PhD Thesis

Towards causal discovery for Earth system sciences

PhD Program “Enginyeria Electrònica” (3131)

Emiliano Diaz Salas Porras

Universitat de València

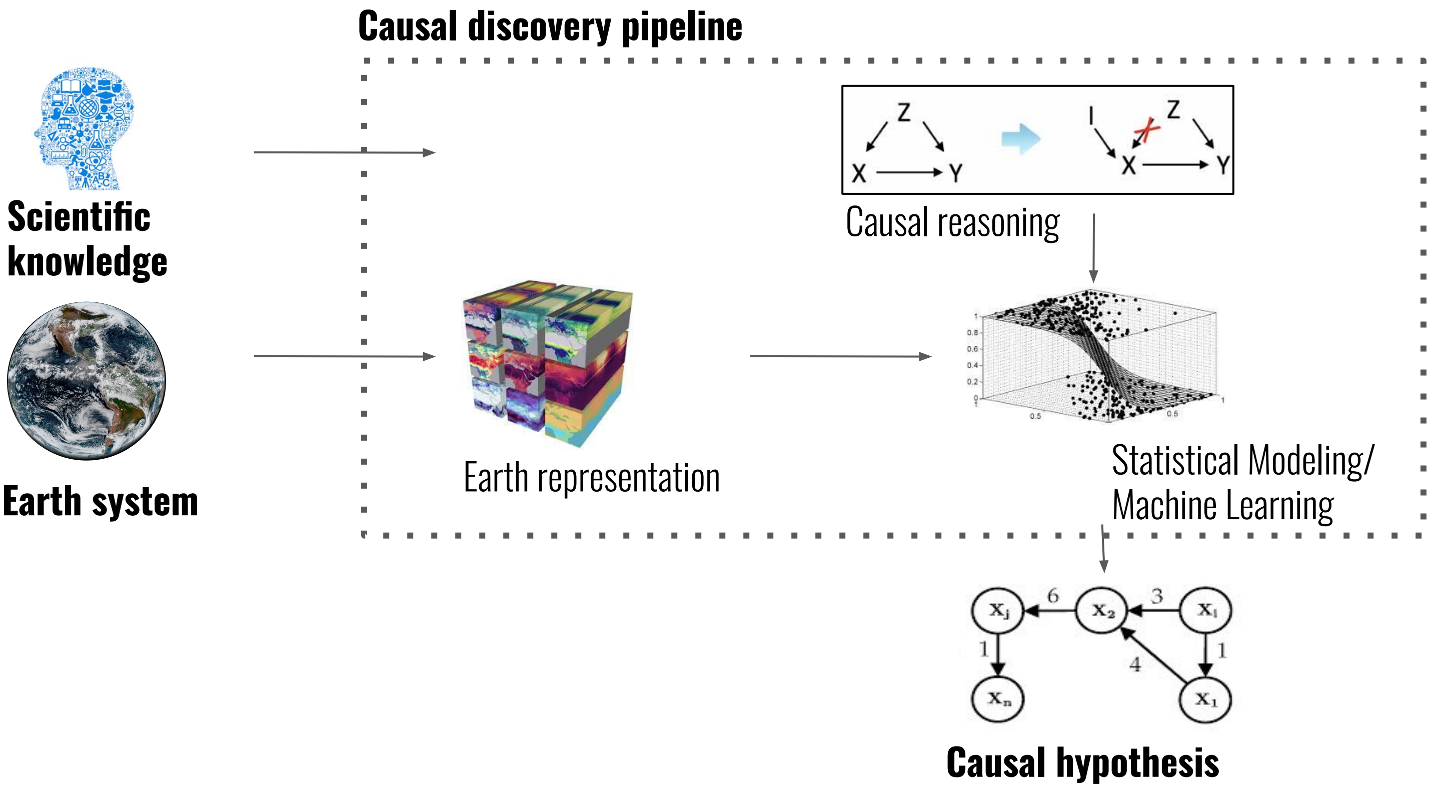
Advisors: Gustau Camps-Valls, Gherardo Varando, Adrián Pérez-Suay

Outline

1. Causal discovery in Earth system sciences
2. Bivariate causal discovery for Non-additive data
    a. PDF approach - model conditionals
    b. FCM approach - latent noise estimation
3. Spatial maps of causal relations
4. Natural interventions
5. Conclusions

Causal discovery for Earth system sciences

Causal discovery in Earth System Sciences



Challenges

1. **Data**: complex structure, dependencies, distributions, high dimensional, causally insufficient, measurement error.

2. **Scientific**: Integrating scientific knowledge into causal discovery
   Eg. data and physical simulation models.

3. **Statistical**:
   a. non-linear conditional independence tests
   b. extrapolation capabilities (failed promise of generalizability)

4. **Causal***:*
   a. discrete optimization problem,
   b. causal representation learning,

Goals

New causal discovery methods for Earth system sciences:

1. Asymmetry bivariate causal discovery for Earth system sciences data:
   a. non-additive data and (weakly) causally-insufficient for i.i.d data
   b. structured data

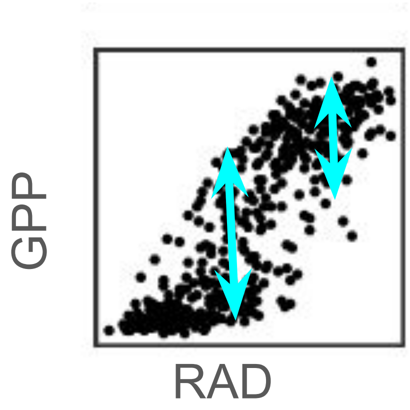
2. Causally heterogenous data:
   a. different interventions have occurred, not fully identified
   b. latent causal representations, circumvent the large discrete space. incorporate physical knowledge

Bivariate causal discovery for non-additive data

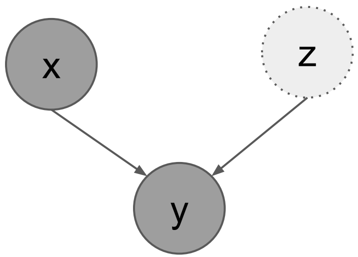
Causal insufficiency widespread in Earth system sciences

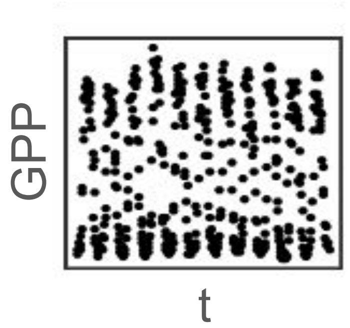
Non-additive data important for Earth system science

1. weak form of non causal sufficiency



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

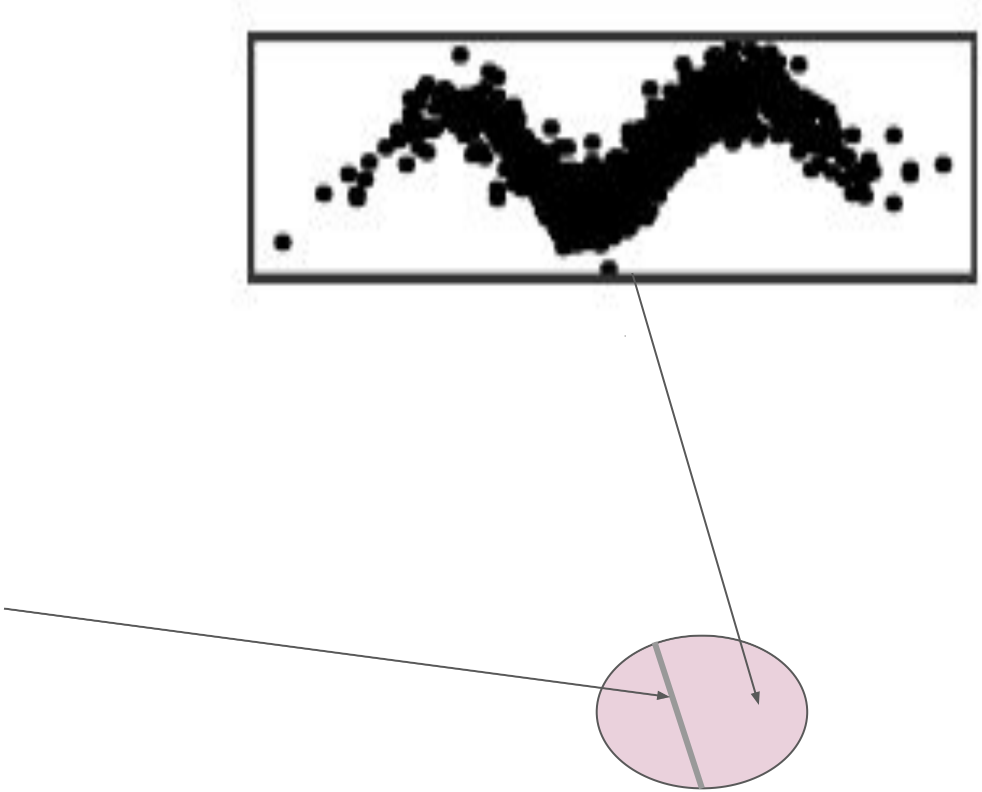




Is causal identifiability possible?

FCM : Additive Noise model :

$$y = f(x) + z$$



Result: (Hoyer et al. 2009)

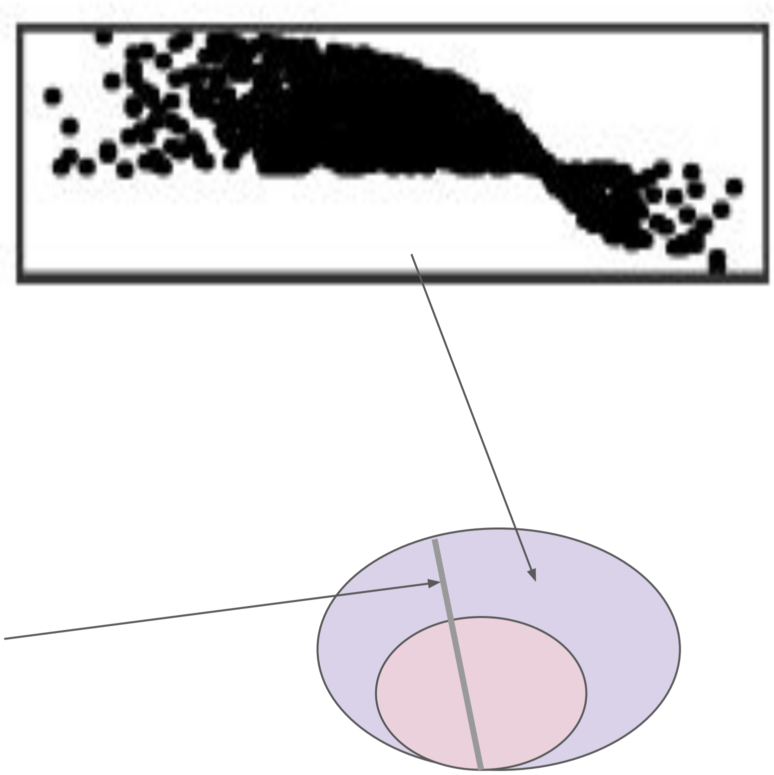
Exceptions: eg. linear gaussian

Hoyer, Patrick et al. "Nonlinear causal discovery with additive noise models.", *Neurips,* (2009).

Is causal identifiability possible?

FCM : Post non-linear model :

$$y = g(f(x) + z)$$



Result: (Zhang and Hyvärinen, 2009)

Exceptions: eg. noise z generalized mixture of exponentials, f(x) two-sided, asymptotically exponential, f,g strictly monotonic

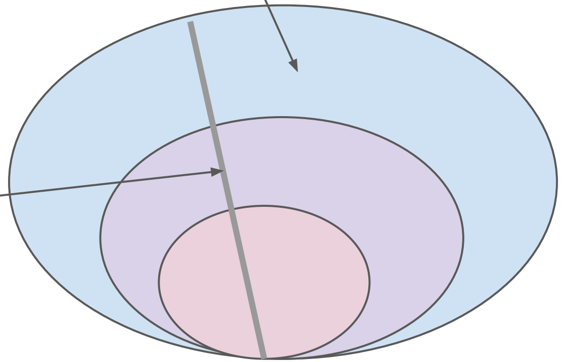
Zhang, Kun, and Aapo Hyvarinen. "On the identifiability of the post-nonlinear causal model.", UAI, (2009).

