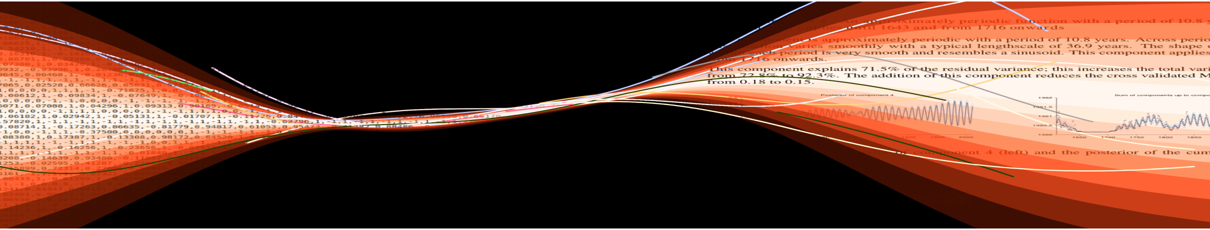
Inferring Causal Relations in EO:
Methods, Applications and a Web-platform



G. Camps-Valls[1], J. Muñoz Marí[1], E. Díaz[1]
A. Pérez-Suay[1], A. Moreno[1],
D. Sejdinovic[2] and J. Runge[3]

[1]IPL, Universitat de València

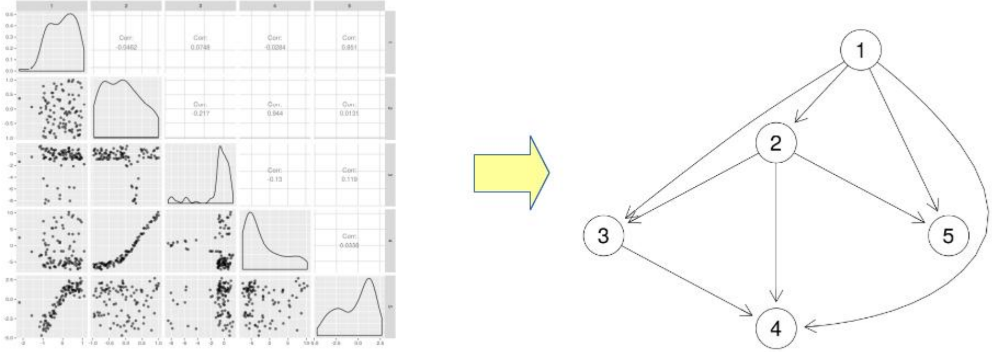
[2]Statistics Dep., Oxford Uni.

[3]Inst. Data Science., DLR

emiliano.diaz@uv.es | http://isp.uv.es

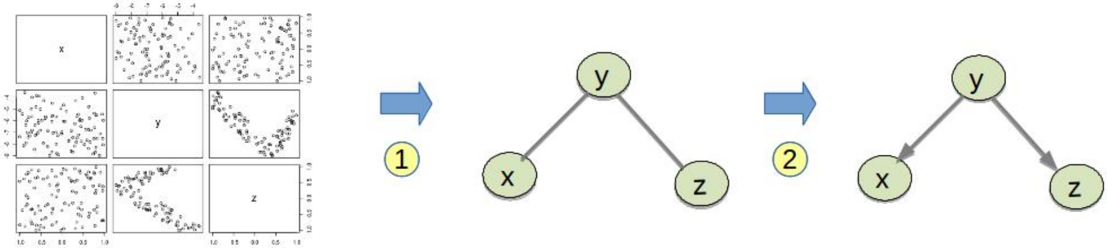
Goal of Causal Discovery

Given a system of p variables, with n observations available for each, learn underlying causal Directed Acyclic Graph (DAG)



