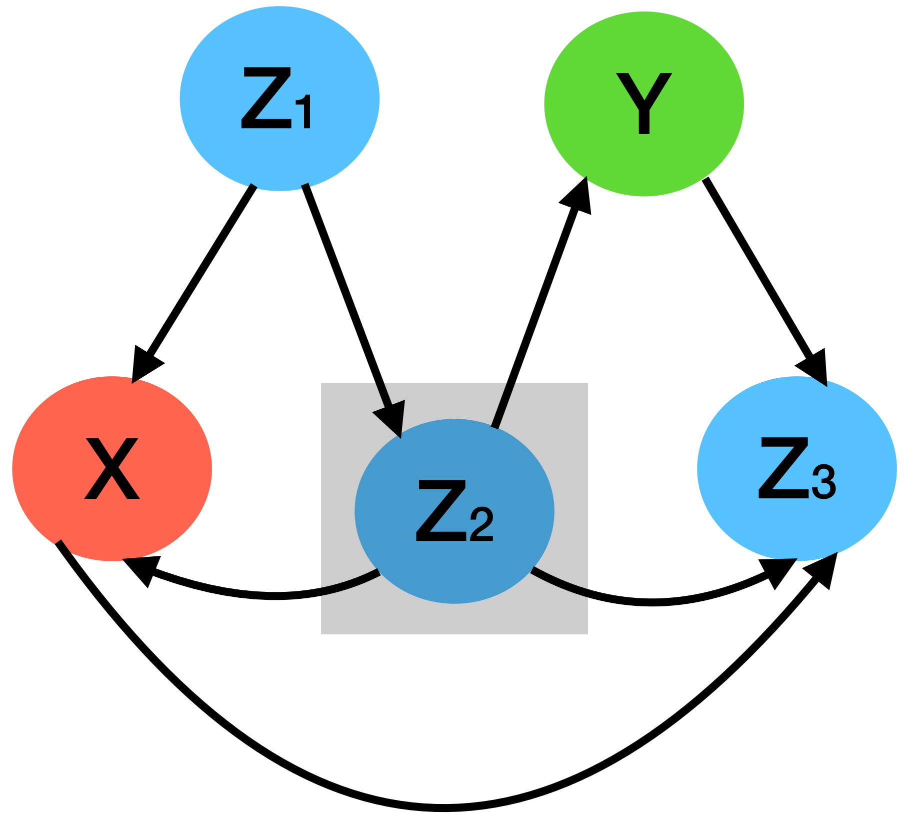
1: Hans Reichenbach, "The Direction of Time", 1956

d-separation and Causal Markov Condition

A graphical criterion to aid causal inference

- A vertex X in a graph is said to be d-separate (⋈) from another vertex Y given a set of vertices *S*, when a set of conditions concerning all paths from X to Y are satisfied.





$X \bowtie Y | Z_2$? (Yes)

- Causal Markov Condition: $X \bowtie Y | Z \Rightarrow X \perp\!\!\!\perp Y | Z$ (Causal graph is Markov relative to $P_{X,Y,\ldots}$)
- ‘The underlying causal graphical structure leaves certain (conditional) independencies as imprints in the observational distribution.’

Causal Faithfulness Assumption

The other side of the coin

- If all the (conditional) independencies implied by the Markov condition are true, and no more, then causal faithfulness is said to hold.

- Causal faithfulness Assumption : $X \perp\!\!\!\perp Y | Z \Rightarrow X \bowtie Y | Z$

- Both the causal Markov and causal faithfulness properties state a relationship between a causal graph and probability distribution over the same set of variables.

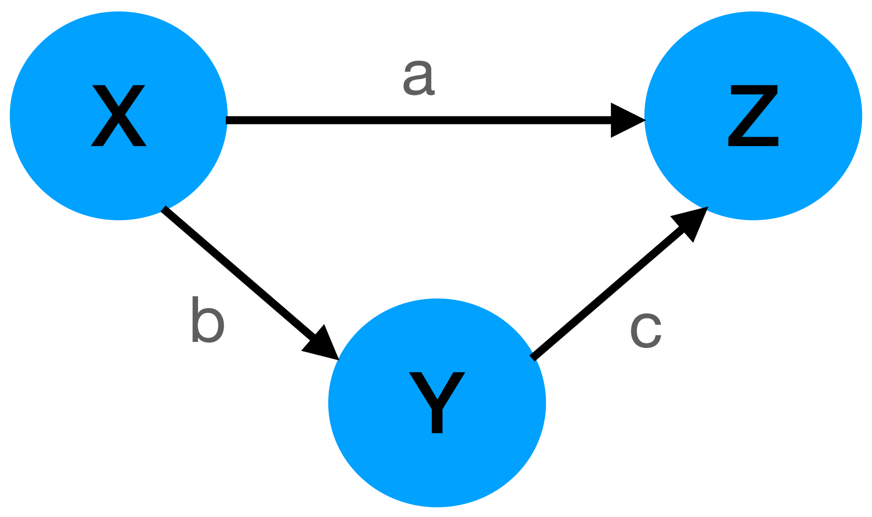
$X := \eta_X$

$Y := b \cdot X + \eta_Y$

$Z := a \cdot X + c \cdot Y + \eta_Z$

Here $\eta_i \sim N(0,1)$ are independent noise terms.

(The ':=' denotes that these are *causal* assignments, and the set of equations together is called a *structural causal model*)

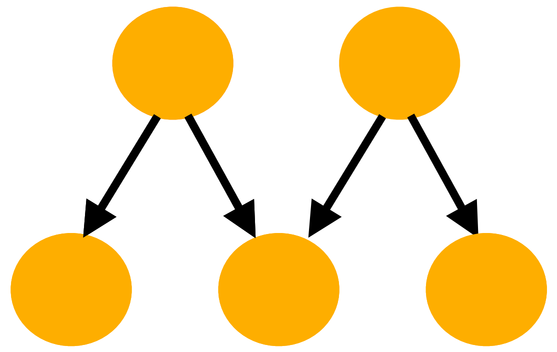


- If $a = -b \cdot c$, then causal faithfulness is violated. Therefore, we rule out such fine-tuning of causal influences from different paths when we assume faithfulness.

The PC Algorithm for Causal Graph Discovery

- The PC algorithm[1] has become the standard example for the success of causal (graph) discovery using conditional independence testing.
- It assumes causal Markov property, faithfulness, no cycles and no hidden common causes

Ground Truth $\mathscr{G}$



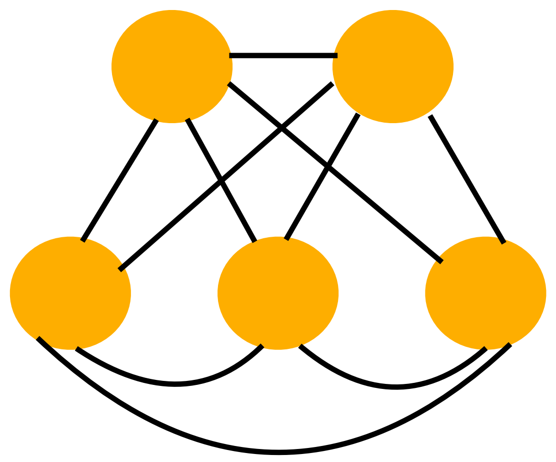
1: Spirtes, Glamour, Scheines, “Causation, Prediction and Search”, 2000
2. Meek, “Complete Orientation Rules for Patterns”, ‘95

The PC Algorithm for Causal Graph Discovery

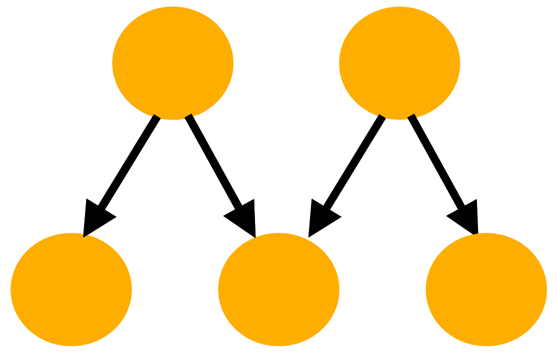
- The PC algorithm[1] has become the standard example for the success of causal (graph) discovery using conditional independence testing.

- It assumes causal Markov property, faithfulness, no cycles and no hidden common causes

1. To discover a causal graph $\mathscr{G}$ over variables $X_1, \ldots, X_n$, start with a fully-connected undirected graph G.



Ground Truth $\mathscr{G}$



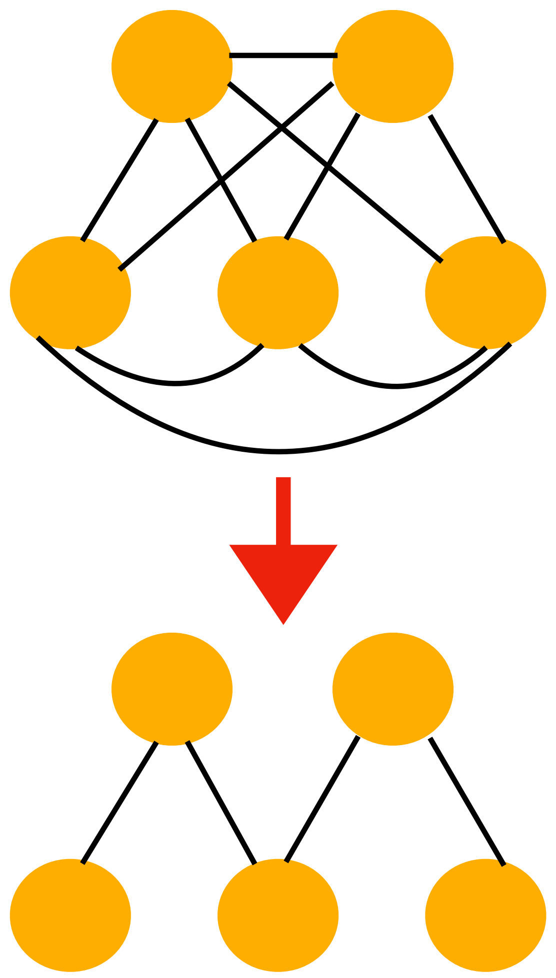
1: Spirtes, Glamour, Scheines, “Causation, Prediction and Search”, 2000
2. Meek, “Complete Orientation Rules for Patterns”, ‘95

The PC Algorithm for Causal Graph Discovery

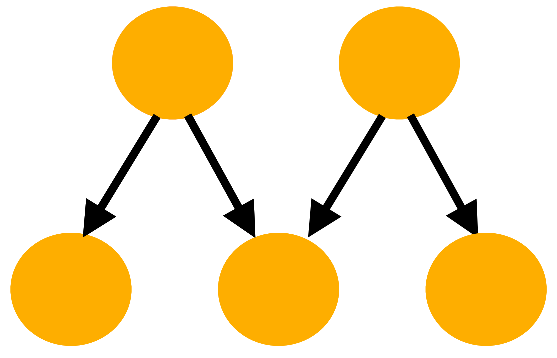
- The PC algorithm[1] has become the standard example for the success of causal (graph) discovery using conditional independence testing.
- It assumes causal Markov property, faithfulness, no cycles and no hidden common causes

1. To discover a causal graph $\mathscr{G}$ over variables $X_1, \ldots, X_n$, start with a fully-connected undirected graph G.
2. Progressively remove edges to get skeleton graph:

$p = 0$
For $(X_i, X_j) \in \mathbf{X}$:
    For $S \subset adj(X_i)$ or $S \subset adj(X_j)$:   ($adj(X)$:= adjacencies of X)
        If $X_i \perp\!\!\!\perp X_j \mid S$ and $|S| = p$: Remove $X_i - X_j$ edge
$p = p + 1$



Ground Truth $\mathscr{G}$



1: Spirtes, Glamour, Scheines, “Causation, Prediction and Search”, 2000
2. Meek, “Complete Orientation Rules for Patterns”, ‘95

The PC Algorithm for Causal Graph Discovery

- The PC algorithm[1] has become the standard example for the success of causal (graph) discovery using conditional independence testing.

- It assumes causal Markov property, faithfulness, no cycles and no hidden common causes

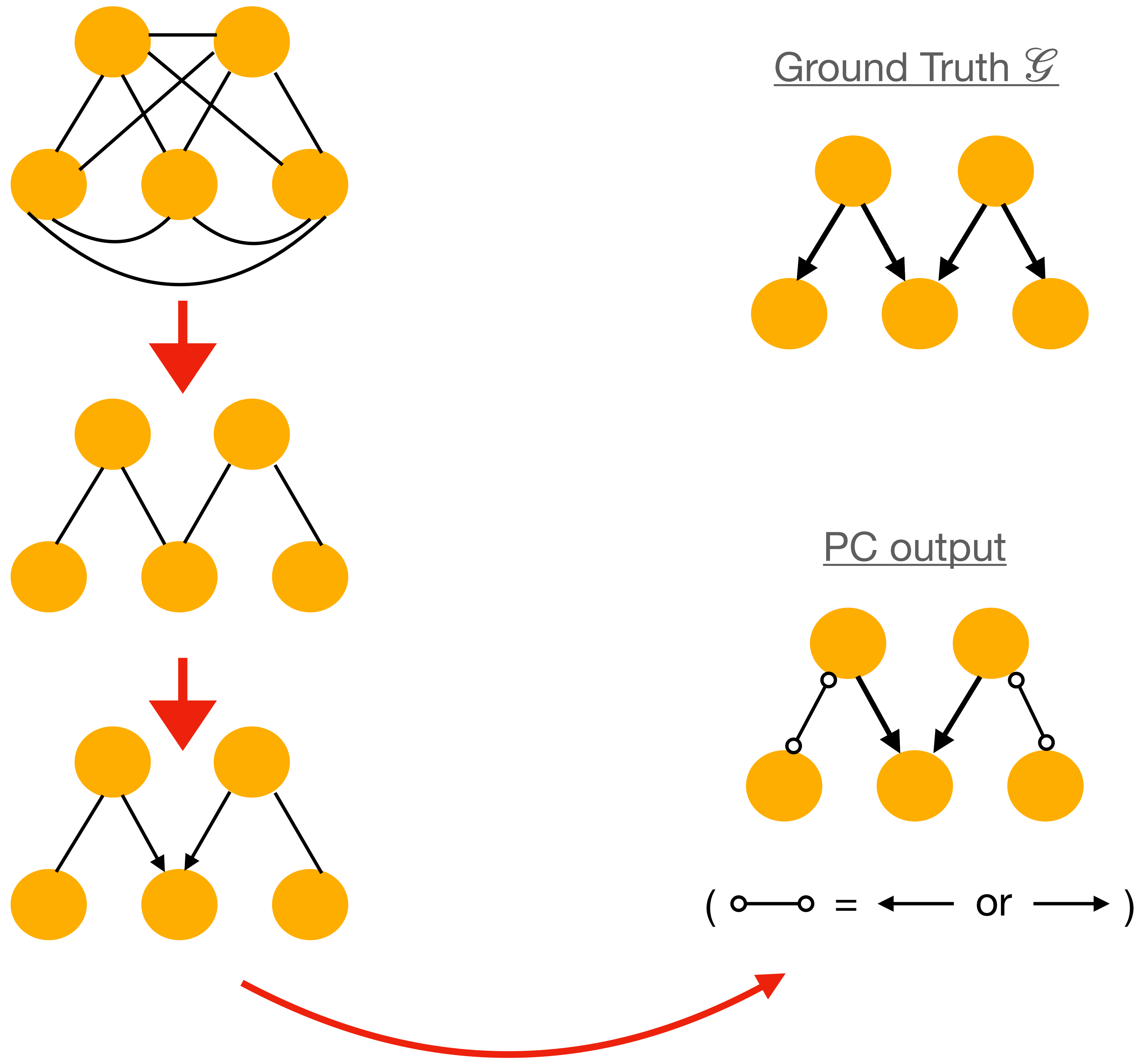
1. To discover a causal graph $\mathscr{G}$ over variables $X_1, \ldots, X_n$, start with a fully-connected undirected graph G.

2. Progressively remove edges to get skeleton graph:

   $p = 0$
   For $(X_i, X_j) \in \mathbf{X}$:
   For $S \subset adj(X_i)$ or $S \subset adj(X_j)$: ($adj(X)$:= adjacencies of X)
   If $X_i \perp\!\!\!\perp X_j \mid S$ and $|S| = p$: Remove $X_i - X_j$ edge
   $p = p + 1$

3. Orient colliders $X_i - X_j - X_k \Rightarrow X_i \rightarrow X_j \leftarrow X_k$
   When $X_i \perp\!\!\!\perp X_k \mid S$ and $X_j \notin S$

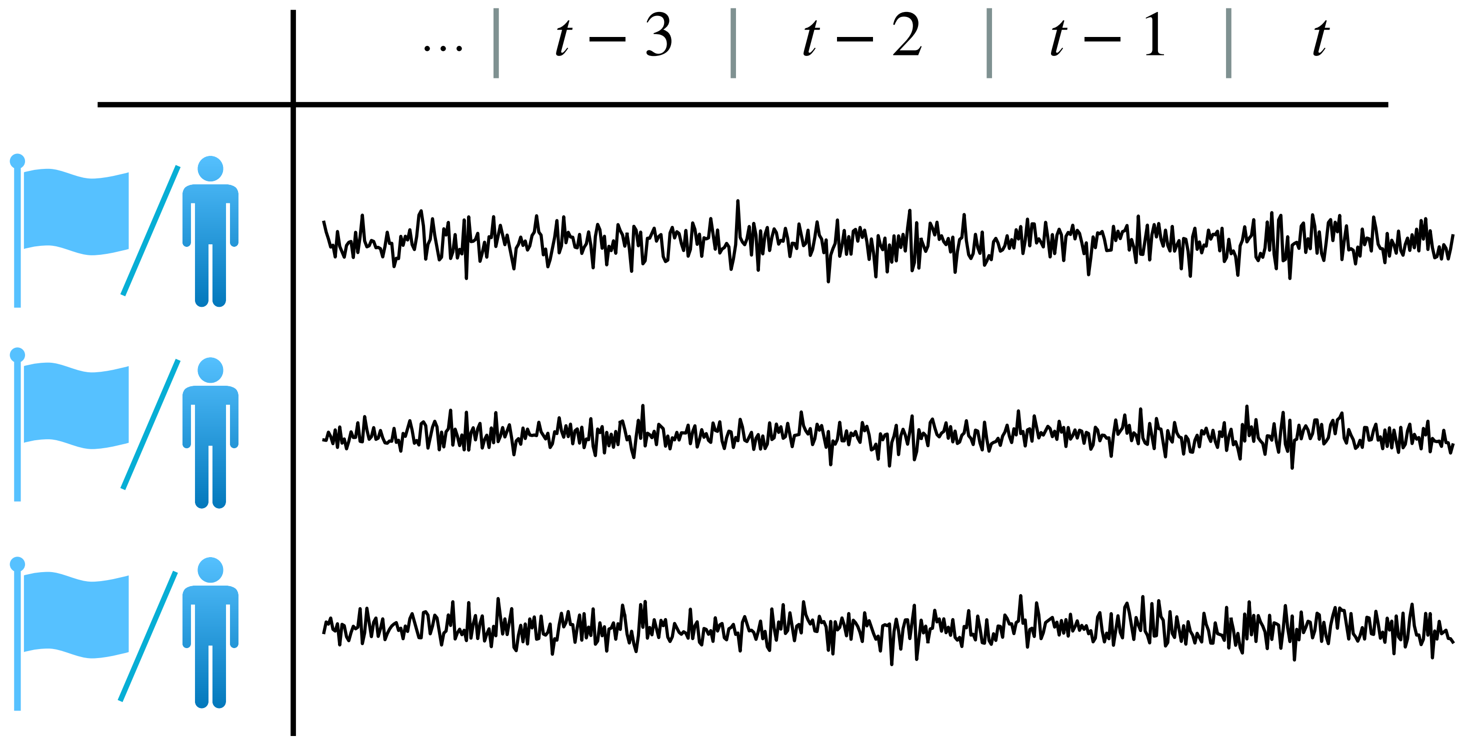
4. Orient as many remaining edges as possible (using orientation rules[2])