Is causal identifiability possible?

FCM : Location scale model :

$$y = f(x) + g(x) * z$$

Result: (Immer et al. 2023)



Exceptions: eg. noise z gaussian, x log-mix-rational log, f and g functions of polynomials of degree 2 or less.

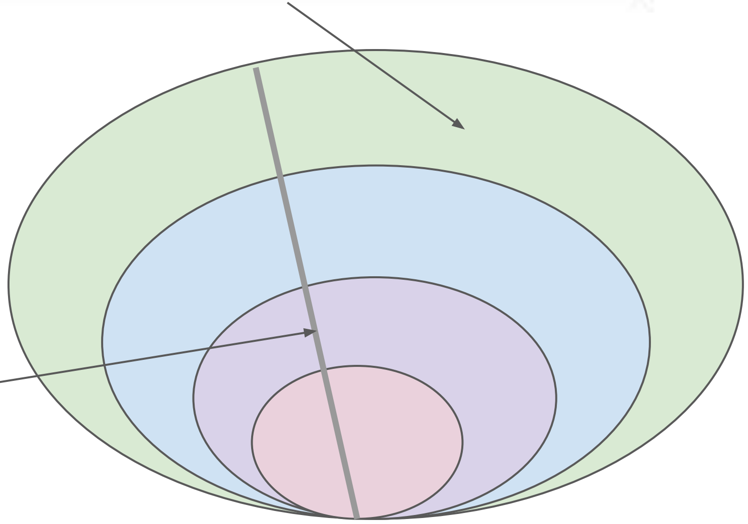
Immer, Alexander et al. "On the identifiability and estimation of causal location-scale noise models." *ICML,* (2023).

Is causal identifiability possible?

FCM : general modular model :

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

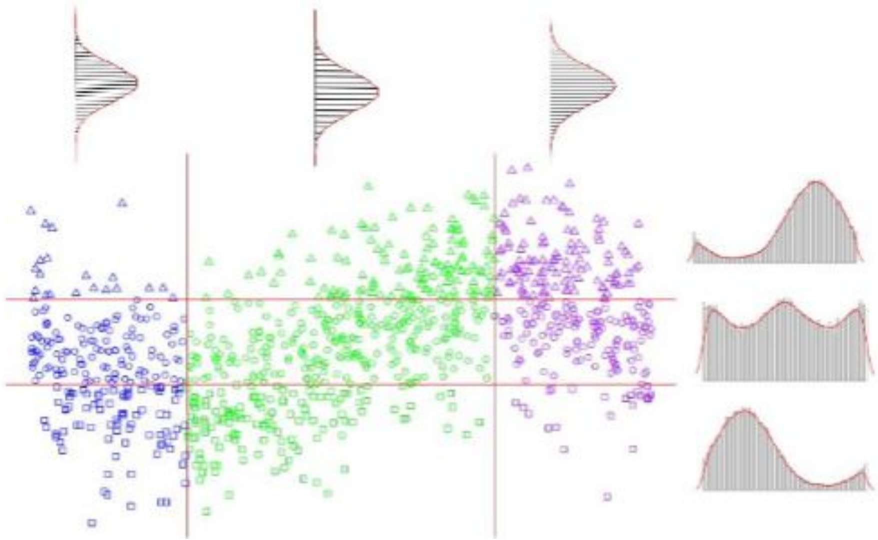
~~**Result**~~ **Principle**:
Independence of cause and mechanism (ICM) (Daniusis et al, 2010)
Equivalent to minimal complexity factorization



Exceptions: Both directions algorithmically independent. ie both factorizations have equal complexity

Daniussi, Povilas, et al. "Inferring deterministic causal relations." *UAI* (2010).

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



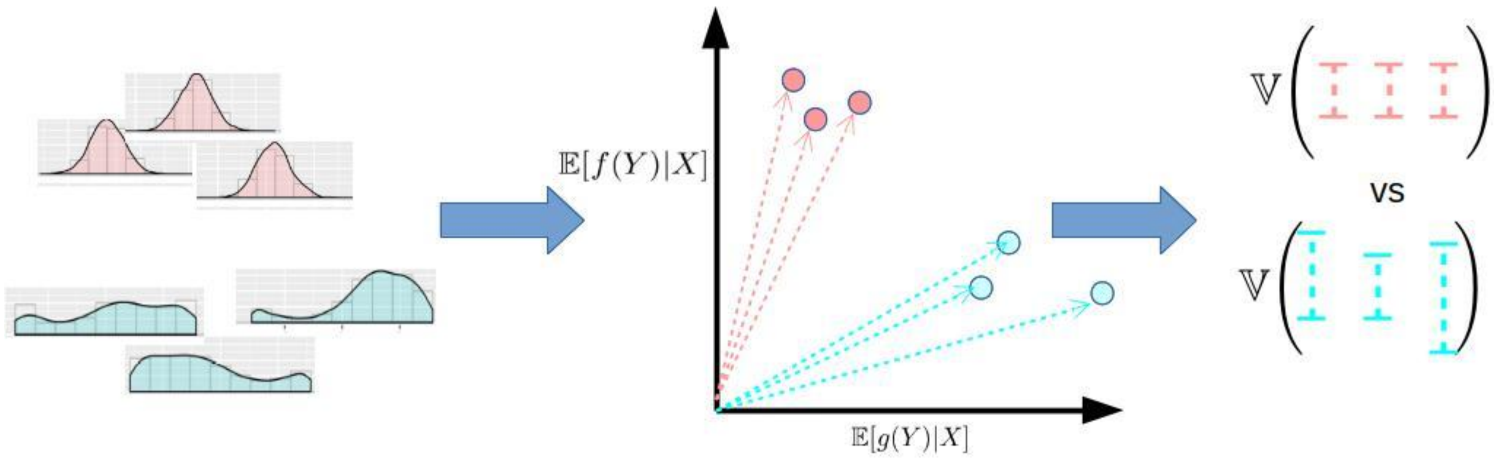
- modularity assumption
- $p(x)$ algorithmically independent from $p(y|x)$
- ie no info about $p(y|x)$ in $p(x)$

Daniussi, Povilas, et al. "Inferring deterministic causal relations." *UAI* (2010).

PDF approach: modeling conditionals

Causal inference in Geosciences with multidimensional kernel deviance measures
Emiliano Diaz Salas Porras, Adrian Perez Suay, Valero Laparra, and Gustau Camps-Valls
EGU General Assembly, Geophysical Research Abstracts, Vol. 21 2019x

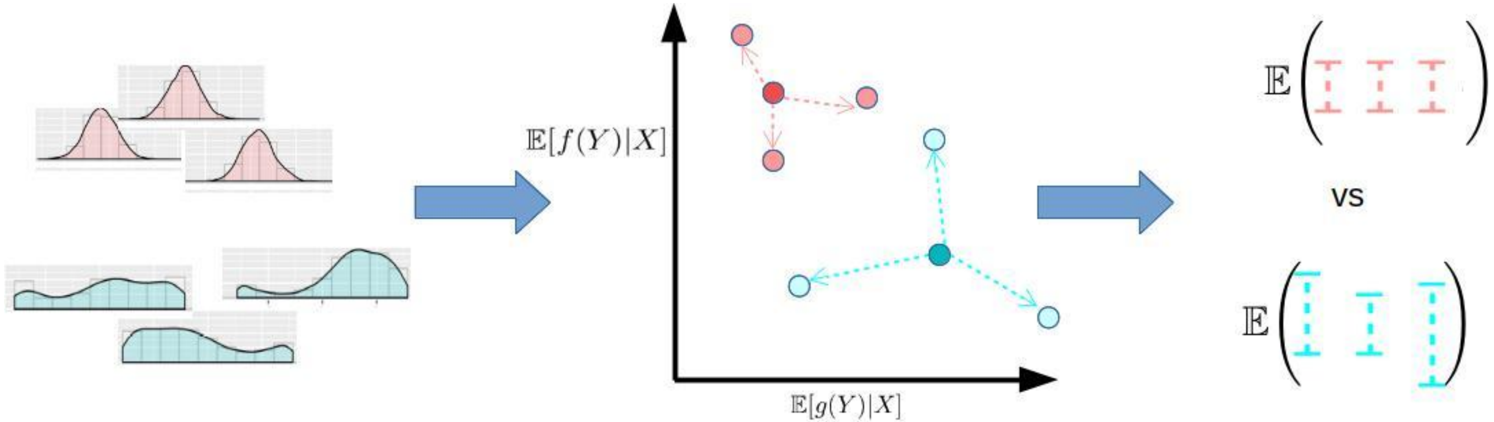
KCDC (Mitrovic et al, 2018): use conditional mean embedding (CME)



Represent $p(y|x)$ using (possibly infinite) moments estimated implicitly (CME) or explicitly (multi-output regression)

Mitrovic et al, "Causal Inference via Kernel Deviance Measures", NeurIPS 2018

KCMC (Díaz et al 2019): change in complexity



- Norm of differences instead of differences in the norm
- More info about complexity of change and not only change in complexity

Diaz et al. “Causal inference in Geosciences with multidimensional kernel deviance measures” EGU General Assembly, Geophysical Research Abstracts, Vol. 21 2019

How well does the CME represent $p(y|x)$?

- CME

$$\hat{\mu}_{Y|X=x}(y) = \alpha(x)^\top k_y = l_x^\top B_\lambda k_y$$

- Multi output regression on kernel similarities k_y

$$\hat{f}(X) = L(L + n\lambda I)^{-1} K = L B_\lambda K$$

- Multi output regression equivalent to estimating CME and evaluating at data points.

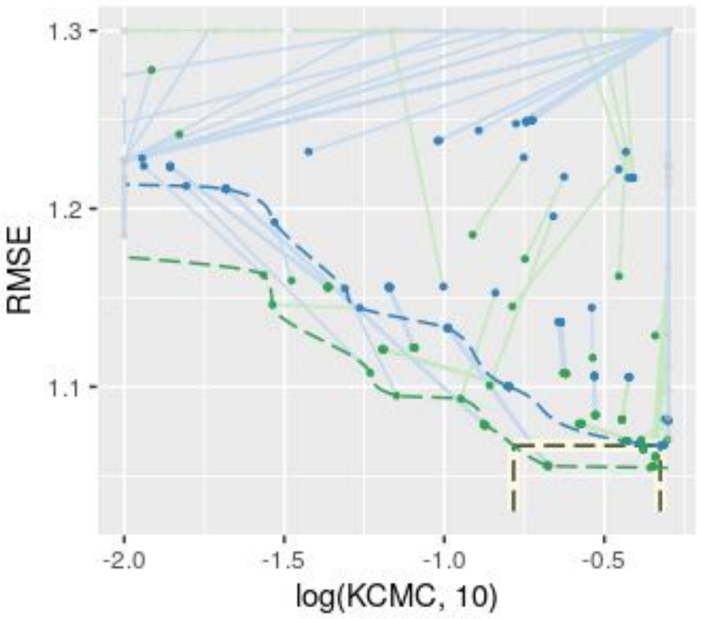
How well does the CME represent $p(y|x)$?

