A comparison of two causal methods in the context of climate analyses

David Docquier[1] (david.docquier@meteo.be), Giorgia Di Capua[2,3], Reik V. Donner[2,3], Carlos A. L. Pires[4], Amélie Simon[4,5], Stéphane Vannitsem[1]

[1] Royal Meteorological Institute of Belgium, Brussels, Belgium, [2] Magdeburg-Stendal University of Applied Sciences, Magdeburg, Germany,
[3] Potsdam Institute for Climate Impact Research, Potsdam, Germany, [4] Instituto Dom Luiz, Universidade de Lisboa, Lisbon, Portugal, [5] IMT Atlantique, Lab-STICC, Brest, France

Background

Identifying causes of specific processes is crucial in order to better understand our climate system. Traditionally, correlation analyses have been used to identify cause-effect relationships in climate studies. However, correlation does not imply causation, which justifies the need to use causal methods.

2 Causal methods:

1) Liang-Kleeman information flow (**LKIF**; Liang, 2021): derived from first principles of information theory; computation of the rate of information transfer from one variable to another.
2) Peter and Clark momentary conditional independence (**PCMCI**; Runge et al., 2019): use of partial correlations to iteratively test conditional dependencies in a set of actors.

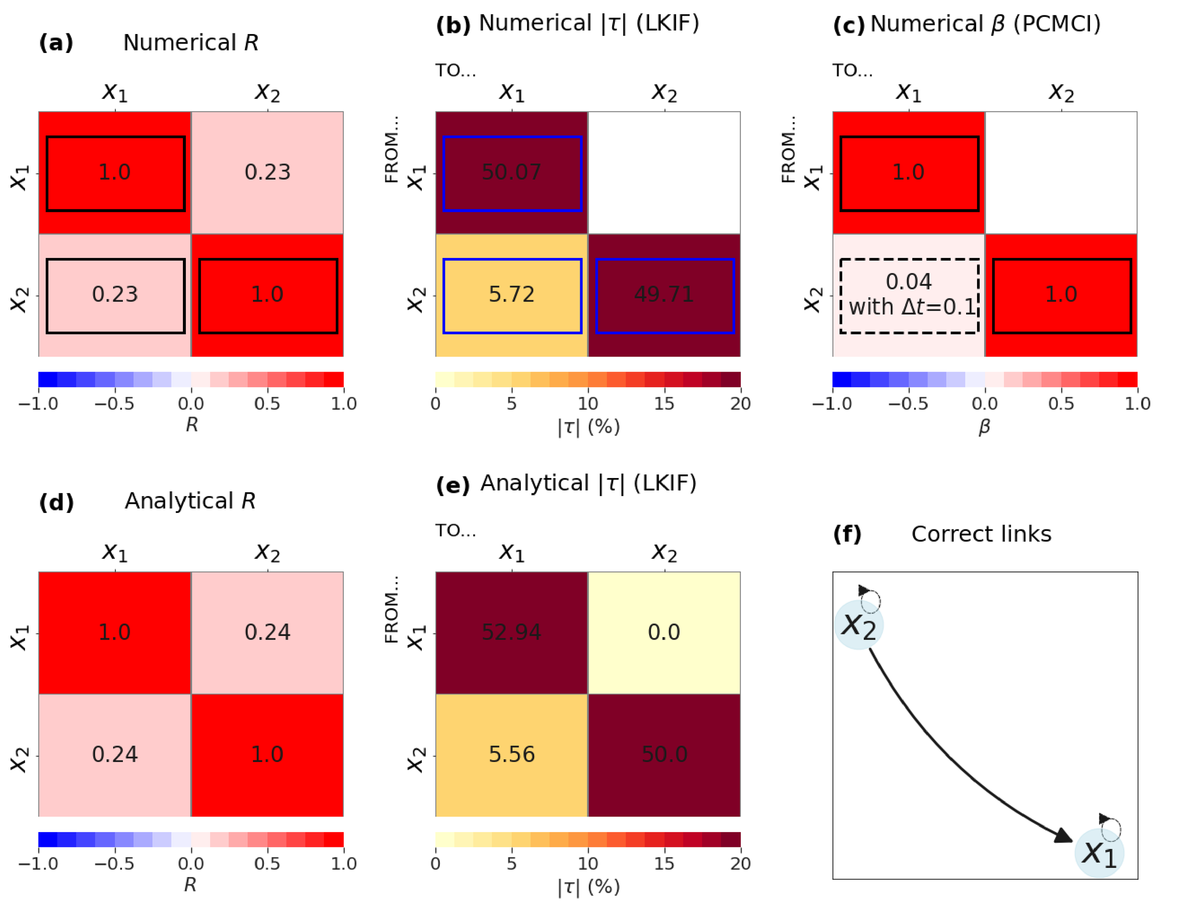
- Applied in a "linear setup" to 4 artificial models of increasing complexity and 1 real-world case study based on climate indices in the Atlantic and Pacific regions.