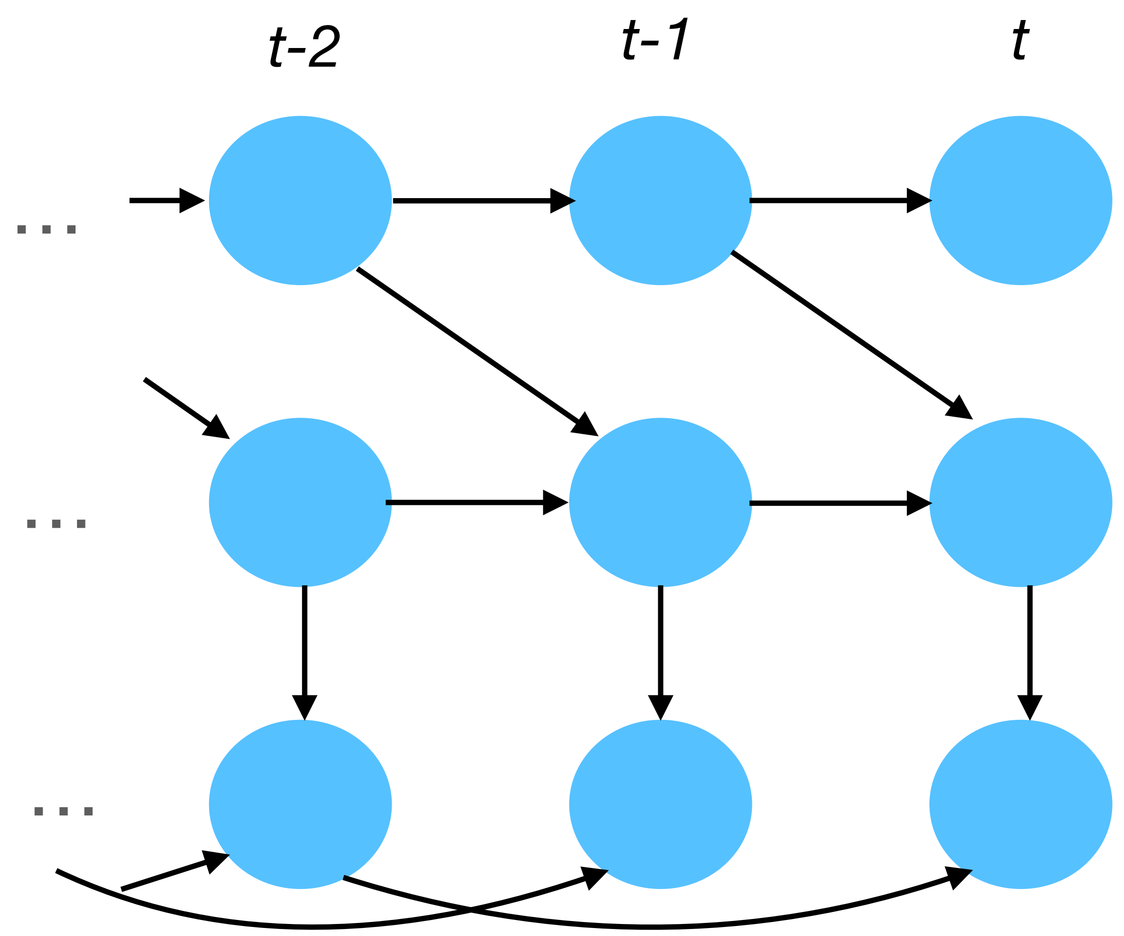


1: Spirtes, Glamour, Scheines, “Causation, Prediction and Search”, 2000
2. Meek, “Complete Orientation Rules for Patterns”, ‘95

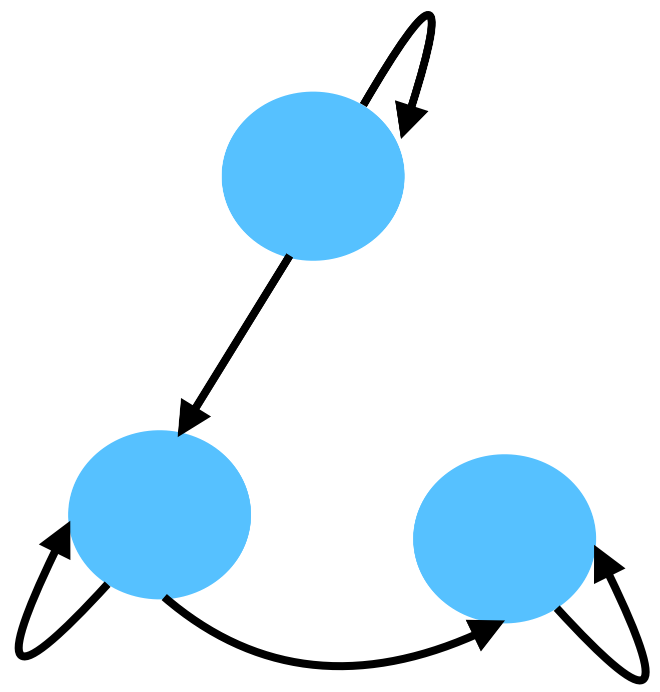
Causal Inference for Time Series

Possible Target Graphs



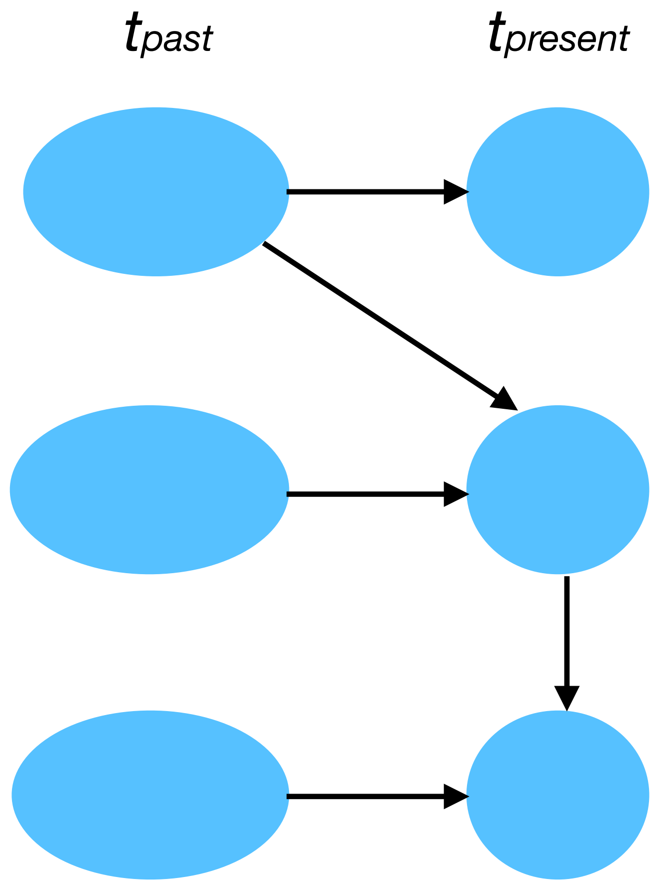


Full-time Graph



Summary Graph

Extended Summary Graph

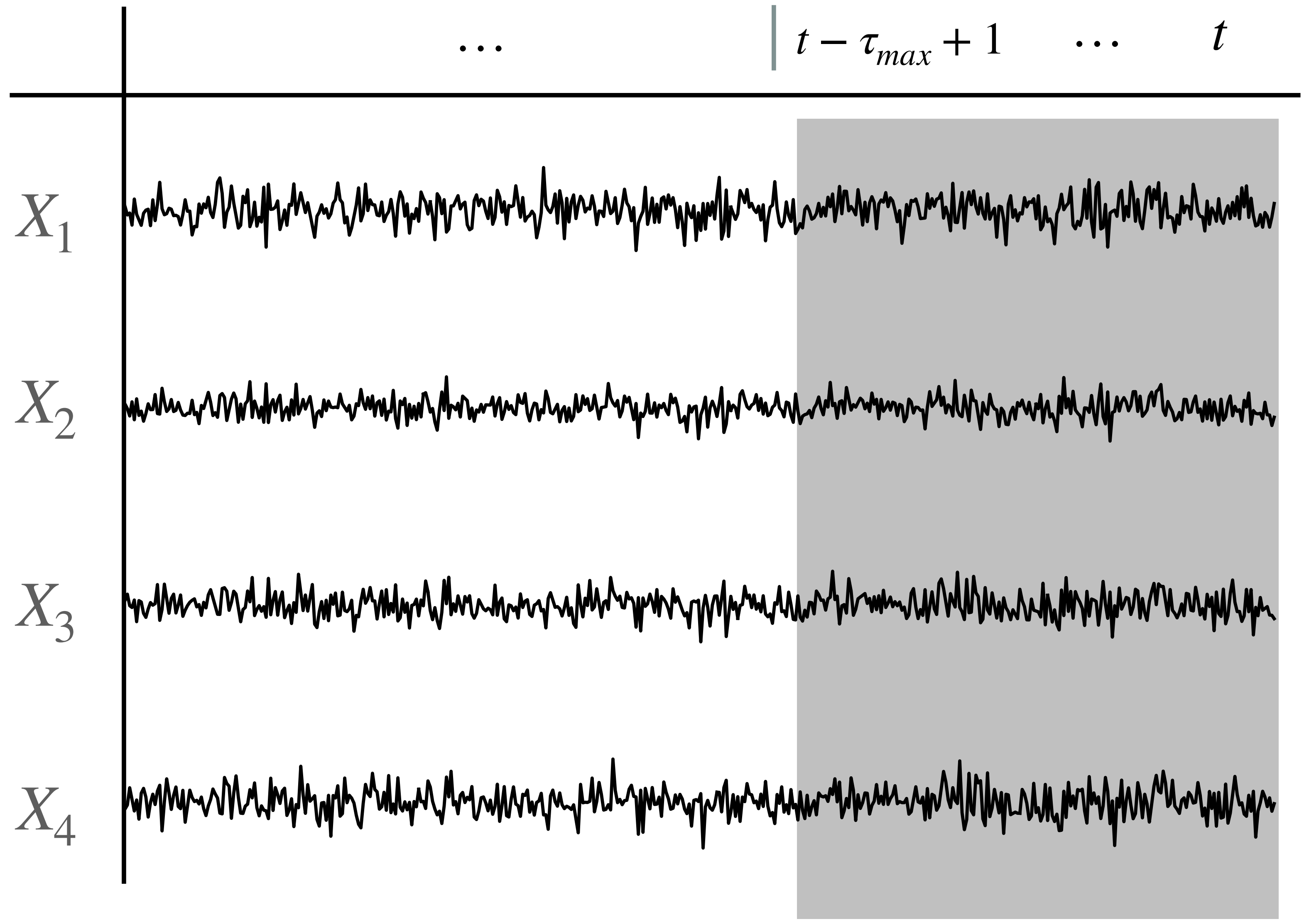


- Full-time Graph stretches infinitely into the past and the future
- Summary Graph is a finite graph that does not retain information about time-lags and time-indices
- Extended Summary Graph goes midway between the former two: it is a finite graph which distinguishes between lagged and contemporaneous links

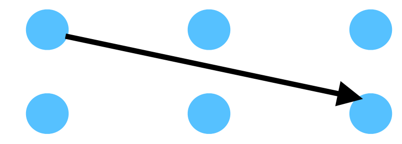
1: 'Discovering Causal Relations and Equations from Data', Camps-Valls et al, 2023
2: 'Discovery of Extended Summary Graphs in Time Series', Assaad et al, 2022

Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

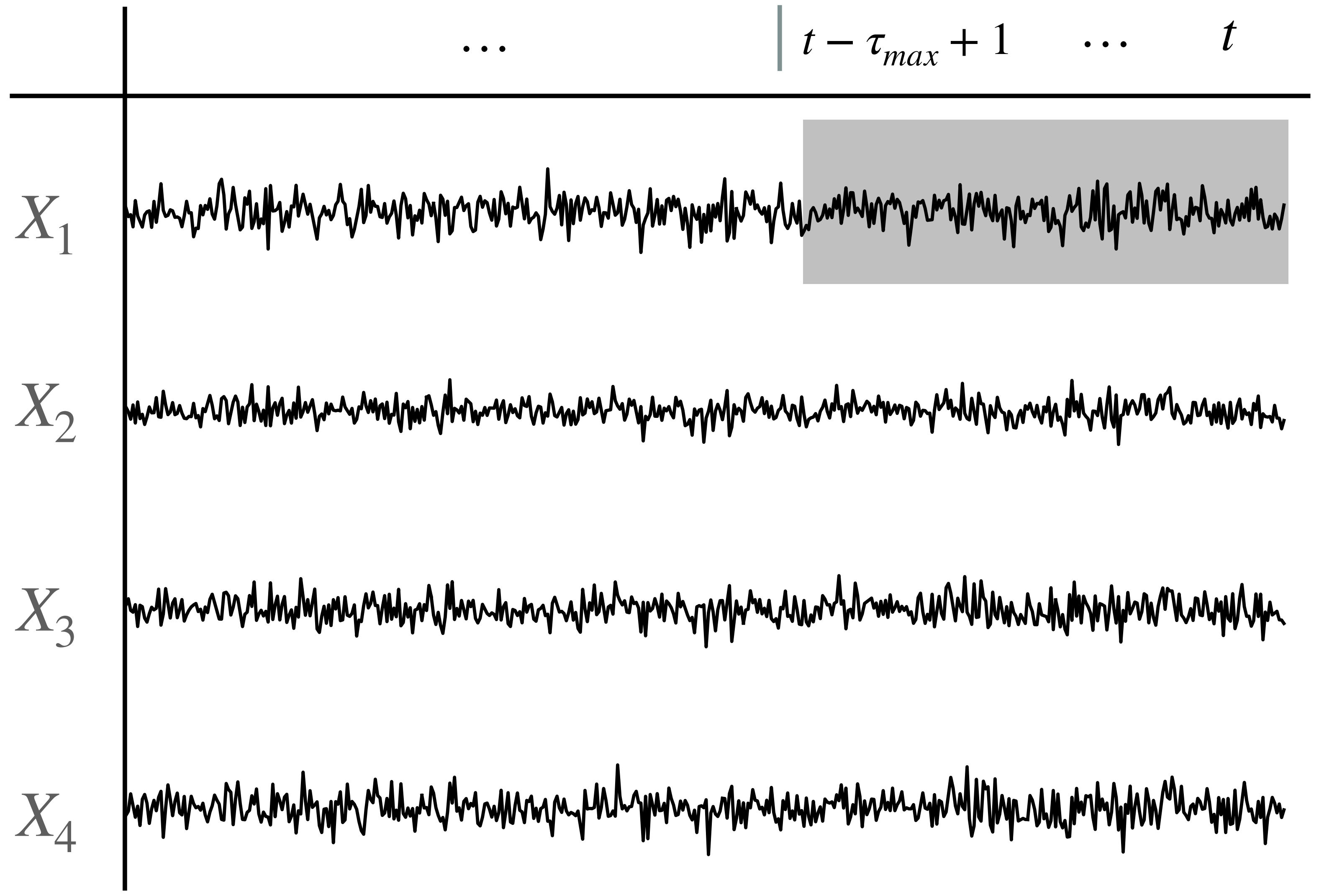


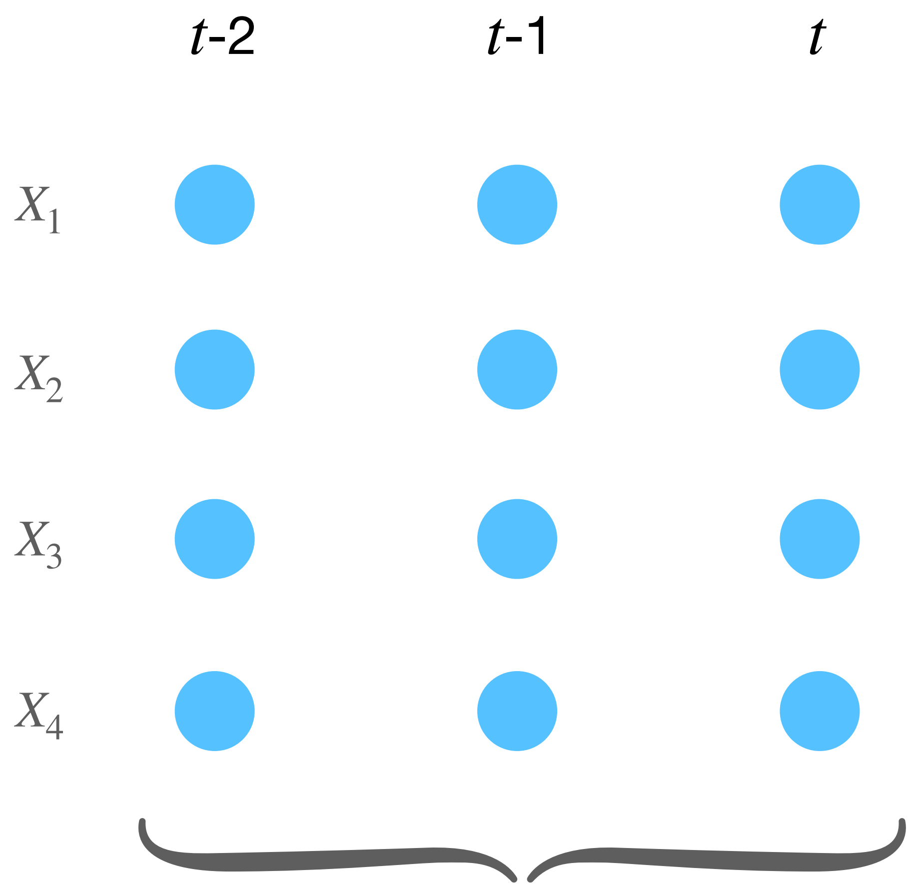
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



Causal Inference for Time Series

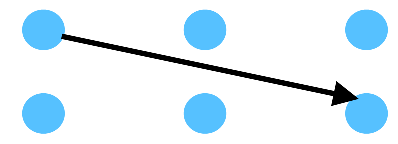
Basic tenets of time-series causal graph discovery



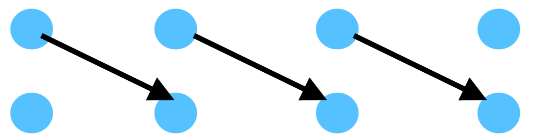


$[t, t - \tau_{max} + 1]$

- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.