- CME

$$\hat{\mu}_{Y|X=x}(y) = \alpha(x)^\top k_y = l_x^\top B_\lambda k_y$$

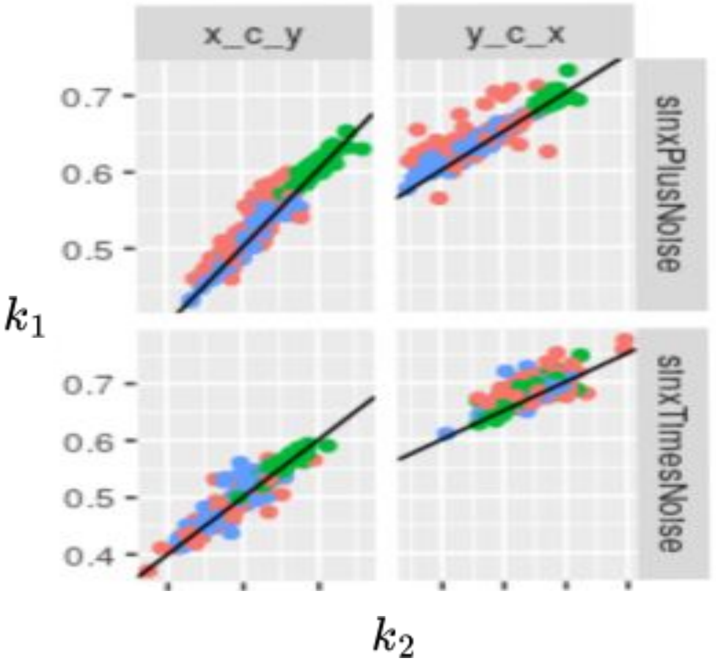
- Two ingredients:
  - Output: which moments to select to represent
  - Input: which features to select to estimate those moments

Input feature selection: Pareto front of fit and complexity



- different hyperparameters result in different causal directions
- use pareto front to decide which direction is more efficient
- but can only do this for fixed output parameter (different scales)

Output feature: noise contrastive estimation (NCE)

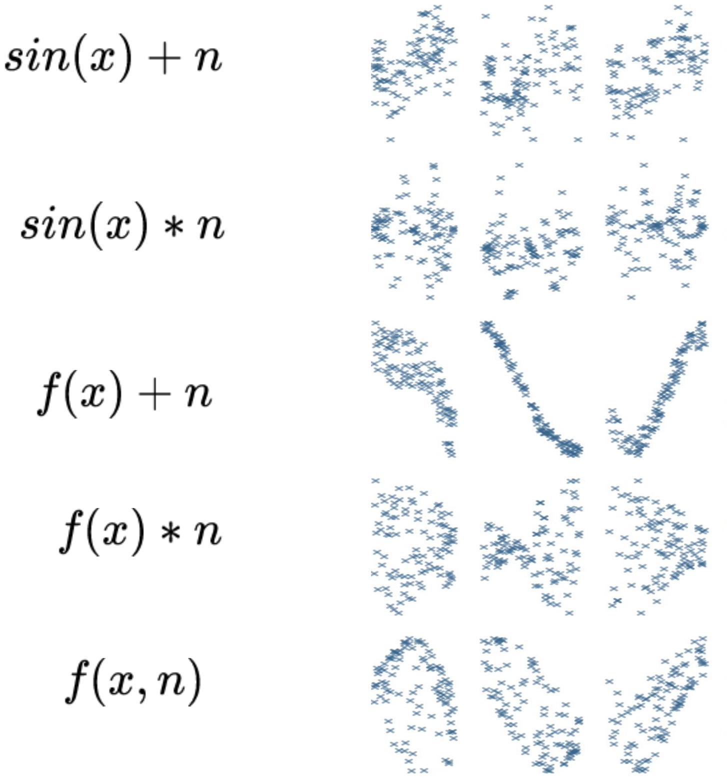


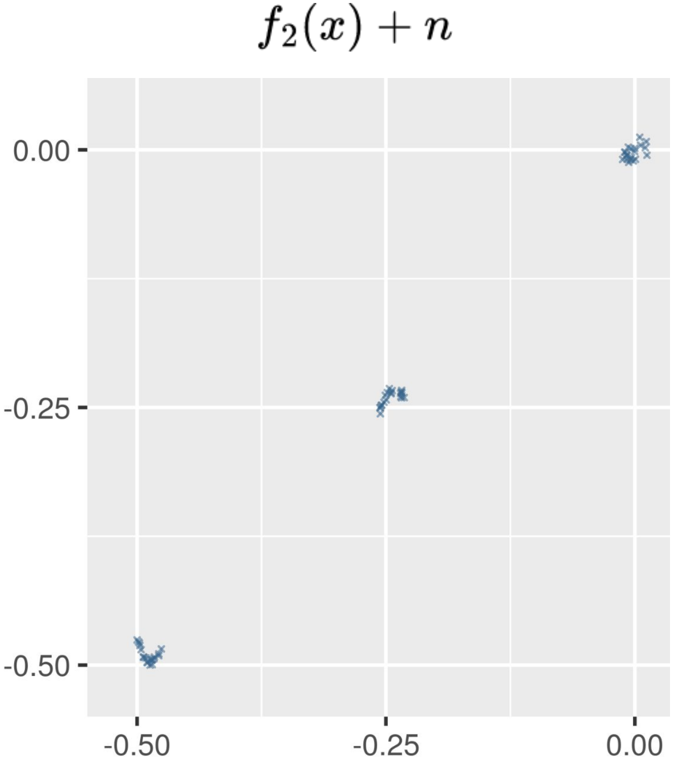
- generate real pairs with $p(x)$ and $p(y|x)$
- generate fake pairs with marginals $p(x)$ and $p(y)$
- discriminate between two by using CME based similarity score:

$$s(x, y) = \langle \psi, \psi \rangle_{\mathcal{H}_y} \qquad \hat{\psi}(x) = \hat{\mu}_{Y|X=x}$$

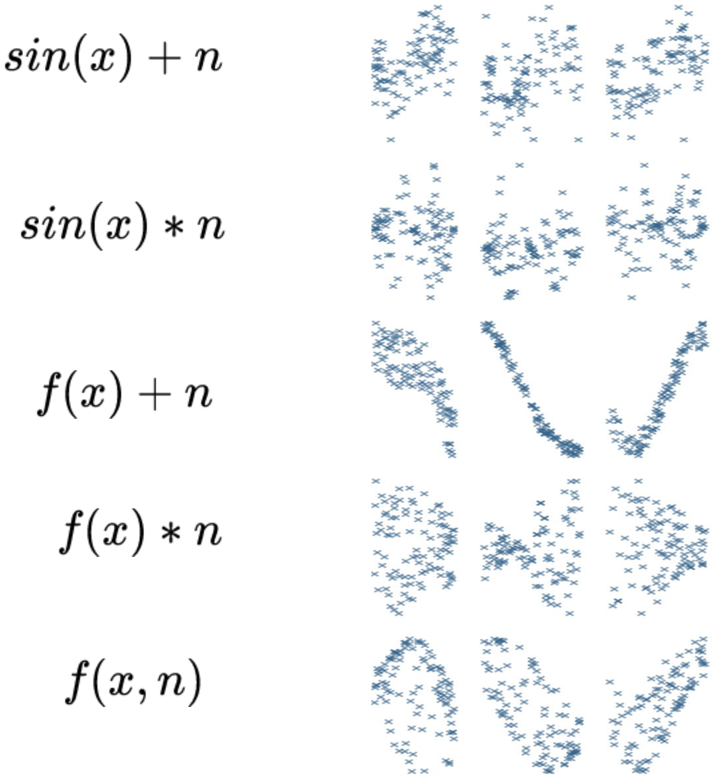
	k1	k2
y=f(x)+n	100	100
y= f(x)*n	73	90

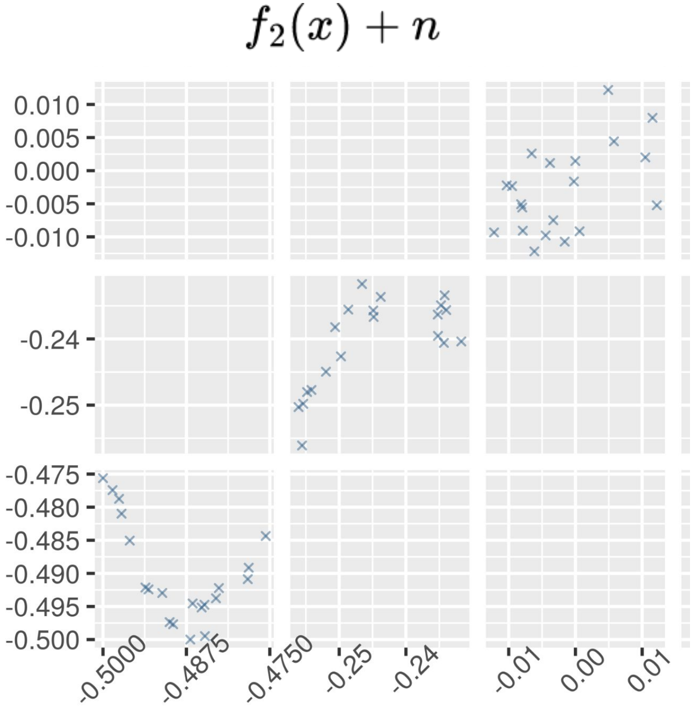
Data: artificially generated





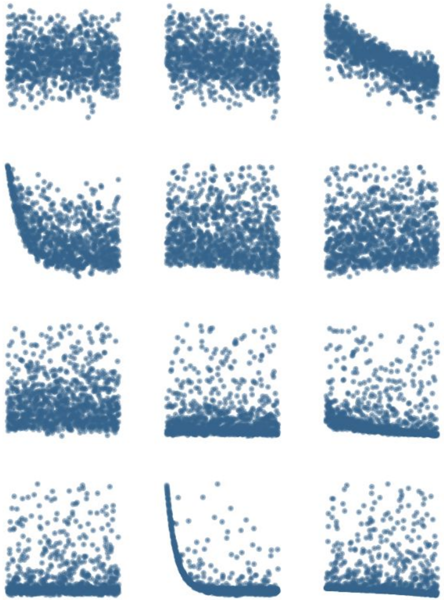
Data: artificially generated



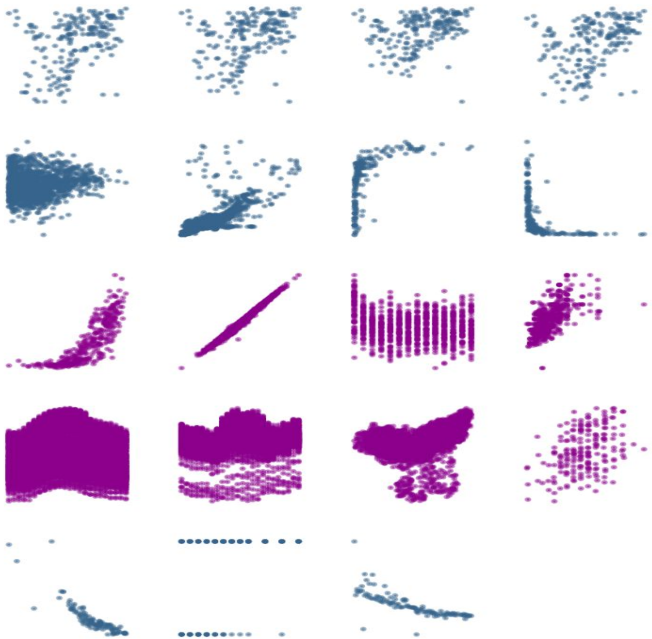


Data: physical model and real-world

$g(x, n)$ RTM: Prosail



$TCEP$



Results

	Causal measure			
$\mathcal{D}$	ANM	KCDC	KCMC med	KCMC NCE
$sin(x)+n$	**1**	**1**	**1**	**1**
$sin(x)*n$	0.96	0.68	**1**	**1**
$f(x)+n$	**1**	0.45	0.98	0.99
$f(x)*n$	0.74	0.88	**0.96**	**0.96**
$f(x,n)$	0.68	0.88	0.81	**0.9**
$f_2(x)+n$	0.57	0.42	0.62	**0.82**
$g(x,n)$	0.6	0.81	**1**	**1**
TCEP	0.57	0.65	**0.71**	0.66

Benchmarks

1. Additive noise models (ANM)
2. KCDC
3. KCMC- median heuristic
4. KCMC- NCE

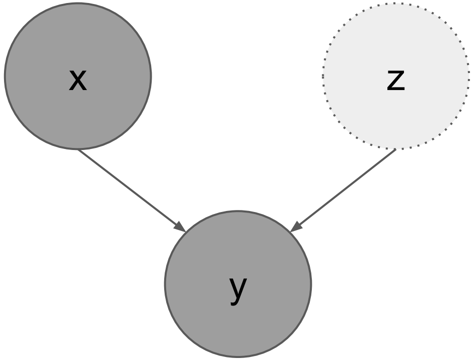
FCM approach latent noise estimation

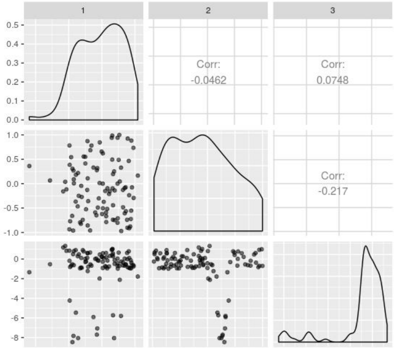
Learning latent functions for causal discovery
Diaz, A., Johnson, J.E., Varando, G. and Camps-Valls, G.
Machine Learning: Science and Technology IOP Science 2023

Taking a step back: modeling the inducing FCM

$$\boxed{y := f(x, z)}$$

Latent Noise approach





CME approach

z

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

Stegle, Oliver et al. "Probabilistic latent variable models for distinguishing between cause and effect.", Neurips, (2010).

