Causal Inference and Causal Discovery for Studying Teleconnection Pathways

Marlene Kretschmer[1]

Ted Shepherd[1], Elena Saggioro[1], Alberto Arribas[2], Rachel Prudden[2], Niall Robinson[2], Sam Adams[2]

[1]University of Reading, [2]Met Office Informatics Lab

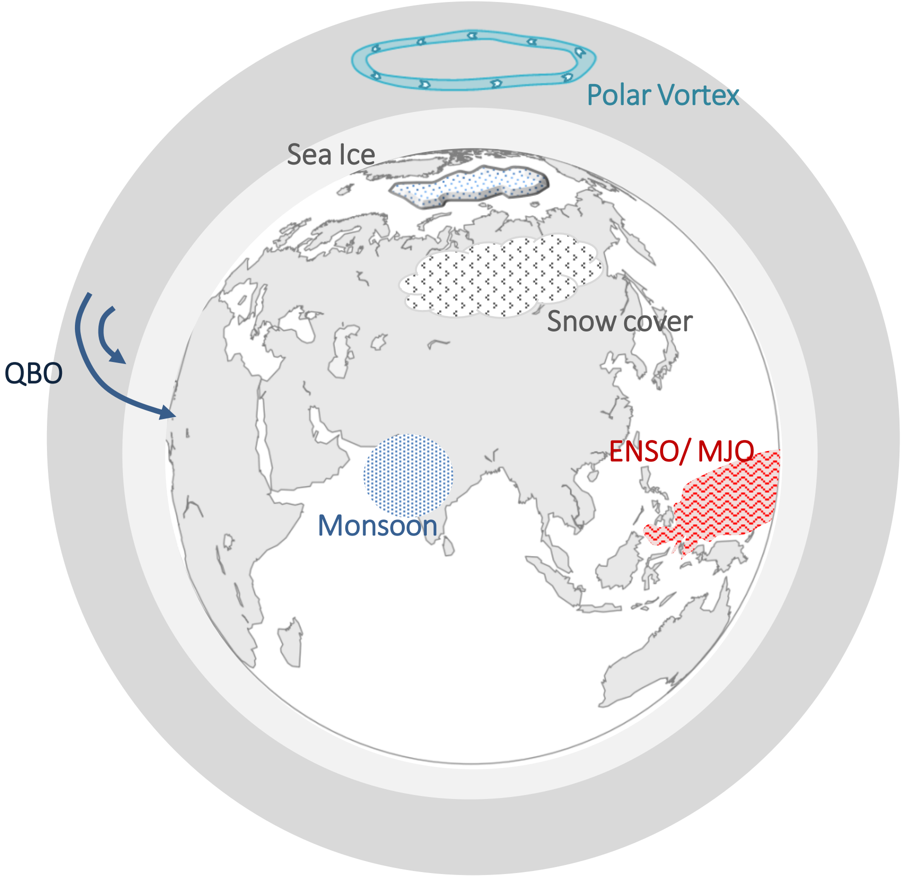
m.j.a.kretschmer@reading.ac.uk @Marlene_Climate

