Inferring the connectivity and the community structure of a complex system from observed data

J. I. Deza, G. Tirabassi and C. Masoller

Universitat Politecnica de Catalunya
Cristina.masoller@upc.edu

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

Dynamics Days Latin America and the Caribbean, Puebla, Mexico, October 2016

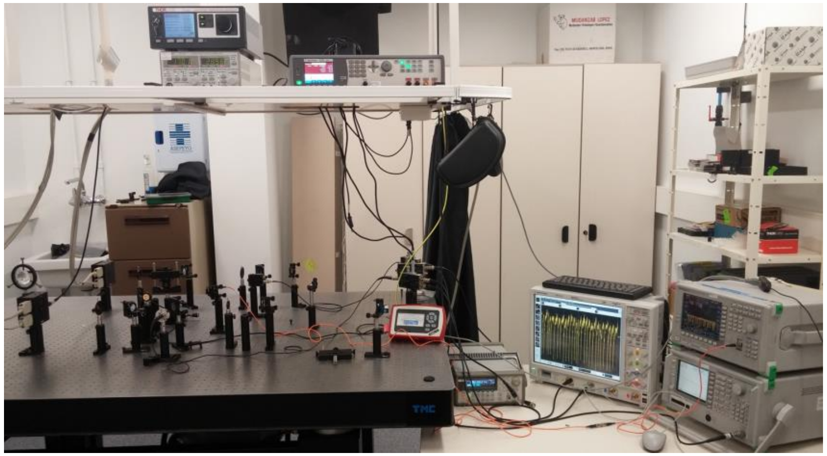
ICREA

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Our research: nonlinear and stochastic phenomena in complex systems

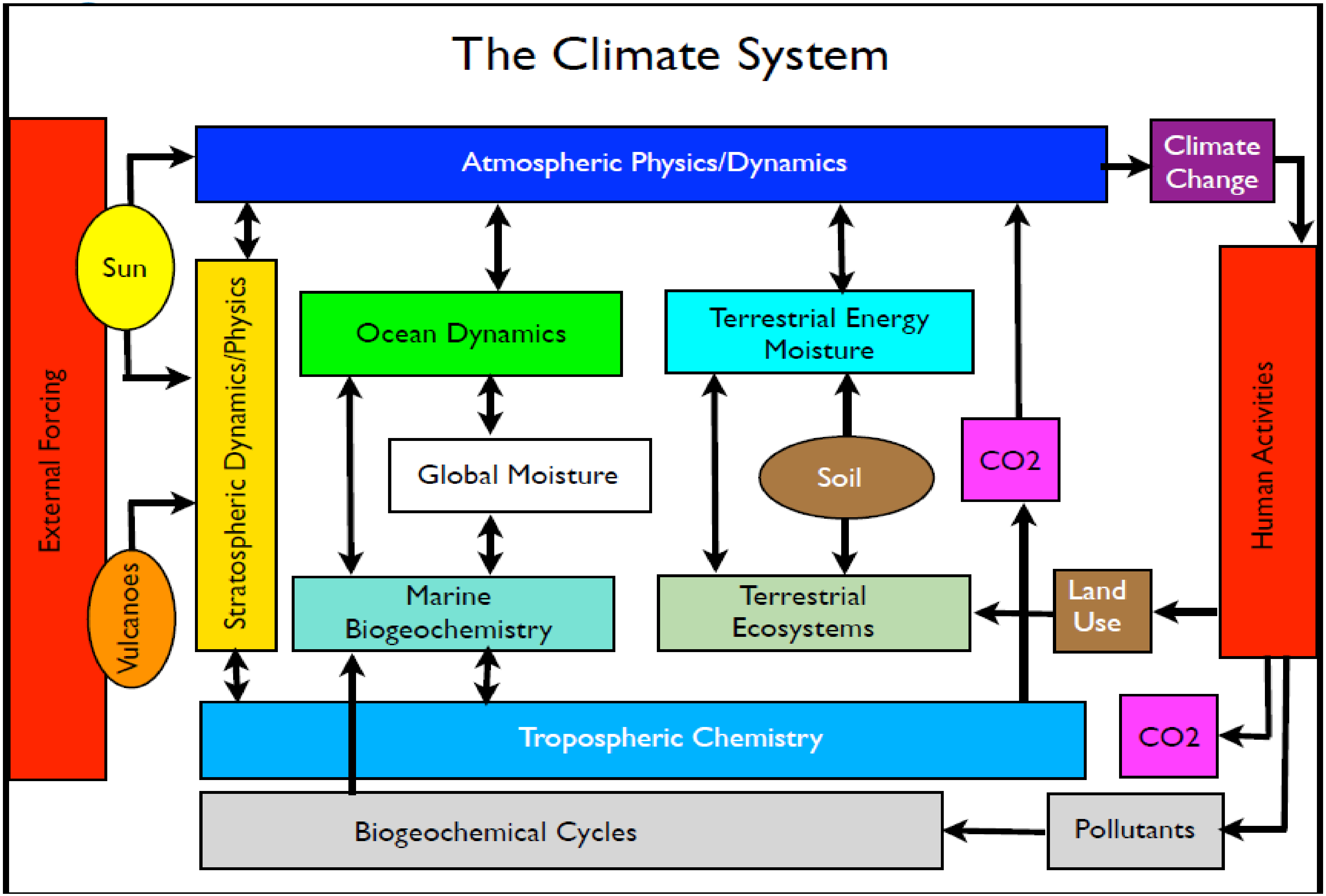


- Photonics (dynamics of laser, time-delay feedback),
- Biophysics (neuronal excitability, synchronization),
- Time series analysis (extreme events, tipping points and regime transitions, climate data analysis, biomedical signals),
- Complex networks (network inference, climate networks and communities).

WHAT DO NETWORKS HAVE TO DO WITH CLIMATE?

BY ANASTASIOS A. TSONIS, KYLE L. SWANSON, AND PAUL J. ROEBBER

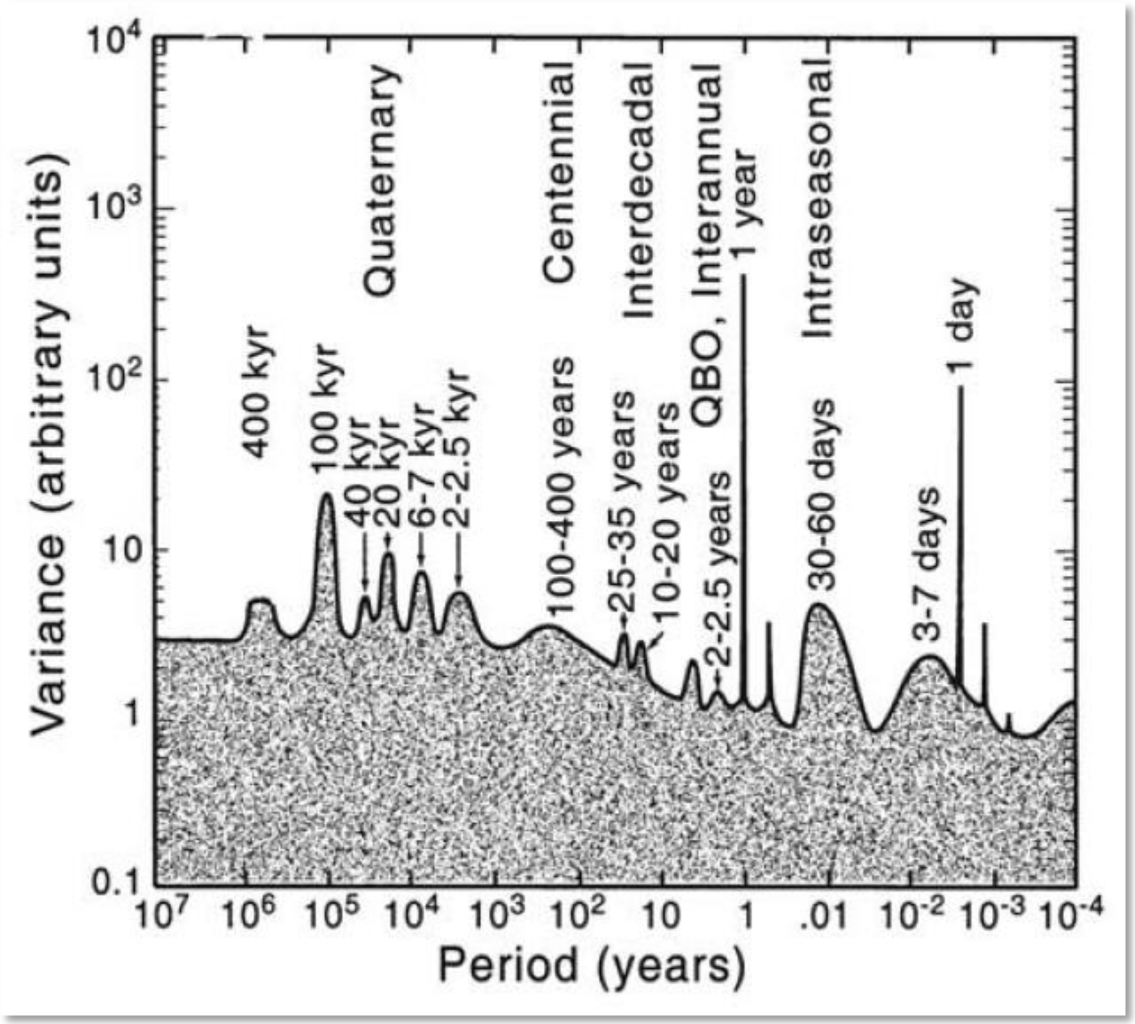
Advances in understanding coupling in complex networks offer new ways of studying the collective behavior of interactive systems and already have yielded new insights in many areas of science.