Method idea: estimate noise, approximate with ANM

- Adapt bayesian model selection approach (GPI,Stegle et al, 2010) which uses ICM, to frequentist, kernel non parametric

- Why?
    - Doesn't work very well for certain classes
    - Approach is nice since you get more ingredients from the FCM
    - If we obtain point-wise estimates of z, can turn into additive noise model: good methods in this case

Stegle, Oliver et al. "Probabilistic latent variable models for distinguishing between cause and effect.", Neurips, (2010).

Enforce soft assumptions

Advantages

- Traverse model space more efficiently (asymmetry generation vs optimality)
- Rank relative importance of assumptions
- Relaxing determinism assumption: non-additvity as causal signal:
  - model misspecification (anti causal direction)
  - estimation error (both directions)
  - asymmetry assumption: model misspecification generates more non-additivity

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:

Deterministic process

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

Loss function for finding z penalizes assumption violations

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

where:

Exogenous noise

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

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

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

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

Loss function for finding z penalizes assumption violations

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

where:

Gaussian noise

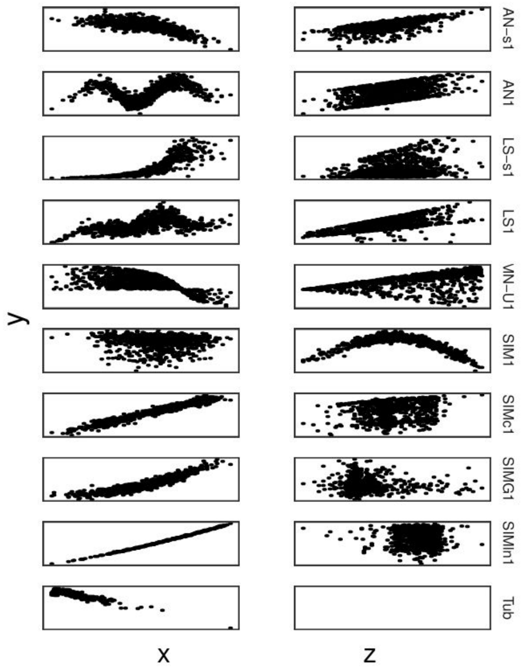
$$\mathcal{X}_a := \{(x_i, z_i)\}_{i=1}^n$$

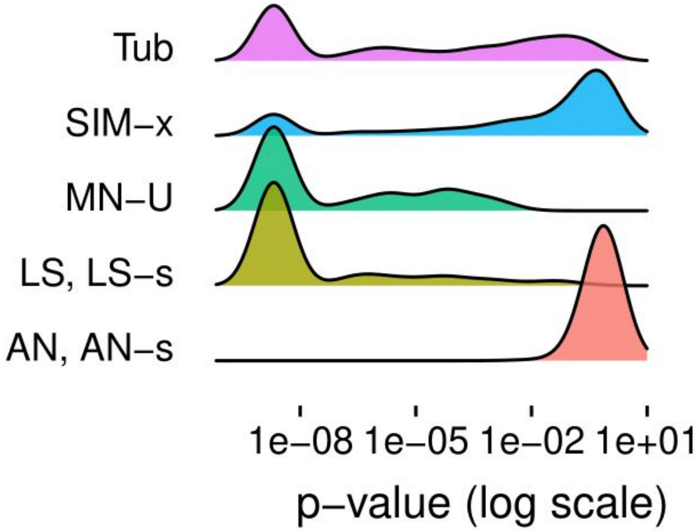
$$\mathcal{R}_{x\to y} := \{y_i - f(x_i, z_i)\}_{i=1}^n$$

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

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

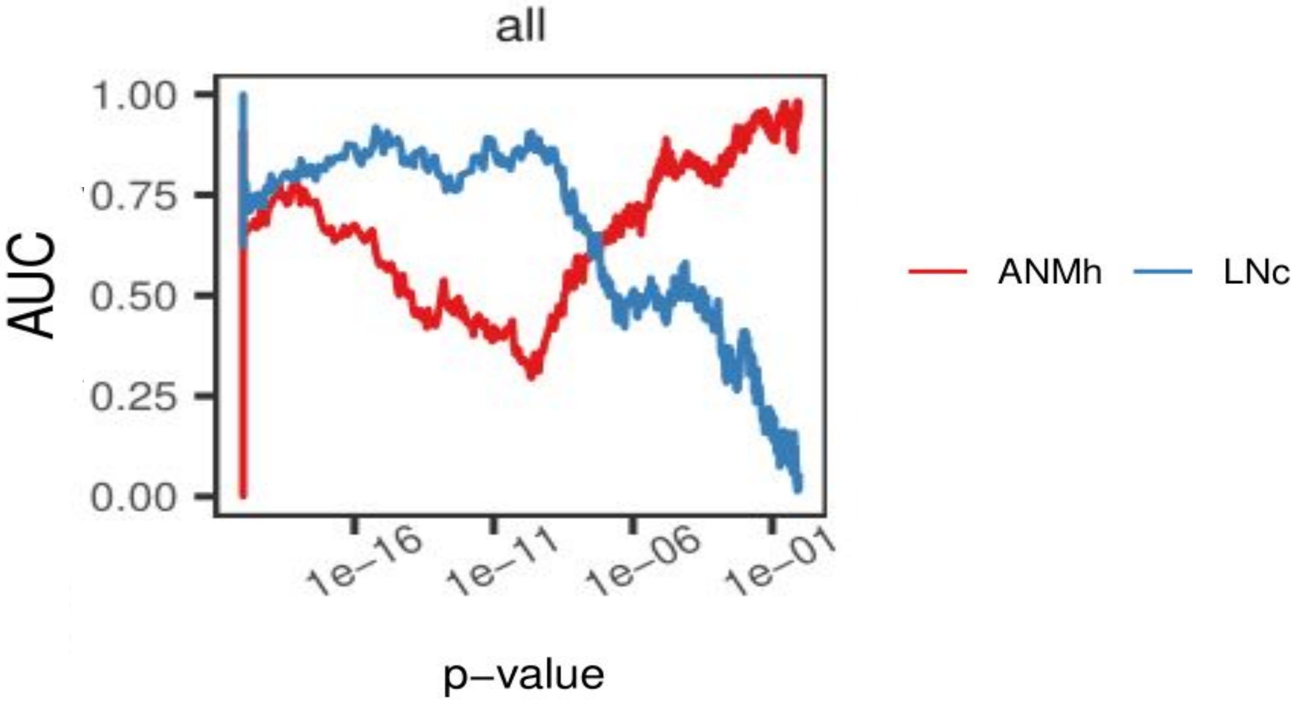
IID Data



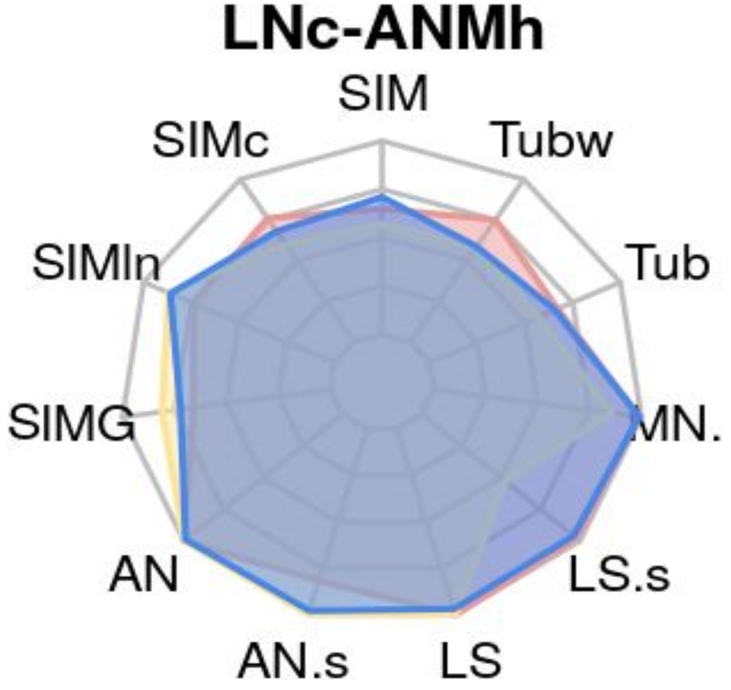


$$S = \min\{nHSIC(\mathcal{X}, \mathcal{R}_{x \to y}), nHSIC(\mathcal{Y}, \mathcal{R}_{y \to x})\}$$

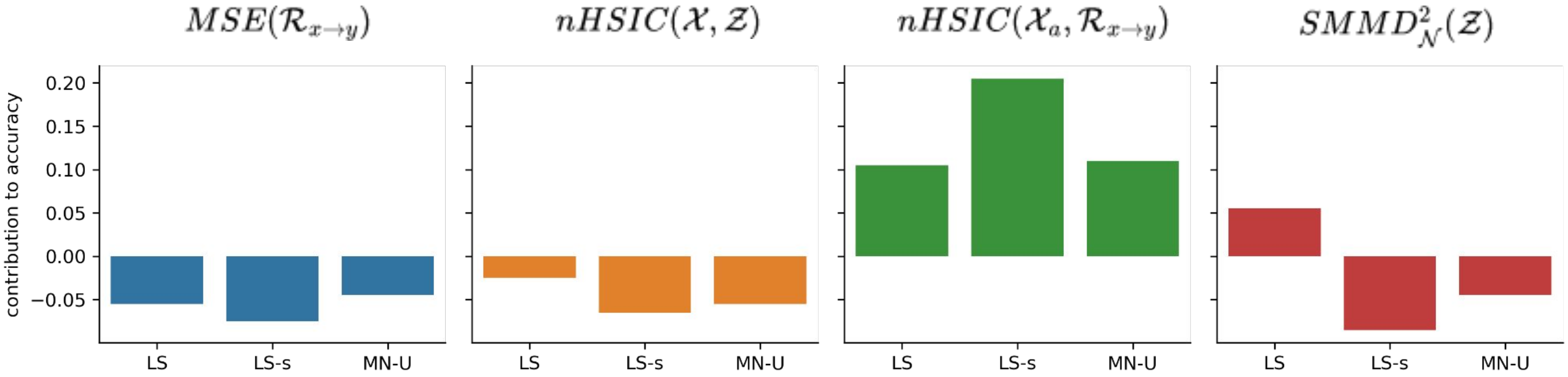
LNc method's accuracy improves with non-additivity