INTRODUCTION

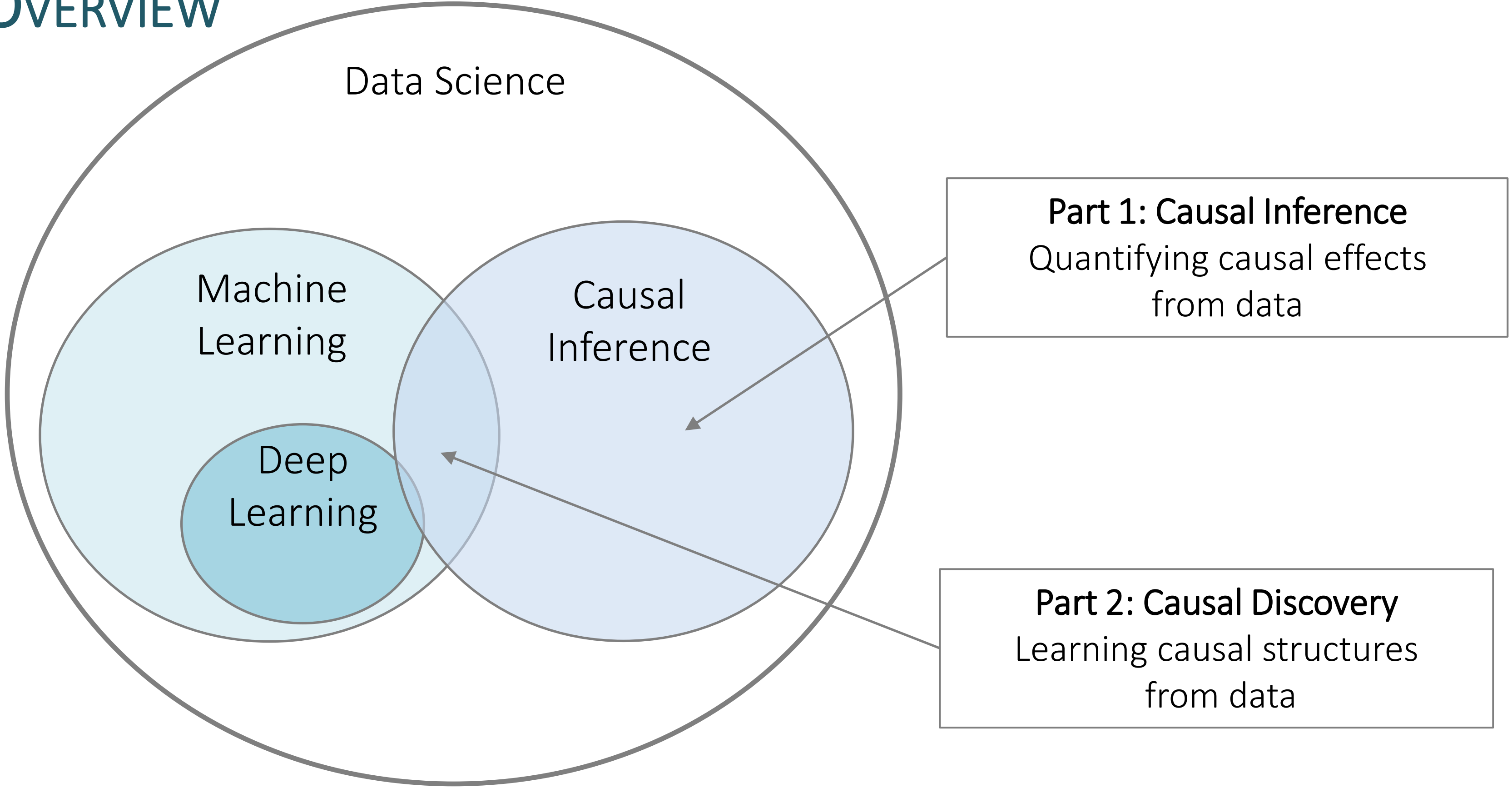
- Motivation:
  Improved understanding of teleconnections is key to reduce uncertainties about regional weather and climate predictions

- Typical questions:
  How much does ENSO contribute to temperature variability in region A?
  Which processes drive precipitation in region B?

- Challenge:
  Extracting the (causal) information from data



OVERVIEW


Data Science
Machine Learning
Deep Learning
Causal Inference
Part 1: Causal Inference
Quantifying causal effects from data
Part 2: Causal Discovery
Learning causal structures from data

TELECONNECTIONS

AMS Definition

A significant [...] correlation in the fluctuations of a field at widely separated points.

[...] the name refers to the fact that such correlations suggest that information is propagating between the distant points through the atmosphere.

Large gap between our physical understanding (= causal) and our statistical description (= correlational) of teleconnections.

