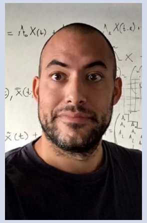
Learning latent functions for causal discovery

.*Machine Learning: Science and Technology* IOP Science 2023

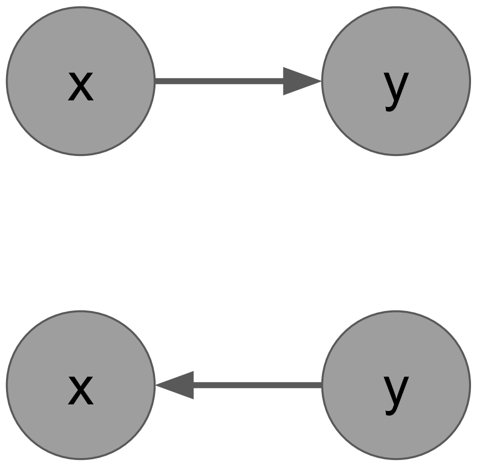
Emiliano Diaz Salas Porras, Gherardo Varando, J. Emmanuel Johnson, Gustau Camps-Valls

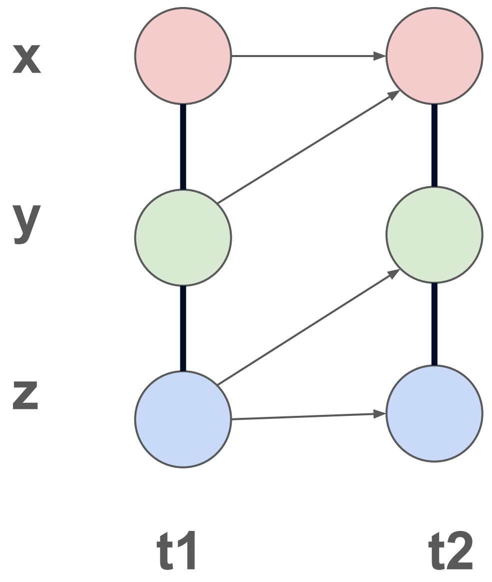



ISP · Image & Signal Processing
Universitat de València

Bivariate causal discovery algorithm for non–additive data

A method to "plug-in" after the Markov equivalence class has been estimated

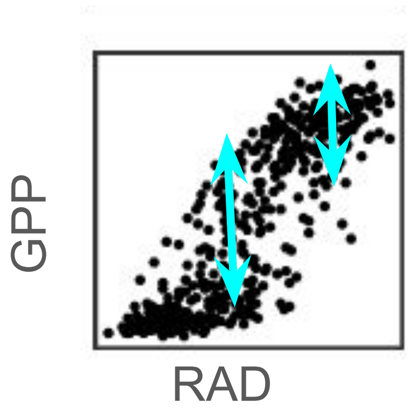




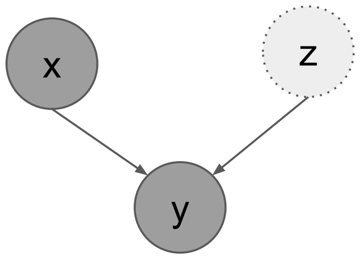
Causal insufficiency widespread in Earth system sciences

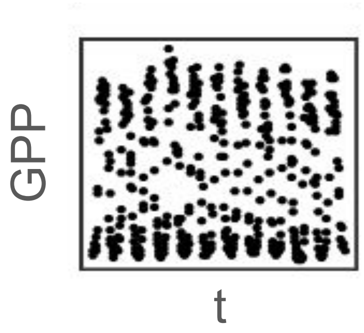
Non-additive data important for Earth system science