Combining LNc and ANMh obtains SOTA performance

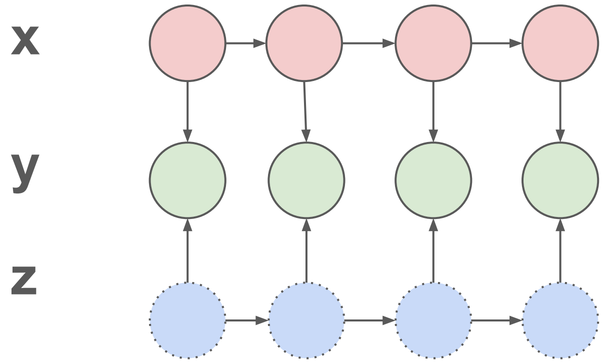


Relative importance of assumptions



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

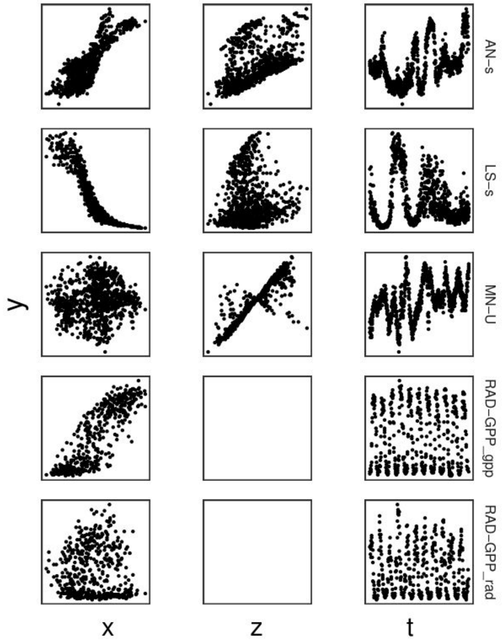
Time series extension

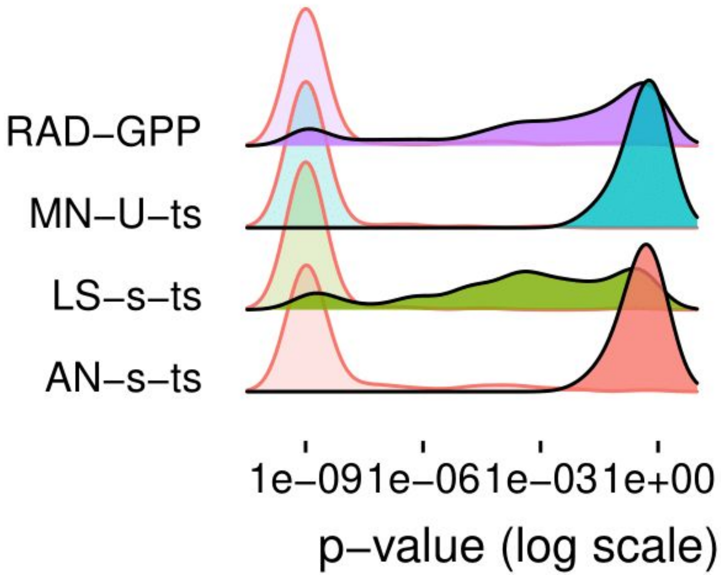


$$L'(\mathcal{Z}) = L(\mathcal{Z}) + \ln\left(nHSIC(\mathcal{R}, \mathcal{T})\right)$$

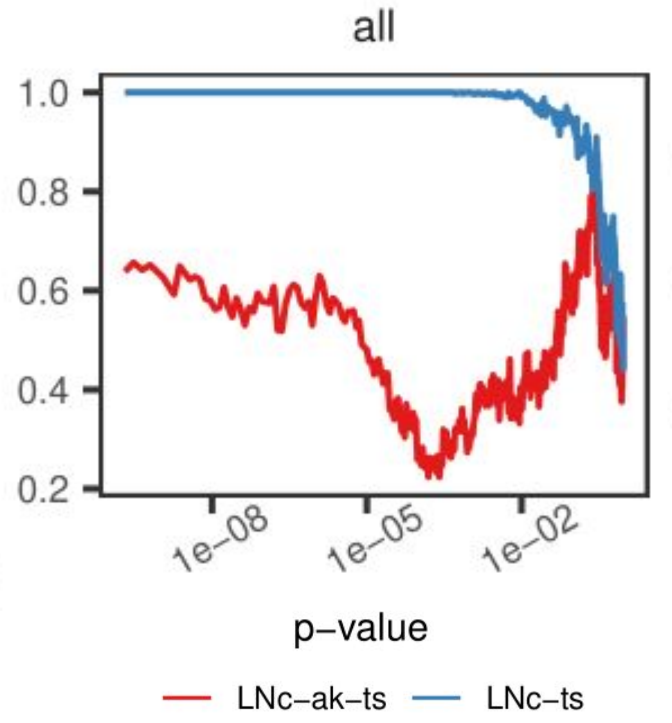
- Conditioning on x and z removes temporal structure of time-series y
- Regularizer to favor solutions z which result in residuals without temporal structure

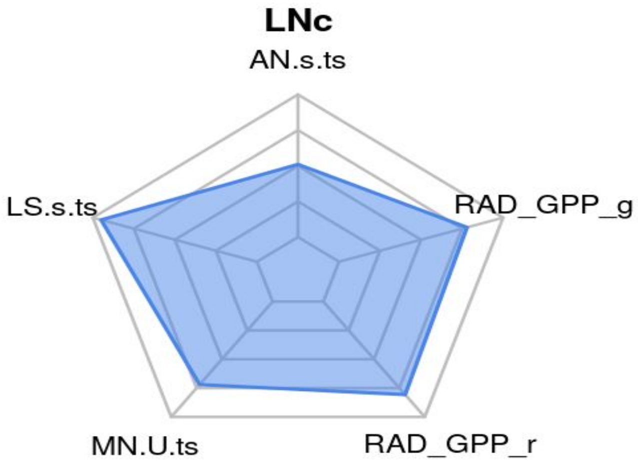
Time series data





LNc method's accuracy improves with non-additivity





Spatial maps of causal relations

Inferring causal relations from observational long-term carbon and water fluxes records
Diaz, E.,Adsuara, J.E., Moreno-Martinez, A., Piles, M. and Camps-Valls, G.
Scientific Reports 12 :1610, 2022

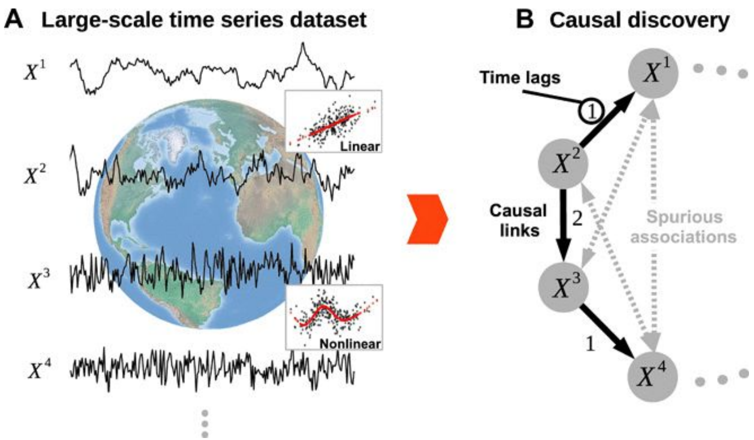
Convergent cross mapping (CCM) (Sugihara et al, 2012)

Context:

- Causal inference method for time series/dynamic systems

Intended for data from:

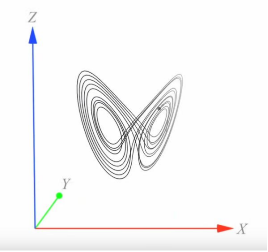
- Deterministic systems
- No strong forcings
- No “instantaneous” processes



Credits:
(Runge, et al, 2019)

Sugihara, George et al. "Detecting causality in complex ecosystems.", Science, (2012).
Runge, Jakob, et al, “Detecting and quantifying causal associations in large nonlinear time series datasets”, (2019)

ODE equations encode causal relations

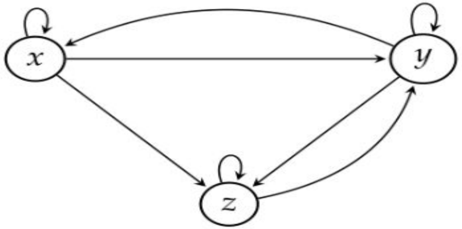


Data

system

$$
\begin{aligned}
x' &= \sigma y - \sigma x \\
y' &= -xz + \rho x - y \\
z' &= xy - \beta z
\end{aligned}
$$

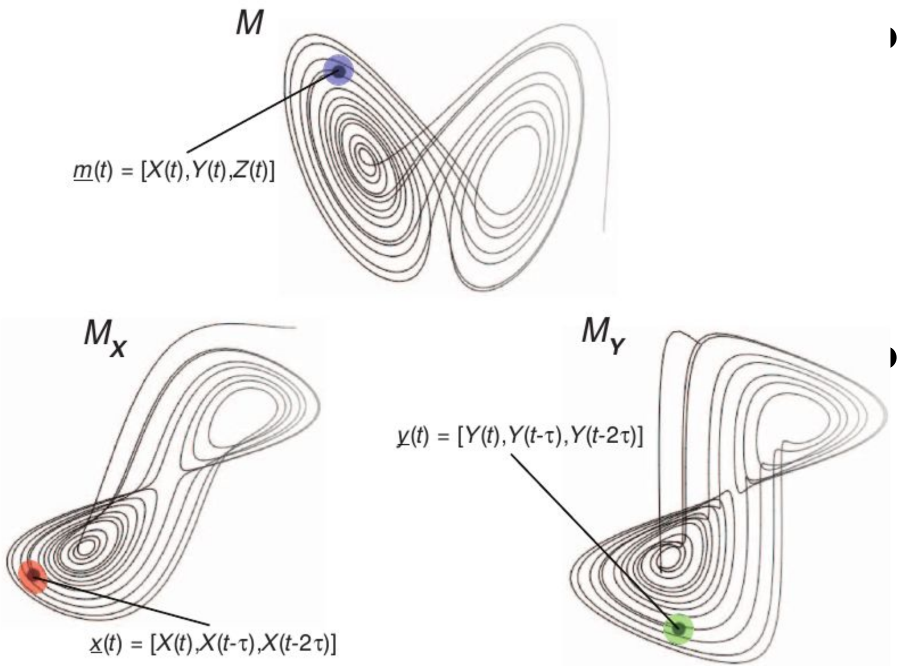
CCM



causality

- X causes Y IFF X's equation expresses its dynamics in terms of Y's state
- CCM: circumvent the ODE

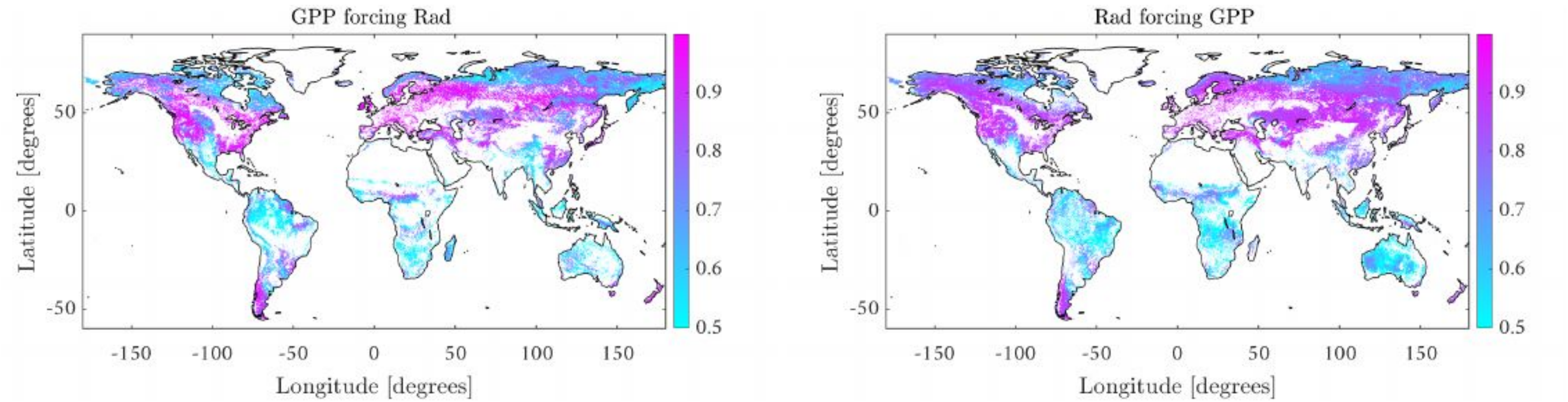
Takens' theorem (Takens, 1981) - informal implications



- "shadow manifold" using time series of one variable retains topology of original Manifold: points close on M are also close on Mx and My
- **CCM in a nutshell**: two variables causally related if you can rebuild the state-space from the variables' (embedded/lagged) individual time series

Takens, Floris et al. "Detecting strange attractors in turbulence.", Dynamical systems and turbulence, (1981).