	LKIF	PCMCI
Full name	Liang–Kleeman information flow	Peter and Clark momentary conditional independence
Type of method	Information flow	Causal discovery algorithm
Use of time lags	Not by default	Always
Use of iterative conditioning	No	Yes
Metric	Rate of information transfer T (absolute) or τ (relative)	Path coefficient β
Unit	T: nat per unit time; τ: %	No unit
Sign meaning	$>0: x_j$ variability $\rightarrow x_i$ variability $\uparrow$ $<0: x_j$ variability $\rightarrow x_i$ variability $\downarrow$	$>0: x_j \uparrow \rightarrow x_i \uparrow$ $<0: x_j \uparrow \rightarrow x_i \downarrow$
Key references	Liang (2014, 2021)	Spirtes et al. (2001); Runge et al. (2019b)

2D model

$$dx_1 = (-x_1 + 0.5\,x_2)\,dt + 0.1\,dw_1,$$
$$dx_2 = -x_2\,dt + 0.1\,dw_2,$$

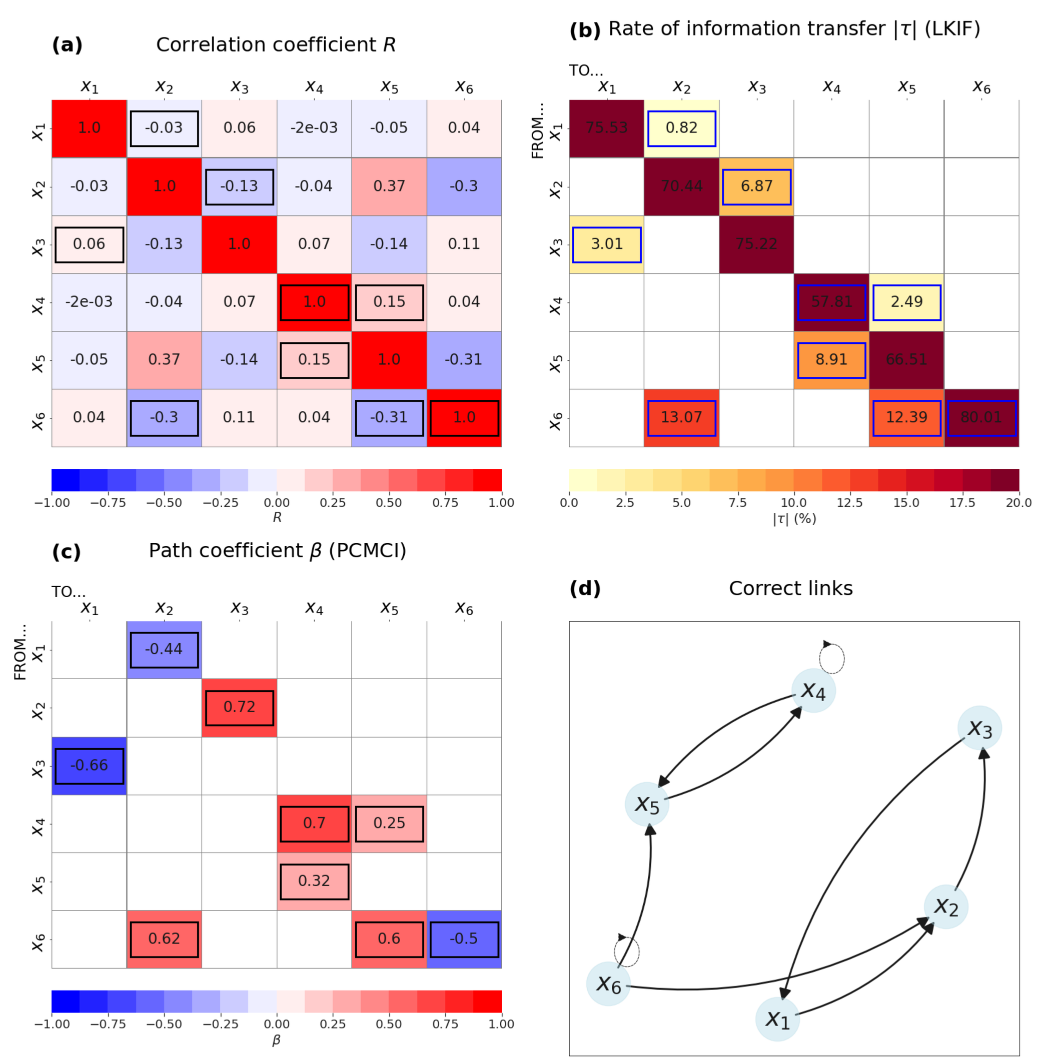
Liang (2014)