Two step learning process

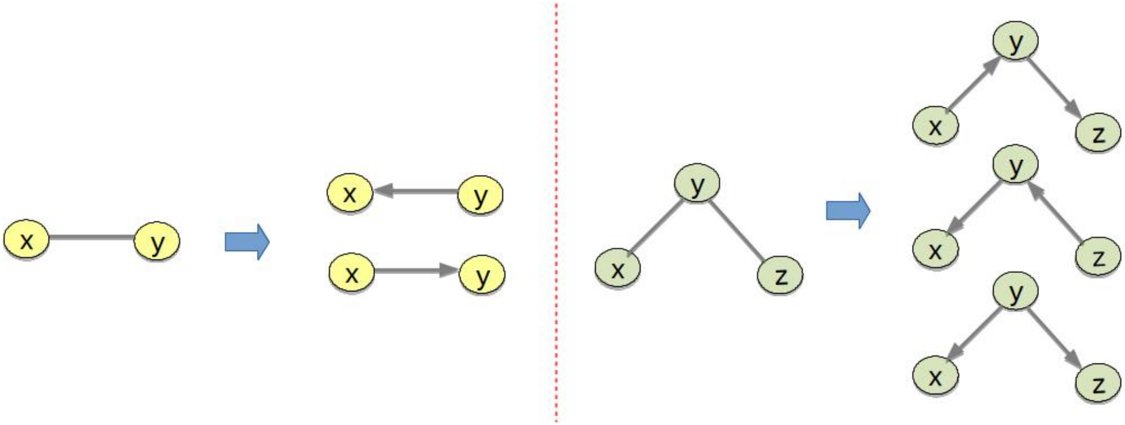
Learning a DAG can be separated into two steps:



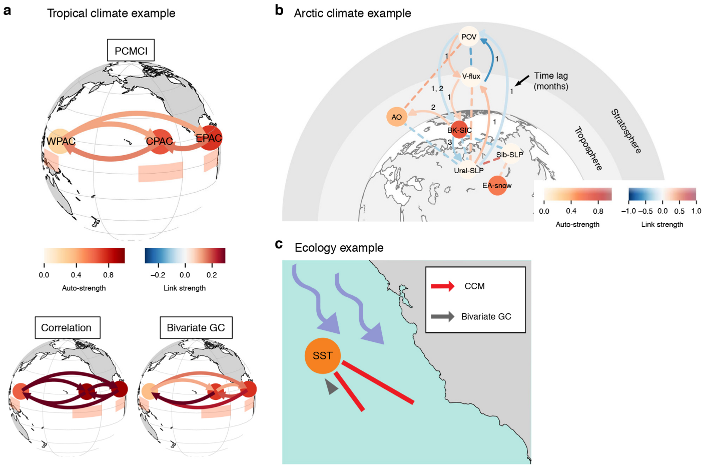
1. Learn conditional independencies (learn dag skeleton and colliders)
2. Learn directions (learn undetermined causal relations)

Some methods only do first others do both tasks.

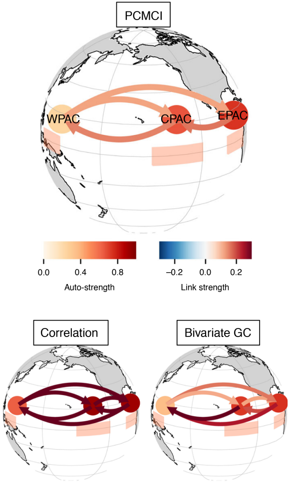
Second task for two and three variable examples



- Given that we know x and y dependent ($x \not\perp\!\!\!\perp y$): choose between $x \rightarrow y$ or $y \rightarrow x$
- Given that we know x and z conditionally independent given y ($x \perp\!\!\!\perp z | y$): choose between $x \rightarrow y \rightarrow z$, $x \leftarrow y \leftarrow z$ or $x \leftarrow y \rightarrow z$.


a
Tropical climate example
PCMCI
WPAC
CPAC
EPAC
0.0 0.4 0.8
Auto-strength
−0.2 0.0 0.2
Link strength
Correlation
Bivariate GC
b
Arctic climate example
POV
V-flux
AO
BK-SIC
Sib-SLP
Ural-SLP
EA-snow
Time lag (months)
Stratosphere
Troposphere
0.0 0.4 0.8
Auto-strength
−1.0 −0.5 0.0 0.5 1.0
Link strength
c
Ecology example
SST
CCM
Bivariate GC