1. weak form of non causal sufficiency



2. can generate structured data e.g. spatial, temporal



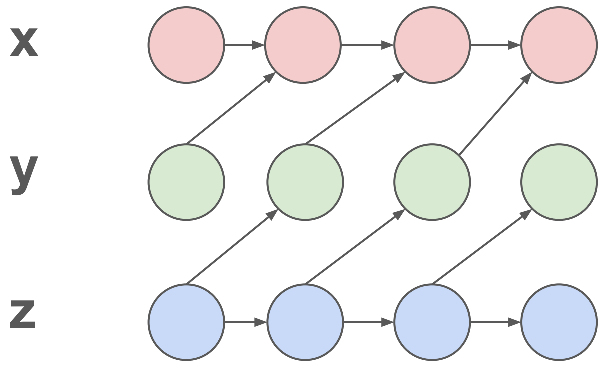


Instantaneous relationships widespread in Earth system sciences

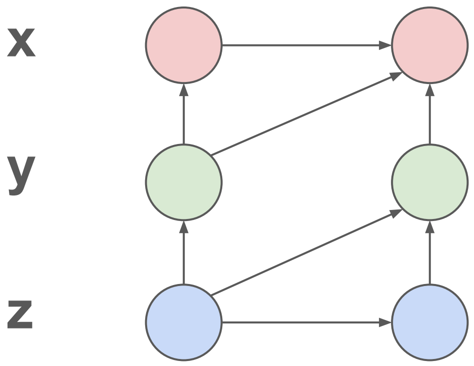
- “Instantaneous” relationships often occur in practice due to systems observed at lower resolution than the fundamental mechanisms.

- In this case additional assumptions necessary to identify the causal structure.

native resolution



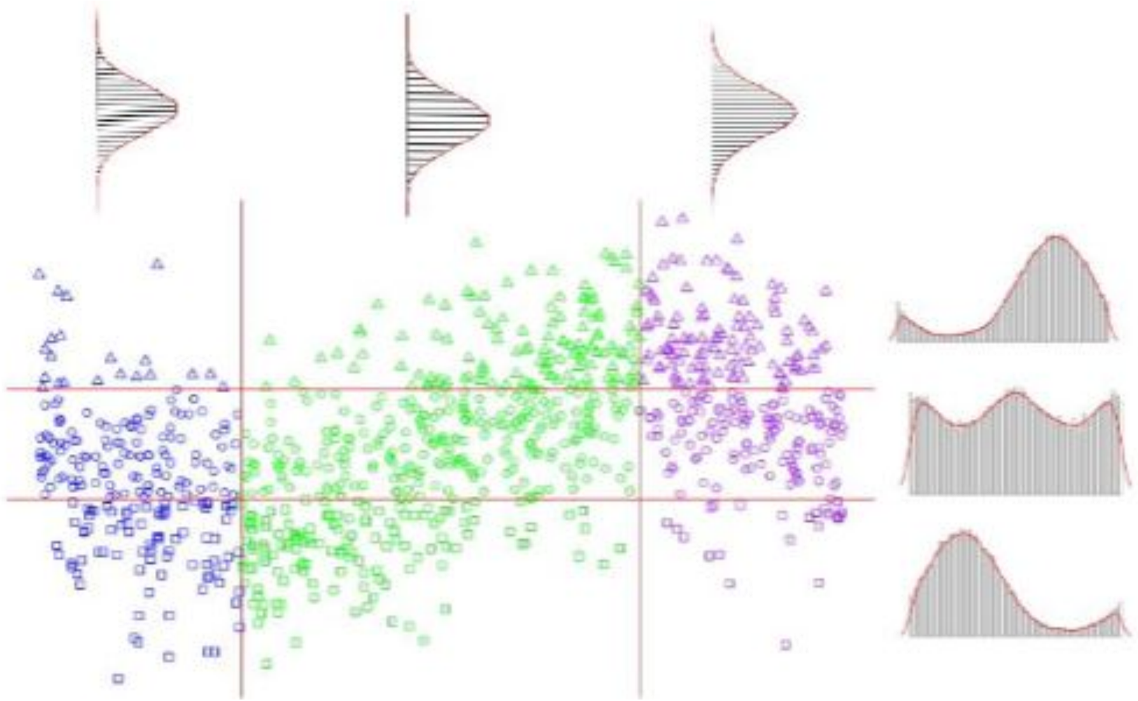
aggregated low resolution



Examples

- Daily ERA5 data
- Weekly satellite data
- harmonizationoof different products to lower temporal resolution

Independence of cause and mechanism (ICM) (Daniusis et al, 2010)



To identify causal structure when instantaneous relationships exist we need extra assumption.