Courtesy of Henk Dijkstra (Ultrech University)

The climate system: a complex system with a wide range of time-scales

- hours to days,
- months to seasons,
- decades to centuries,
- and even longer...

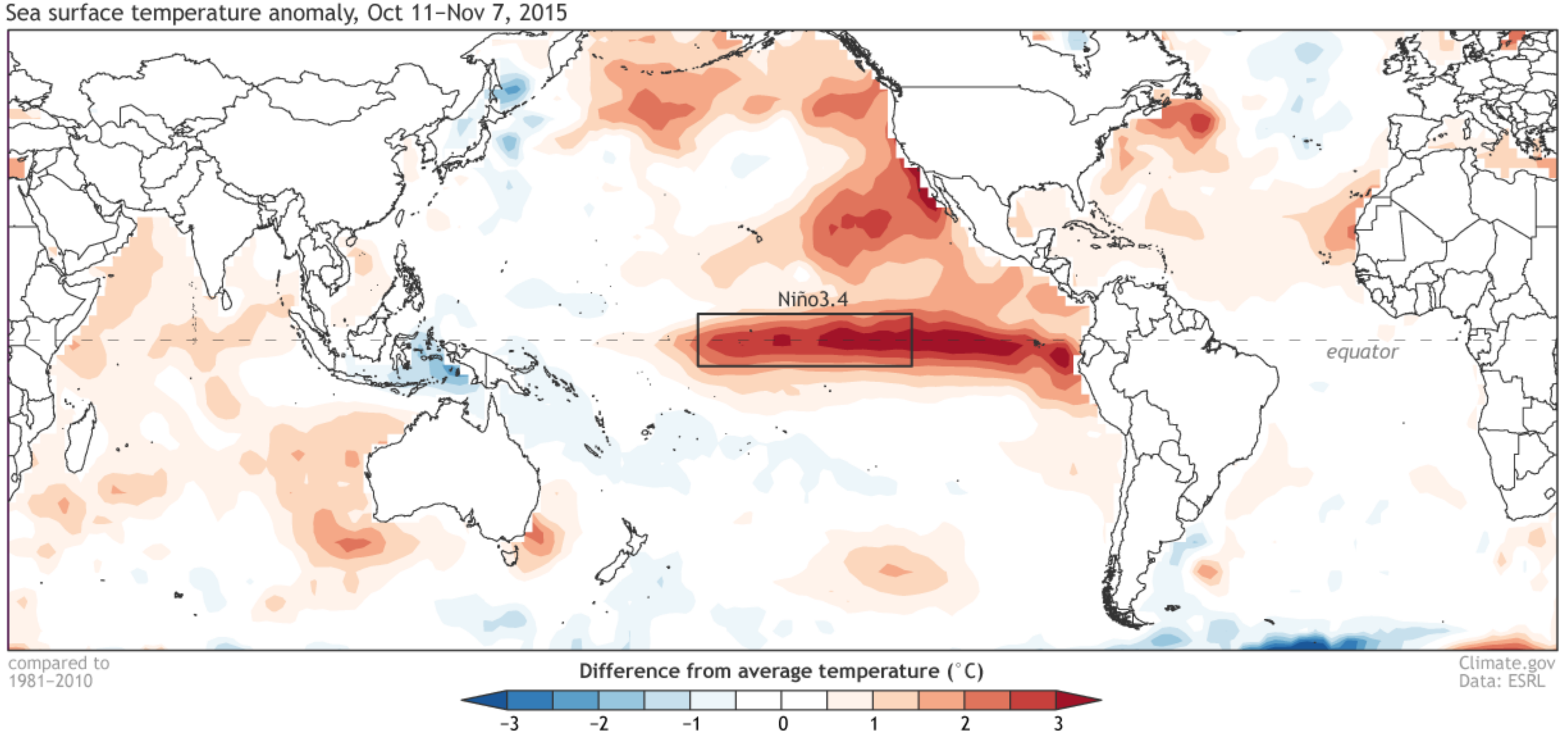


An “artist’s representation” of the power spectrum of climate variability (Ghil 2002).

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

And a wide range of spatial modes of variability

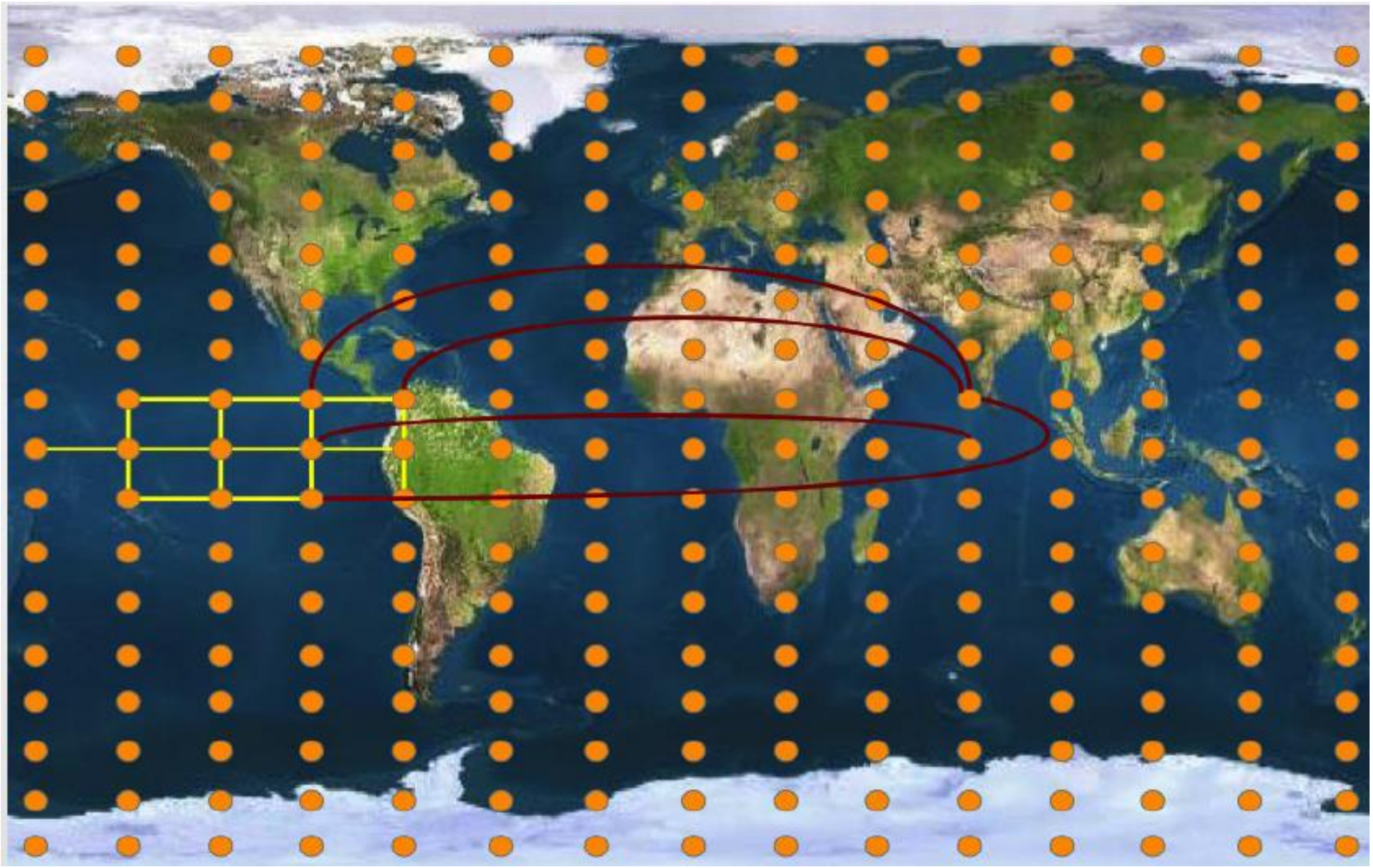


- ENSO
- The Atlantic multi-decadal oscillation
- The Indian Ocean Dipole
- The Madden–Julian oscillation
- The North Atlantic oscillation
- The Pacific decadal oscillation
- Etc.