CAUSALITY IN STATISTICS

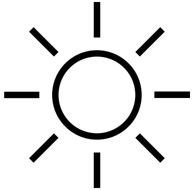
X causes Y, if intervening in X
(while keeping everything else fixed)
changes Y

$P(Y \mid \mathit{do}(X) = x) \neq P(Y)$

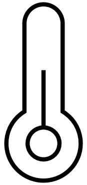
Causal Inference:
Predict the effect of an intervention from observed data
(without actually doing the intervention)

Example

X: Temperature



Y: Thermometer



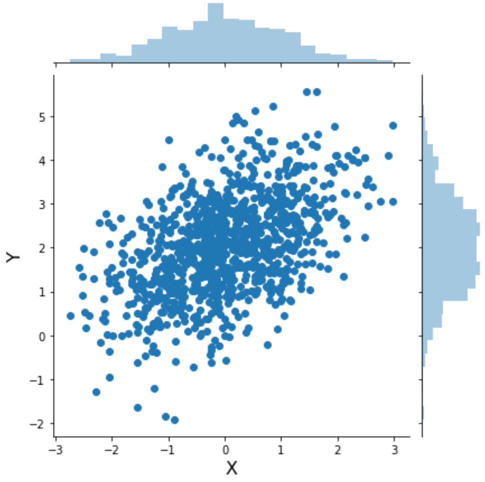
Changes in X will cause changes in Y:
$P(Y) \neq P(Y \mid \mathit{do}(X) = x)$

Intervening in Y will not change X:
$P(X) = P(X \mid \mathit{do}(Y) = y)$

Toy Example

Question: What happens to Y if X is changed to x =1?

Answer only possible if we have knowledge about the underlying causal mechanisms that generated the data



We can compute the observational cond. distribution
$P(Y \mid X = 1)$

We want the interventional cond. distribution
$P(Y \mid \mathbf{do(X)} = 1)$

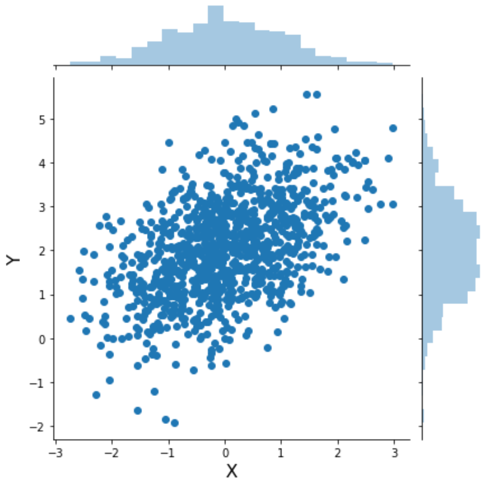
But these are usually not the same
$P(Temp \mid Thermo = 1) \neq P(Temp \mid do(Thermo) = 1)$

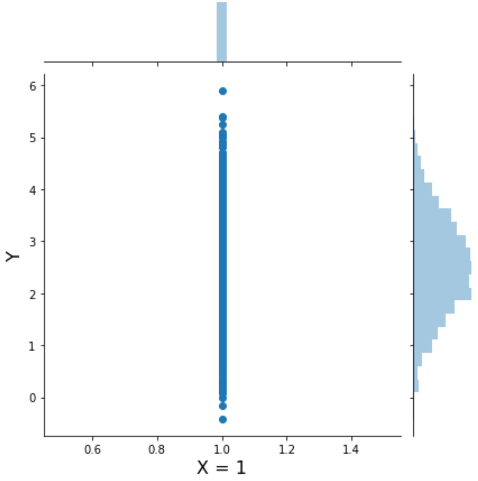
Source: https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

TOY EXAMPLE

Question: What happens to Y if X is changed to x =1?

Expert knowledge	Estimate from data	Answer causal questions
$X \rightarrow Y$	$X = \varepsilon x$ $Y = \mathbf{0.5} * X + 2 + \varepsilon y$	$P(Y \mid do(X) = 1) = P(Y \mid X = 1)$

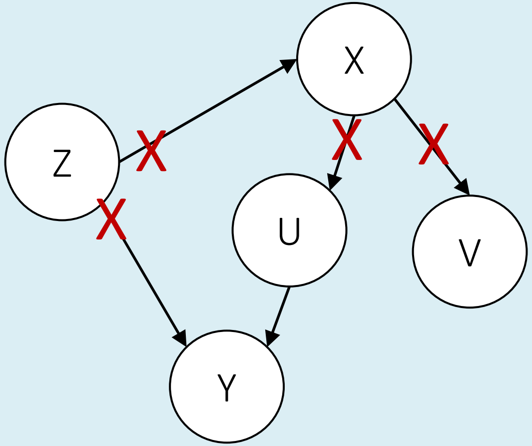




To answer causal questions, we need causal knowledge (i.e. hypotheses)