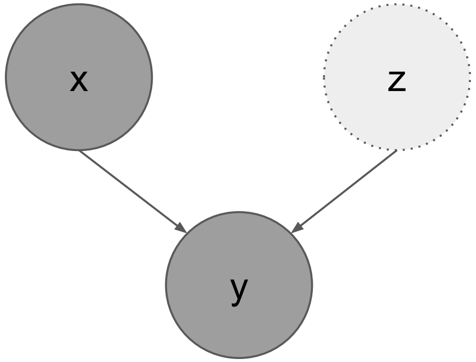
- modularity assumption

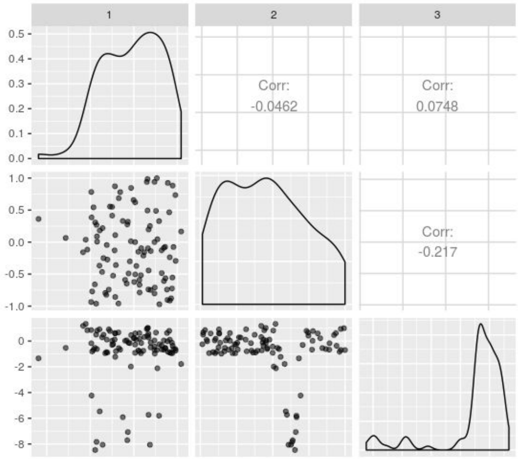
- ie no info about $p(y|x)$ in $p(x)$

Taking a step back: modeling the inducing FCM

$$y := f(x, z)$$

Latent Noise approach





conditional pdf approach

An extended ICM assumption

For a data generating mechanism $y = f(x, z)$ we make the following assumptions, following (Stegle, et al 2010)

1. Deterministic process
2. Exogenous noise z
3. Gaussian noise z
4. Algorithmic independence

Loss function for finding z penalizes assumption violations

$$L(\mathcal{Z}) = \boxed{\ln\left(nHSIC(\mathcal{X}_a, \mathcal{R}_{x\to y})\right) + \zeta \ln\left(MSE(\mathcal{R}_{x\to y})\right)}$$
$$+\eta \ln\left(nHSIC(\mathcal{X}, \mathcal{Z})\right) + \nu \ln\left(SMMD^2_{\mathcal{N}}(\mathcal{Z})\right)$$

where:

$$\mathcal{X}_a := \{(x_i, z_i)\}_{i=1}^n$$
$$\mathcal{R}_{x\to y} := \{y_i - f(x_i, z_i)\}_{i=1}^n$$
$$\hat{y} = f(x, z) = \sum_{i=1}^n \alpha_i k\big((x, z), (x_i, z_i)\big) \in \mathcal{H}_{xz}$$
$$\boldsymbol{\alpha} = (K_{xz} + n\lambda I)^{-1}\mathbf{y}$$

Deterministic process

Loss function for finding z penalizes assumption violations

$$L(\mathcal{Z}) = \ln\left(nHSIC(\mathcal{X}_a, \mathcal{R}_{x\to y})\right) + \zeta \ln\left(MSE(\mathcal{R}_{x\to y})\right)$$

$$+\eta \boxed{\ln\left(nHSIC(\mathcal{X}, \mathcal{Z})\right)} + \nu \ln\left(SMMD^2_{\mathcal{N}}(\mathcal{Z})\right)$$

where:

$$\mathcal{X}_a := \{(x_i, z_i)\}_{i=1}^n$$

$$\mathcal{R}_{x\to y} := \{y_i - f(x_i, z_i)\}_{i=1}^n$$

$$\hat{y} = f(x, z) = \sum_{i=1}^n \alpha_i k\big((x, z), (x_i, z_i)\big) \in \mathcal{H}_{xz}$$

$$\boldsymbol{\alpha} = (K_{xz} + n\lambda I)^{-1}\mathbf{y}$$

Exogenous noise

Loss function for finding z penalizes assumption violations

$$L(\mathcal{Z}) = \ln\left(nHSIC(\mathcal{X}_a, \mathcal{R}_{x\to y})\right) + \zeta \ln\left(MSE(\mathcal{R}_{x\to y})\right)$$
$$+\eta \ln\left(nHSIC(\mathcal{X}, \mathcal{Z})\right) + \nu \boxed{\ln\left(SMMD^2_{\mathcal{N}}(\mathcal{Z})\right)}$$