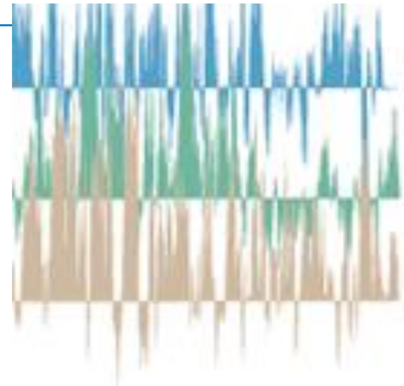
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

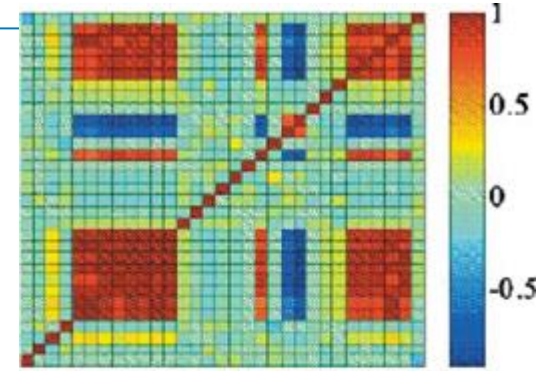
Nodes and links of the climate network



Time series of a climate variable (air temperature, wind, precipitation, etc.)

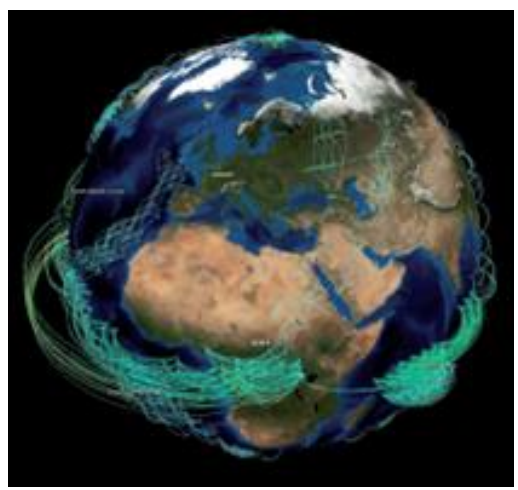


Similarity measure (correlation, mutual information, etc.)

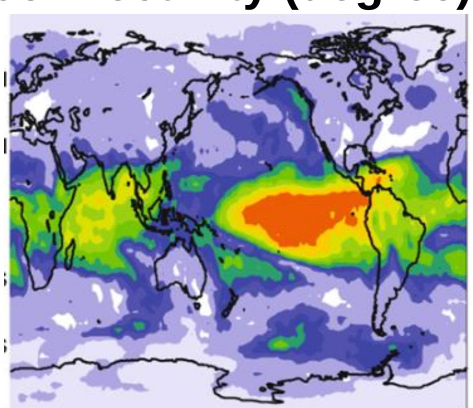


Thresholded matrix (adjacency matrix)

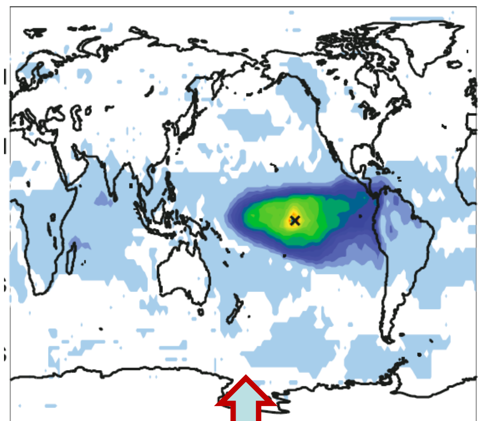
Climate network



Area-weighted connectivity (degree)



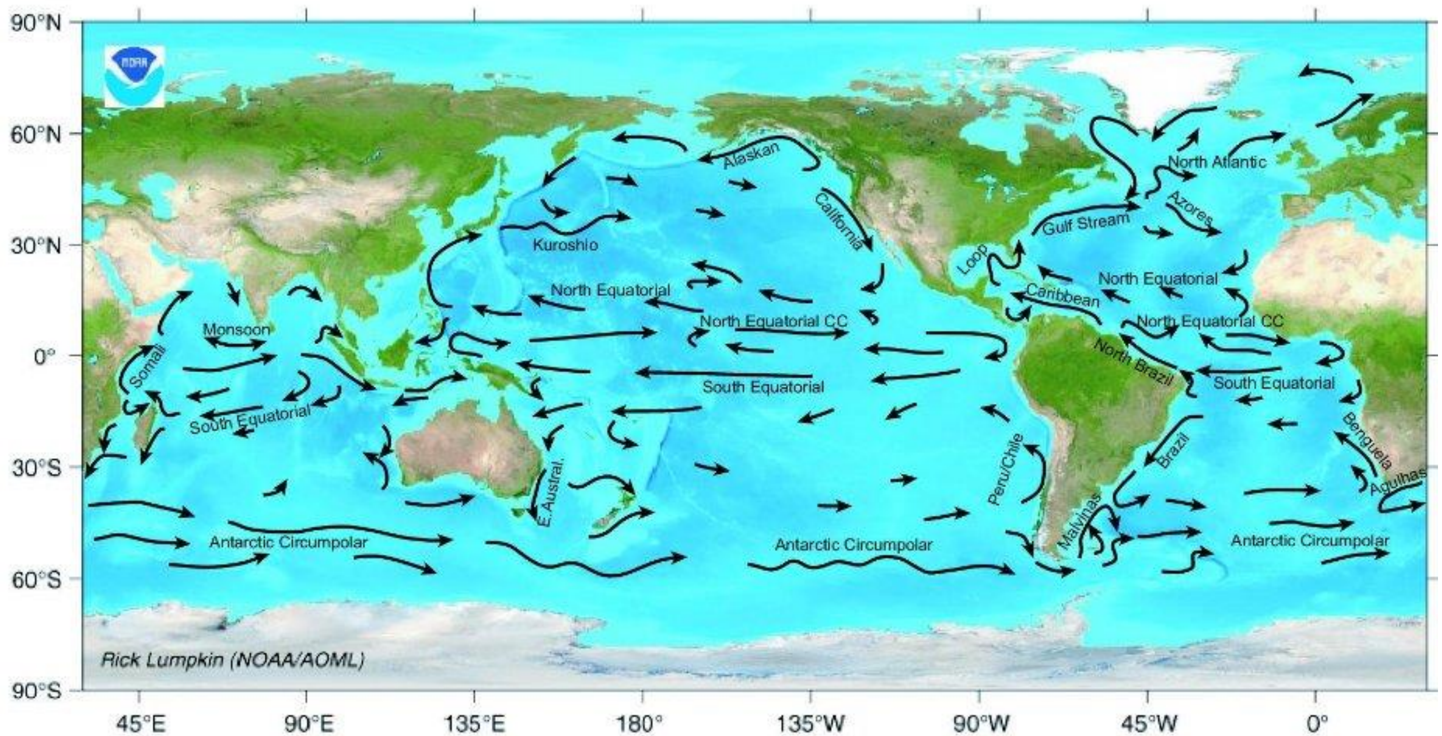
Statistically significant similarity values (links)



Eguiluz et al, PRL 2005
Deza et al, Chaos 2013
Donges et al, Chaos 2015

Physical mechanisms responsible for teleconnections

Winds, ocean currents and solar forcing.



http://www.aoml.noaa.gov

The data: surface air temperature

- Anomalies = annual solar cycle removed
- Spatial resolution 2.5 x 2.5 $\Rightarrow$ 10226 nodes
- Daily / monthly 1949 - 2013 $\Rightarrow$ 23700 / 700 data points

Where does the data come from?