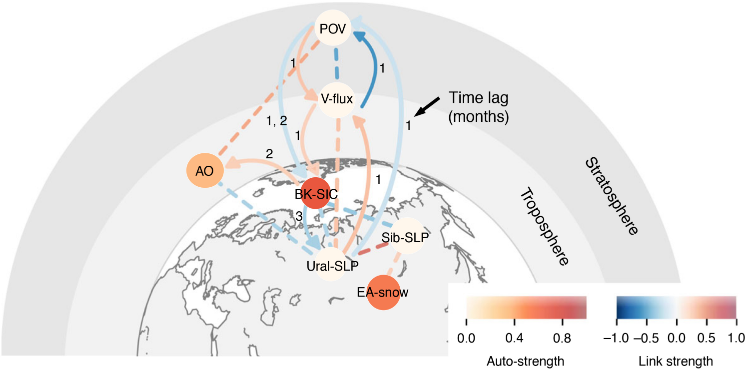
Tropical climate



Walker Circulation

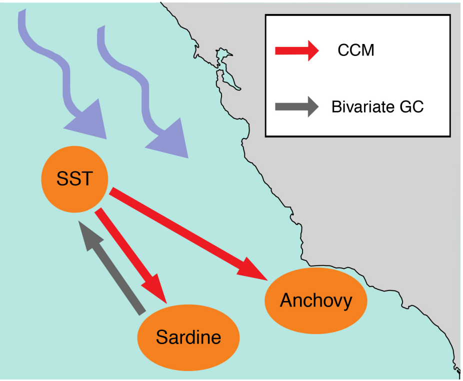
- EPAC → CPAC: trade winds carry anomolous warm surface air in EPAC to CPAC.
- CPAC → WPAC: warm moist air over CPAC rises to troposphere over WPAC.
- WPAC → CPAC, WPAC → EPAC: cool dry air cools and drops eastward.
- **Challenges**: Spurious correlations need to be detected.

Arctic climate



- Sea ice concentration (SIC) important driver of mid-latitude winter circulation.
- **Challenges**: strong autocorrelation, feedback loops, spurious correlation, time lags, Time Space (3D) aggregation and subsampling.

Ecology



- Sea Surface Temperature (SST) common driver of anchovy and sardine populations.
- Sardine $\rightarrow$ SST not plausible yet detected by traditional correlation method.

Methodological challenges for EO: complex spatio-temporal systems

Challenges

Process:

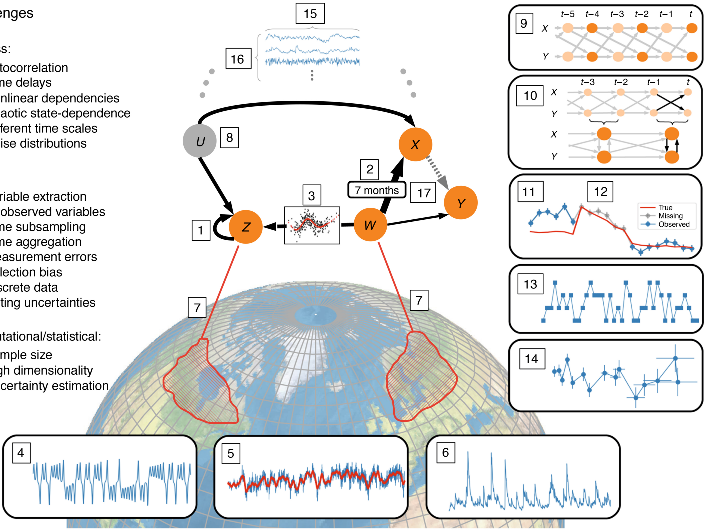
1 Autocorrelation
2 Time delays
3 Nonlinear dependencies
4 Chaotic state-dependence
5 Different time scales
6 Noise distributions

Data:

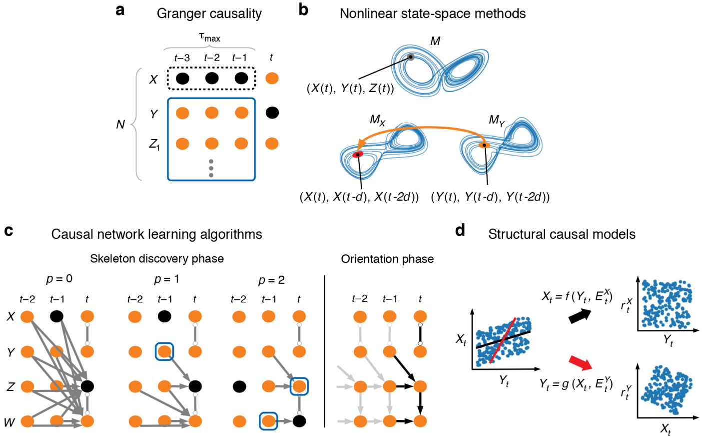
7 Variable extraction
8 Unobserved variables
9 Time subsampling
10 Time aggregation
11 Measurement errors
12 Selection bias
13 Discrete data
14 Dating uncertainties

Computational/statistical:

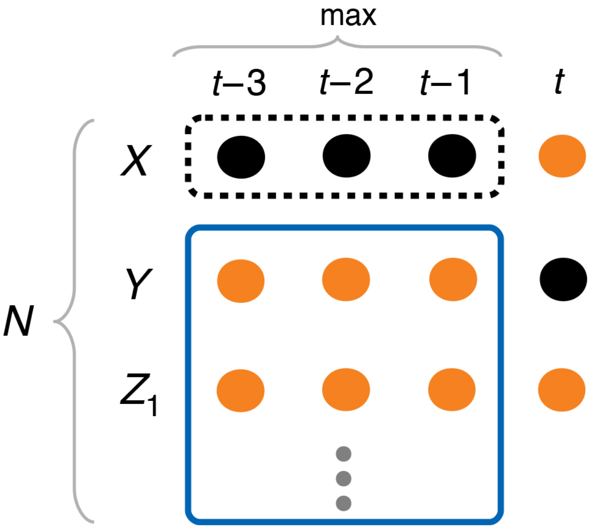
15 Sample size
16 High dimensionality
17 Uncertainty estimation



Overview of causal inference methods

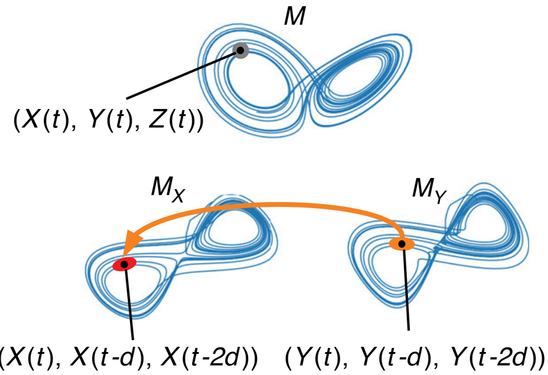


Granger Causality



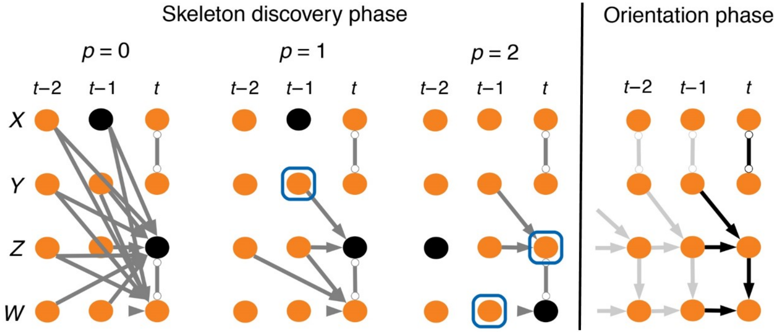
- **Method**: based on prediction error. Omitting past of X increases prediction error for Y?
- **Advantages**: Simple VAR framework, non-linear multivariate extensions, significance assessment.
- **Disadvantages**: No contemporaneous links (sub-sample series may lead to issues), high-dimensional, deterministic, unobserved variables cause problems.