Gaussian noise

where:

$$\mathcal{X}_a := \{(x_i, z_i)\}_{i=1}^n$$
$$\mathcal{R}_{x\to y} := \{y_i - f(x_i, z_i)\}_{i=1}^n$$
$$\hat{y} = f(x, z) = \sum_{i=1}^n \alpha_i k\big((x, z), (x_i, z_i)\big) \in \mathcal{H}_{xz}$$
$$\boldsymbol{\alpha} = (K_{xz} + n\lambda I)^{-1}\mathbf{y}$$

USMILE

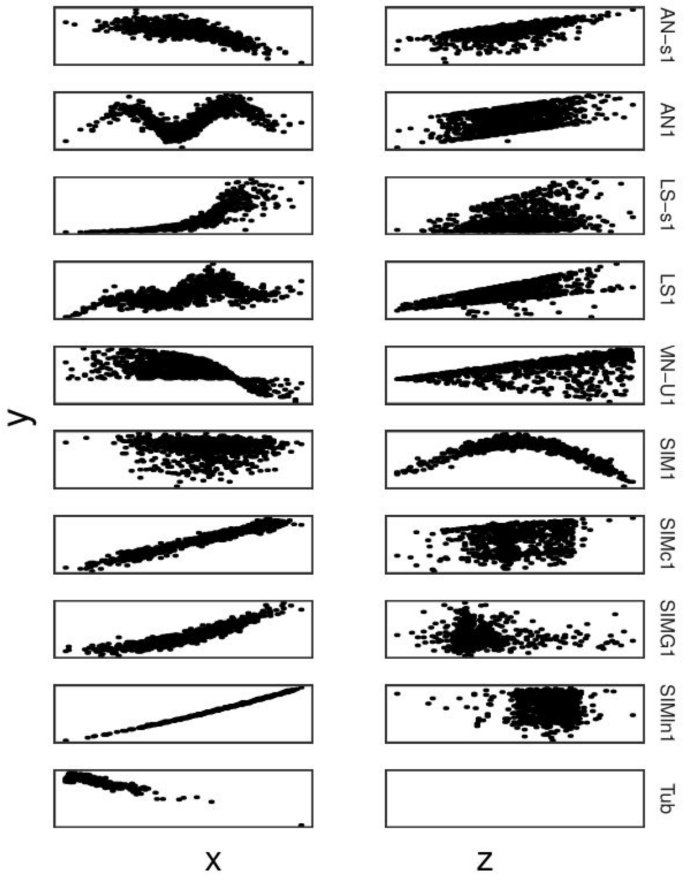
Enforce soft assumptions

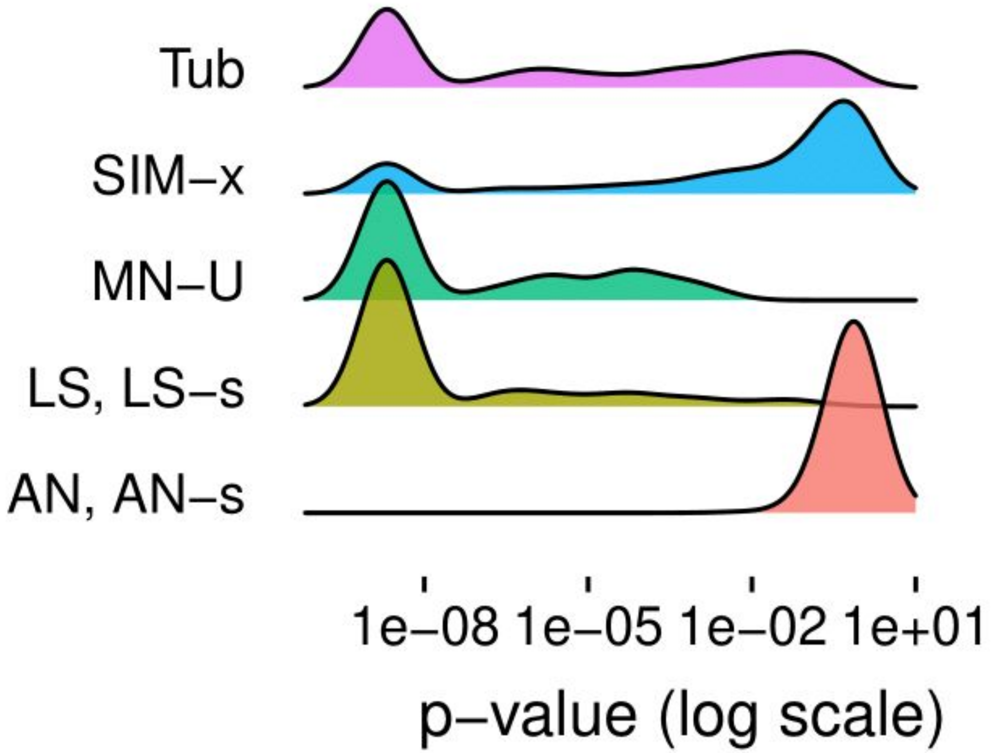
Advantages

- Rank relative importance of assumptions

- Relaxing determinism assumption: use non-additvity as causal signal:
  - model misspecification (anti causal direction)
  - estimation error (both directions)
  - asymmetry assumption: model misspecification generates more non-additivity