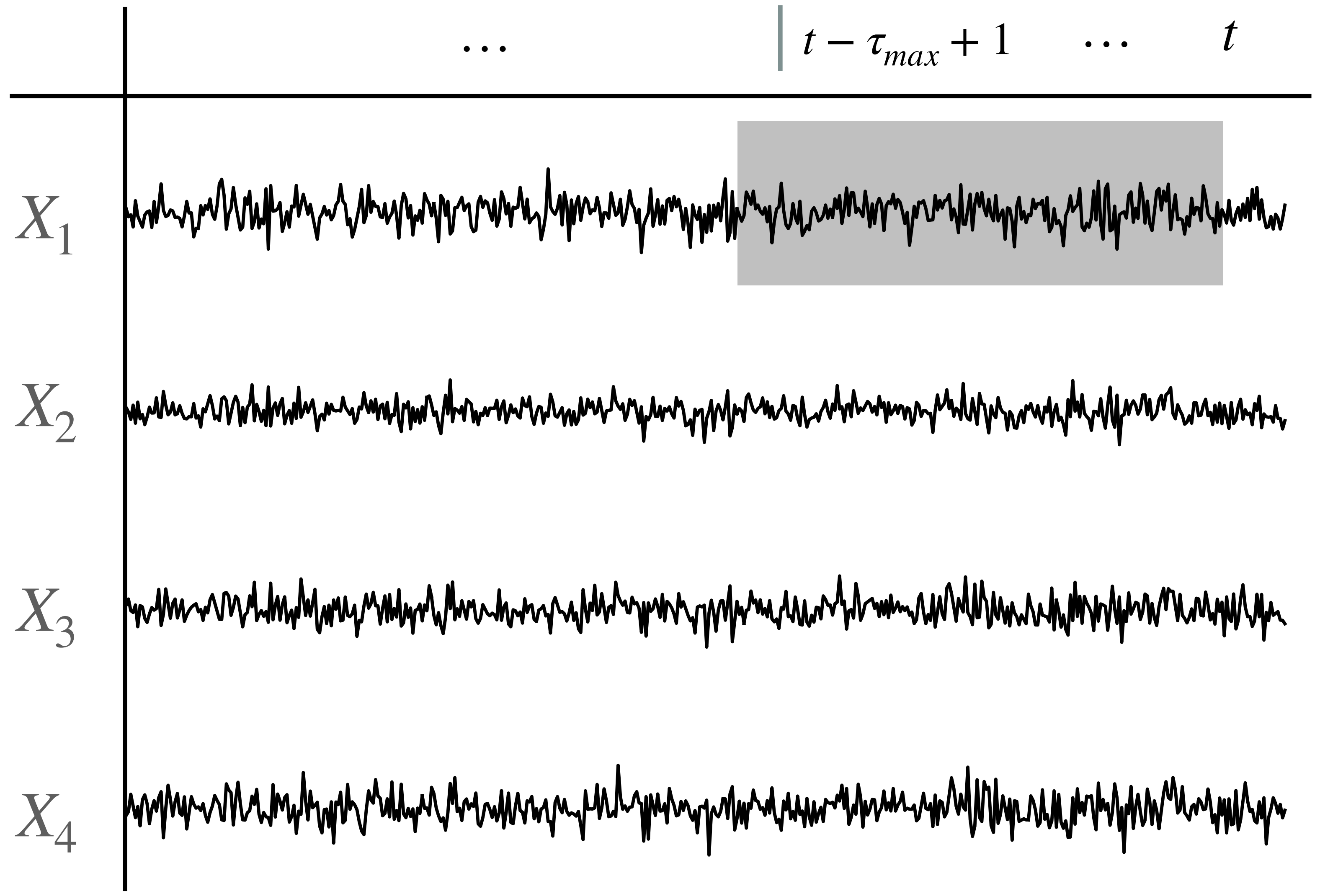


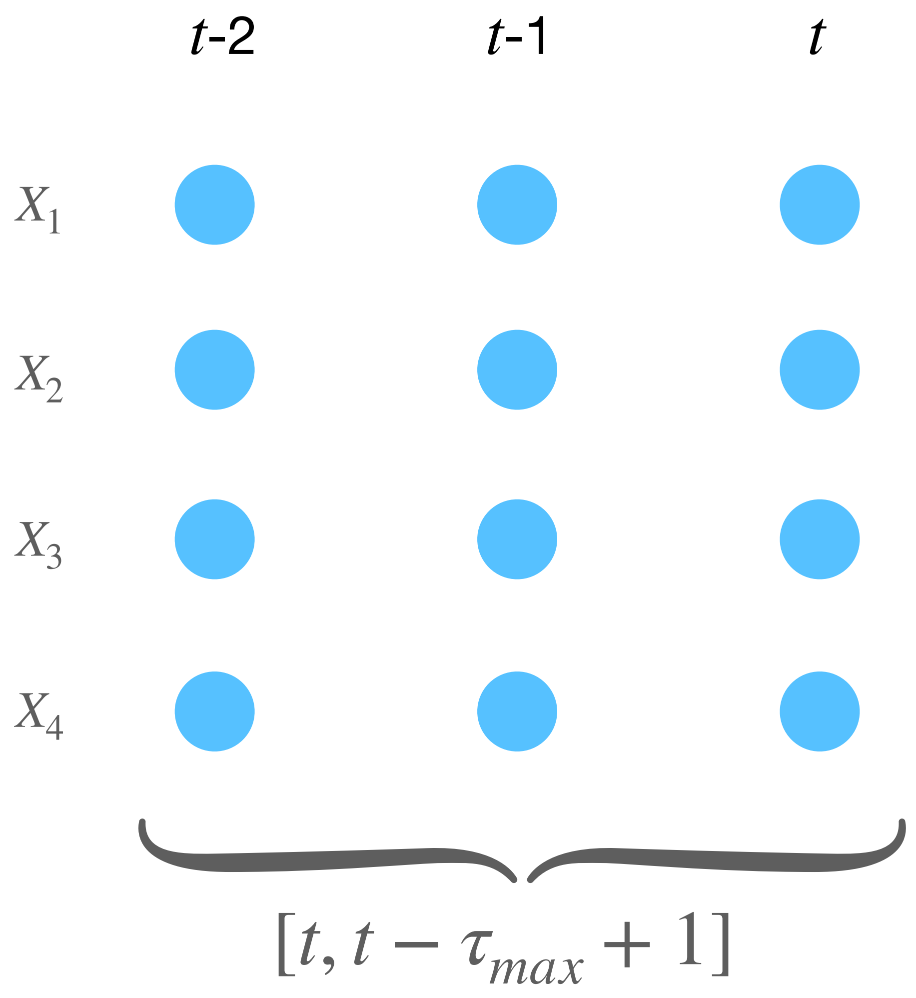
- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$



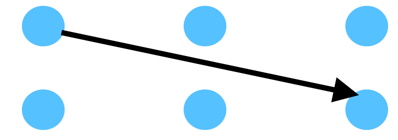
Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

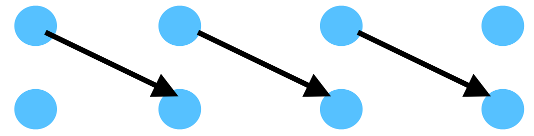




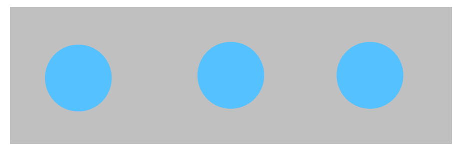
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$

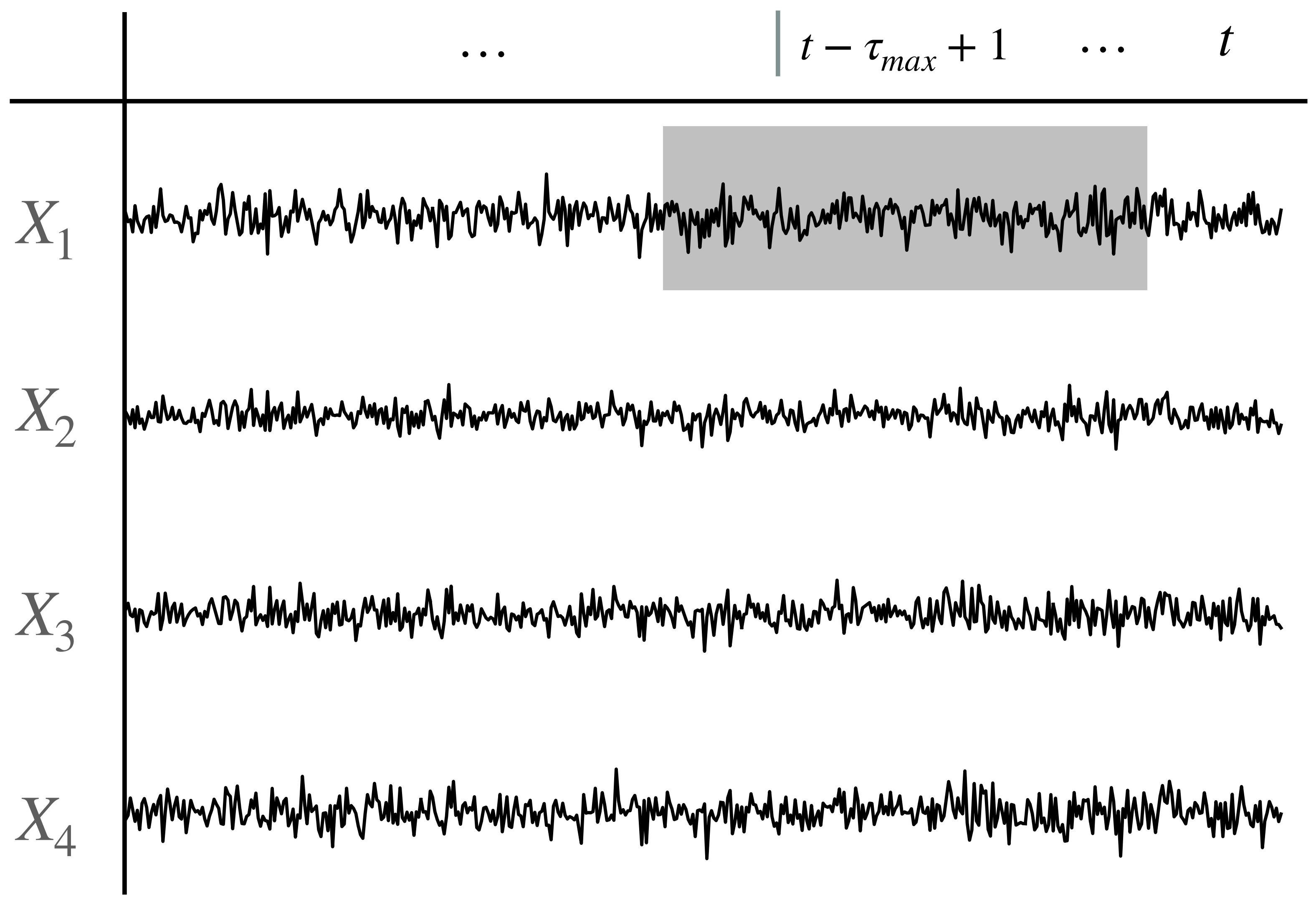


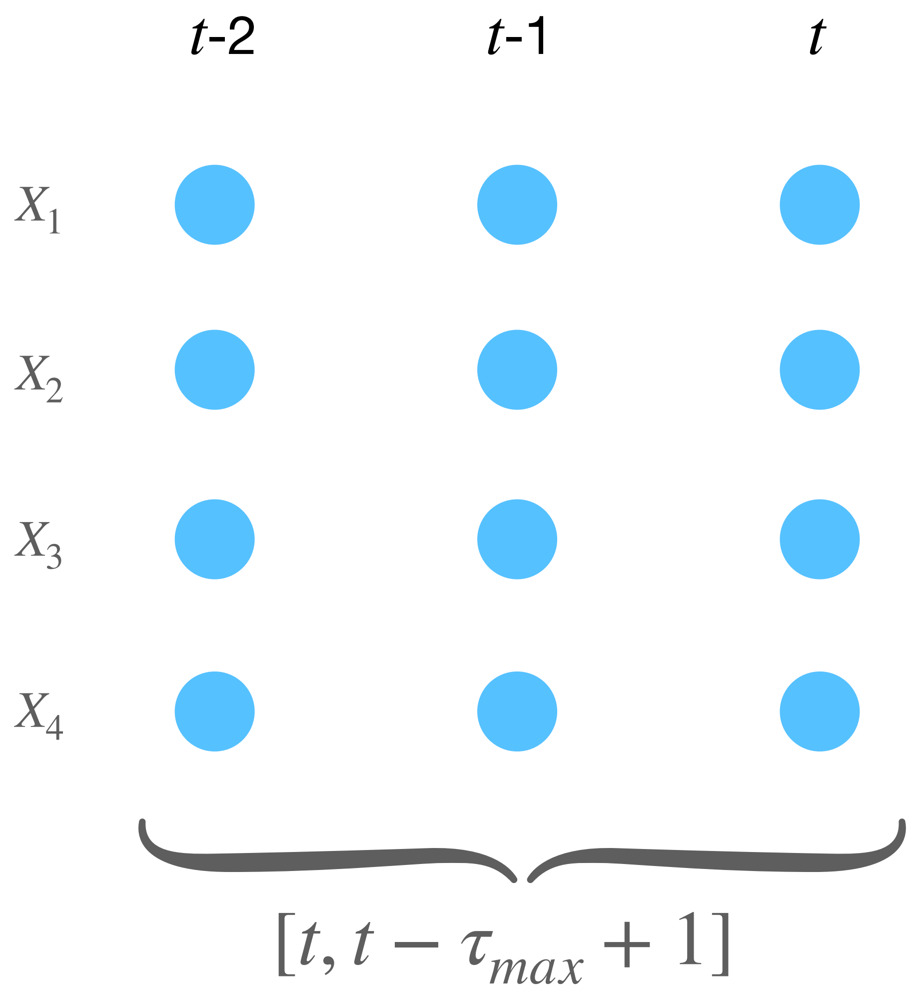
- Slide the time window $[t, t - \tau_{max} + 1]$ to generate samples for all X_i's



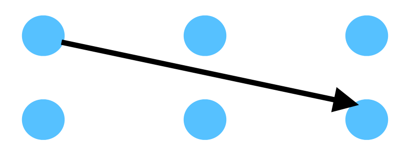
Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

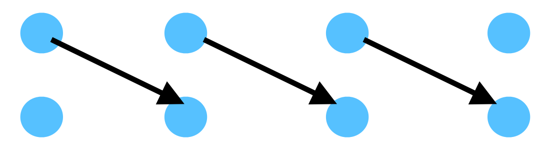




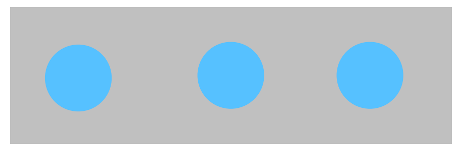
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$