Strong unidirectional forcing



- CCM wrongly infers that GPP causes Radiation because of strong unidirectional forcing

Our contribution - Robust CCM (RCCM)

- Detect strong "instantaneous" unidirectional forcing using extended CCM

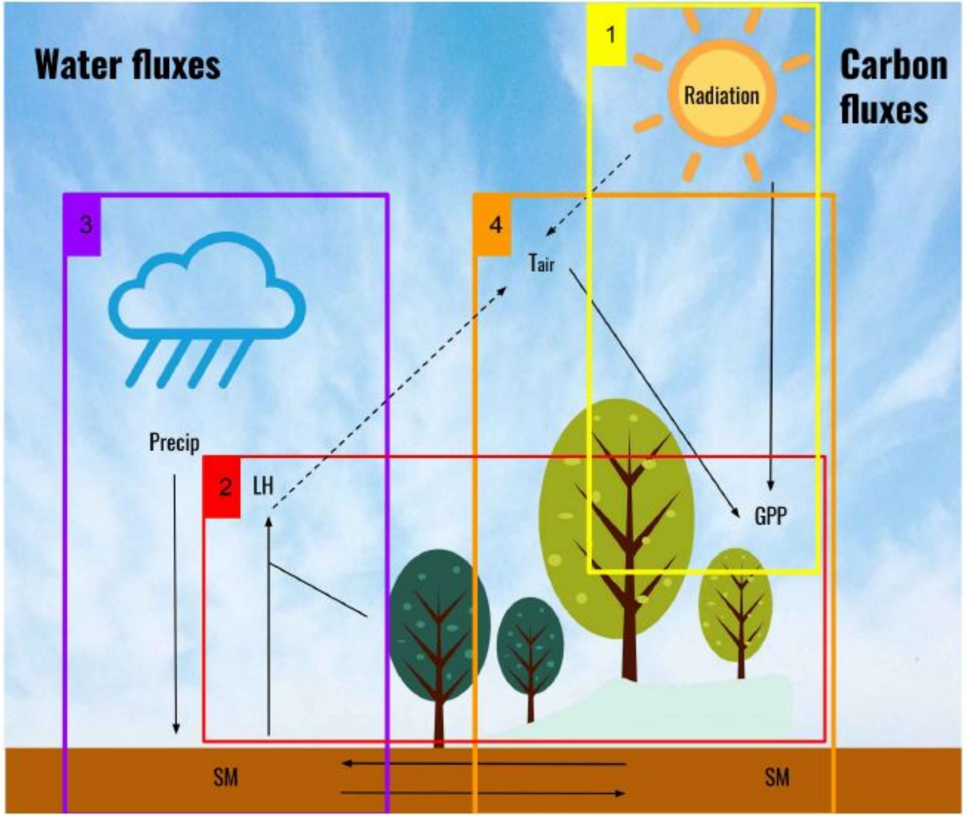
- Information-Geometric Causal Inference (IGCI) (Janzing et al, 2012):
  - method for instantaneous causal relationships.
  - works well in deterministic data: compatible with CCM.

- **RCCM: CCM + IGCI**

- **Systematic robust estimation of embedding dimension E:**
  - pool data across time
  - apply algorithm "pixel-wise" to obtain spatial maps of causality

Janzing, Dominik et al. " Information-geometric approach to inferring causal directions..", Artificial Intelligence , (2012).

RCCM to understand carbon and water cycle spatial patterns



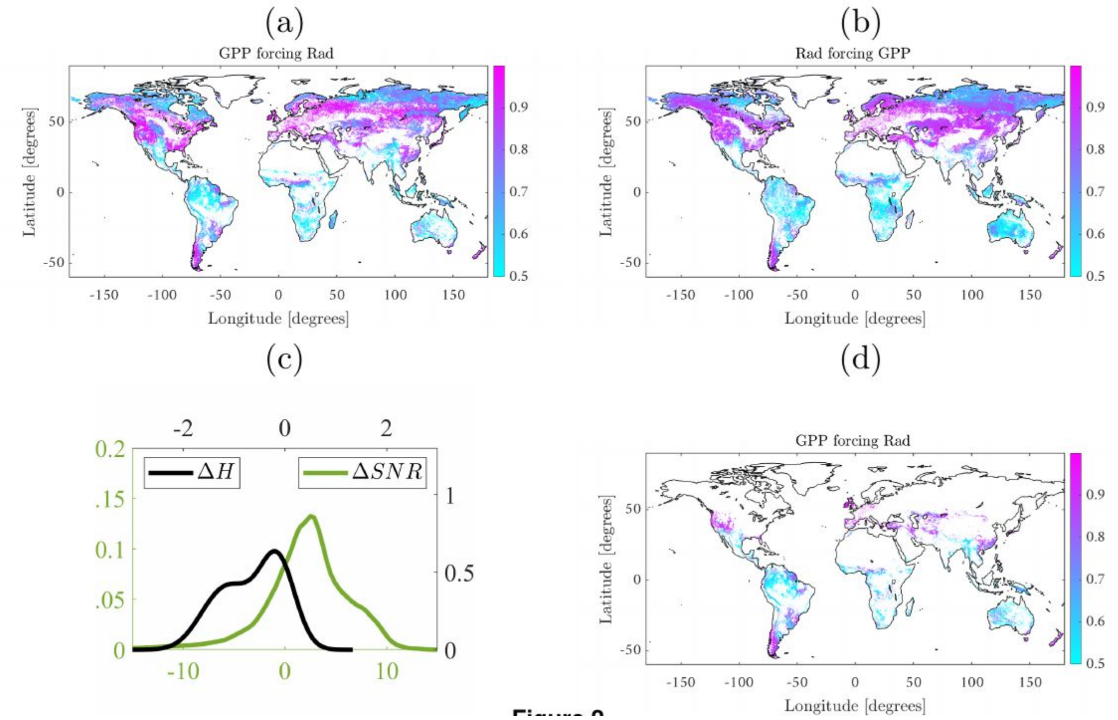
Spatio-temporal data cubes

- Earth System Data Lab (Mahecha et al., 2020)
  - 6 biosphere & atmosphere global gridded products
  - GPP, SM, Tair, LH, Precip, Rad
- 8-daily temporal resolution
- 2001-2011
- 0.0833 degrees spatial resolution

Credits:
Mahecha 2020
ESA ESDL project

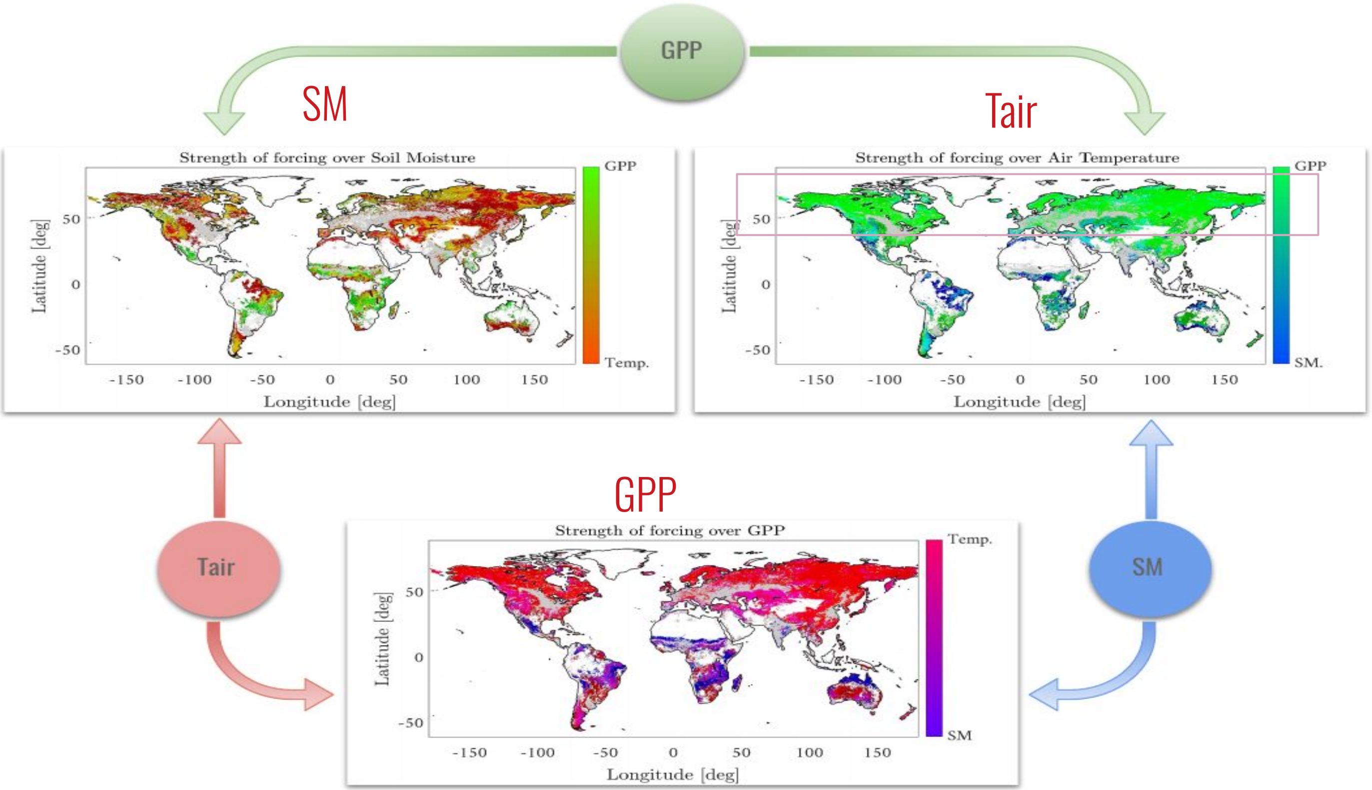
Mahecha, Miguel et al. "Earth system data cubes unravel global multivariate dynamics", Earth system dynamics , (2020).