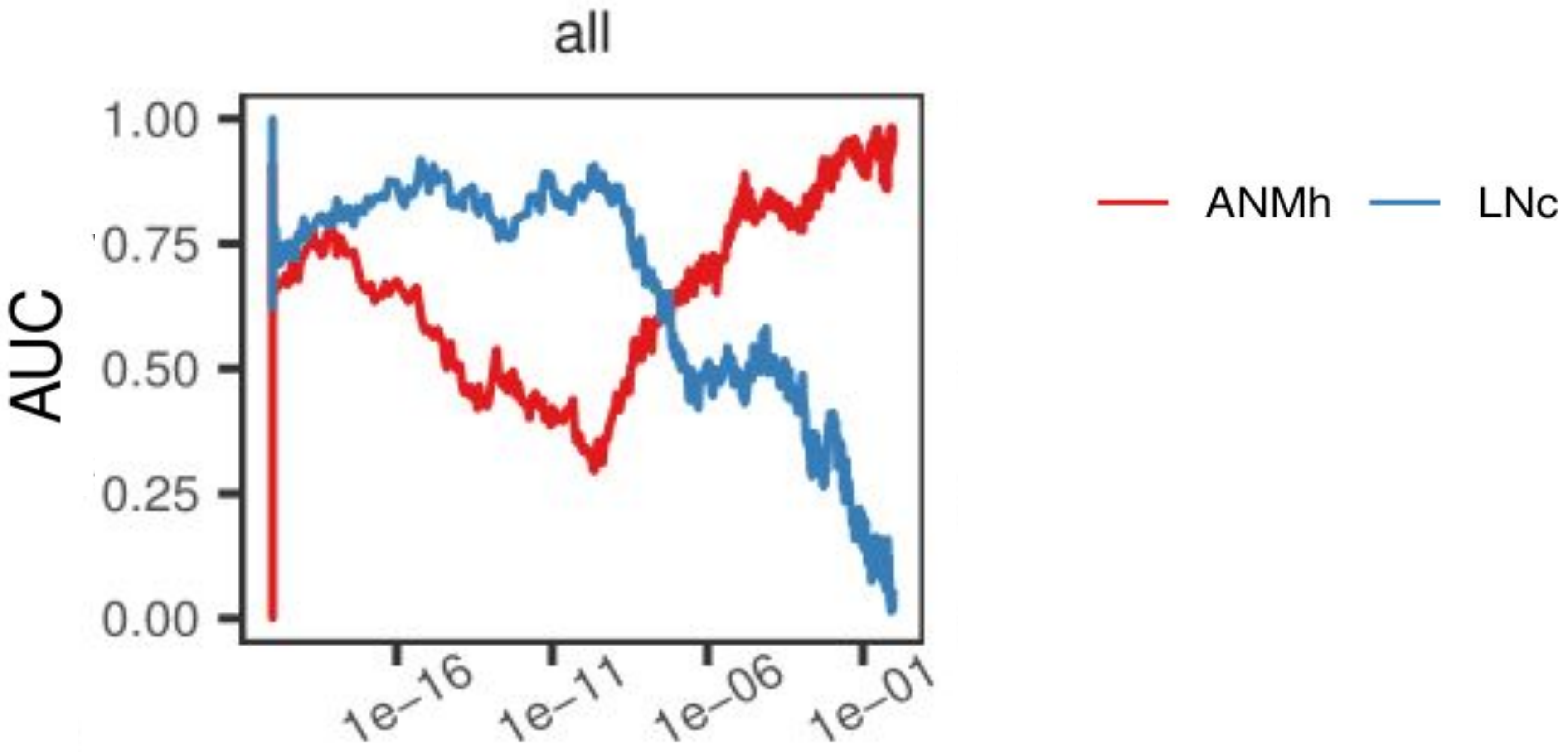
IID data & additivity hypothesis test



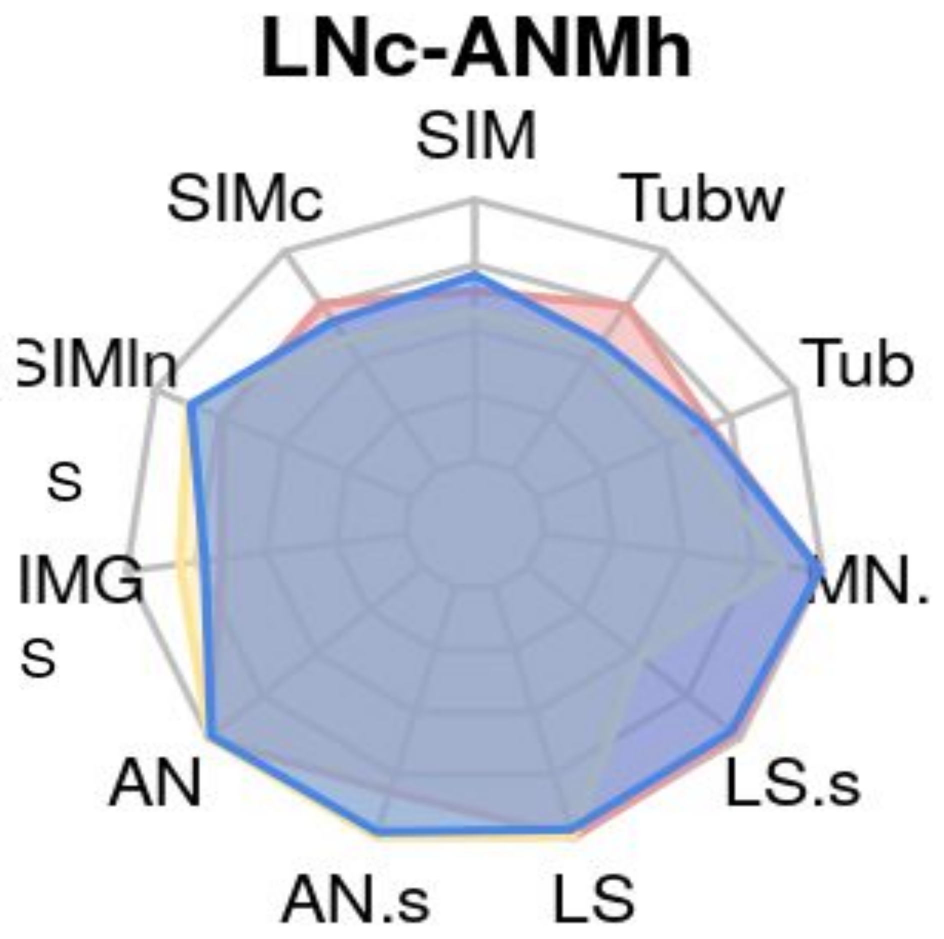


$$S = \min\{nHSIC(\mathcal{X}, \mathcal{R}_{x\to y}), nHSIC(\mathcal{Y}, \mathcal{R}_{y\to x})\}$$