Non-linear state space methods



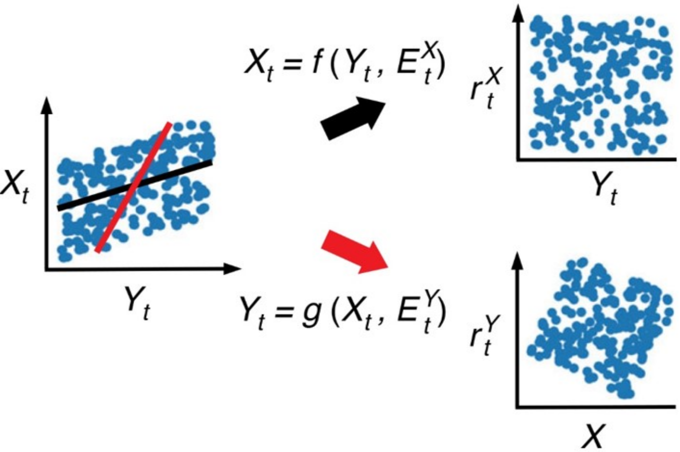
- **Method**: Dynamical system perspective: X and Y causally linked if they belong to same deterministic dynamic system.
- By Takens Theorem this dynamic system can be recovered from time delay embedding of each time series.
- Use time restriction: do you need past or future of Y to reconstruct state X?
- **Advantages**: Works for deterministic , state dependent, non-linear systems
- **Disadvantages**: No significance assessment, doesn't work for stochastic, high-dimensional systems with unobserved variables.

Causal network learning algorithms



- **Method**: Conditional independencies arise from causal structure. Use statistical independence tests to learn this causal structure.
- **Advantages**: High dimensionality, unobserved variables, nonparametric tests, time delays, strong autocorrelation
- **Disadvantages**: Multiple hypothesis testing problem, deterministic effects, time scales, contemporaneous feedback cycles

Structural causal methods

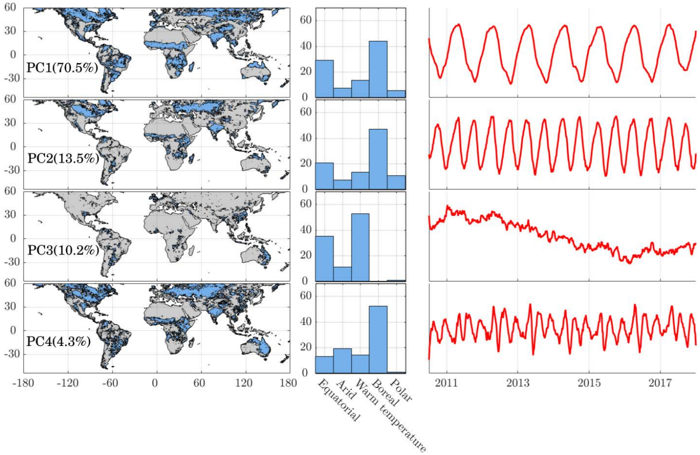


- **Method**:Make assumptions about the type of causal mechanisms creating systems. Model under this assumptions; check assumptions satisfied; use model fit and simplicity of model.
- **Advantages**: Contemporaneous effects, hig dimensionality, score-based methods available.
- **Disadvantages**: contemporaneous feedback cycles, deterministic effects, unobserved variables.

4 methods from 3 of the families

1. ROCK PCA + Cross-Information Kernelized Granger Causality (GC family)
2. Unbiased CCM (Non-linear state space method family)
3. Sensitivity measures for ANMs (Structural Causal Model family)
4. Multivariate conditional kernel deviance measures (Structural Causal Model family)

ROCK PCA + XKGC: ROCK PCA



Rock-PCA tackles spatio-temporal sub-sampling and aggregating issues by finding latent drivers and their *share* at each location.