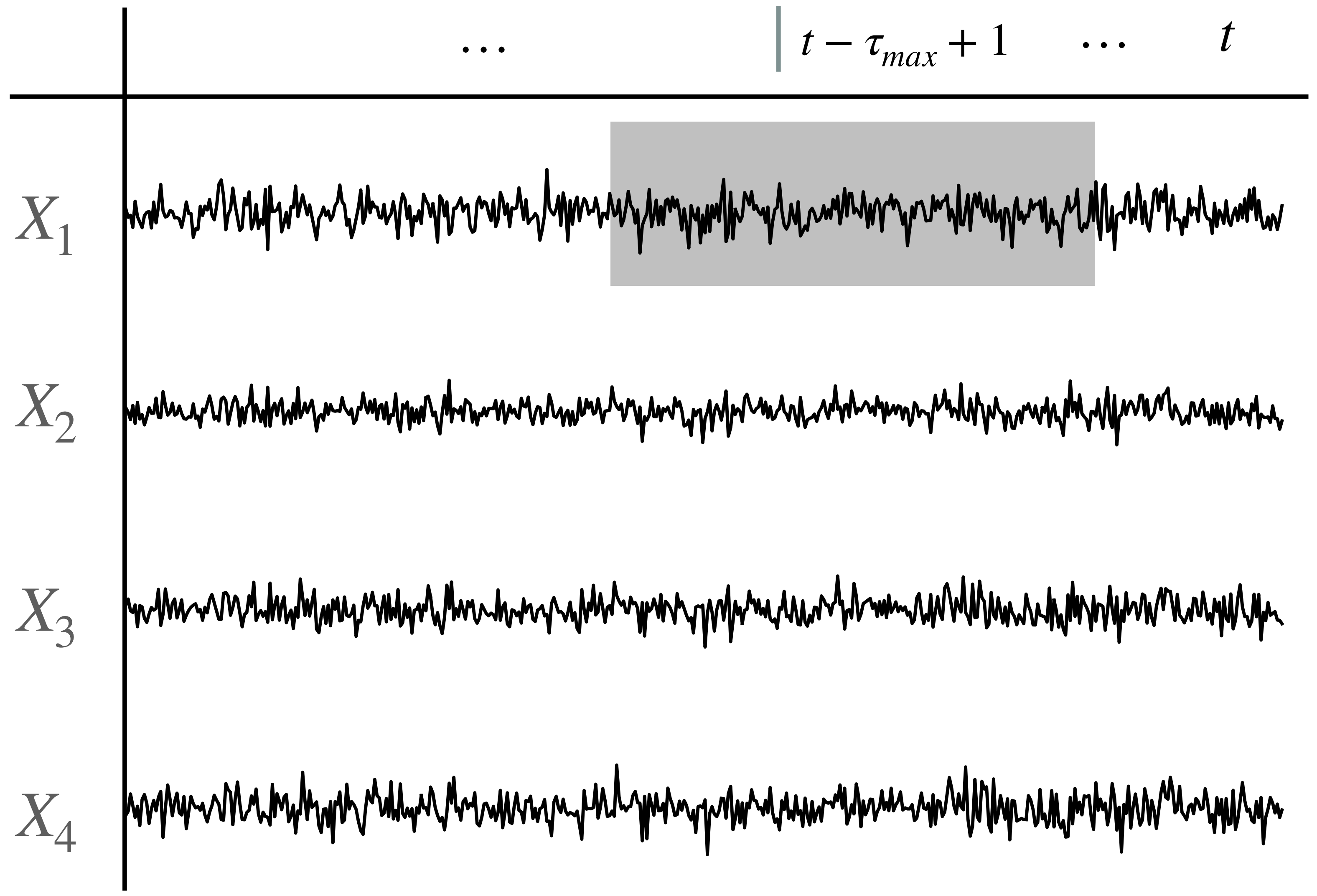


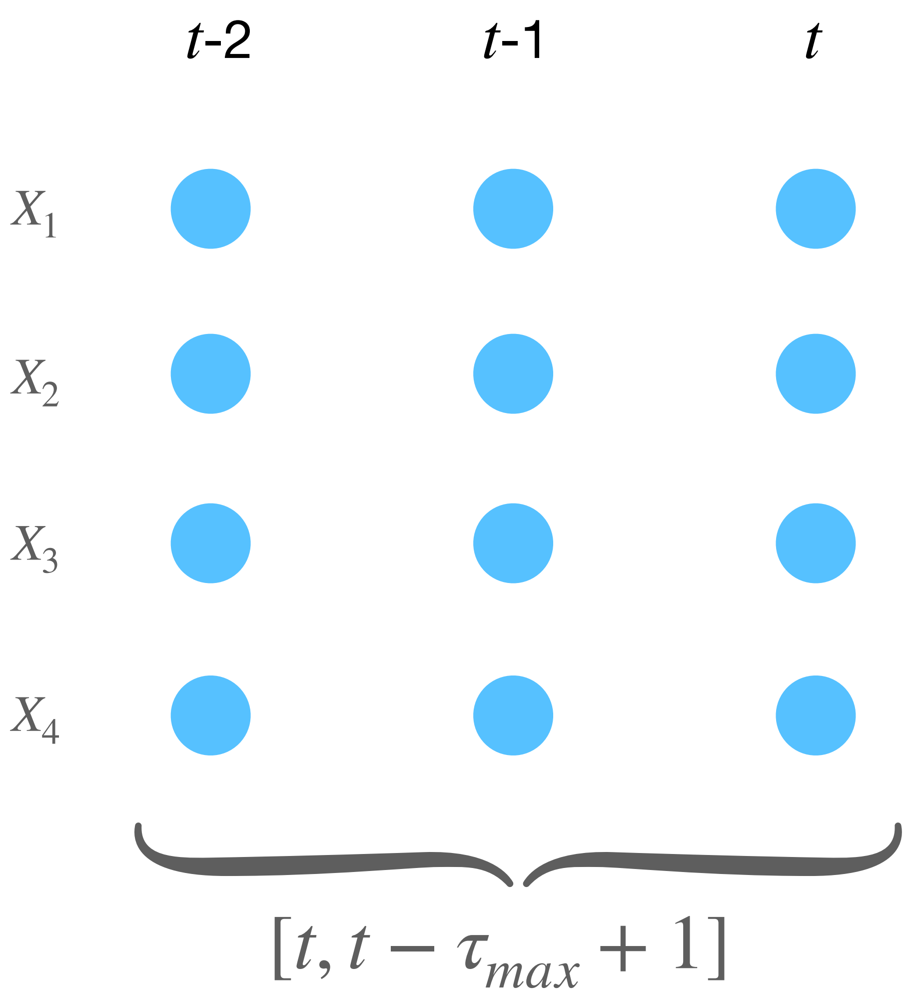
- Slide the time window $[t, t - \tau_{max} + 1]$ to generate samples for all X_i's



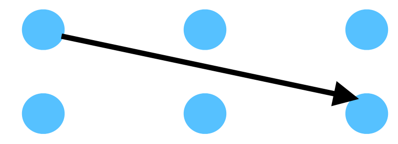
Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

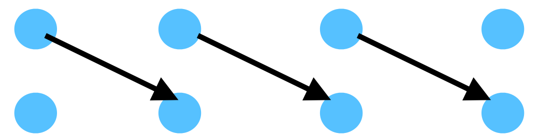




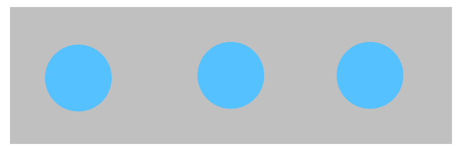
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$

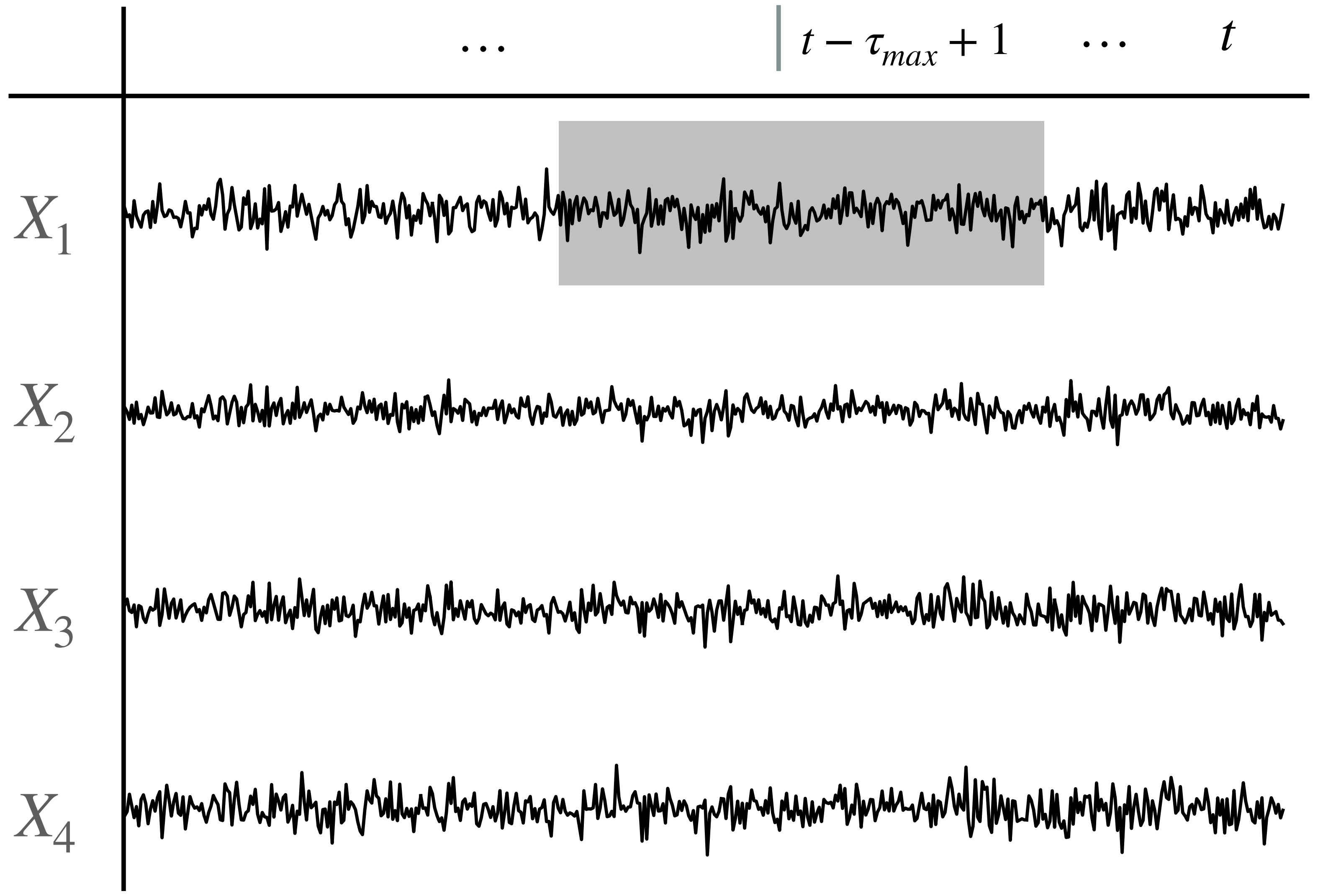


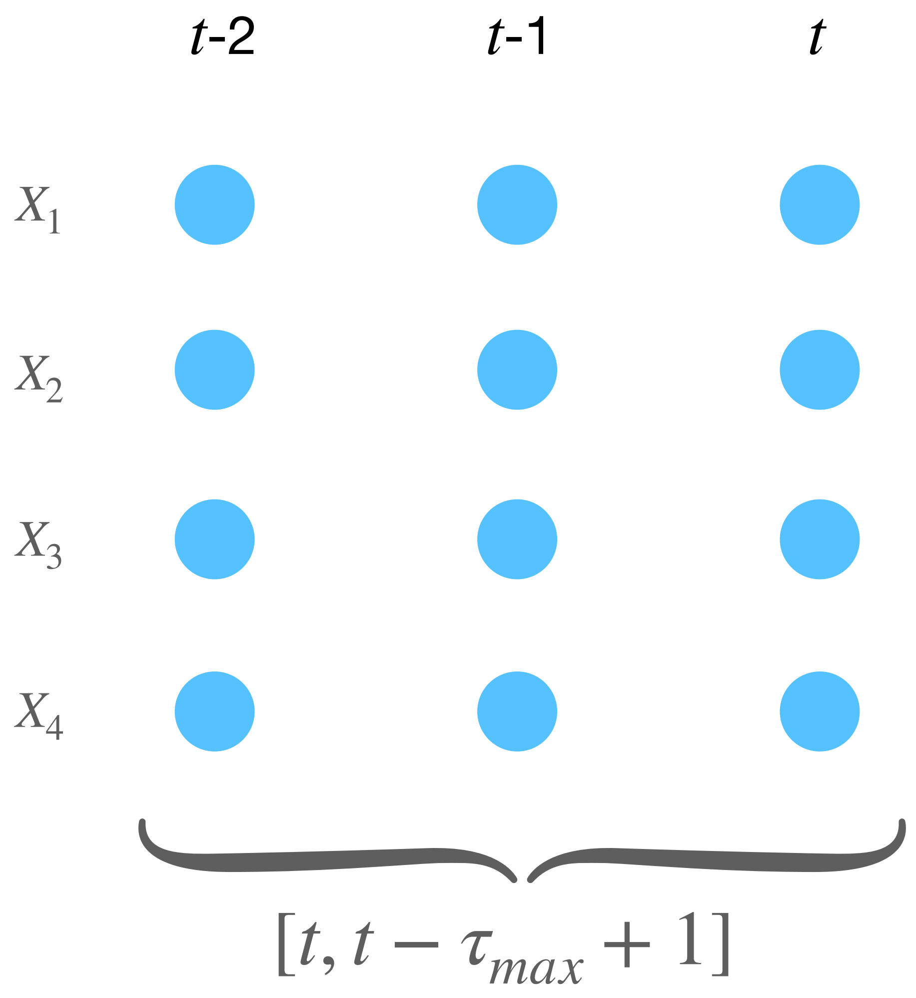
- Slide the time window $[t, t - \tau_{max} + 1]$ to generate samples for all X_i's



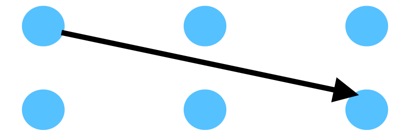
Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

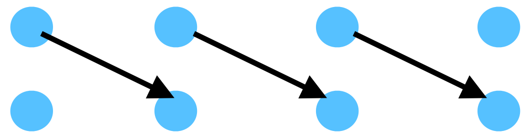




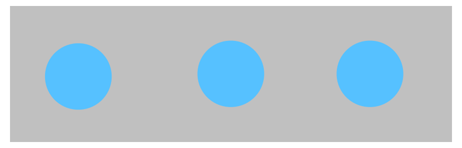
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$

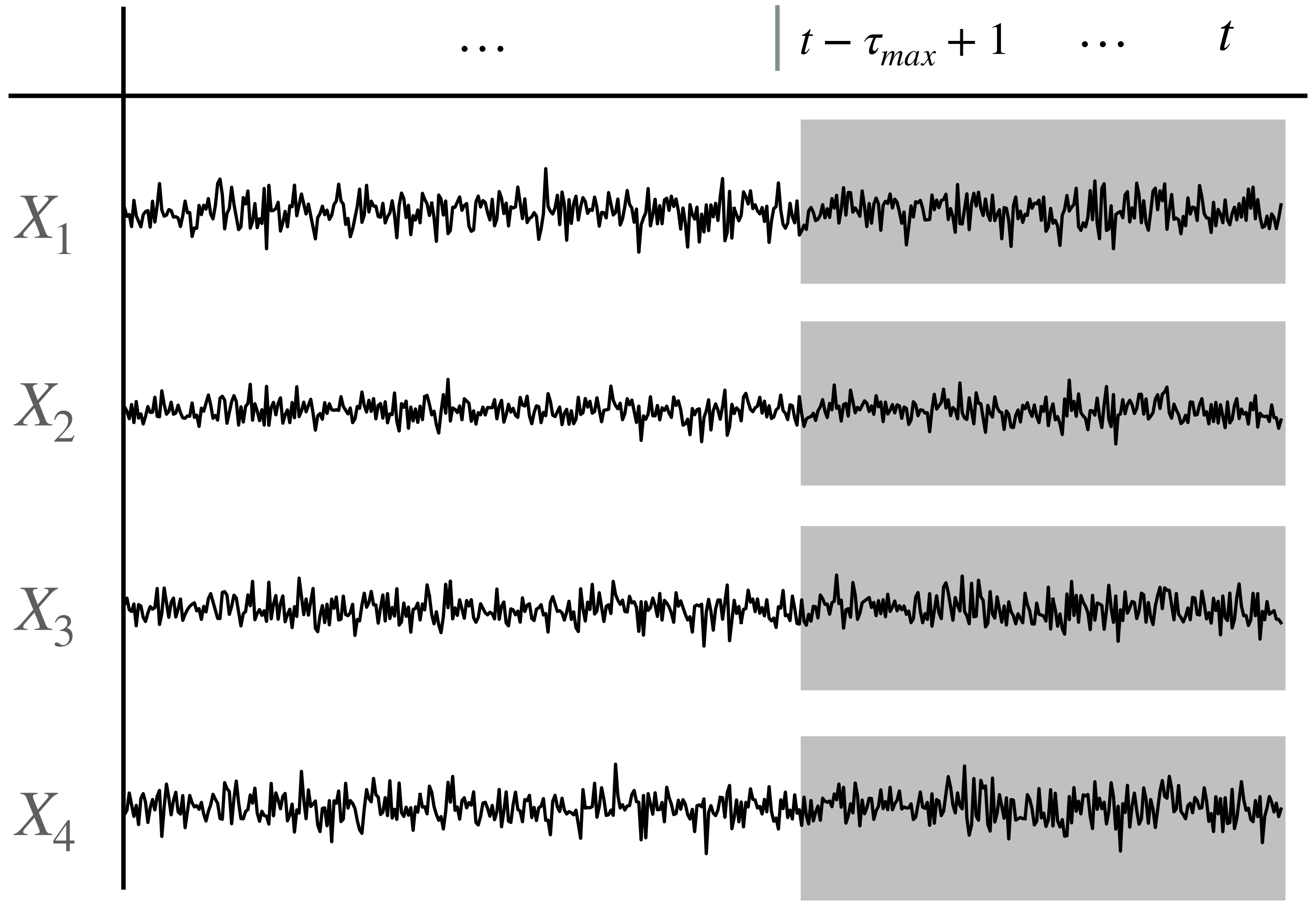


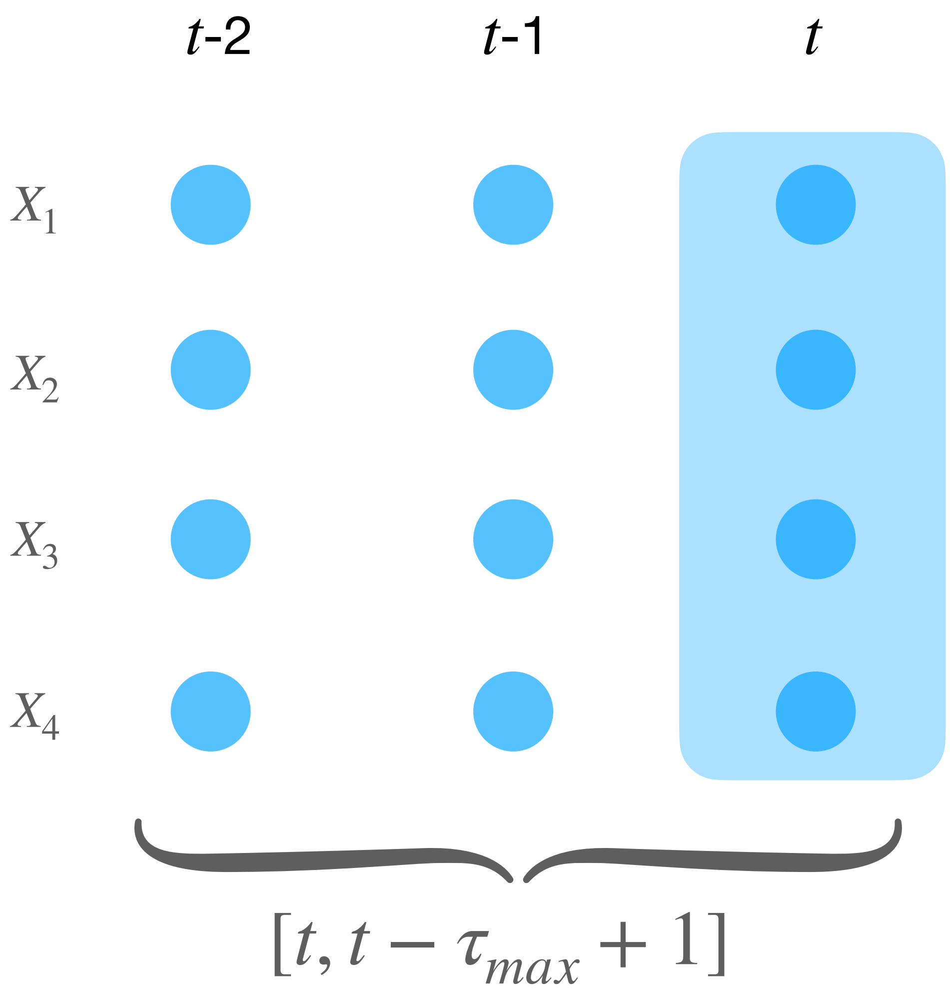
- Slide the time window $[t, t - \tau_{max} + 1]$ to generate samples for all X_i's