6D model

$$x_{1,t+1} = 0.1 - 0.6\,x_{3,t} + u_{1,t+1},$$
$$x_{2,t+1} = 0.7 - 0.5\,x_{1,t} + 0.8\,x_{6,t} + u_{2,t+1},$$
$$x_{3,t+1} = 0.5 + 0.7\,x_{2,t} + u_{3,t+1},$$
$$x_{4,t+1} = 0.2 + 0.7\,x_{4,t} + 0.4\,x_{5,t} + u_{4,t+1},$$
$$x_{5,t+1} = 0.8 + 0.2\,x_{4,t} + 0.7\,x_{6,t} + u_{5,t+1},$$
$$x_{6,t+1} = 0.3 - 0.5\,x_{6,t} + u_{6,t+1},$$

Liang (2021)



Results

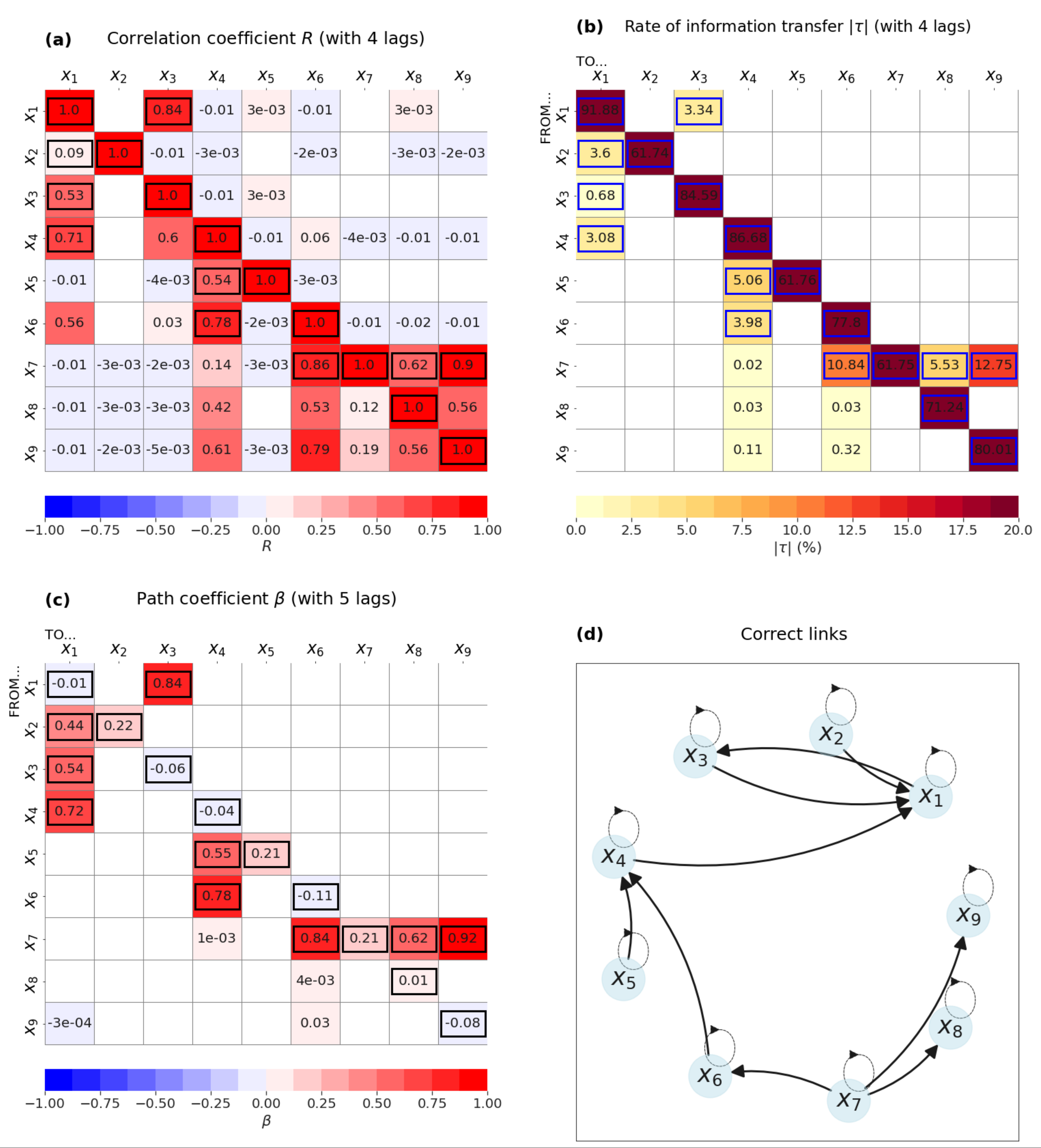
- **Both methods are superior to the classical correlation analysis**, especially in removing spurious links.
- Considering the three simplest models (2D, 6D and 9D models), LKIF performs better with a smaller number of variables (2D model) and PCMCI is best with a larger number of variables (9D model).
- Results with the Lorenz (1963) model are more challenging as the system is nonlinear and chaotic. Both methods only detect obvious causal links with original variables; other links appear when nonlinear variables (e.g. x^2) are used.
- For the real-world case study with climate indices, both methods present some similarities and differences at monthly timescale. One of the key differences is that LKIF identifies the Arctic Oscillation (AO) as the largest driver, while the El Niño-Southern Oscillation (ENSO) is the main influencing variable for PCMCI.