- National Center for Environmental Prediction, National Center for Atmospheric Research (NCEP-NCAR).
- Free!
- Reanalysis = run a sophisticated model of general atmospheric circulation and feed the model (data assimilation) with empirical data, where and when available.

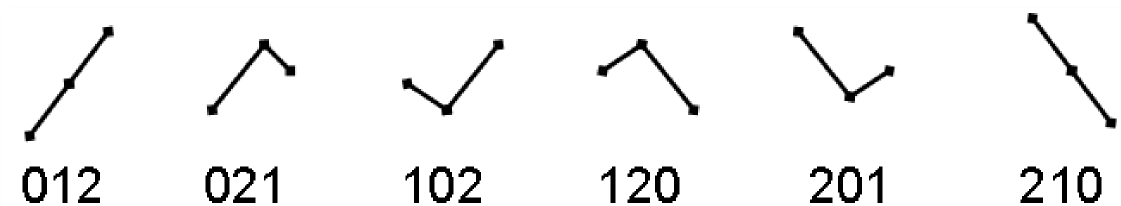
Statistical similarity measure

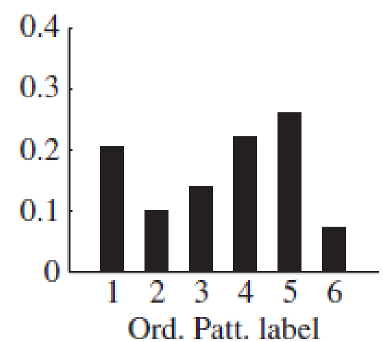
- Mutual information

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

$$p_{ij}(m,n) = p_i(m)p_j(n) \Leftrightarrow M_{ij} = 0$$

- We use **ordinal symbolic** time-series to compute the probabilities. $X = \{\ldots x_i, x_{i+1}, x_{i+2}, \ldots\}$





  - Drawback: we lose information about the actual values.
  - Advantage: we can select the time scale of the analysis.

Brandt & Pompe, PRL 88, 174102 (2002)

Ordinal analysis allows selecting the time scale of the analysis

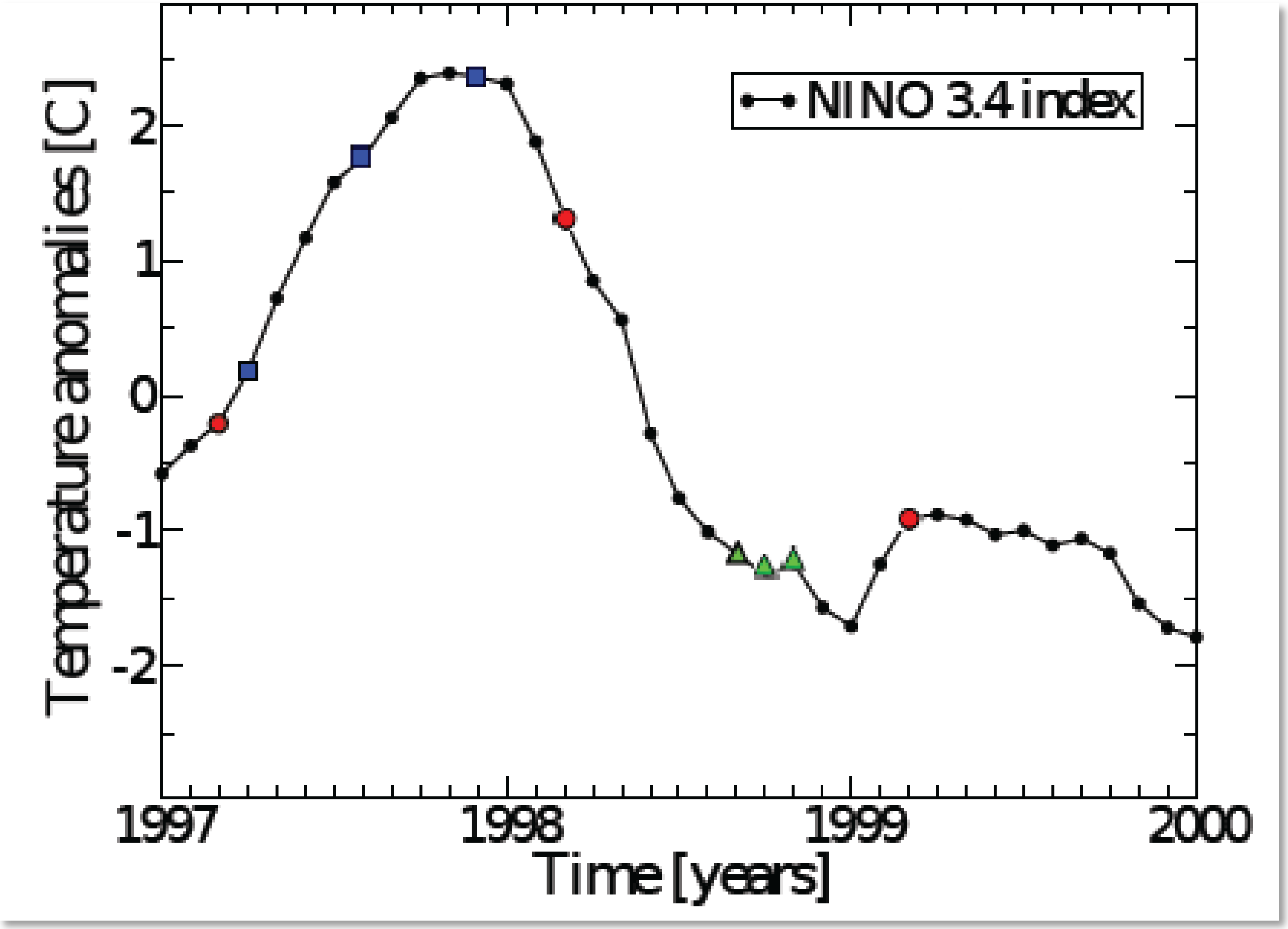
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Intra-season 102

Intra-annual 012

Inter-annual 120



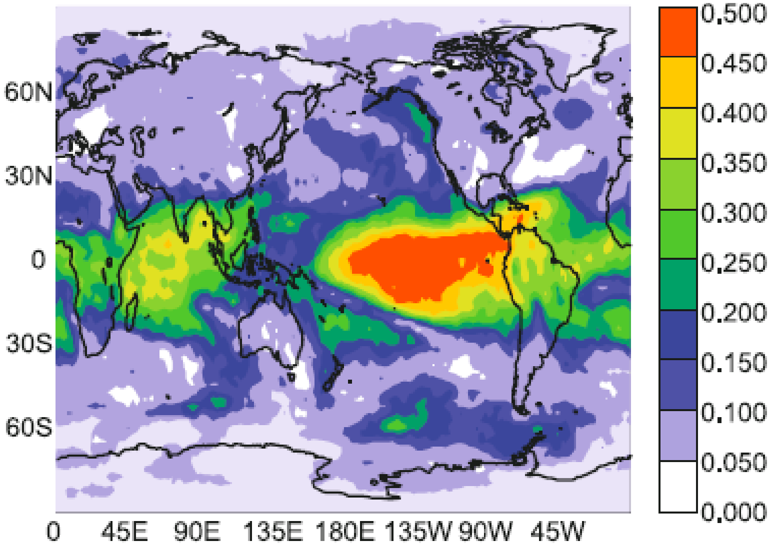
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