CAUSAL INFERENCE (= "KNOWLEDGE GUIDED STATISTICS")

Use knowledge to set a (plausible) causal model



Draw logical implications and test if data support them

U and V must be independent conditioned on X

Corr(U, V | X) = 0

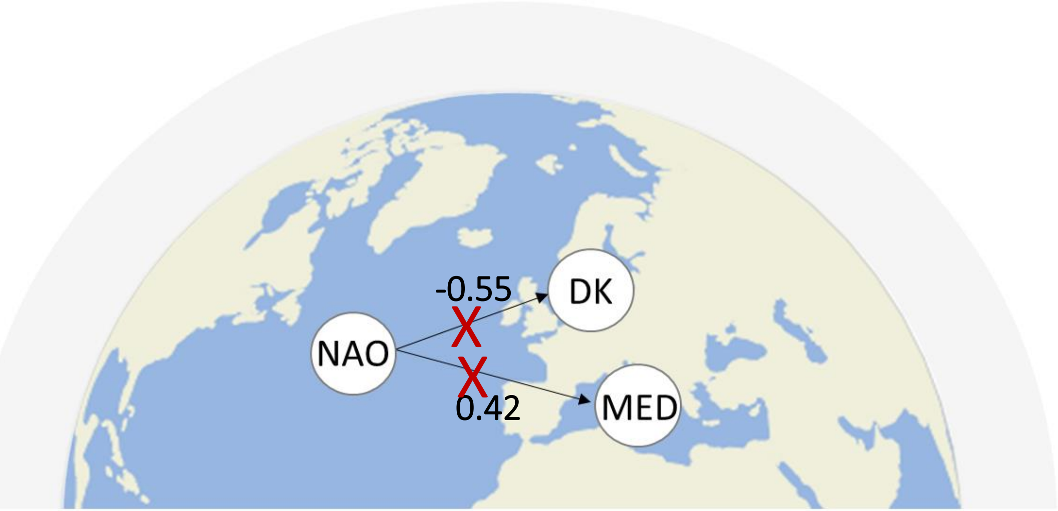
Use CI rules to estimate causal effects

To estimate the causal effect from X to Y, need to control for Z

P(Y |do(X)) = P(Y | X, Z)

Pearl (2000), Pearl & Mackenzie (2018), Pearl (2013)

EXAMPLE 1: COMMON DRIVER



Precipitation in Denmark and the Mediterranean are significantly correlated

Corr(DK, MED) = -**0.25**

DK and MED are independent conditional on NAO

$\text{DK} = 0.01\ \text{MED}\ \mathbf{-\ 0.55}\ \text{NAO} + \varepsilon$

$\text{MED} = 0.0\ \text{DK} + \mathbf{0.42}\ \text{NAO} + \varepsilon$

The causal effect explains the association

$-0.55 * 0.42 \approx -0.25$

(JJA mean, NCEP)

EXAMPLE 2: MEDIATOR

What is the effect of ENSO on California precipitation?