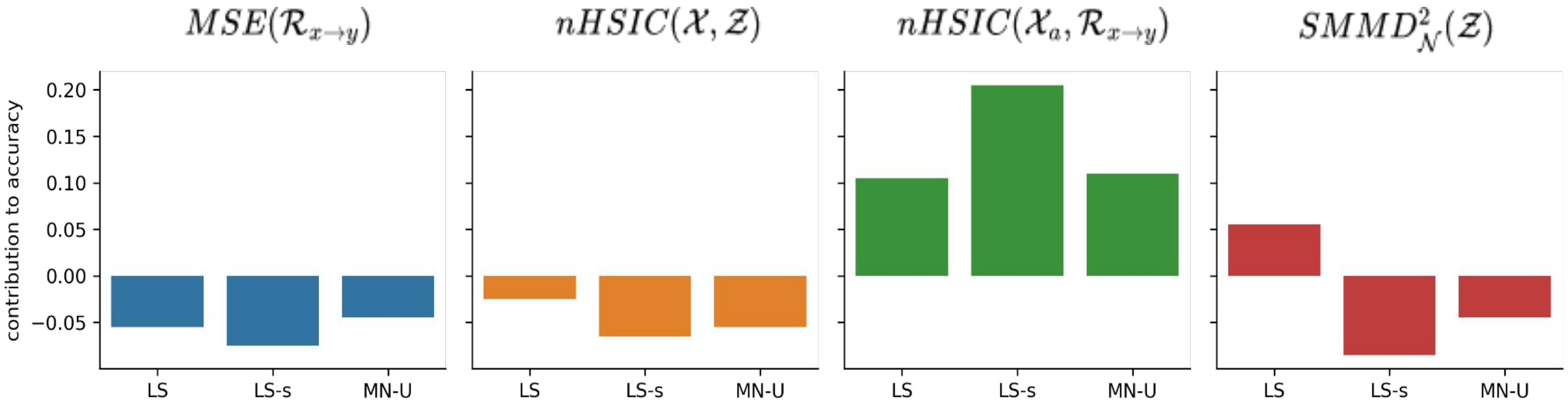
LNc method's accuracy improves with non-additivity