		Correlation	LKIF	PCMCI
2D model	True positives (1) [%]	100	100	0 (100)
	True negatives (1) [%]	0	100	100 (100)
	False positives [%]	100	0	0 (0)
	False negatives [%]	0	0	100 (0)
	ϕ coefficient	0	1	0 (1)
6D model	True positives (7) [%]	100	100	100
	True negatives (23) [%]	0	100	100
	False positives [%]	100	0	0
	False negatives [%]	0	0	0
	ϕ coefficient	0	1	1
9D model without lag	True positives (9) [%]	100	89	–
	True negatives (63) [%]	60	79	–
	False positives [%]	40	21	–
	False negatives [%]	0	11	–
	ϕ coefficient	0.40	0.50	–
9D model with lags	True positives (9) [%]	100	100	100
	True negatives (63) [%]	27	92	94
	False positives [%]	73	8	6
	False negatives [%]	0	0	0
	ϕ coefficient	0.21	0.77	0.81

9D model

$$x_{1,t} = 3.4\,x_{1,t-1}(1 - x_{1,t-1}^2)e^{-x_{1,t-1}^2} + 2.5\,x_{2,t-4} + 1.8\,x_{3,t-2} + 1.5\,x_{4,t-2} + 0.4\,u_{1,t},$$
$$x_{2,t} = 3.4\,x_{2,t-1}(1 - x_{2,t-1}^2)e^{-x_{2,t-1}^2} + 0.4\,u_{2,t},$$
$$x_{3,t} = 3.4\,x_{3,t-1}(1 - x_{3,t-1}^2)e^{-x_{3,t-1}^2} + 0.25\,x_{1,t-1} + 0.4\,u_{3,t},$$
$$x_{4,t} = 3.4\,x_{4,t-1}(1 - x_{4,t-1}^2)e^{-x_{4,t-1}^2} + 1.5\,x_{5,t-3} + 1.2\,x_{6,t-1} + 0.4\,u_{4,t},$$
$$x_{5,t} = 3.4\,x_{5,t-1}(1 - x_{5,t-1}^2)e^{-x_{5,t-1}^2} + 0.4\,u_{5,t},$$
$$x_{6,t} = 3.4\,x_{6,t-1}(1 - x_{6,t-1}^2)e^{-x_{6,t-1}^2} + 1.5\,x_{7,t-3} + 0.4\,u_{6,t},$$
$$x_{7,t} = 3.4\,x_{7,t-1}(1 - x_{7,t-1}^2)e^{-x_{7,t-1}^2} + 0.4\,u_{7,t},$$
$$x_{8,t} = 3.4\,x_{8,t-1}(1 - x_{8,t-1}^2)e^{-x_{8,t-1}^2} + 0.8\,x_{7,t-1} + 0.4\,u_{8,t},$$
$$x_{9,t} = 3.4\,x_{9,t-1}(1 - x_{9,t-1}^2)e^{-x_{9,t-1}^2} + 1.8\,x_{7,t-1} + 0.4\,u_{9,t},$$