CA = **0.0** ENSO + 0.81 Jet + ε

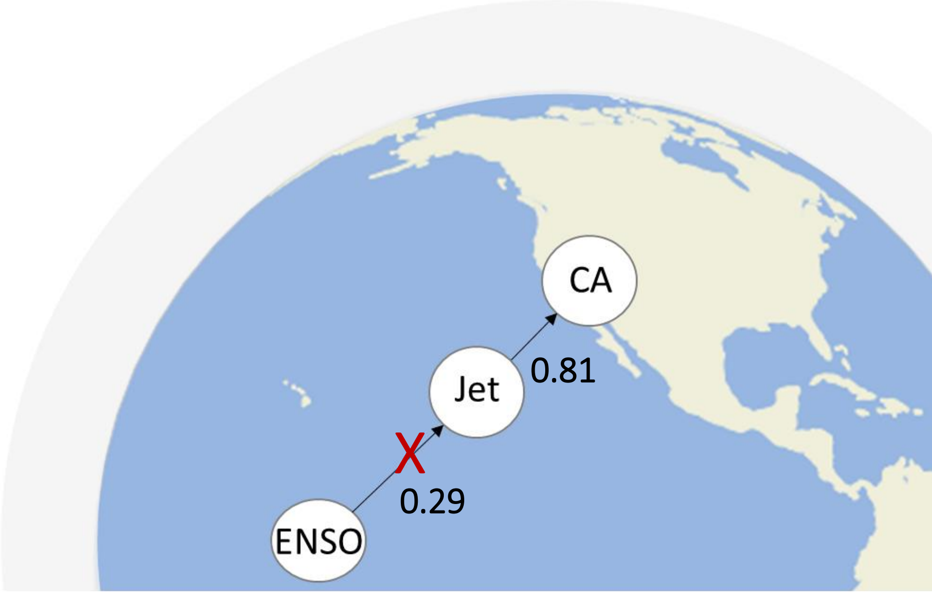
Correct way:

CA = **0.24** ENSO + ε

Or via product along pathway:

Jet = **0.29** ENSO + ε   CA = **0.81** Jet + ε

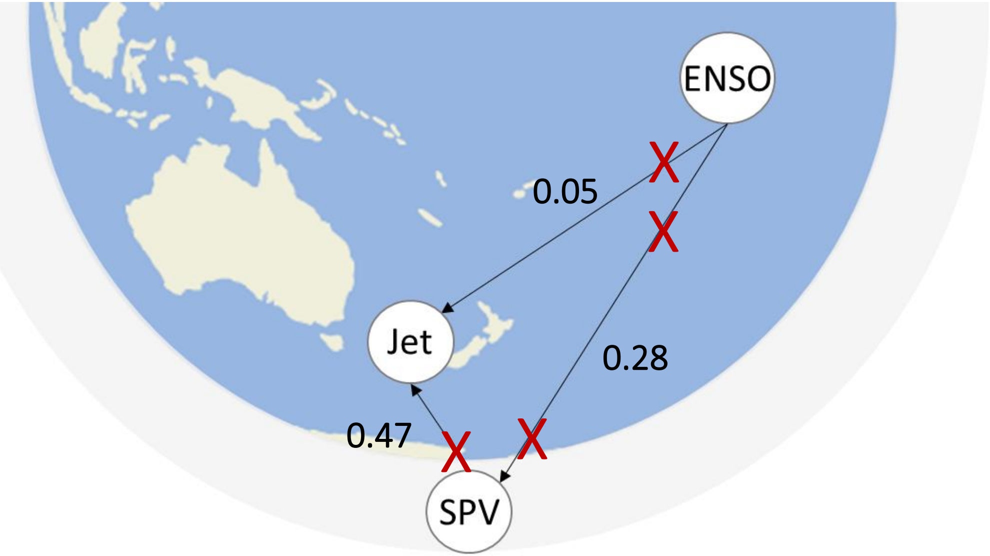
0.29 * 0.81 = 0.24



(DJF mean, NCEP)

Kretschmer et al. (in preparation for *BAMS*)

Example 3: Indirect + Direct Effects



(OND mean, NCEP)