ROCK PCA + XKGC: XKGC

Granger Causality (GC)

$$Y_{t+1} = a^\top X_t + \varepsilon_t^Y$$
$$Y_{t+1} = b_1^\top Y_t + b_2^\top X_t + \varepsilon_t^{Y|X}$$
$$X \to Y \leftrightarrow \mathbb{V}[\varepsilon_t^Y] \ll \mathbb{V}[\varepsilon_t^{Y|X}]$$

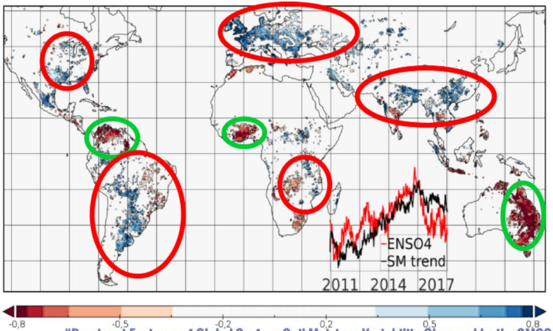
Cross-Information Kernelized GC (XKGC)

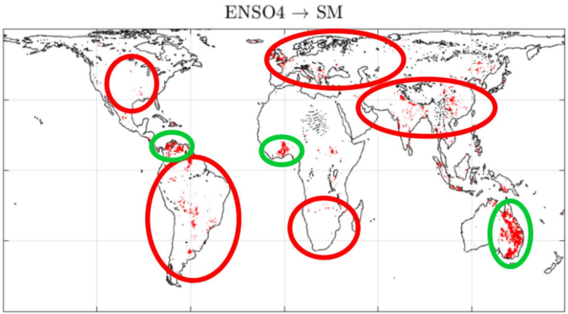
$$a_H = (K(X_t, X_t') + \varepsilon_t^Y)^{-1} Y_{t+1}$$
$$b_H = (L([Y_t, X_t], [Y_t', X_t']) + \varepsilon_t^{Y|X})^{-1} Y_{t+1}$$
$$X \to Y \leftrightarrow \mathbb{V}_H[\varepsilon_t^Y] \ll \mathbb{V}_H[\varepsilon_t^{Y|X}]$$

XKGC extends GC:

- Allows non-linear relationships.
- Generalized flexible kernel design that allows weight of features to be fitted.
- GP formulation improves hyperparameter fitting.

ROCK PCA + XKGC: XKGC

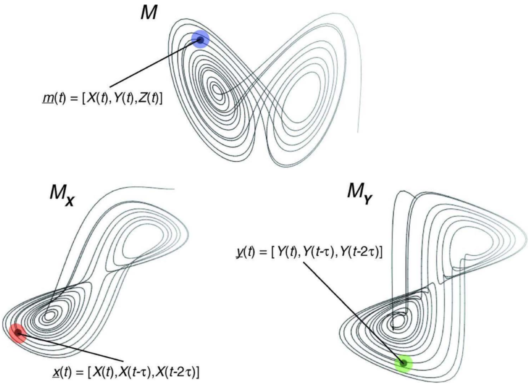


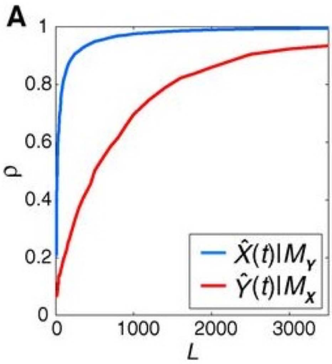


"Dominant Features of Global Surface Soil Moisture Variability Observed by the SMOS Satellite" M. Piles et al. Remote Sensing, 2019

- Causality is sharper than mere correlation! Some relationships causal other not.
- ENSO4 causes SM in very dry (Sahel) and very wet (tropical rain forest) regions.

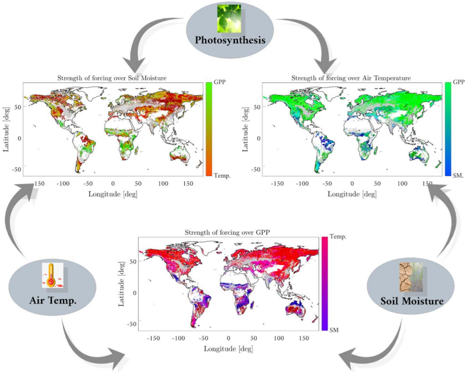
Unbiased CCM

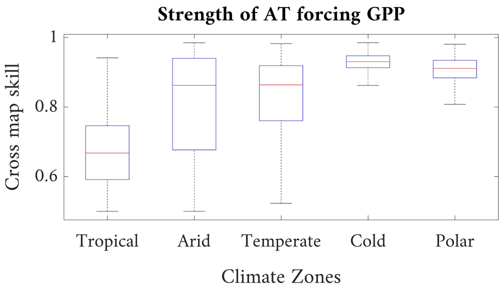




CCM pipeline was implemented in a more robust way by using bootstrap techniques to obtain estimates of time embedding dimension at each location.