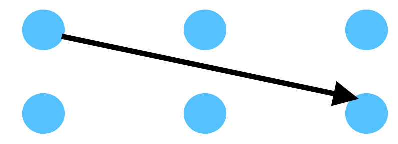
Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

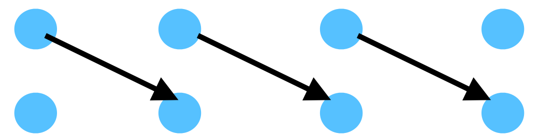




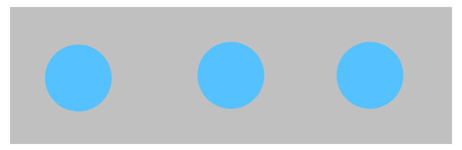
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



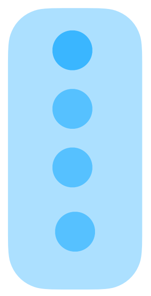
- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$



- Slide the time window $[t, t - \tau_{max} + 1]$ to generate samples for all X_i's

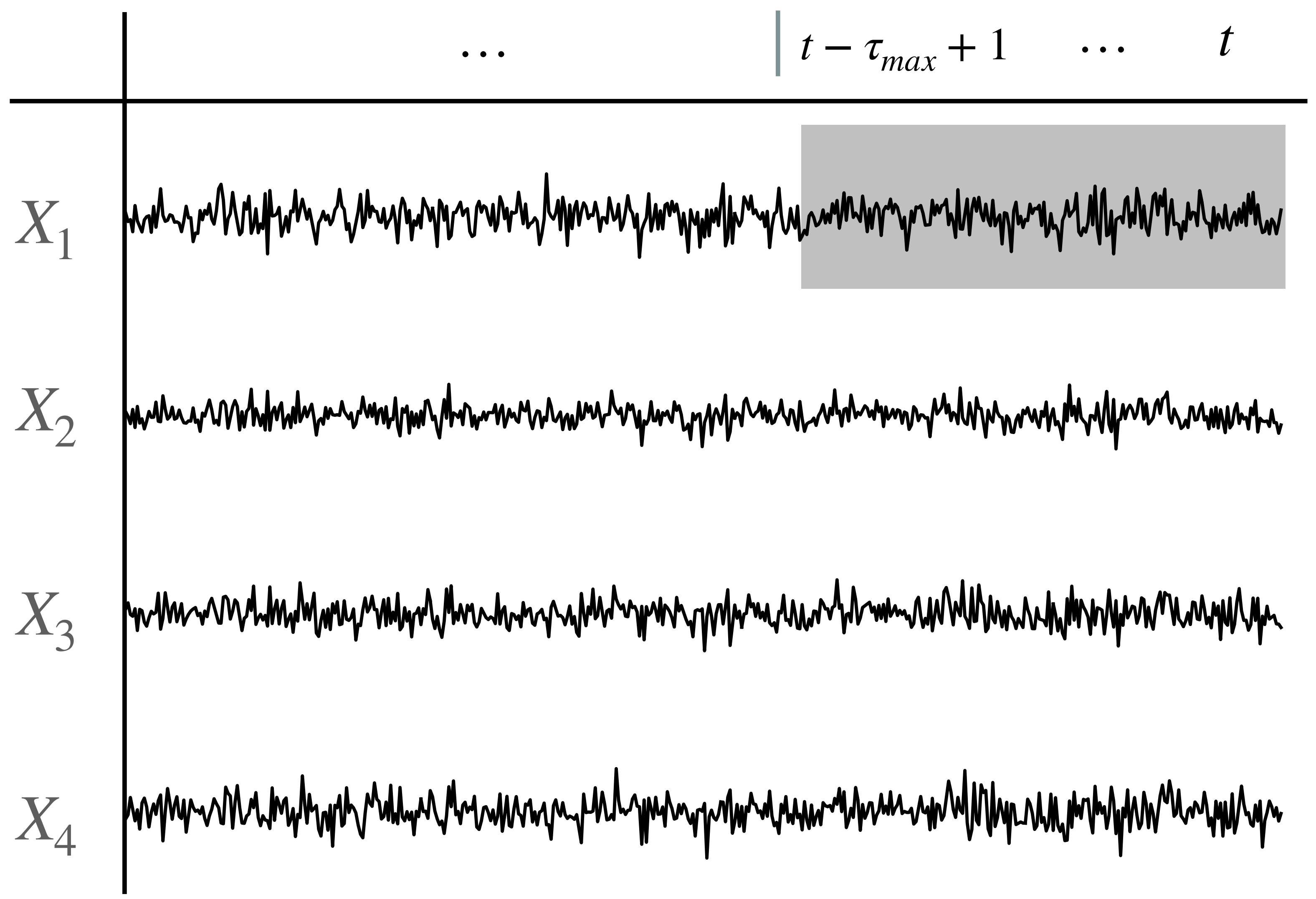


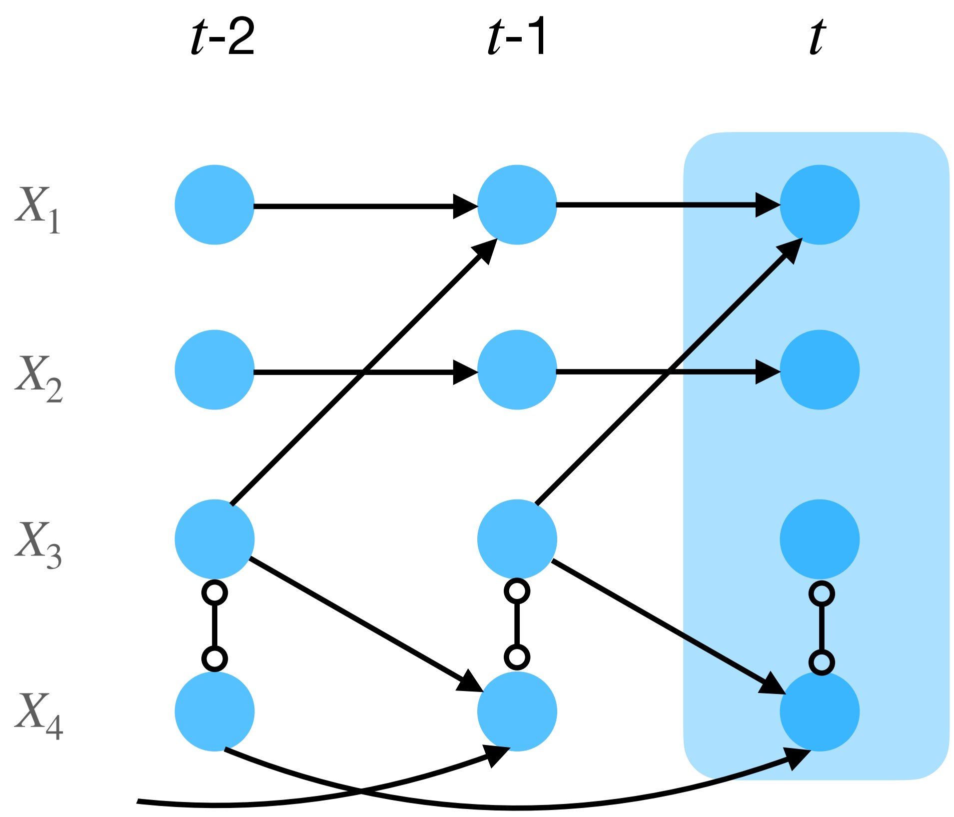
- Devise an algorithm to learn causal parents of the variables at time t, i.e. $Pa(X_i^t)$.
(rest of the graph by stationarity)



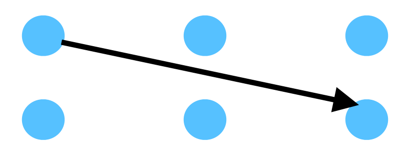
Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

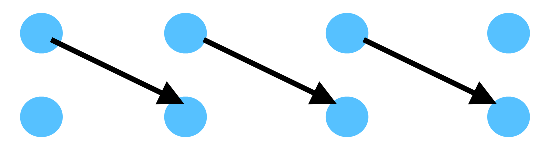




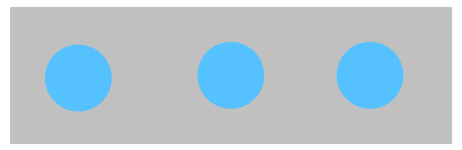
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



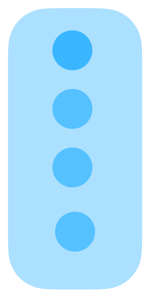
- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$



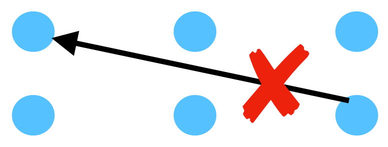
- Slide the time window $[t, t - \tau_{max} + 1]$ to generate samples for all X_i's



- Devise an algorithm to learn causal parents of the variables at time t, i.e. $Pa(X_i^t)$. (rest of the graph by stationarity)

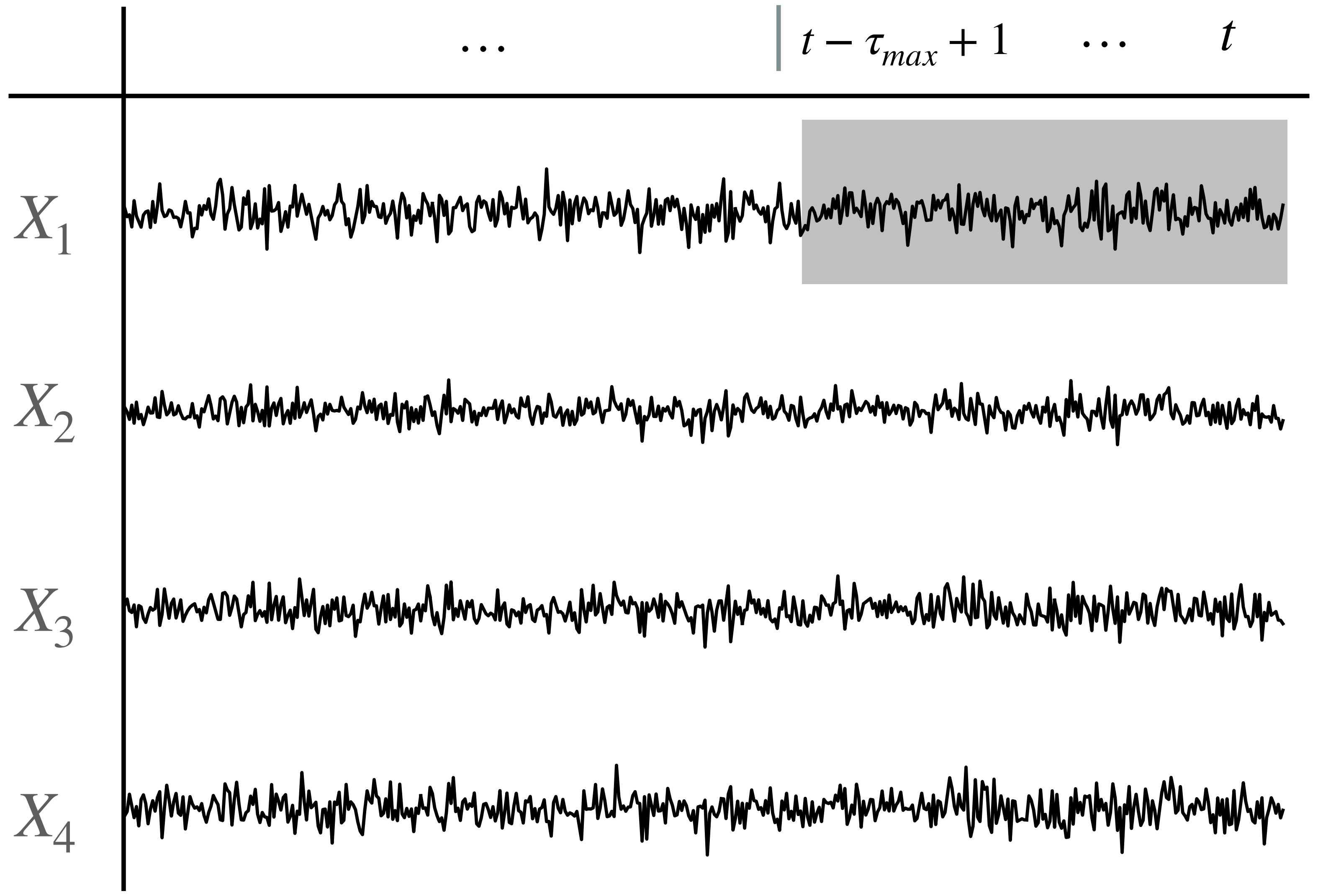


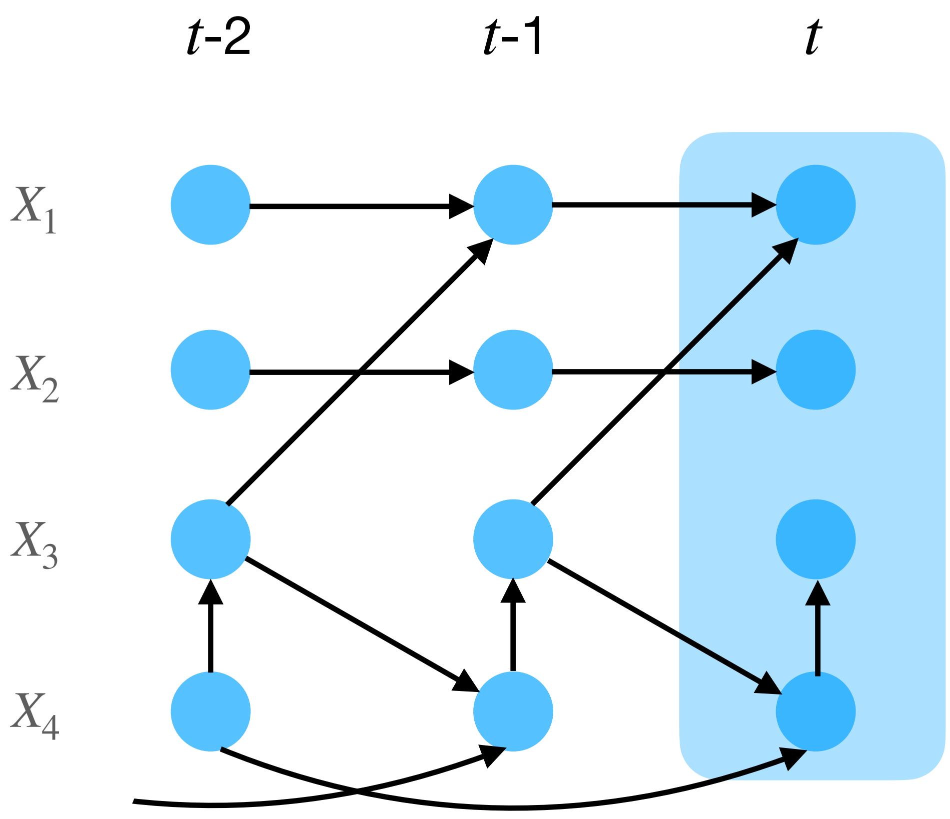
- Time order helps orient all but contemporaneous edges



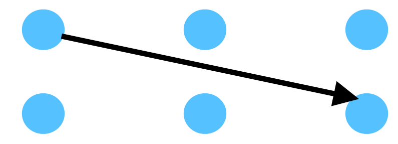
Causal Inference for Time Series

Basic tenets of time-series causal graph discovery

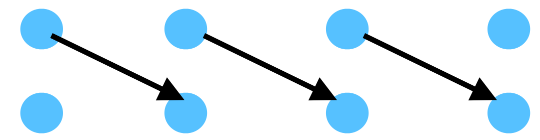




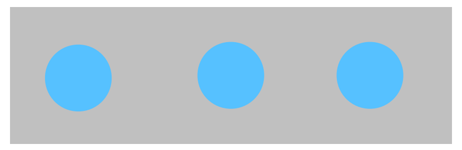
- Start with maximum time lag τ_{max} , provided by domain expert or intuition, eg. $\tau_{max} = 3$.