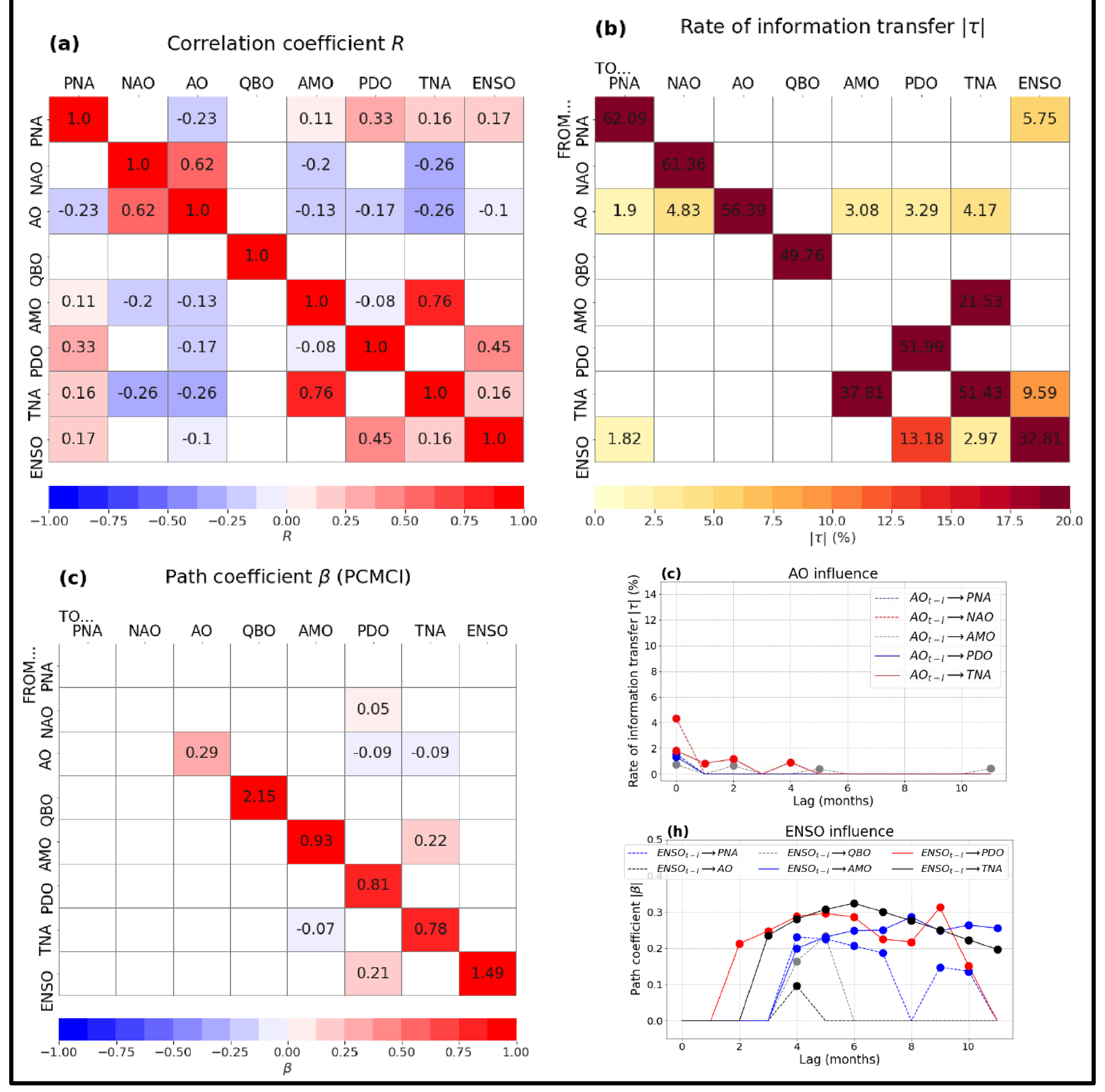
Subramaniyam et al. (2021)



Atmospheric indices:
- PNA: Pacific-North American
- NAO: North Atlantic Oscillation
- AO: Arctic Oscillation
- QBO: Quasi-Biennial Oscillation

Oceanic indices:
- AMO: Atlantic Multidecadal Oscillation
- PDO: Pacific Decadal Oscillation
- TNA: Tropical North Atlantic
- ENSO: Niño3.4 index