Radiation and photosynthesis

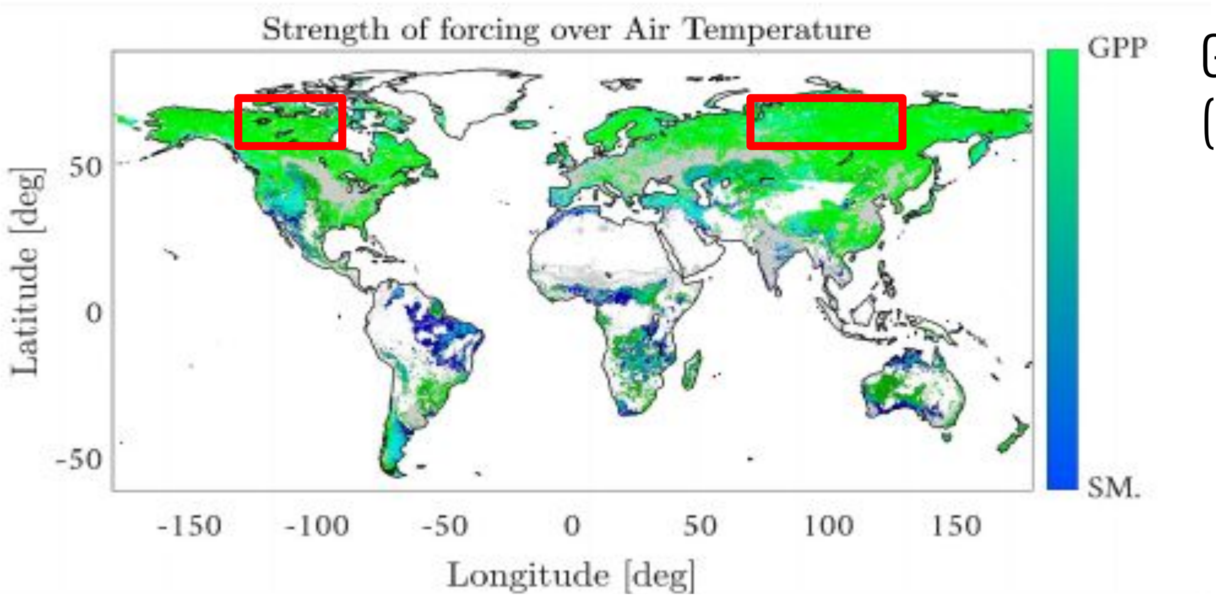


- RCCM mostly removes GPP→ Rad inference
- GPP → Rad in tropical and cloudy regions: increase in GPP increases Latent Heat, moistens atmosphere affecting cloud cover

Photosynthesis, temperature & soil moisture



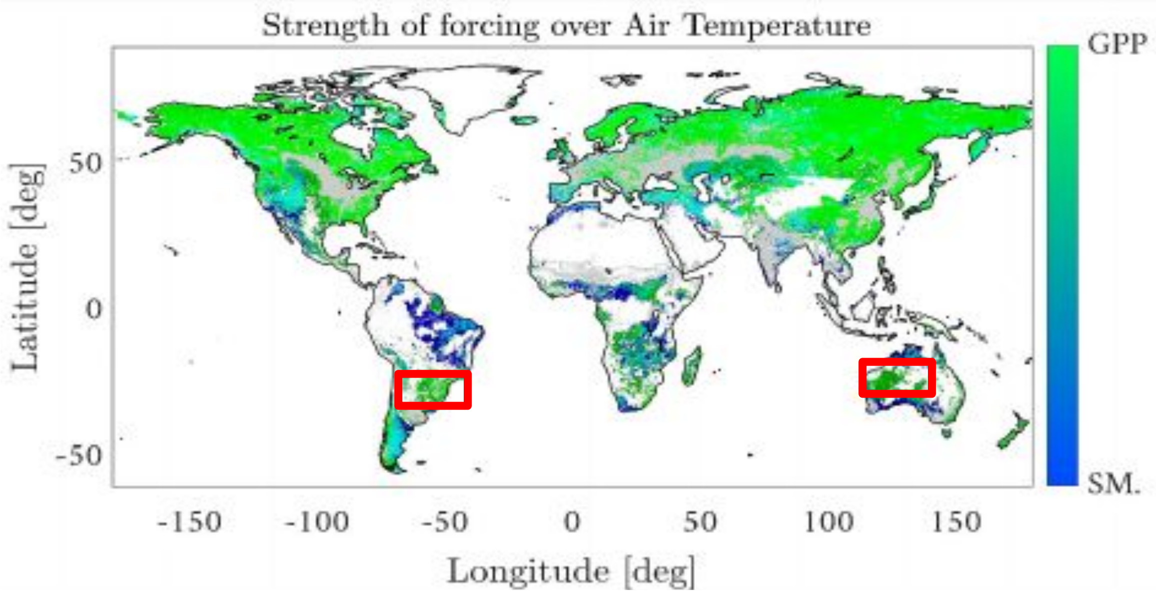
Photosynthesis, temperature & soil moisture



GPP drives Tair in many areas (green).

- Cold ecosystems: changes in land surface albedo such as snow/ice & vegetation.

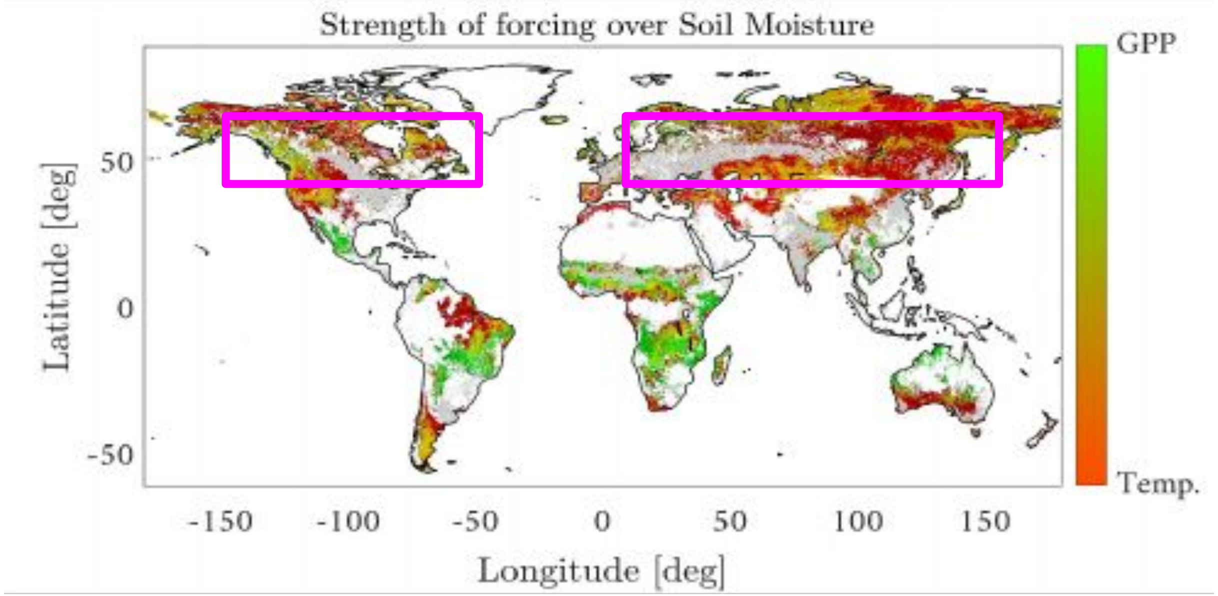
Photosynthesis, temperature & soil moisture



GPP drives Tair in many areas (green).

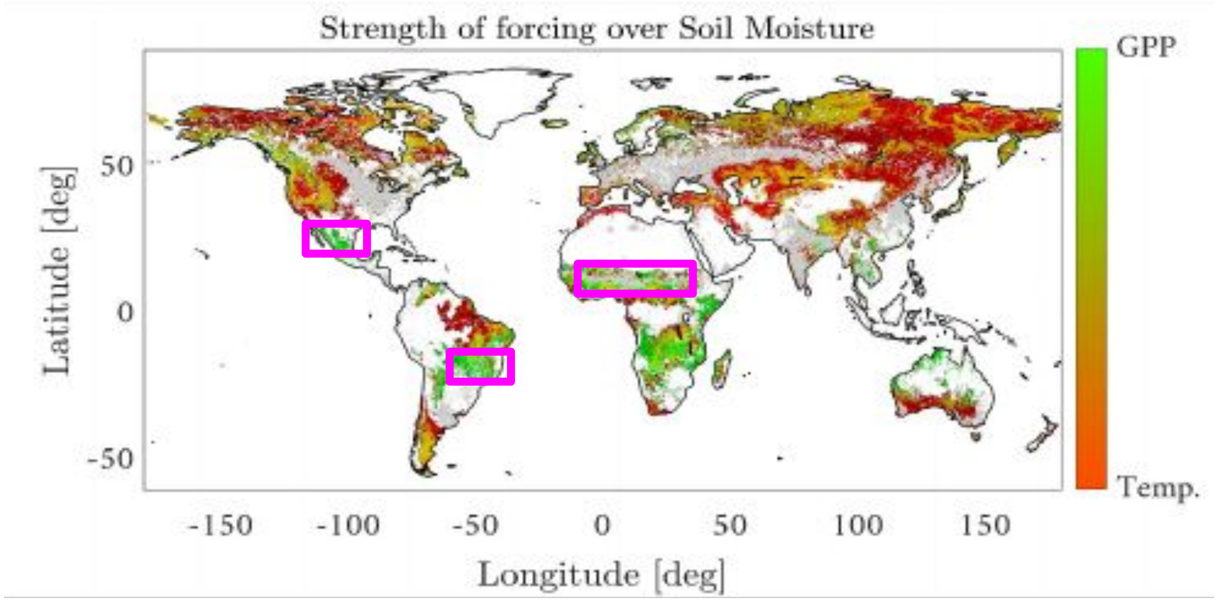
- Warmer and drier ecosystems: turbulent energy fluxes (enhancement of latent exchange and subsequent cooling effect)

Photosynthesis, temperature & soil moisture



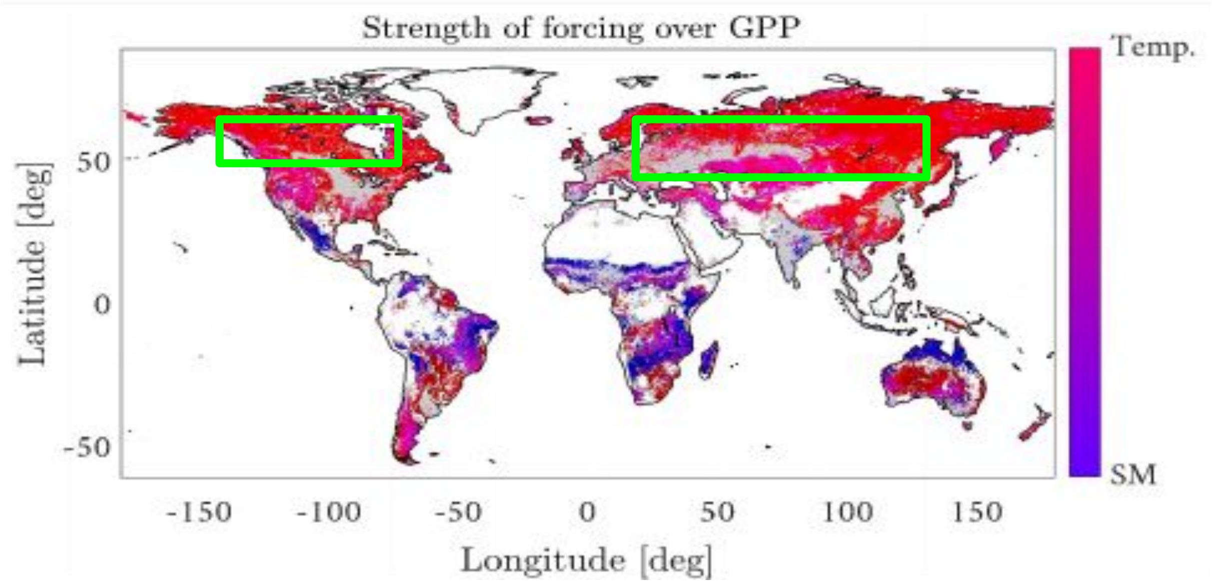
- SM mostly controlled by Tair (red)