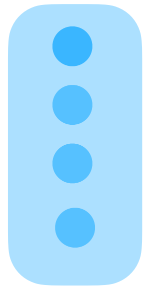
- Assuming causal stationarity, the graph of interest can be obtained by focusing on the window $[t, t - \tau_{max} + 1]$



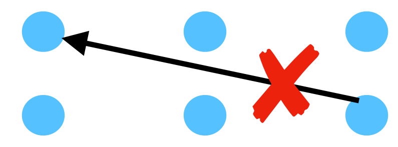
- Slide the time window $[t, t - \tau_{max} + 1]$ to generate samples for all X_i's



- Devise an algorithm to learn causal parents of the variables at time t, i.e. $Pa(X_i^t)$. (rest of the graph by stationarity)

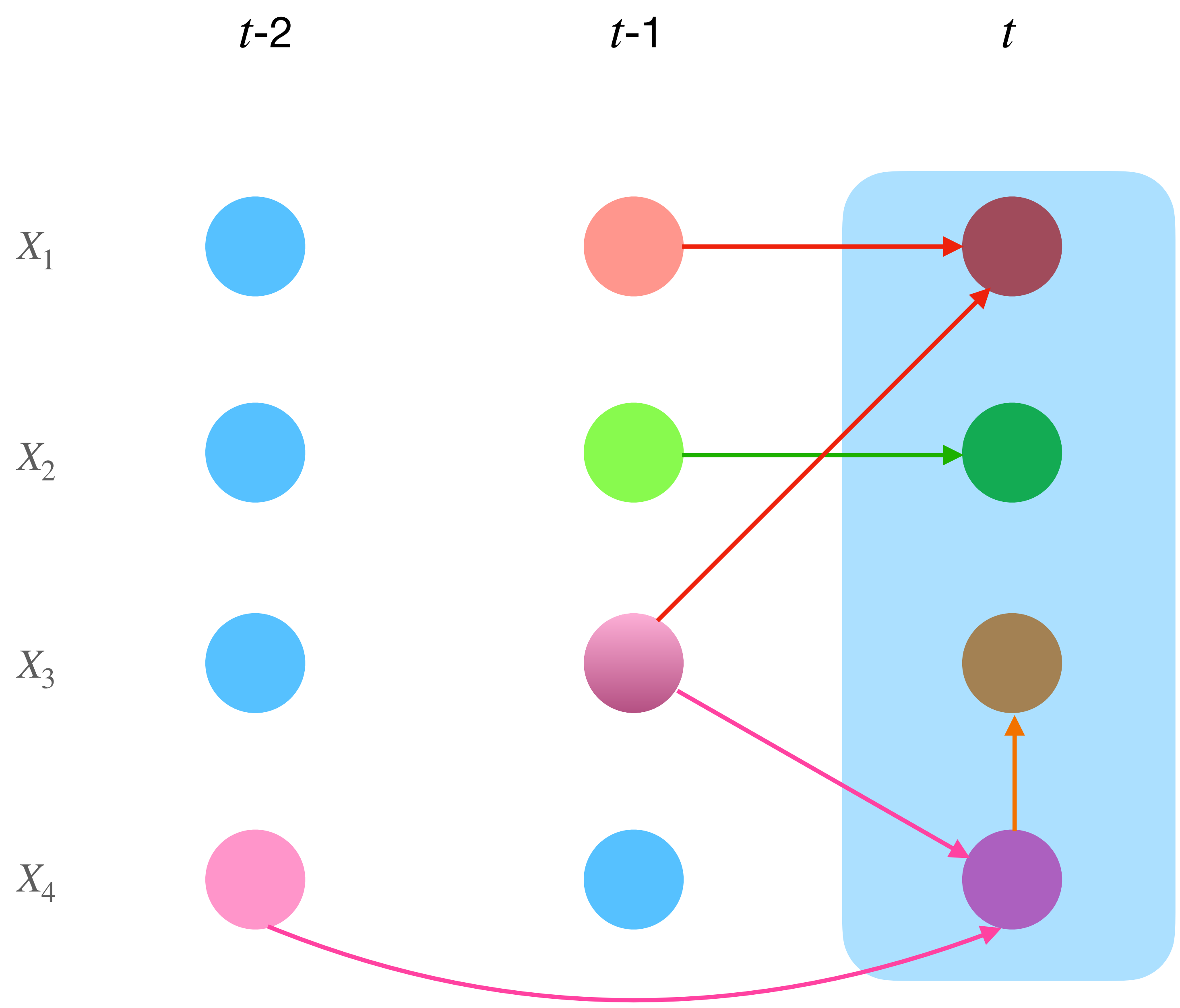


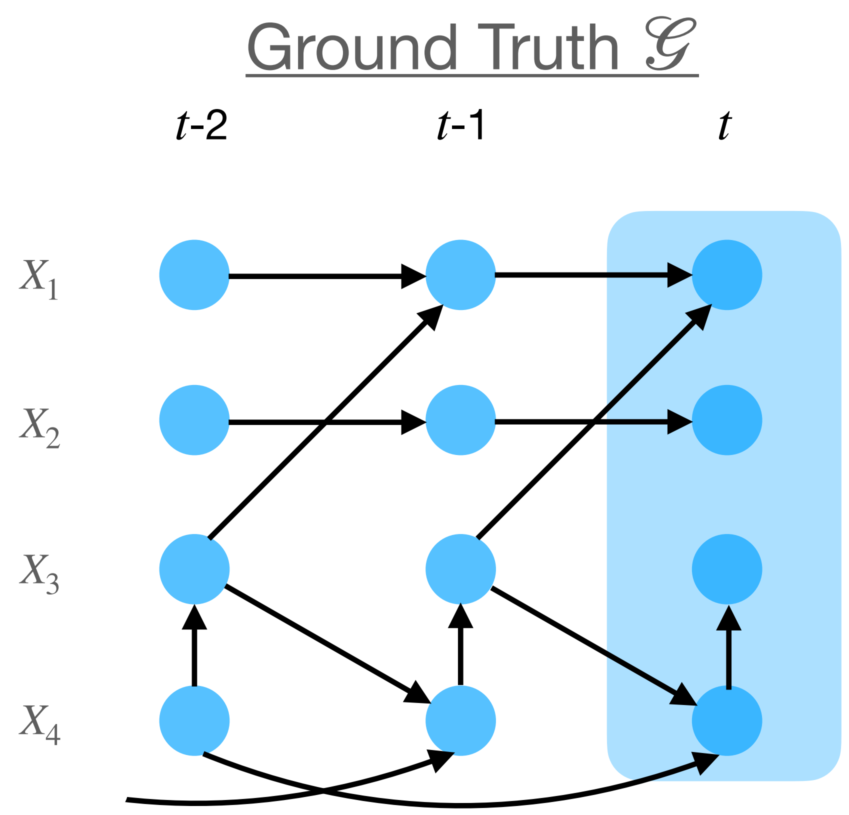
- Time order helps orient all but contemporaneous edges



Causal Inference for Time Series

PCMCI Algorithm





- **Problem**: Samples are autocorrelated so detection power of links is low and samples are non-iid so tests not well-calibrated.

- **Idea**: A momentary conditional independence (MCI) test, instead of a usual conditional independence tests makes samples iid.

  Require: Parents of all variables, then conduct CI tests.

Step 1: Find *superset* of lagged parents of $X_1, \ldots, X_4$ using PC algorithm (and tricks to avoid unnecessary deletion of links)

Step 2: Perform MCI test to discover true parents

"Detecting and quantifying causal associations in large nonlinear time series datasets", Runge et al, Science, 2019

Challenges of Time Series Causal Discovery for the Earth System

Challenges

Process:

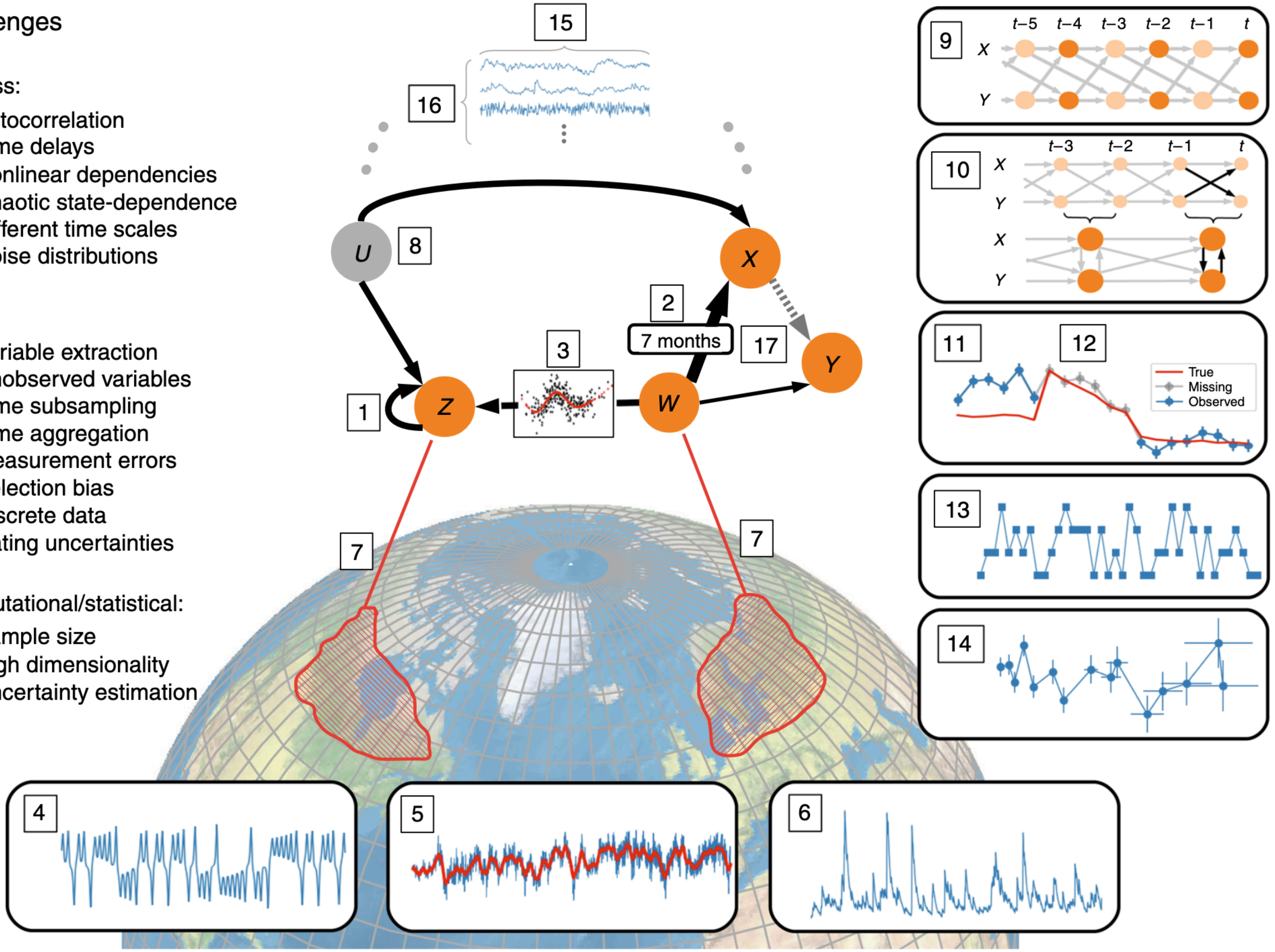
1 Autocorrelation
2 Time delays
3 Nonlinear dependencies
4 Chaotic state-dependence
5 Different time scales
6 Noise distributions

Data:

7 Variable extraction
8 Unobserved variables
9 Time subsampling
10 Time aggregation
11 Measurement errors
12 Selection bias
13 Discrete data
14 Dating uncertainties

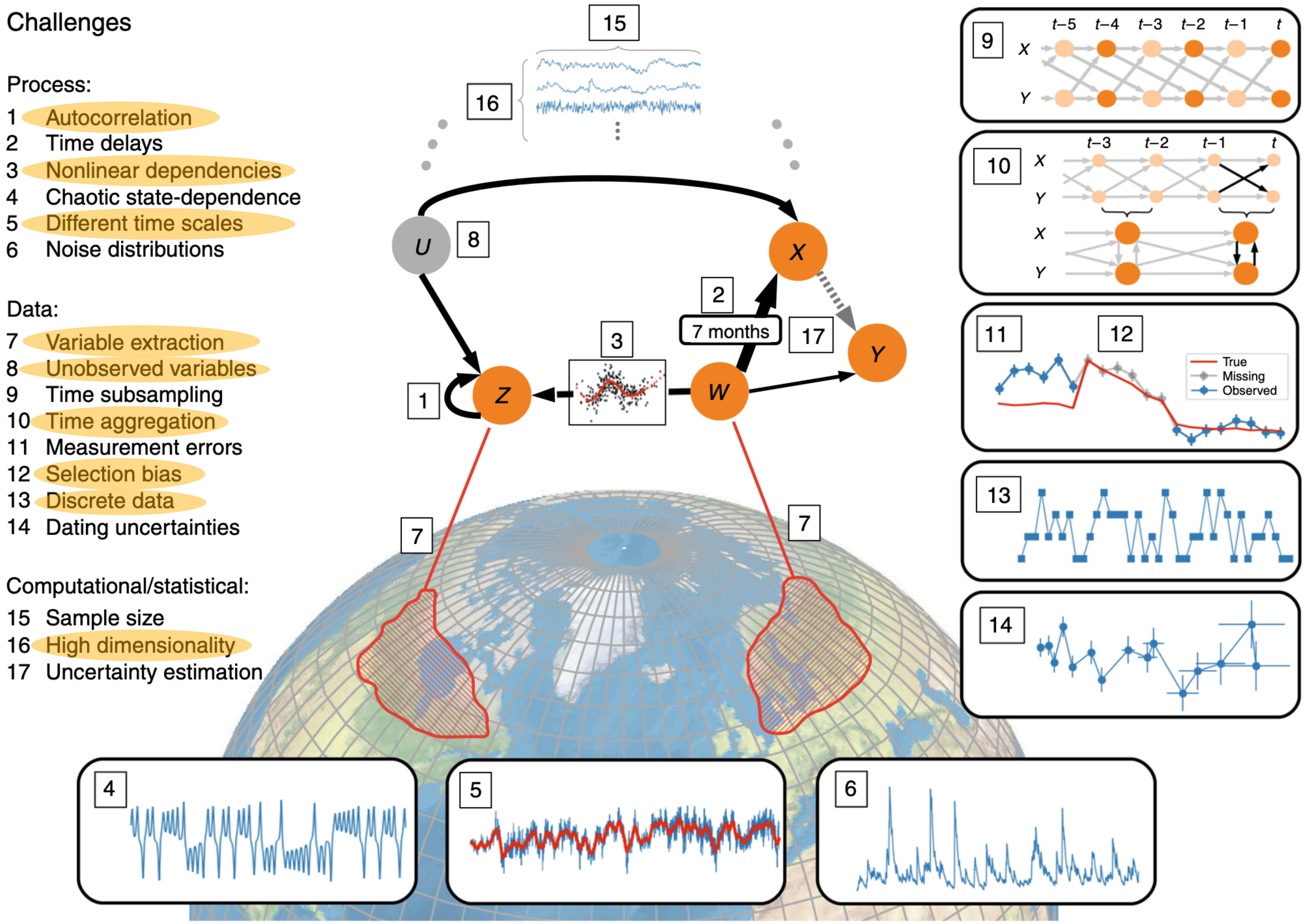
Computational/statistical:

15 Sample size
16 High dimensionality
17 Uncertainty estimation



Source: Runge et al, “Inferring Causation from time series in earth system sciences”, Nature Communications 2019

Challenges of Time Series Causal Discovery for the Earth System

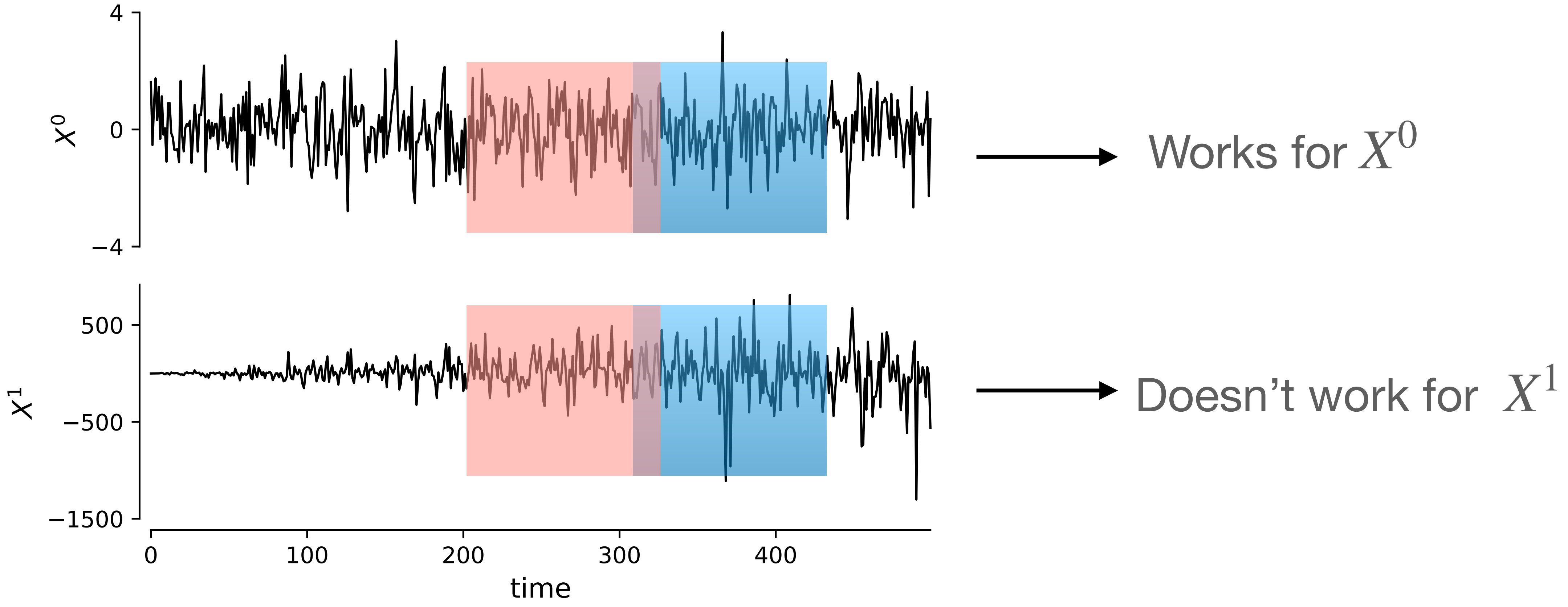


1,3: Detecting and quantifying causal associations in large nonlinear time series datasets, Runge et al, Science, 2019
5: Causal inference for temporal patterns, Domenic-Reiter et al, 2022
5: Causal discovery for time series from multiple datasets with latent contexts, Günther et al, UAI 2023
7: Identifying Linearly-Mixed Causal Representations from Multi-Node Interventions, Bing et al, Clear 2024
8: High-recall causal discovery for autocorrelated time series with latent confounders, Gerhardus et al, Neuritis 2020
10: Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets, Runge, UAI 2020
12: Endogenous Regimes and Causal Discovery, Rabel et al, in prep.
13: Non-parametric Conditional Independence Testing for Mixed Continuous-Categorical Variables: A Novel Method and Numerical Evaluation, Popescu et al, 2023
15: Increasing effect sizes of pairwise conditional independence tests between random vectors, Hochsprung et al, UAI 2023
16: Vector causal inference between two groups of variables, Wahl*, Ninad* et al, AAAI 2023
16: Spatiotemporal Causal Effect Estimation, Herman et al, EGU 2024

The Non-Stationarity Problem in ts-Causal Discovery

"Causal relationships that change over time"

- Roughly, stationarity implies that causal mechanisms do not change overtime
- Most causal discovery algorithms for time series assume *stationarity*
- Reason for assuming stationarity: In a sliding window approach to generating samples, we need to assume that the samples are identically distributed in order to make statistical inferences



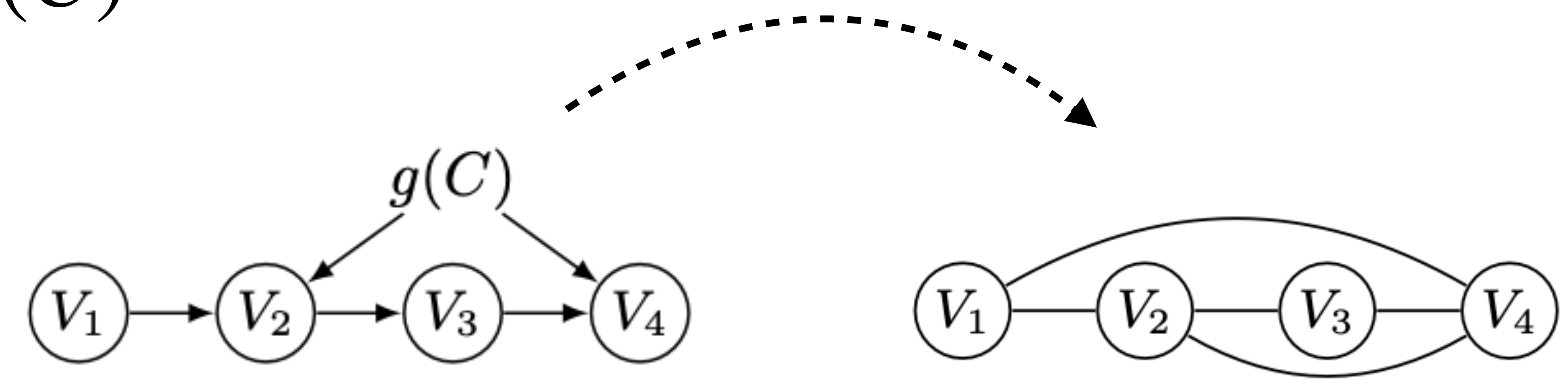
- However, in certain examples, the stationarity assumption becomes unrealistic. Eg. Seasonal variations in climate data.
  Moreover, relaxing this assumption might even aid in orienting links[1].