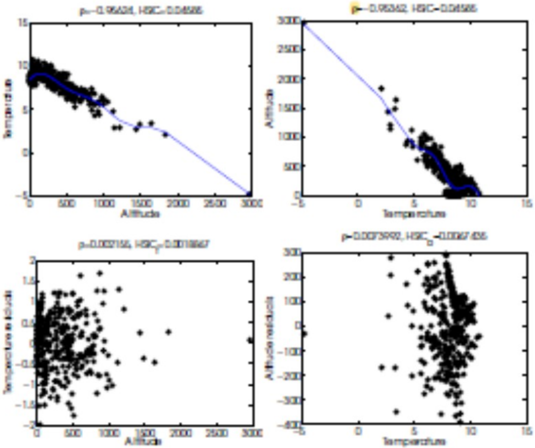
Unbiased CCM





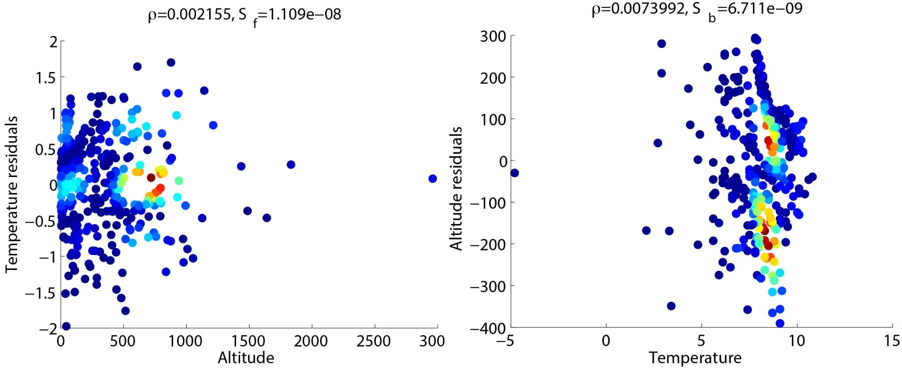
Inferred forcings of soil moisture (SM) and air temperature (AT) on gross primary production (GPP) reproduced some results in the literature, for example in cold and polar climate zones where AT limits GPP

Sensitivity measures in ANMs



Additive Noise Models are method from Structural Causal Model (SCM) model where additive model fitted in both directions ($x \to y$ and $y \to x$) and check which model

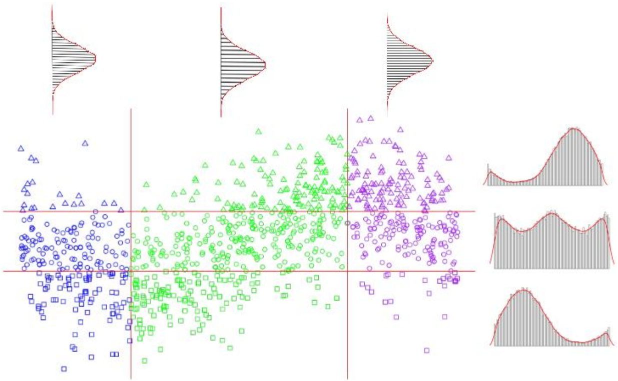
Sensitivity measures in ANMs



Instead of measuring independence of noise to predictor we look at the sensitivity of the overall independence measure to changes in the predictor and residuals. This is more **robust** measure as outliers may affect independence measures severely.