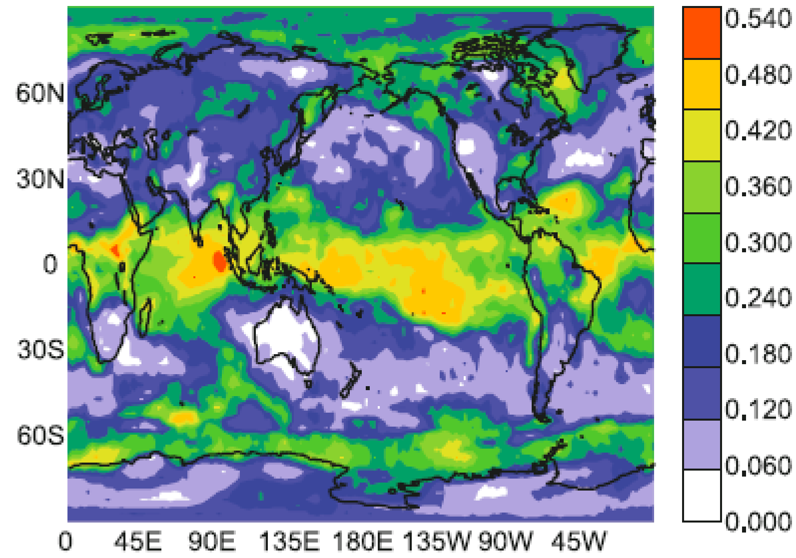
Contrasting two methods for inferring the climate network

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

Network when the probabilities are computed with <u>ordinal analysis</u>



Network when the probabilities are computed with <u>histogram of values</u>

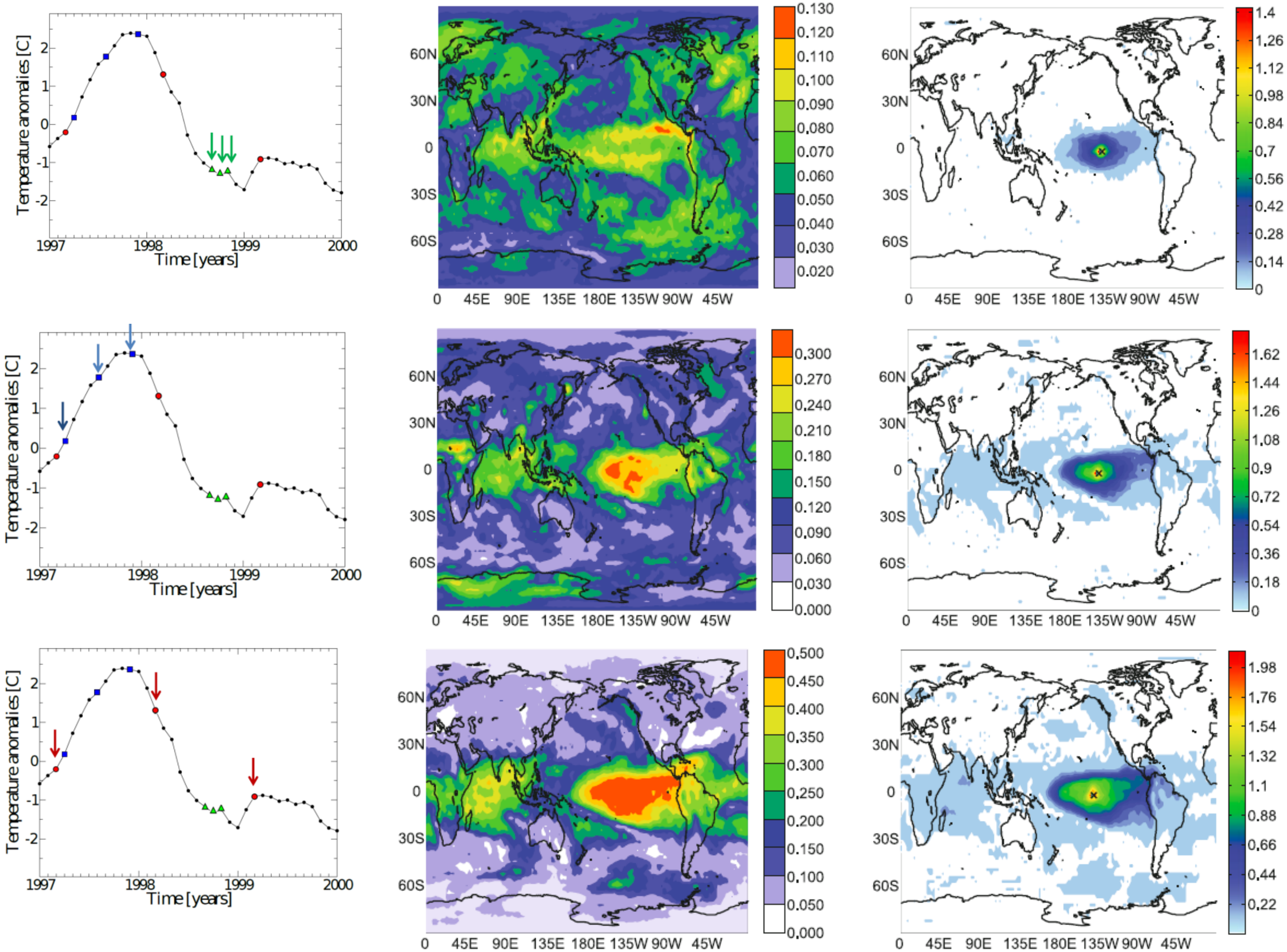


J. I. Deza, M. Barreiro, and C. Masoller, Eur. Phys. J. Special Topics 222, 511 (2013)

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Influence of the time-scale of the symbolic ordinal pattern



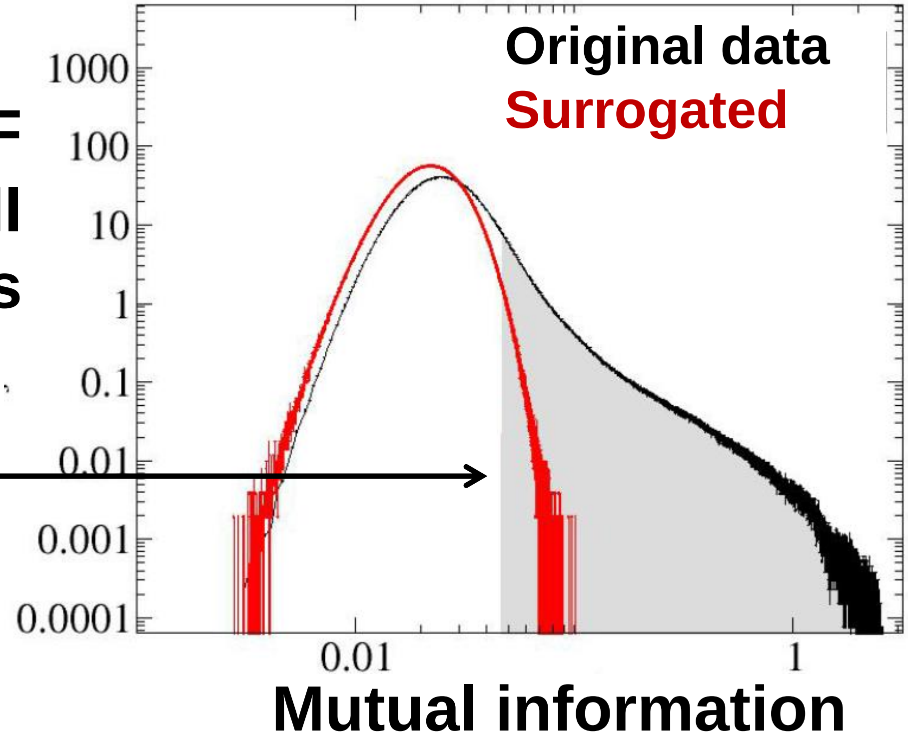
Longer time-scale ⇒ increased connectivity

J. I. Deza et al (2013)

How do we assess the significance of the links?

PDF MI values

TH = <MI> + 3σ



99.87% confidence level that the links have MI values that are not consistent with random values.