1: Huang*, Zhang* et al'20, JMLR

Dealing with Non-Stationarity: The CD-NOD Idea[1]

Leveraging changing probability distributions

- The time index is interpreted as a special random variable C
- **Pseudo causal sufficiency assumption** : Any latent confounder can be written as a smooth function of time, i.e., $g(C)$.
- Then, the source of non-stationarity is the causal variable $g(C)$
- **Problem**: Spurious edges:



Example from [1]: When $g(C)$ is latent, causal discovery may yield spurious edges

- **Solution**: 1. Consider the union of variables $V_i \cup C$
  2. Test $V_i \perp C$ to detect non-stationarity
  3. Test $V_i \perp V_j \,|\, \mathbf{V}_k \cup C$
  $\Rightarrow$ Yields the correct skeleton graph

1: Huang*, Zhang* et al'20, JMLR

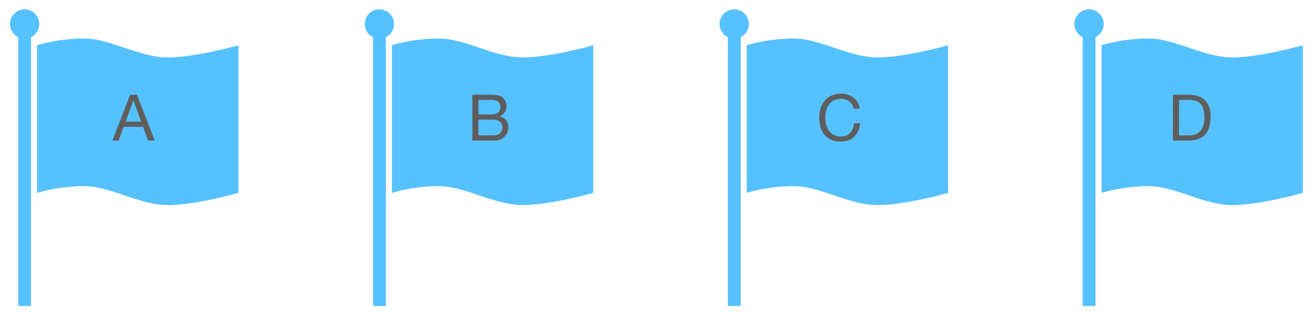
The Multiple Dataset Problem in Causal Discovery

"Data from multiple environments: boon or bane?"

- Data sets of the same variables can come from different *environments/domains/contexts*[1,2,3,4].

  Eg. Data for different individuals (in health, econometrics), data from different countries (in sociology, macroeconomics). . .

X :Time spent playing video games
Y : Aggressive behaviour



Example from [1]

- Given such 'heterogenous' data, we can make different kinds of causal queries:

  1. What is the causal structure *within* each data set?
  2. What is the causal structure *across* data sets?[1,3]
  3. Given the causal structure of one data set, what (if anything) can we say about the causal structure of another data set?[2]
  4. How can we leverage the *invariance* of certain causal relationships across data sets?[4]

1: Mooij et al'20, JMLR
2: Bareinboim'16, PNAS
3: Huang*,Zhang* et al'20, JMLR
4. Peters et al'16, JRSS. . .

Application of the Multiple Dataset Problem

A River Catchment Example

- A *catchment* is an area of land where water collects when it rains, often bounded by hills.
- The characteristics of catchments are highly heterogeneous (area, slope, etc.).
  Catchment behaviour also depends on regional climate and other meteorological variables.
- Can we make causal inferences about the causal drivers of catchment behaviour?

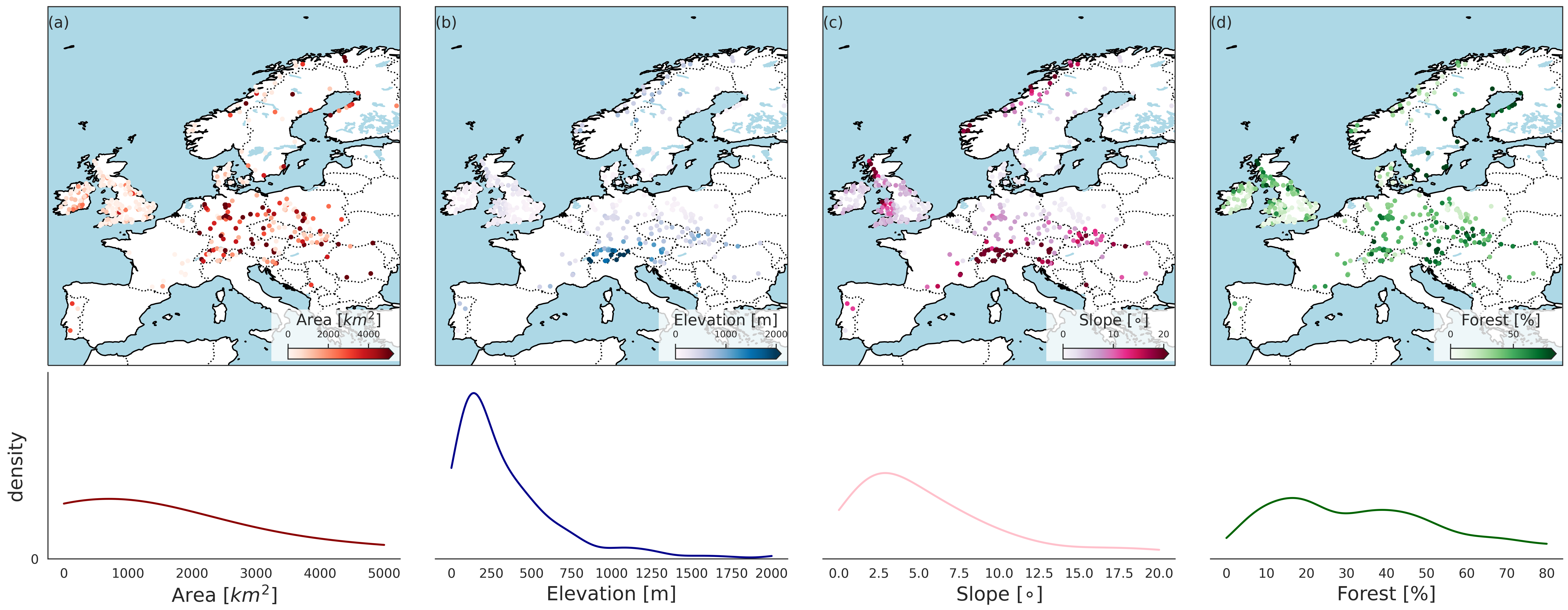


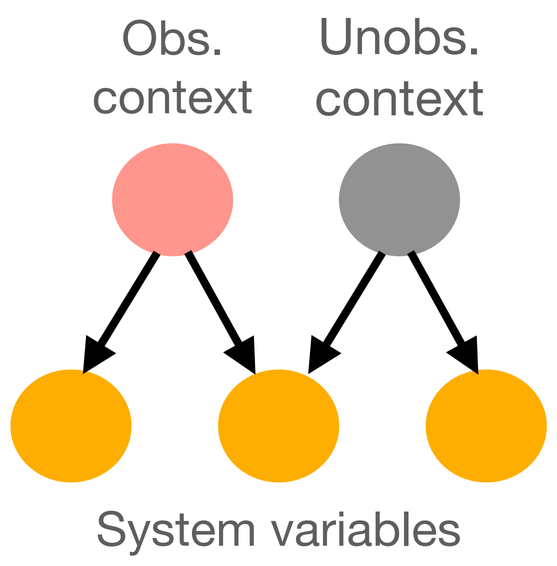
Figure from [1]: An overview of the European catchment characteristics (eg. Area, elevation, etc.)

1: “Clustering of causal graphs to explore drivers of river discharge”, Günther et al ’23, Environmental Data Science

J(oint)-PCMCI+[1]
Schematic Idea:

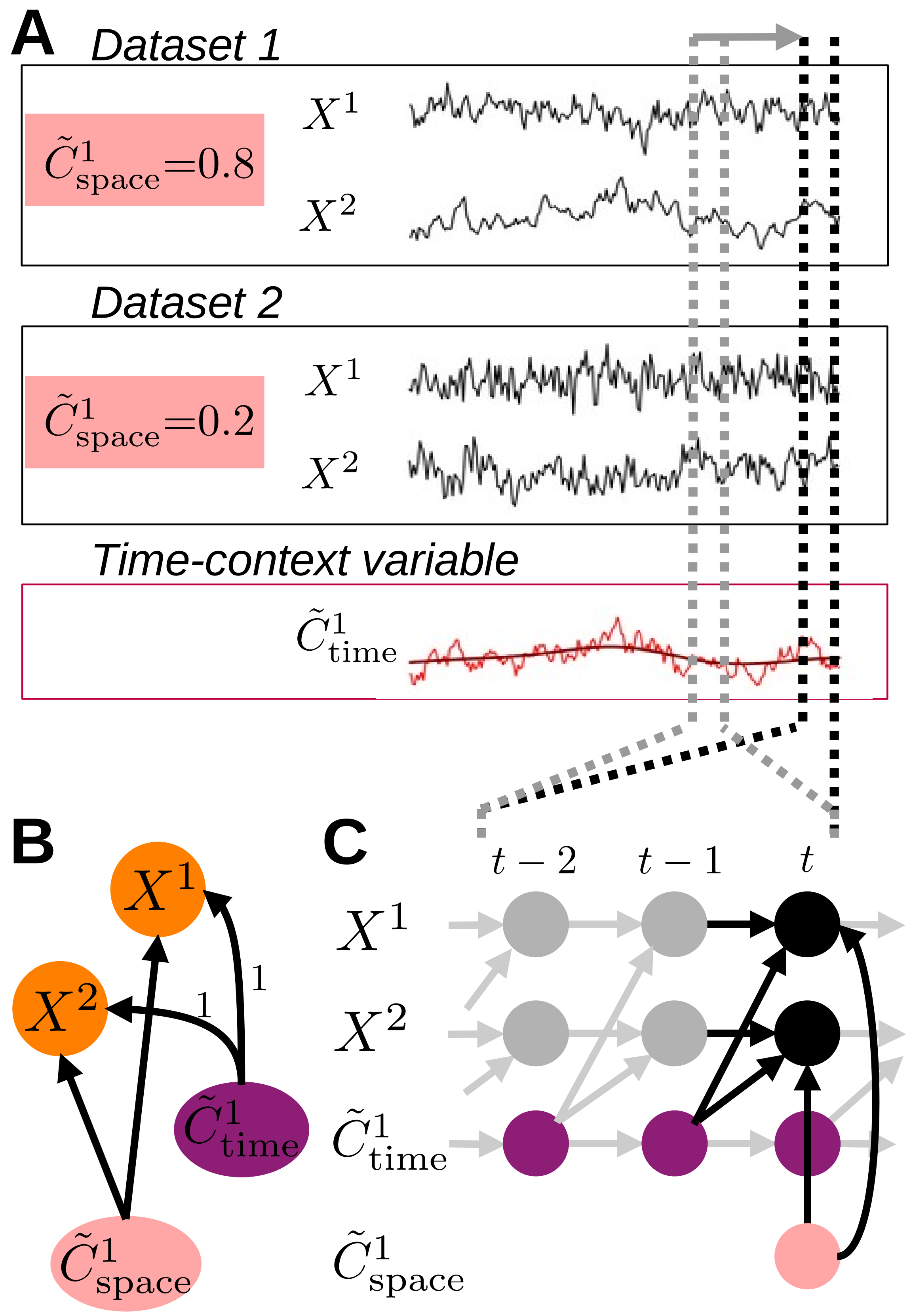
- Introduce ‘spatial’ and ‘temporal’ context variables (Space indicates the data set label dimension, not necessarily physical space)

- Generalise DAGs on system variables to the case with observed as well as unobserved context variables



- For the unobserved contexts introduce a ‘dummy’ variable to keep track of the time index or the data set label.
  Spatial contexts ⇒ Space dummy
  Temporal contexts ⇒ Time dummy

- Apply the fixed effect regression idea (from econometrics) to de-confound system variables (while taking care of faithfulness violations due to determinism)



$$\mathbf{X}_t^d := \mathbf{f}(Pa_X(\mathbf{X}_t^d), Pa_{\tilde{C}_{\text{time}}}(\mathbf{X}_t^d), Pa_{\tilde{C}_{\text{space}}}(\mathbf{X}_t^d), \eta_t^d)$$
$$\tilde{\mathbf{C}}_{\text{time},t} := \mathbf{g}(Pa_{\tilde{C}_{\text{time}}}(\tilde{\mathbf{C}}_{\text{time},t}), \eta_{\text{time},t})$$
$$\tilde{\mathbf{C}}^d_{\text{space}} := \mathbf{h}(Pa_{\tilde{C}_{\text{space}}}(\tilde{\mathbf{C}}_{\text{space}}), \eta^d_{\text{space}})$$

1: “Causal discovery for time series from multiple datasets with latent contexts”, Günther, U.N., Runge’ 23, UAI

Causal Effect Estimation

"How *much* does X cause Y?"

- Causal effect of X on Y is defined as $CE(X \to Y) = \frac{\partial}{\partial x} E(Y \,|\, do(X = x))$, where $E(X) := \int x \cdot p(x)\, dx$.

  Example:

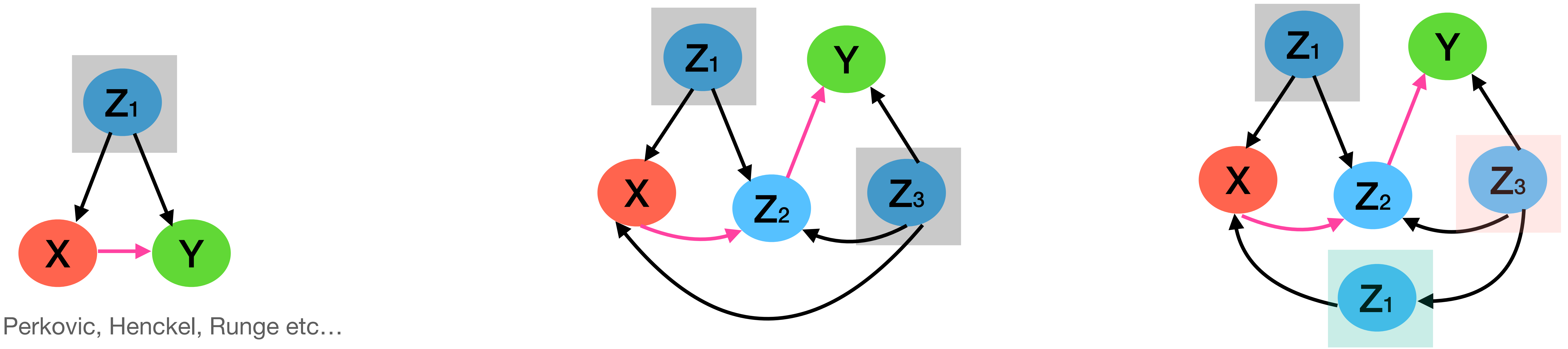
  $X := \eta_X$

  $Y := b \cdot X + \eta_Y \quad \Longrightarrow CE(X \to Y) = b$ (Linear causal effects can be read off from structural causal model)

- Given the causal graph, how to determine the causal effect by statistical methods (i.e. without interventions)?

$\Rightarrow$ Adjustment Set[1] $\mathbf{Z}$ for $CE(X \to Y)$ is defined by $P(Y \,|\, do(X)) = \int P(Y \,|\, X, \mathbf{Z}) P(\mathbf{Z})\, d\mathbf{Z}$

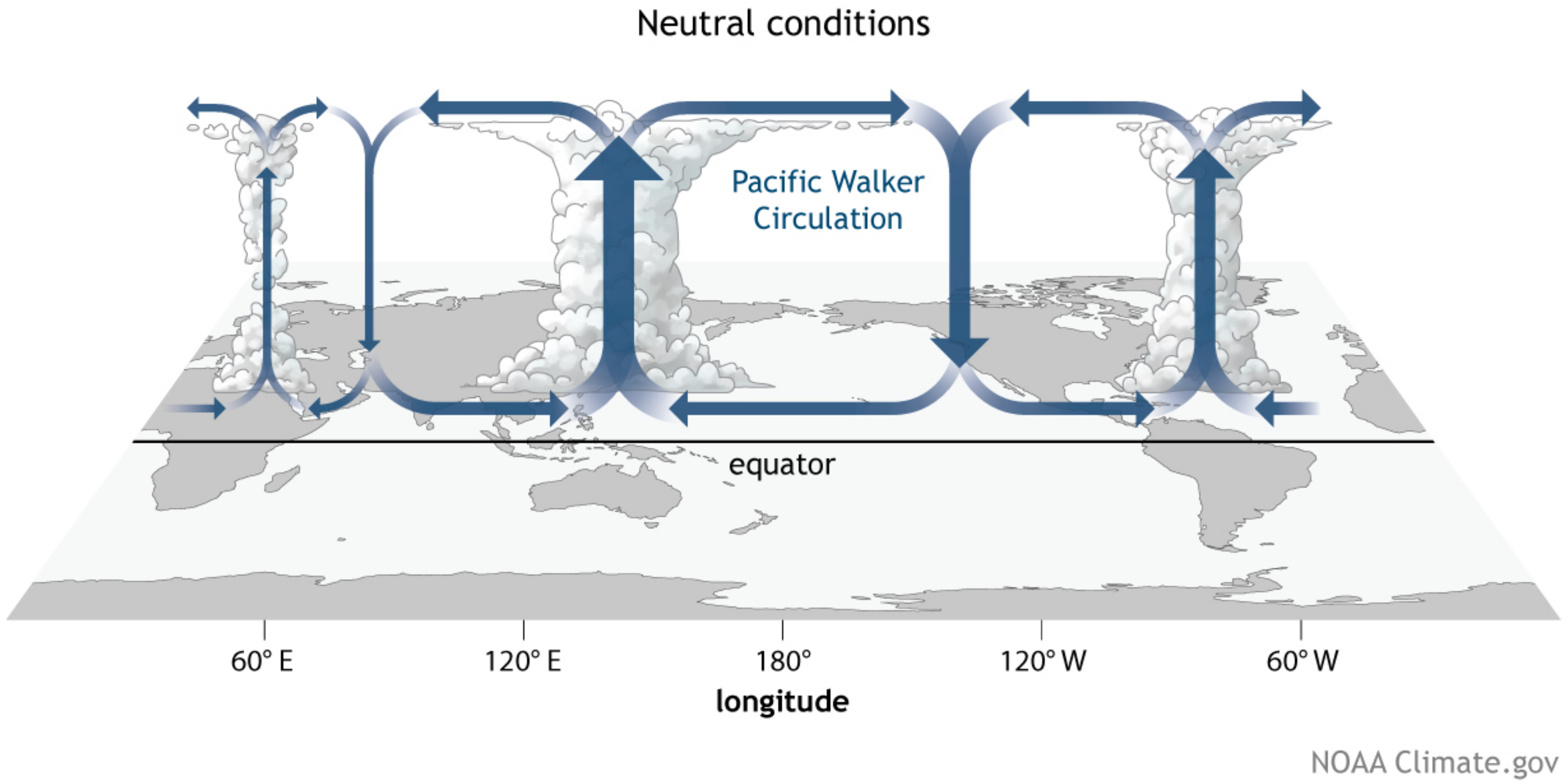
- Many adjustment sets satisfy the defining property (i.e. are unbiased), but which is the optimal set (i.e. has least variance)? $\Longrightarrow$ Optimal adjustment set theory[2]