Multivariate conditional kernel deviance measures

Following figure shows observations from model $y = sin(x) + n$ where $n \sim N(0,1)$

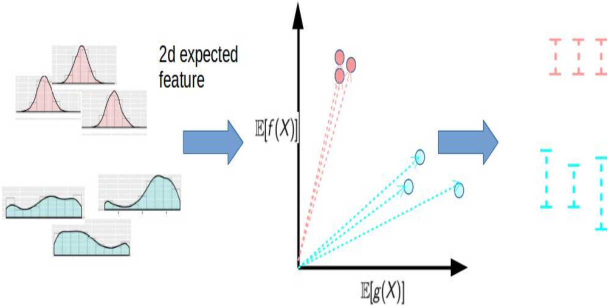


In causal direction $(x \rightarrow y)$ complexity of $p(y|x)$ does not depend on x whereas in anticausal direction $(y \rightarrow x)$ complexity of $p(x|y)$ varies more.

Multivariate conditional kernel deviance measures

Use the norm of vector of expected features as a proxy for complexity.

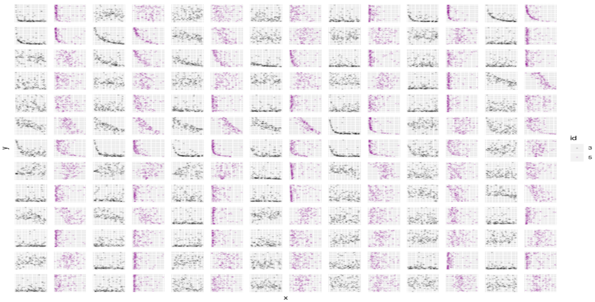
$$\mu = \mathbb{E}\,\phi(X) = \mathbb{E}\begin{bmatrix} f(X) \\ X^2 \\ \sin(X) \\ \vdots \\ g(X) \end{bmatrix}$$



Intuition:

1. Expected feature vector represents distribution if adequate features chosen.
2. Expected feature vector of similar distributions constrained to subspace of feature space and so have similar norms.

Multivariate conditional kernel deviance measures



RTM Prosail Simulated Pairs

- 182 data sets with 1000 pairs of points each
- max 100 points used
- causes consist of **7** biological parameters
- effects consist of reflectances for **13** different bands

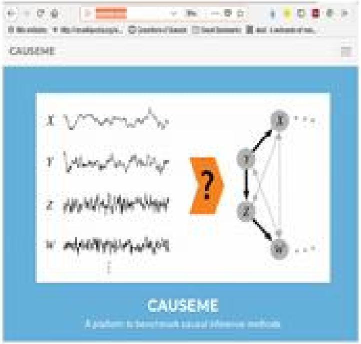
measure	ccr	auc
ANM	62.6 %	60.2 %
KCDC	97.8 %	99.3 %
SHSIC	-	65.0 %

Cause me: test your causal discovery algorithm online.

Extension to causal discovery

- CauseMe: http://causeme.uv.es
  - Download time series with ground truth
  - Run your causal discovery algorithm offline
  - Upload your causal graph
  - Get your results!

CauseMe: http://causeme.uv.es



"Inferring causation from time series with perspectives in Earth system sciences"
Runge, Bathiany, Bollt, Camps-Valls, et al. Nat Comm (submitted), 2018.
"Causal Inference in Geoscience and Remote Sensing from Observational Data,"
Pérez-Suay and Camps-Valls, IEEE Trans. Geosc. Rem. Sens. 2018

References

- Runge et al., 'Inferring causation from time series in Earth system sciences,' Nature communications, 2019.
- Mitrovic et al., 'Causal Inference via Kernel Deviance Measures,' NIPS 2018.
- Pérez-Suay and Camps-Valls, 'Causal Inference in Geoscience and Remote Sensing From Observationl Data,' IIIE-GRSL, 2019.
- Pérez-Suay and Camps-Valls, 'Sensitivity Maps of the Hilbert-Schmidt Independence Criterion,' Applied Soft Computing, 2017.
- Mooij et al., 'Distinguishing cause from effect using observational data,' JMLR, 17(1), 2016.
- Gretton et al., 'Measuring statistical dependence with Hilbert-Schmidt norms,' ALT, 2005
- Hoyer et al., 'Nonlinear causal discovery with additive noise models,' NIPS 2008.
- Camps-Valls, Mooij and Schölkopf, 'Remote sensing feature selection by kernel dependence measures,' IEEE-GRSL 2010