Photosynthesis, temperature & soil moisture



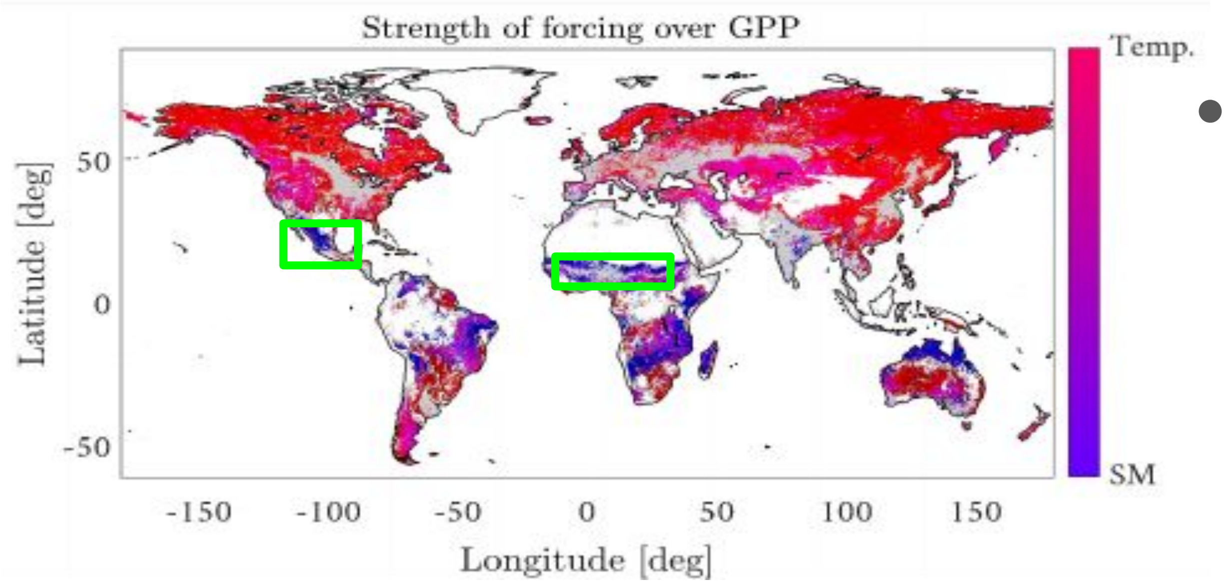
- SM controlled by GPP (green) in water-limited regions

Photosynthesis, temperature & soil moisture



- GPP dominated by Tair (red) in northern ecosystems where cold temp constrains photosynthesis. .

Photosynthesis, temperature & soil moisture



- GPP dominated by SM (blue) in transitional regions from wet to dry climates.

Natural interventions

Identifying the Causes of Pyrocumulonimbus (PyroCb)
Díaz Salas-Porras, E. Tazi, K, Braude, A. Okoh, D. Lamb, K.D. Watson-Parris, D. Harder, P. and Meinert, N.
NeurIPS 2022 Workshop-Causality for Real-world Impact, 2022

Take advantage of "natural" experiments

Causal discovery in Earth System science: **no experiments possible** on global scale, but different regimes act as "natural" interventions to create **experiment like data**.

Goal

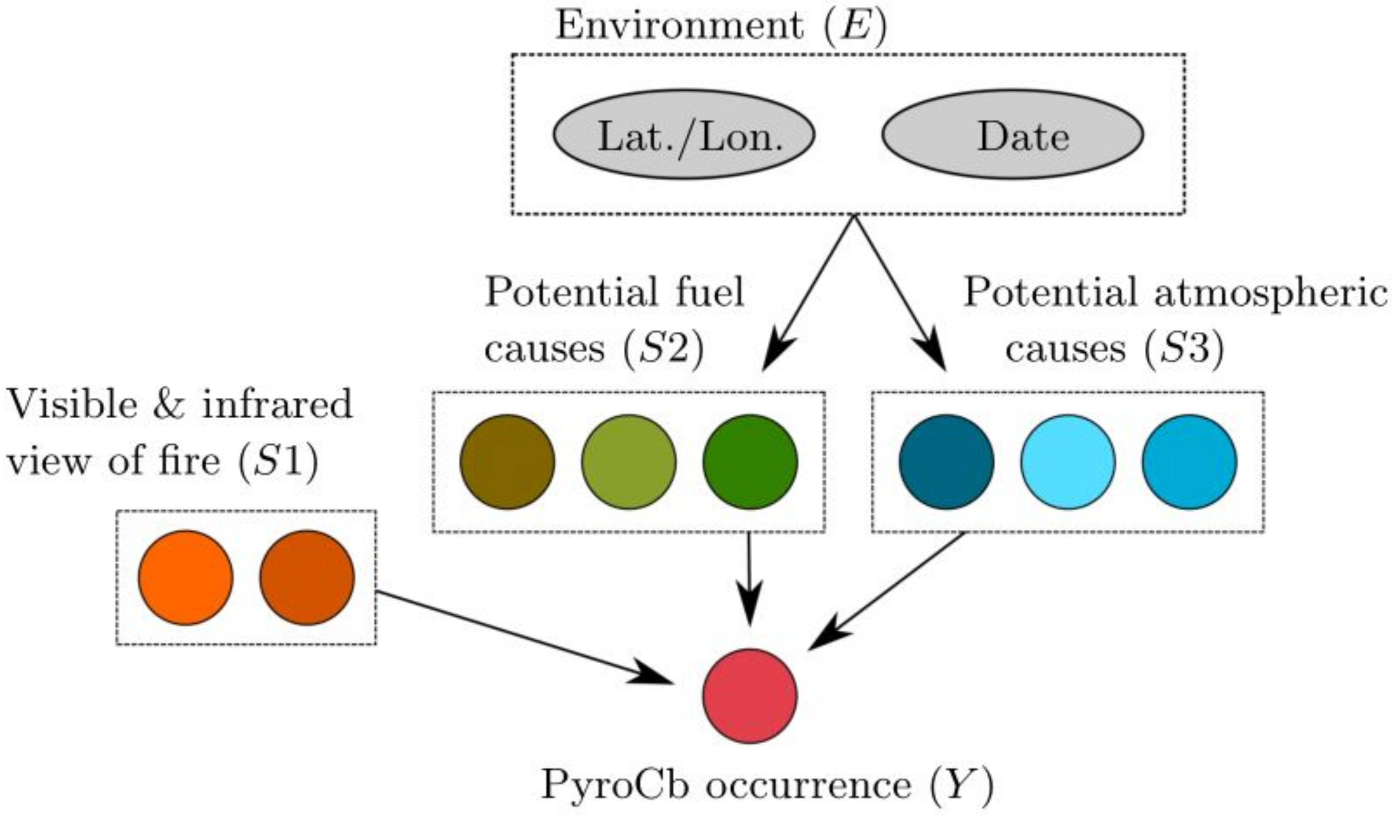
Exploit heterogeneity to find causal drivers of phenomenon, e.g. extreme wildfires (PyroCb)

Research question

pyroCb occurrence: why do some large fires generate pyroCb and others do not?


3. Clouds
4. Thunderstorm
2. Plume clouds
5. Downburst + lightning
1. Smoke plume
6. Unpredictable fire behaviour + new fires

Incorporate our knowledge of the system



Invariant Causal Prediction (ICP) (Peters et al, 2016)

To find the causes of Y :

1. For each subset S_i of candidate predictors do test H_i:

$$\boxed{H_i : Y \perp\!\!\!\perp E \mid S_i}$$

2. Take intersection of S_i, where H_i is not rejected, as causal predictors.

Peters, Jonas et al. "Causal inference by using invariant causal prediction: identification and confidence intervals",Journal of the Royal Society Series , (2016).

Data

28 variables total

atmospheric

fuel

thermal

~ 100 pyroCb events comprising ~6k hourly observations in North America and Australia

Variable	Description	Sensitive to
*ch*1	0.47 µm	smoke, haze
*ch*2	0.64 µm	terrain type
*ch*3	0.86 µm	vegetation
*ch*4	3.9 µm	thermal emissions & cloud ice crystal size
ch{5,6}	{11.2, 13.3} µm	thermal emissions & cloud opacity
{*u*,*v*}	{*u*,*v*} comp. of wind at 250 hPa	upper-level dynamics which influence rising motion
{*u*,*v*}10	10 m {*u*,*v*} component of wind	change in fire intensity and spread
*fg*10	10 m gusts since prev. post-processing	(same as above)
blh	boundary layer height	height of turbulent air at the surface
cape	convective available potential energy	energy for air to ascend into atmosphere
cin	convective inhibition	energy that will prevent air from rising
z	geopotential	energy needed for air to ascend into atmosphere as a function of altitude
{*slhf*, *sshf*}	surface {latent, sensible} heat flux	heat released or absorbed {from, neglecting} phase changes
w	surface vertical velocity	ascent speed of the plume from the wildfire
cv{*h*,*l*}	fraction of {high, low} vegetation	available fuel for the wildfire
type{*H*,*L*}	type of {high, low} vegetation	(same as above)
r{650,750,850}	rel. humidity at {650,750,850} hPa	condensation of vapour into clouds

Tazi, Kenza, et al, " Pyrocast: a machine learning pipeline to forecast pyrocumulonimbus(pyrocb) clouds.", Neurips, Workshop tackling climate change with machine learning (2022)

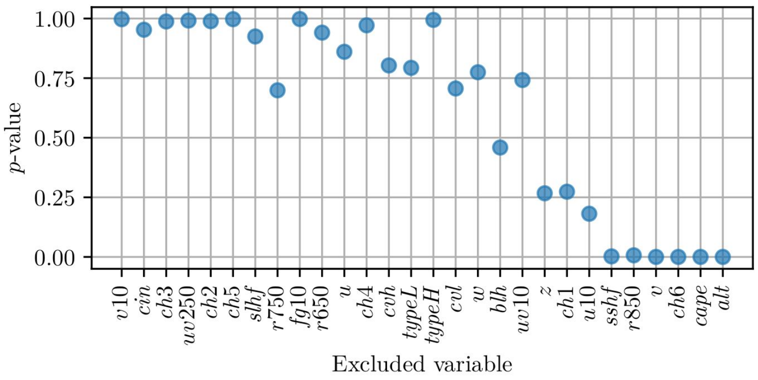
Non-linear conditional independence test for binary target

- **Test** based on difference between **reduced** model (excluding E) and **full** model (including E).
- Random Forest classification models
- Use (DeLong, et al. 1988) test for comparing the AUC of two models.

DeLong, Elizabeth, R, et al, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.", Biometrics. (1988)

ICP algorithm not feasible

- ICP: 28 variables in pyroCb dataset: 250 million tests!
- Greedy ICP: start with all candidate predictors and exclude one at a time: 406 tests
- Plot shows p-value of $\boxed{H_i : Y \perp\!\!\!\perp E \mid S_i}$ as we exclude variables with Greedy ICP



Six plausible causes of PyroCb

	variable	proxy for...
alt	altitude	energy needed to breach atmosphere
sshf	surface sensible heat flux	unstable boundary layer
ch6	13.3 μm reflectance	Very large and intense fire
r850	relative humidity at 850 hPa	Mid-tropospheric moisture source
v	component of wind at 250 hPa	Unstable atmosphere, conditions favorable for thunderstorms
cape	convective available potential energy	

Conclusions

Contribution within Causal Discovery landscape

		methods by causal assumption		
		non causal sufficiency	deterministic	ICM
type of data	i.i.d.		●	●
	time series	●		●
	spatio- temporal		●	
	interventional	●		

Conclusions

- Measuring complexity of CME is an effective way of unveiling causal asymmetries if kernel parameter selection done with care.

- Generative approach is advantageous for extending to spatio-temporal data and in practice needs only relaxed ICM assumptions.

- CCM and IGCI are complementary and can be used to produce spatial maps showing regions where different causal regimes are operating.

- ICP is suited to for Earth system sciences where different environments produce heterogeneity, but an important limitation is the number of candidate variables.

Contributions

- KCMC method including kernel parameter selection for CME.
- Generative LNc method including extension to time series and an additivity hypothesis test.
- Combination of CCM and IGCI to address problem of strong unidirectional forcing.
- Tools for implementing ICP: eg. conditional independence test, greedy ICP algorithm for large sets of predictors

PhD Thesis

Towards causal discovery for Earth system sciences

PhD Program “Enginyeria Electrònica” (3131)

Emiliano Diaz Salas Porras
Universitat de València

VNIVERSITAT DE VALÈNCIA

Advisors: Gustau Camps-Valls, Gherardo Varando, Adrián Pérez-Suay