Combining LNc and ANMh obtains SOTA performance

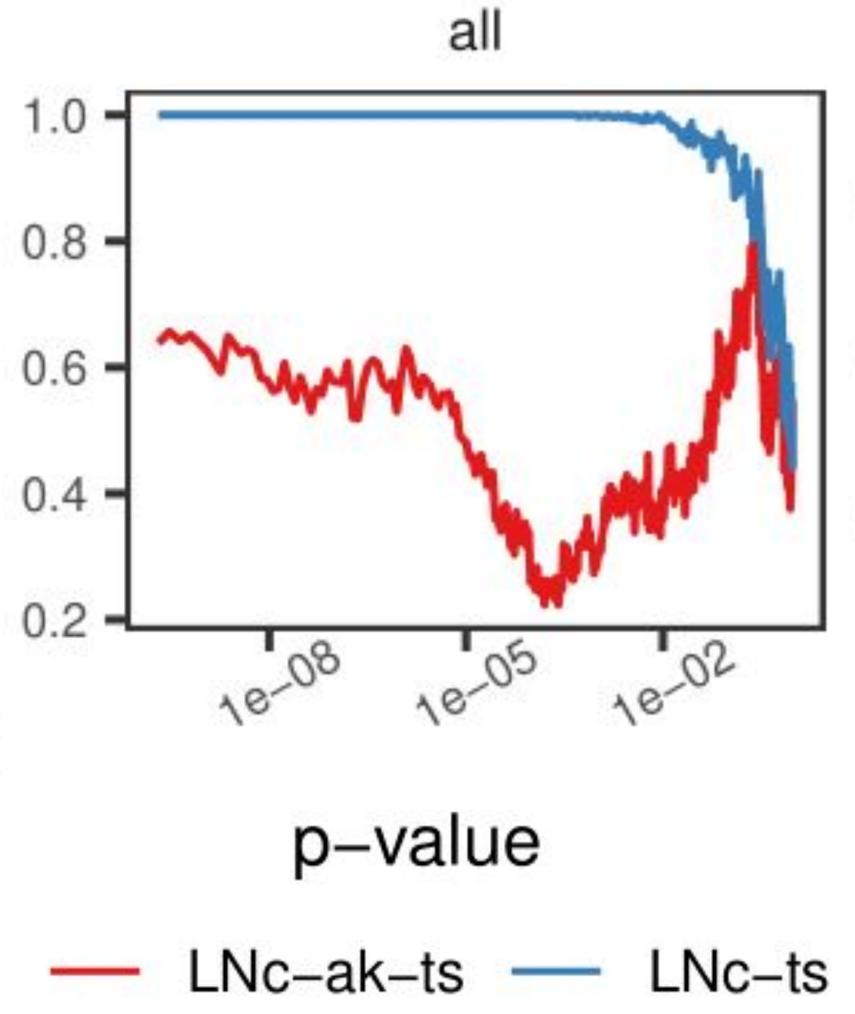


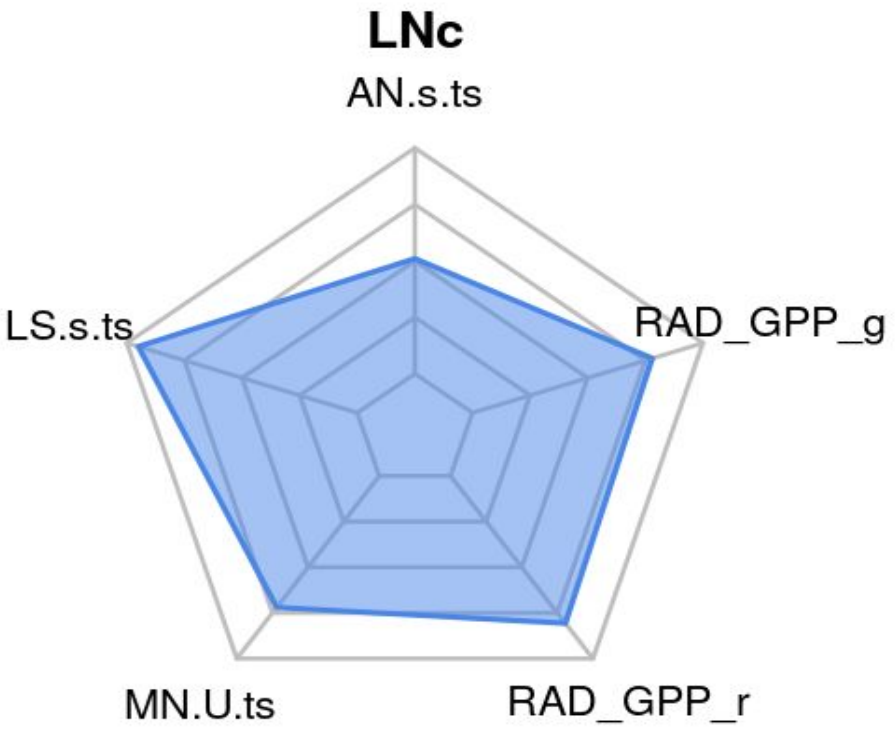
Relative importance of assumptions



- Additive residual assumption only one that needs to be implemented strictly.

LNc method's accuracy improves with non-additivity





Conclusions & contributions

- Generative approach is advantageous for extending to spatio-temporal data.
- Generative LNc method including extension to time series and an additivity hypothesis test.

Limitations

- Time series extension only applies when there is no self-dependence in effect variable
- Only suitable for cases where confounding is ruled out

USMILE

Opportunities

- When causal discovery is applied to time series at an aggregated scale instantaneous effects very common: asymmetry methods suitable.

- A soft/weak form of non-causal sufficiency can be responsible for non-additive data: need asymmetry methods suited to this setting. Non causal sufficiency very common in Earth system science.

- Can be used to complement other causal discovery algorithms that don’t make ICM assumption and can’t fully identify causal structure when “instantaneous” relationships occur.

Learning latent functions for causal discovery

Diaz, A., Johnson, J.E., Varando, G. and Camps-Valls, G.

Machine Learning: Science and Technology IOP Science 2023