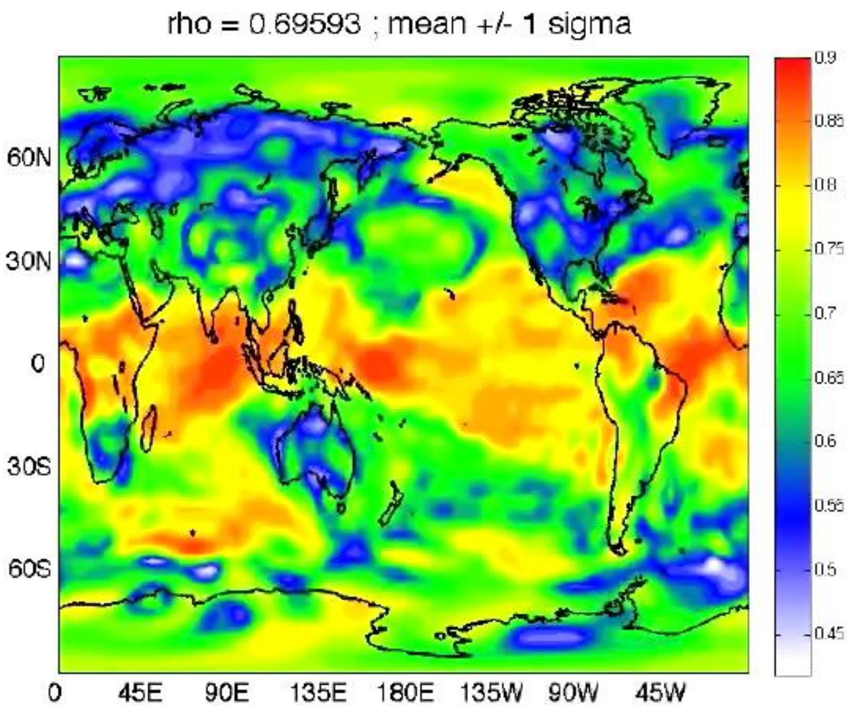
J. I. Deza et al (2013)

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

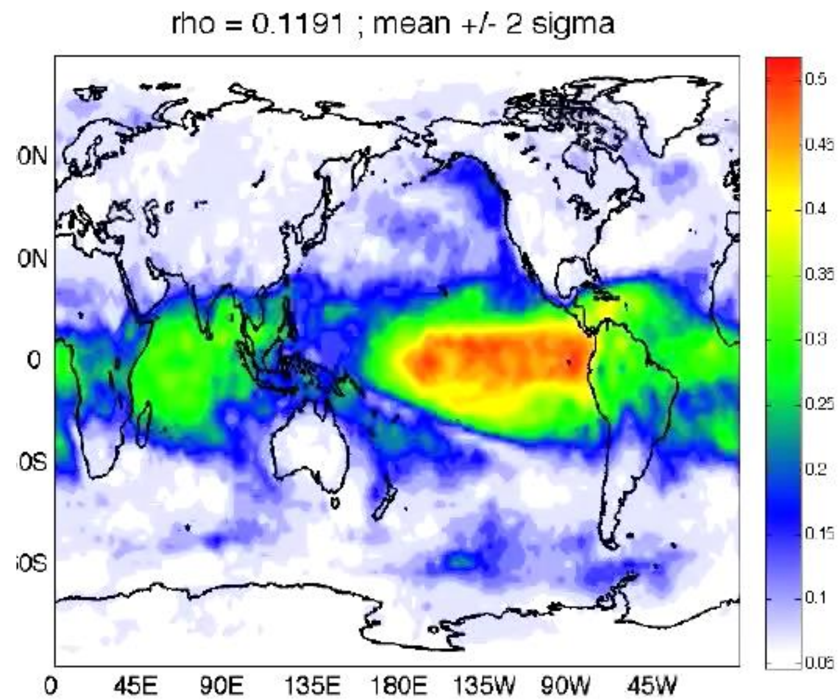
Campus d'Excel·lència Internacional

Influence of the threshold

Network constructed from Cross-Correlation (CC) analysis



Network constructed from Mutual Information (MI ordinal analysis, annual time scale)



J. I. Deza et al (2013)

Can we test the method used to infer the links?

- Simulations of Kuramoto oscillators with known random coupling topology.
- Experiments with chaotic electronic circuits.

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Kuramoto oscillators

$$\mathrm{d}\theta_i = \omega_i \mathrm{d}t + \frac{K}{N}\sum_{j=1}^{N} A_{ij} \sin(\theta_j - \theta_i)\mathrm{d}t + D\,\mathrm{d}W_t^i$$

- A_{ij} is a **known symmetric random matrix**
- N =12 oscillators.
- Performed 100 simulations (10^4 data points each) with different oscillators' frequencies, random matrices, noise realizations and initial conditions.
- For each value of *K*, the threshold was varied to obtain optimal reconstruction.

G. Tirabassi et al, Sci. Rep. ***5*** *10829 (2015)*

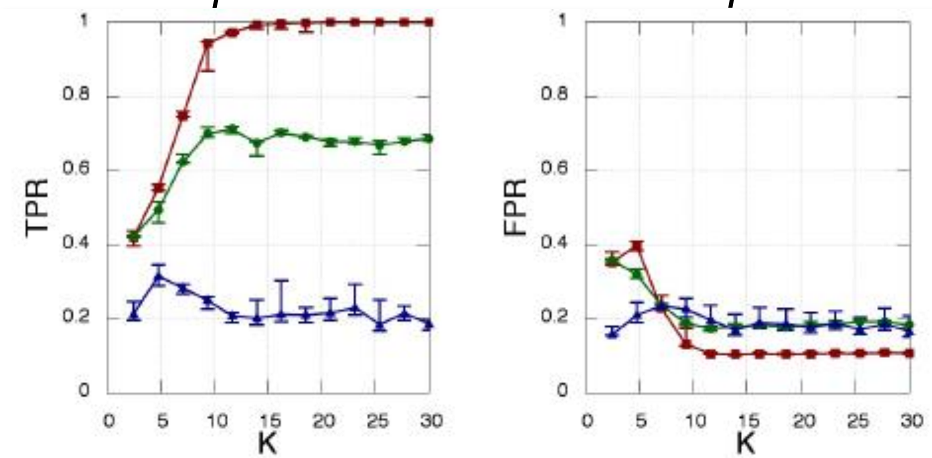
Results

Phases (θ)

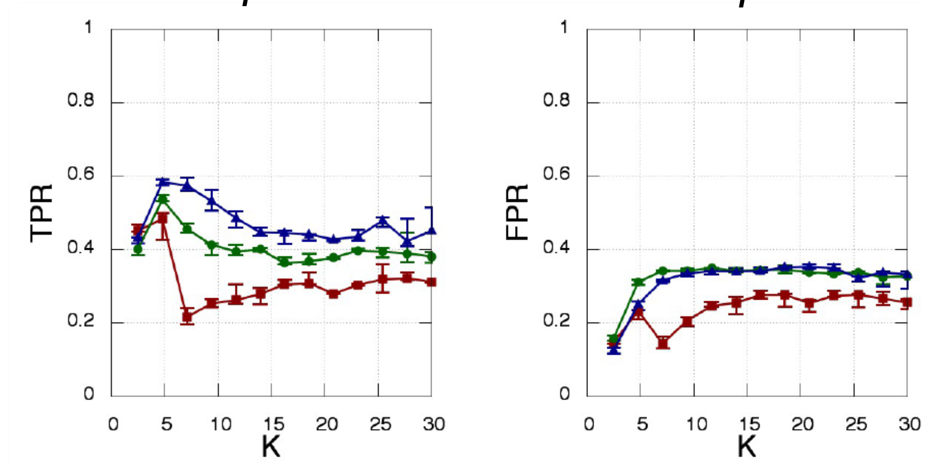
CC MI MIOP

"Observable" Y=sin(θ)