<u>Total effect</u>

Jet = **0.18** ENSO + ε

<u>Direct (tropospheric) pathway:</u>

Jet = **0.05** ENSO + a SPV + ε

<u>Indirect (stratospheric) pathway :</u>

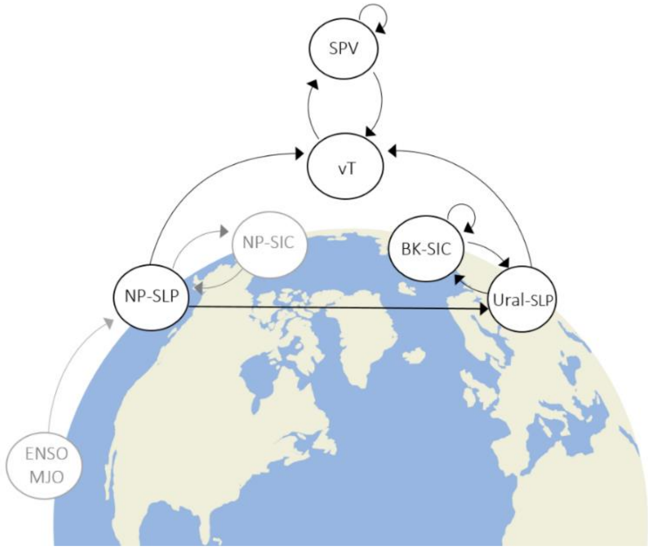
SPV = 0.28 ENSO + ε

Jet = 0.47 SPV + a ENSO + ε

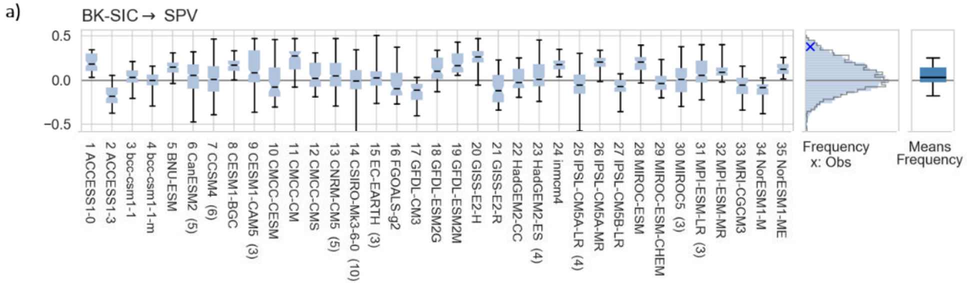
0.13 = 0.28 * 0.47

OPPORTUNITIES & APPLICATIONS

- Easy, transparent and traceable approach to quantify teleconnection pathways
- Process-based model evaluation and bias adjustments



What is the role of Barents and Kara sea ice loss in future polar vortex changes in CMIP5 models?

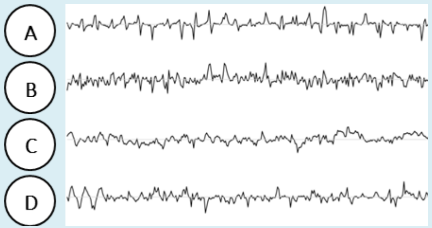


Opportunities & Applications

- Easy, transparent and traceable approach to quantification teleconnection pathways
- Process-based model evaluation and bias adjustments (e.g. Kretschmer et al. (2020))
- Understanding signal to noise issues
- Robust emergent constraints
- Physically self-consistent storylines of regional climate change (e.g. Shepherd.(2018))
- Climate modelling (e.g. Hirt et al.(2020))
- Framework to connect the AI and the climate dynamics communities

PART 2: CAUSAL DISCOVERY

Input: Time-series data



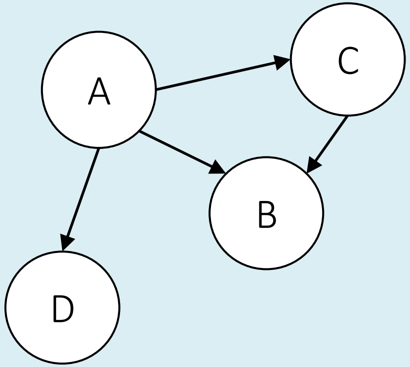
Causal Discovery

<u>PCMCI Algorithm</u>

corr($A_{t-\tau}$, B_t | *Iterate through combinations of conditions*)

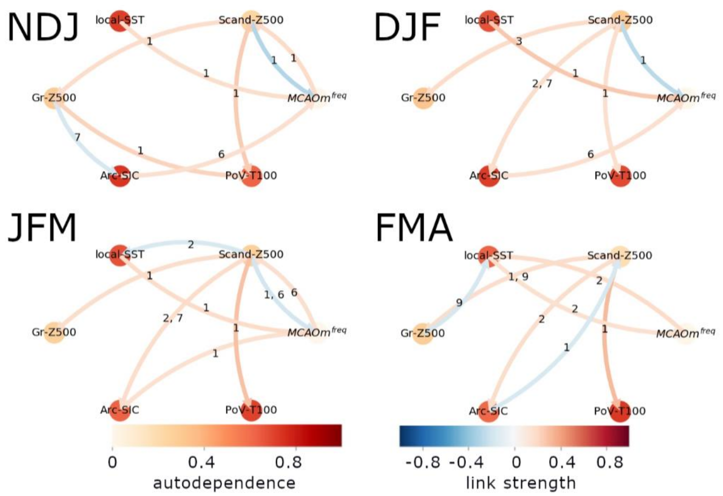
Can deal with auto-correlation, regime-dependence, instantaneous links, ...