Climate indices



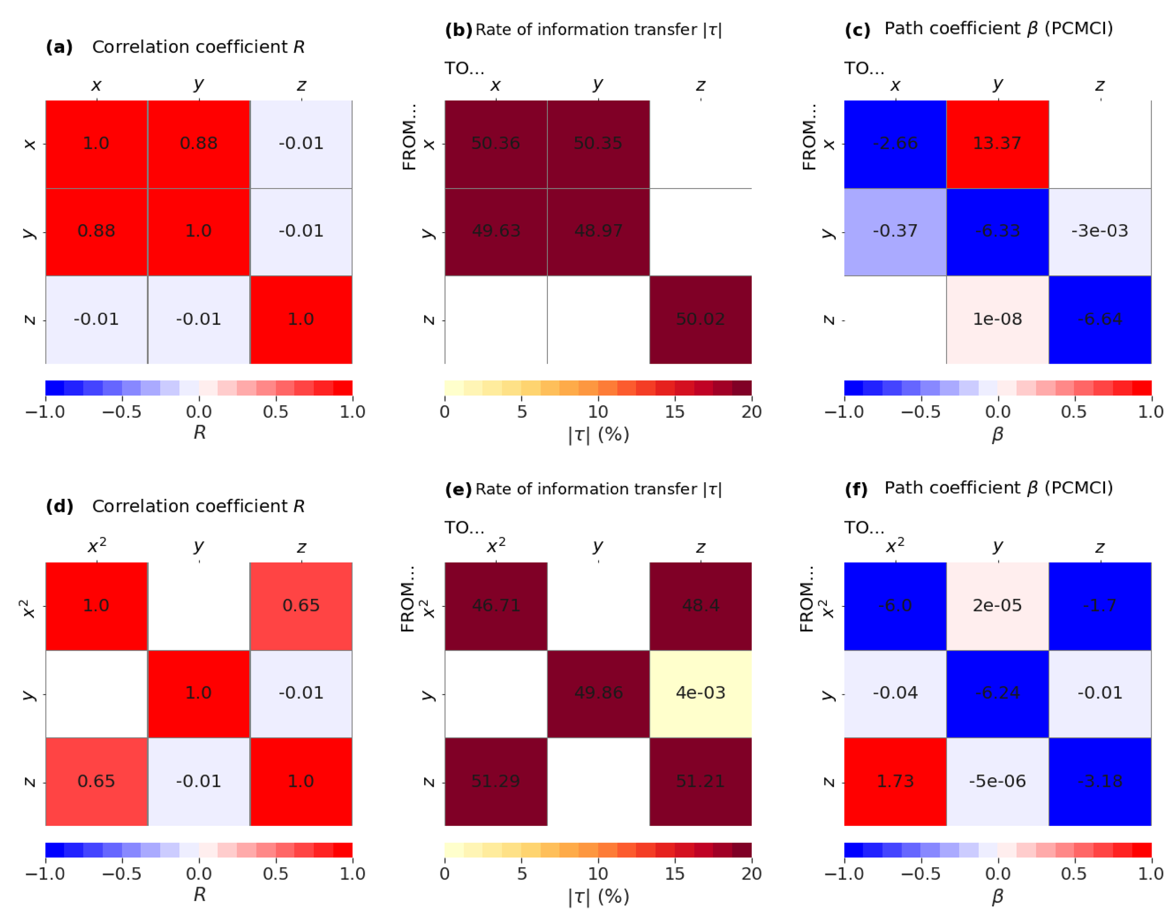
Lorenz (1963) model

$$\frac{dx}{dt} = 10(y - x),$$
$$\frac{dy}{dt} = 28\,x - y - x\,z,$$
$$\frac{dz}{dt} = x\,y - \frac{8}{3}z,$$

Lorenz (1963)

Panels (a)-(c) show results with the original (*x,y,z*) triplet

Panels (d)-(f) show results with the modified (x^2,*y,z*) triplet



Conclusions

- The 2 **causal methods** used here **should be preferred to correlation**, which suffers from e.g. random coincidence, absence of identification of external drivers, and application to 2D cases only.
- As both LKIF and PCMCI display strengths and weaknesses when used with relatively simple models in which correct causal links can be detected by construction, we do not recommend one or the other method but rather **encourage** the climate community **to use several methods whenever possible**.
- Both methods, as used here, **assume linearity**; Pires et al. (2024) have recently developed a nonlinear version of LKIF

- More details can be found in Docquier et al. (2024).