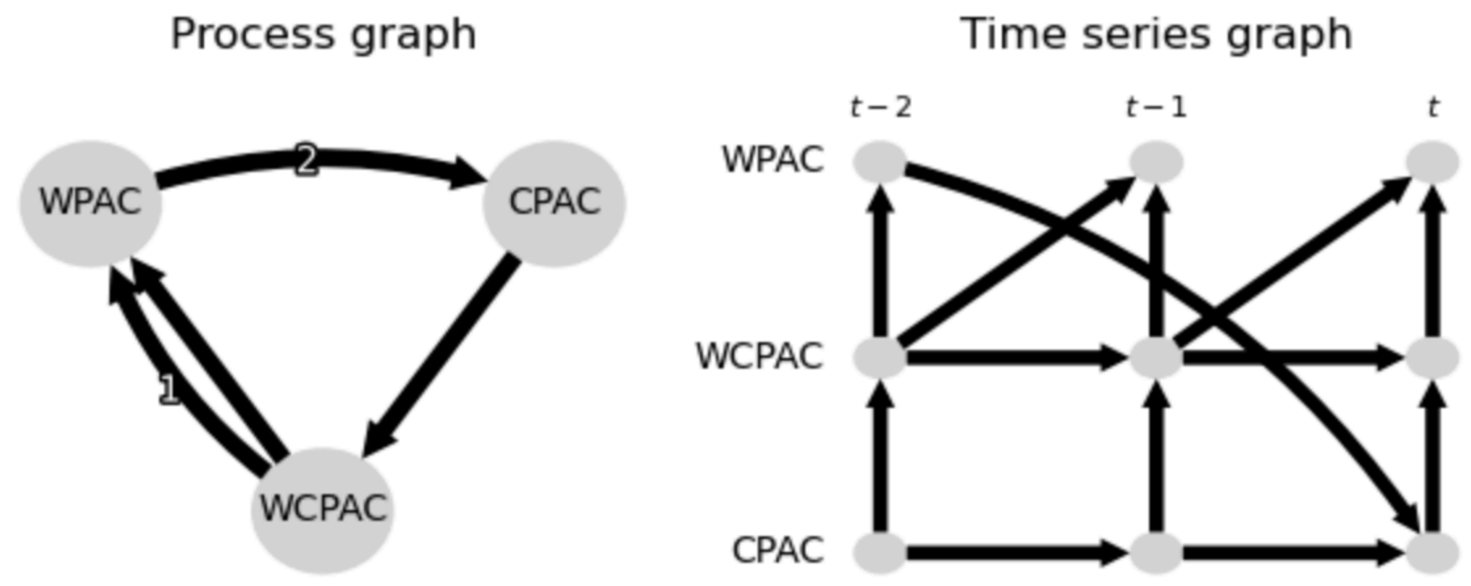


1: Works of Pearl, Shpitser etc…
2. Works of Maathuis, Colombo, Perkovic, Henckel, Runge etc…

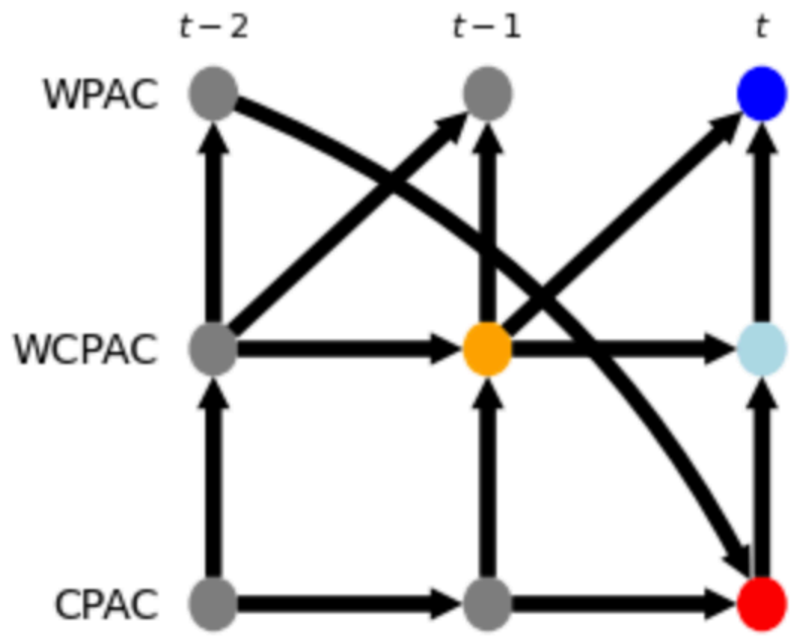
Causal Effect Estimation in the Climate: Walker Circulation

- Walker circulation is a model of air flow in the lower atmosphere in the tropics
- We focus on clock-wise circulation of falling air masses in the Central Pacific (CPAC), westward surface tradewinds in the Western-central pacific (WCPAC), and rising air masses in the western Pacific (WPAC).






```
X = [('CPAC', 0)] -----> Y = [('WPAC', 0)]
Oset =  [('WCPAC', -1)]
```

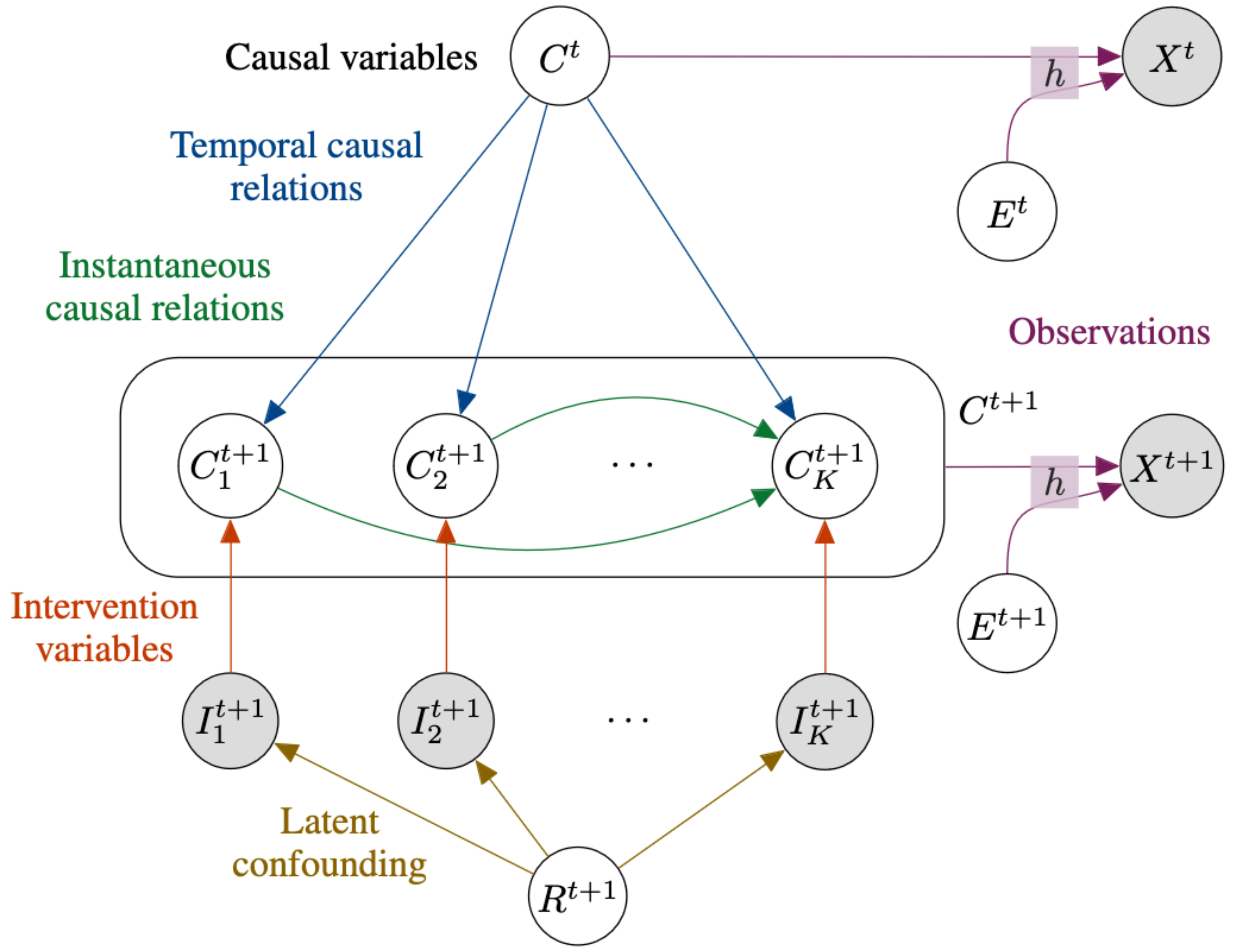


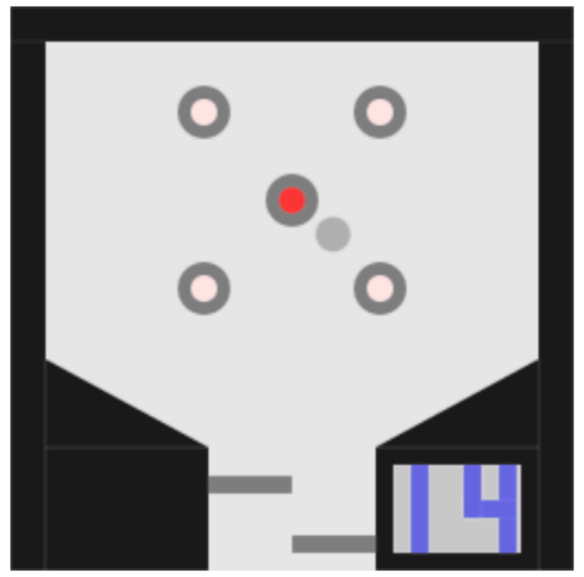
Causal effect of the CPAC on WPAC can be calculated given the optimal adjustment set

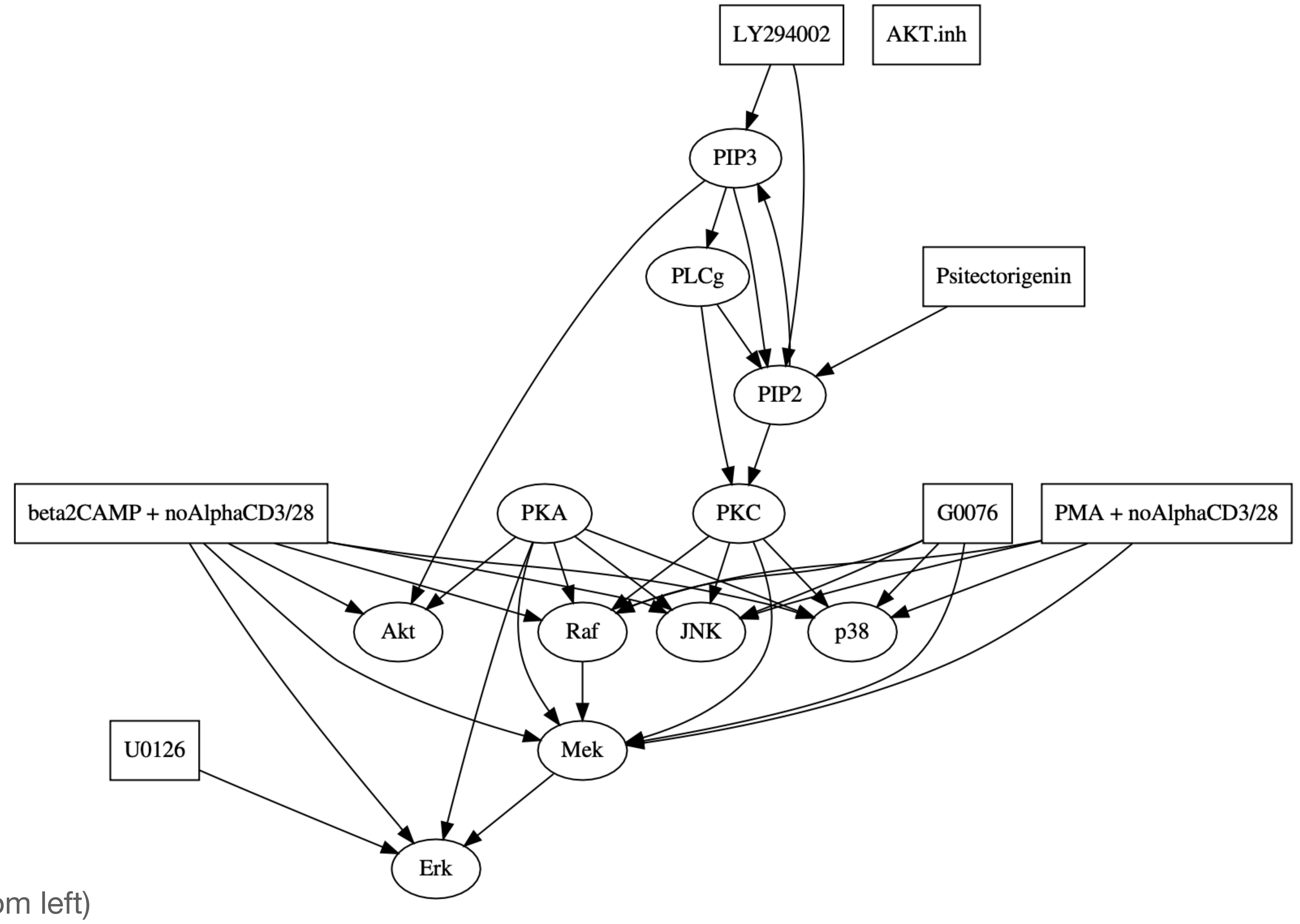
- Example source: Runge et al, "Causal Inference for Time Series", Nature Communications 2023
- Jupyter tutorial available on `github.com/jakobrunge/tigramite`

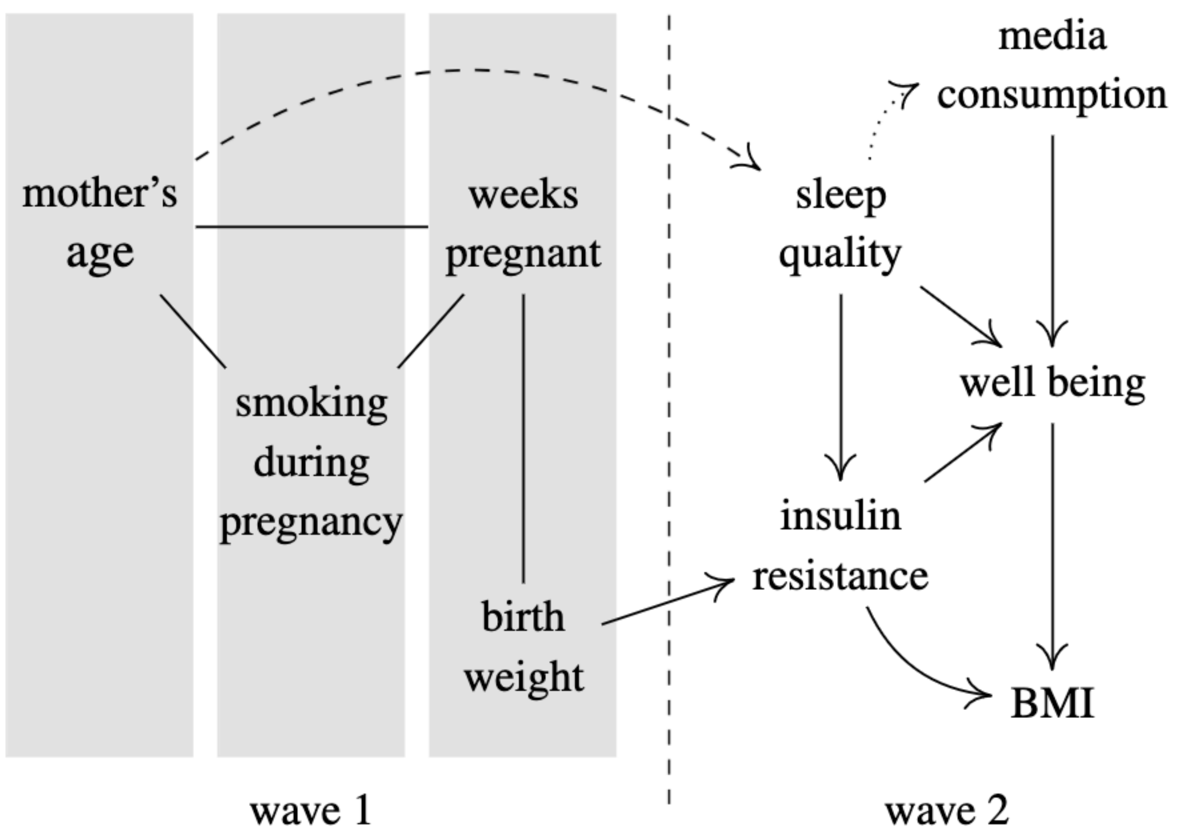
Other Applications of Time Series Causal Discovery

- Gene knockout experiments
- Flow cytometry data
- Cohort studies in epidemiology
- Representation learning in robotics
- Financial markets









(Clockwise from bottom left)
Image: 'Causal protein-signaling networks derived from multiparameter single-cell data', Sachs et al, Science
Image: 'Do we Become Wiser With Time? On causal equivalence with tiered background knowledge', Bang et al, UAI 2023
Image: 'Causal Representation Learning for Instantaneous and Temporal Effects', Lippe et al, ICLR 2023

Thank you!

Urmi.ninad@tu-berlin.de
https://climateinformaticslab.com/about/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 101003469.