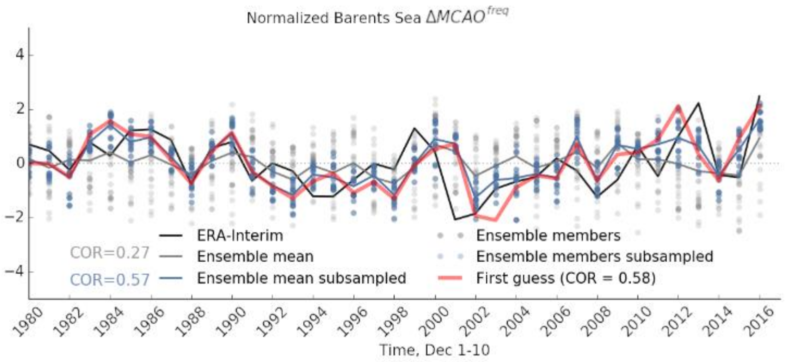
Output: Causal Structure



EXAMPLE: ENSEMBLE SUBSAMPLING

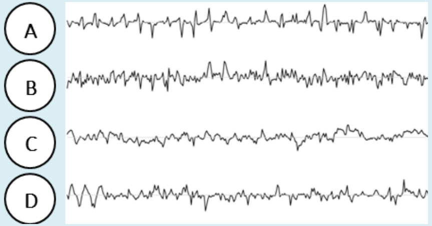
"Predictors and prediction skill for marine cold air outbreaks over the Barents Sea"





Causal Discovery

Input: Time-series data



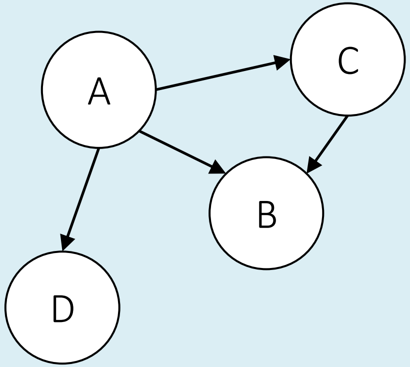
Causal Discovery

PCMCI Algorithm

$\mathrm{corr}(A_{t-\tau}, B_t \mid$ *Iterate through combinations of conditions* $)$

Can deal with auto-correlation, regime-dependence, instantaneous links, ...

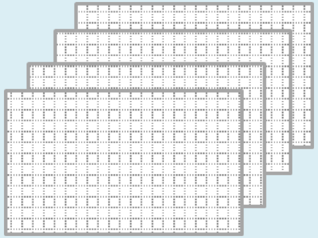
Output: Causal Structure



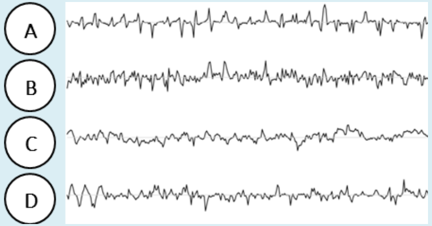
Runge et *al.* (2019), Kretschmer et *al.* (2016, 2018), Saggioro et *al.* (2020), phython code: jakobrunge.github.io/tigramite

Causal Discovery

Input: Gridded Climate data



Machine Learning



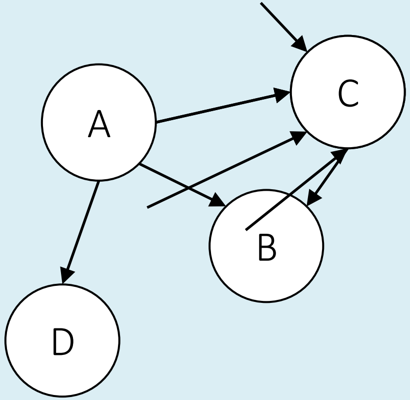
Causal Discovery

PCMCI Algorithm

$\mathrm{corr}(A_{t-\tau}, B_t \mid$ *Iterate through combinations of conditions* $)$

Can deal with auto-correlation, regime-dependence, instantaneous links, ...