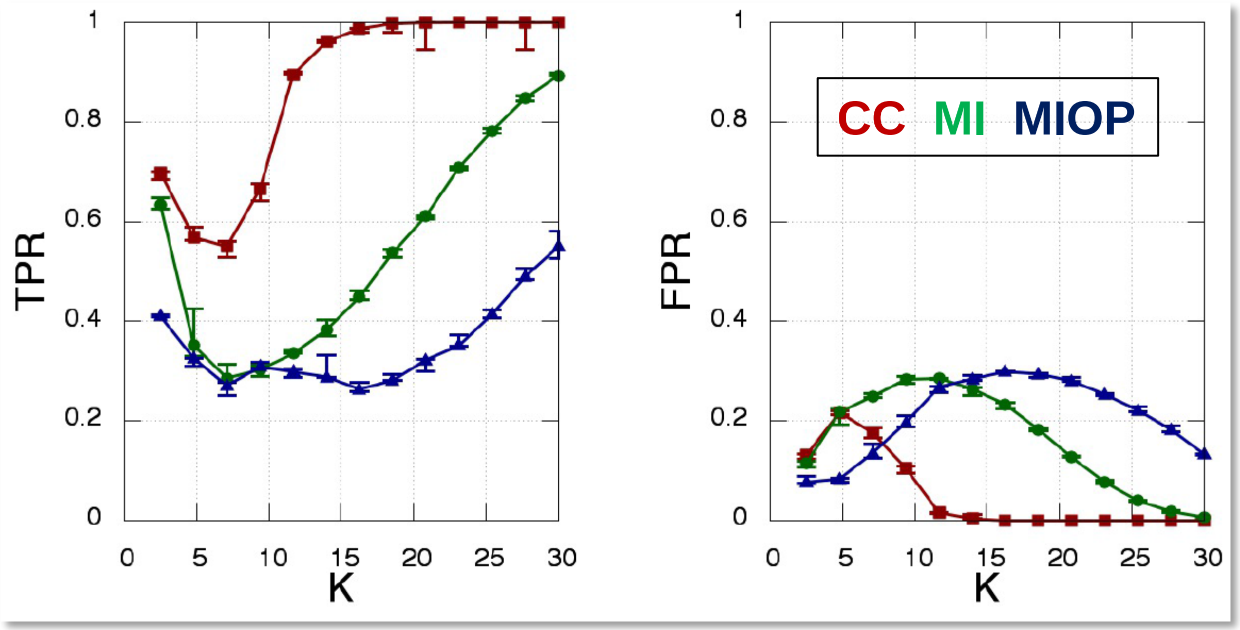
True positives *False positives*



True positives *False positives*



Instantaneous frequencies (dθ/dt)



For strong enough coupling K perfect inference is possible!

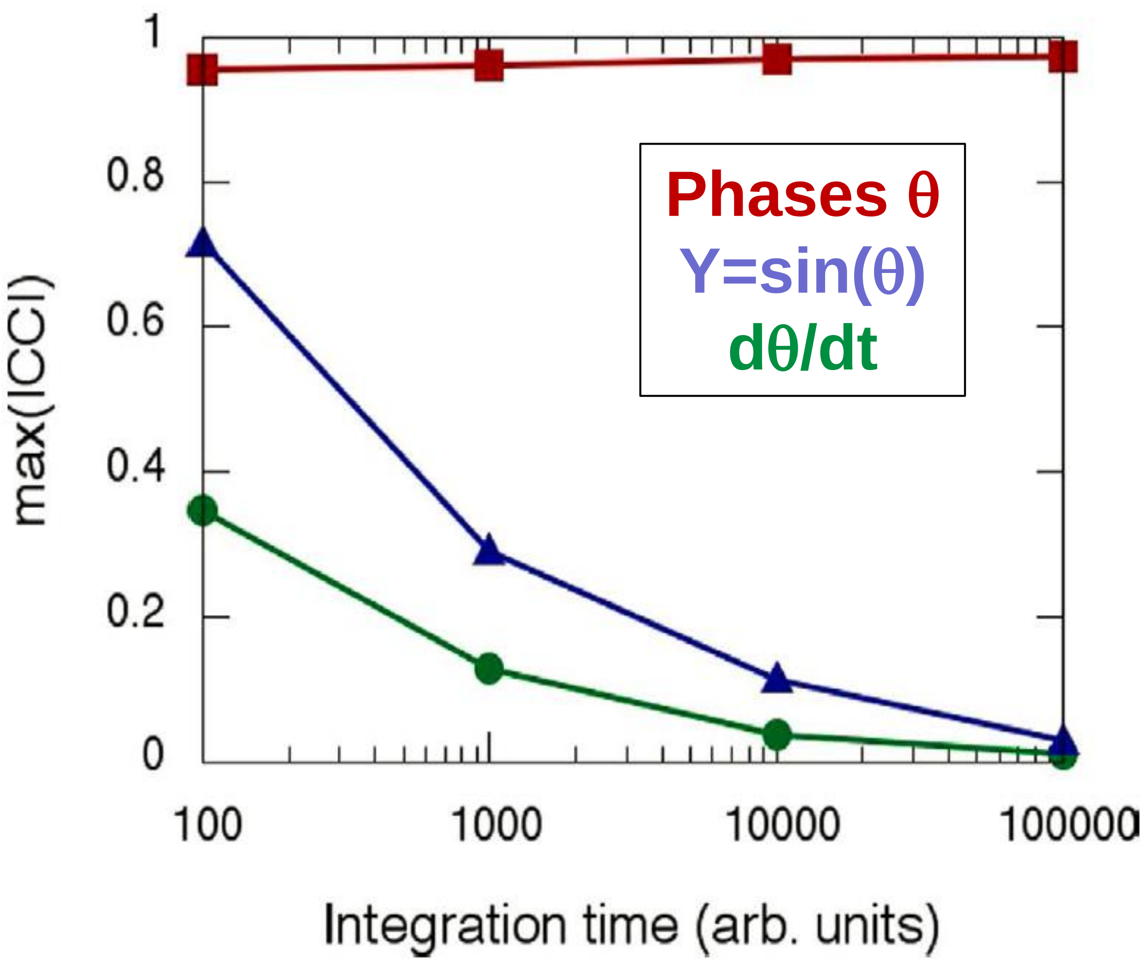
BUT

- the number of oscillators is small (12),
- the coupling is symmetric ($\Rightarrow$ only 66 possible links) and
- the data sets are long (10^4 points)

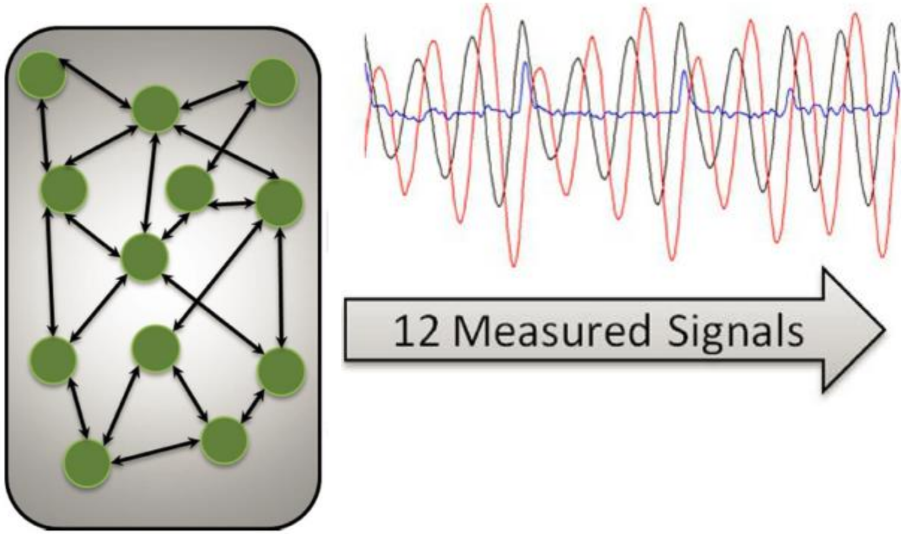
G. Tirabassi et al, Sci. Rep. ***5*** *10829 (2015)*

Why the frequencies are better than phases and "observables"?