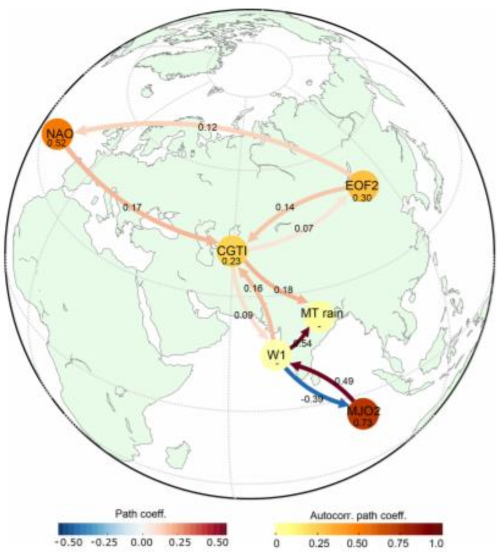
Output: Causal Structure



Applications (Domain Knowledge Required)

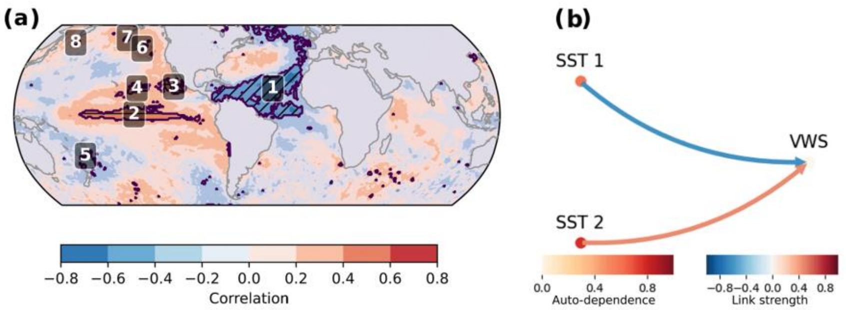
Indian Summer Monsoon

Di Capua et al. (2019), *ESD*



Hurricane Activity

Pfleiderer et al. (2020), *WCD*

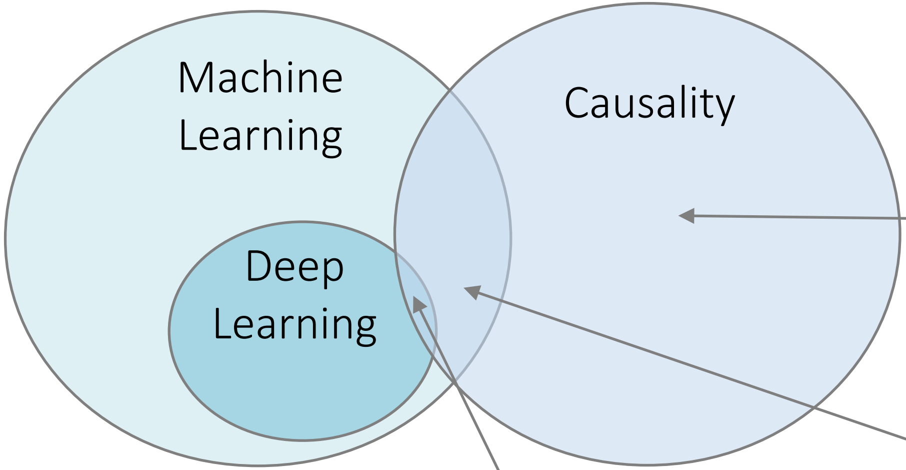


Morocco Crop yield

Lehmann et al. (2020), *GRL*

CONCLUSIONS

THANK YOU!



Including causal knowledge in data analysis is essential to answer causal questions

Causal discovery algorithms can help to close knowledge gaps

Combining these tools with ML/DL/XAI is promising to exploit new sources of predictability

m.j.a.kretschmer@reading.ac.uk @Marlene_Climate