Correlation analysis of two UNCOUPLED oscillators (K=0)



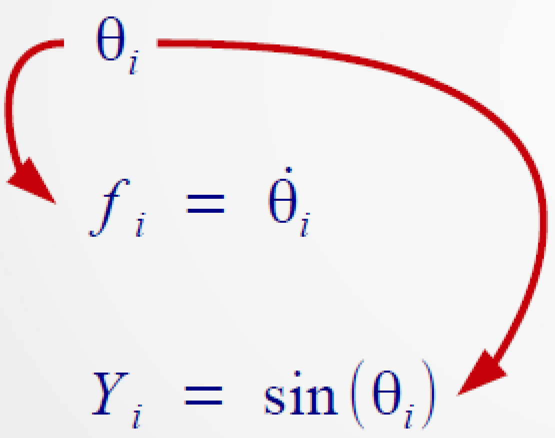
We also analyzed experimental data: 12 chaotic Rössler electronic oscillators (symmetric and known random coupling)



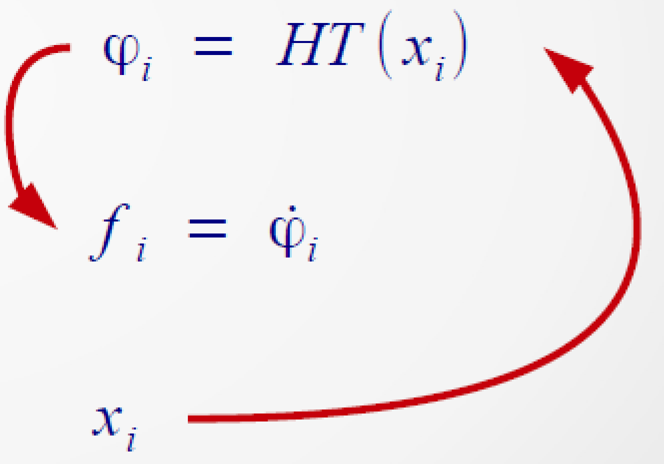
The Hilbert Transform was used to obtain phases from experimental data

Experiments performed by Javier Buldu (Universidad Rey Juan Carlos, Madrid) and Ricardo Sevilla-Escoboza (Universidad de Guadalajara, México)

- Kuramoto Oscillators' Network



- Rössler Oscillators' Network



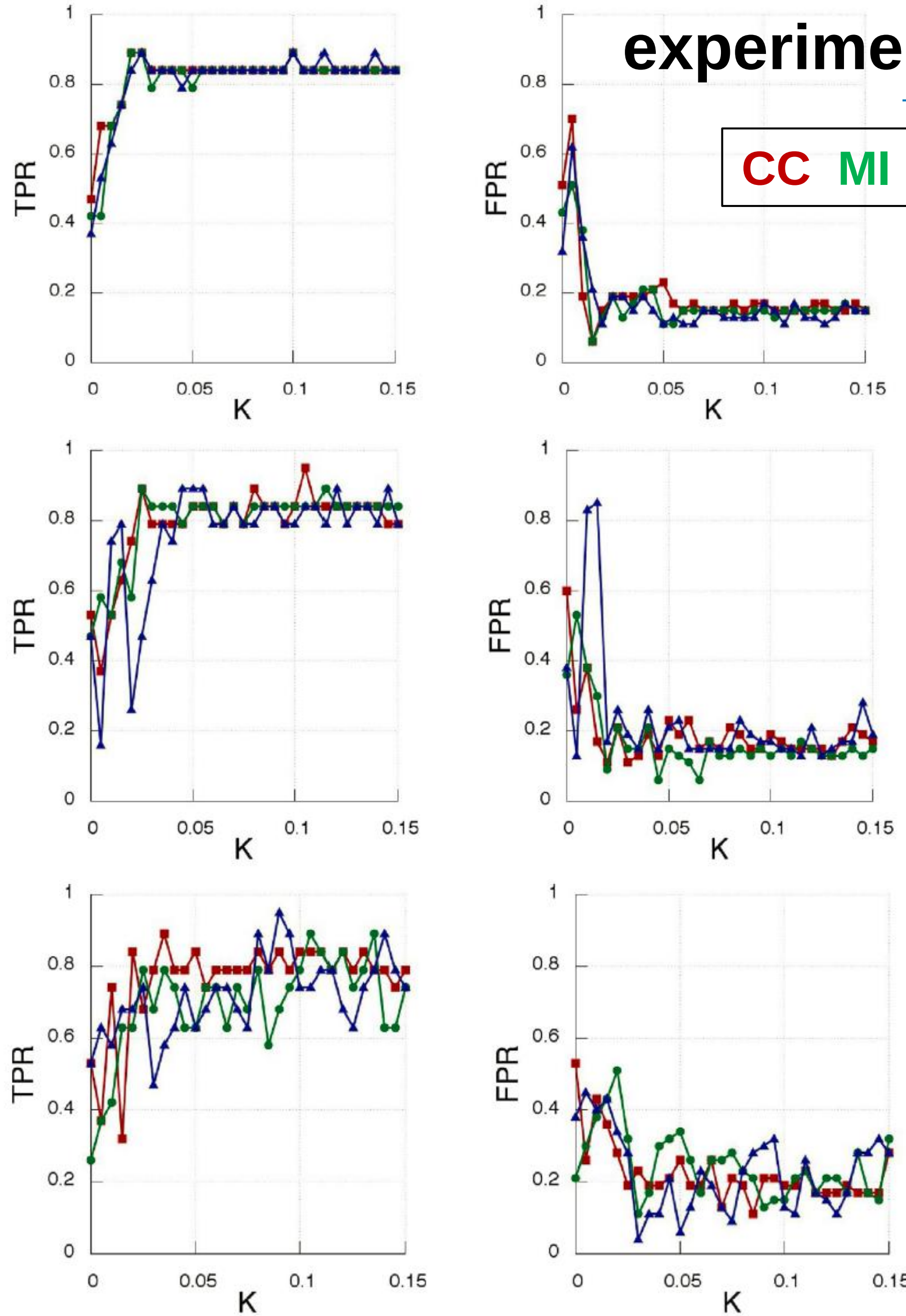
Results obtained with experimental data

CC MI MIOP

Observed variable (x)

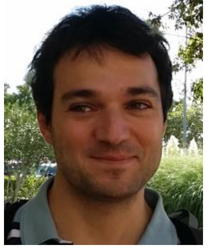
Hilbert phase

Hilbert frequency

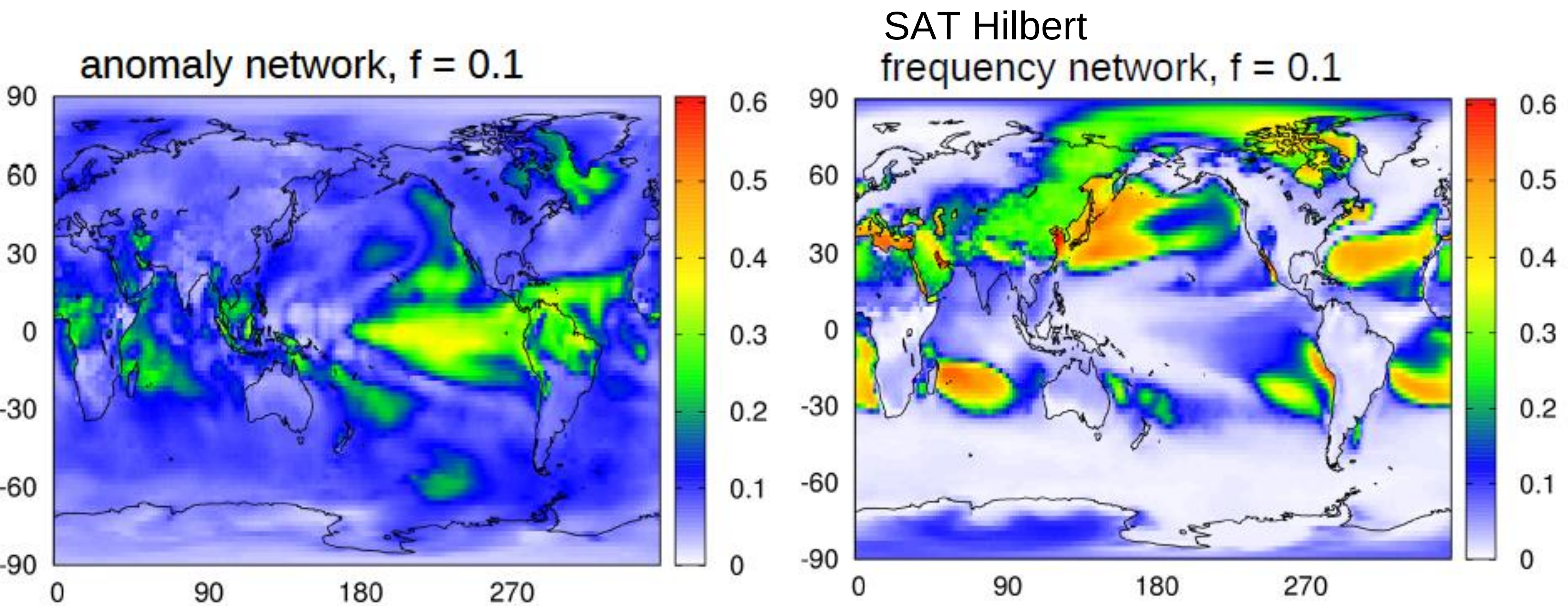


- No perfect reconstruction
- No important difference among the 3 methods & 3 variables

Ongoing work



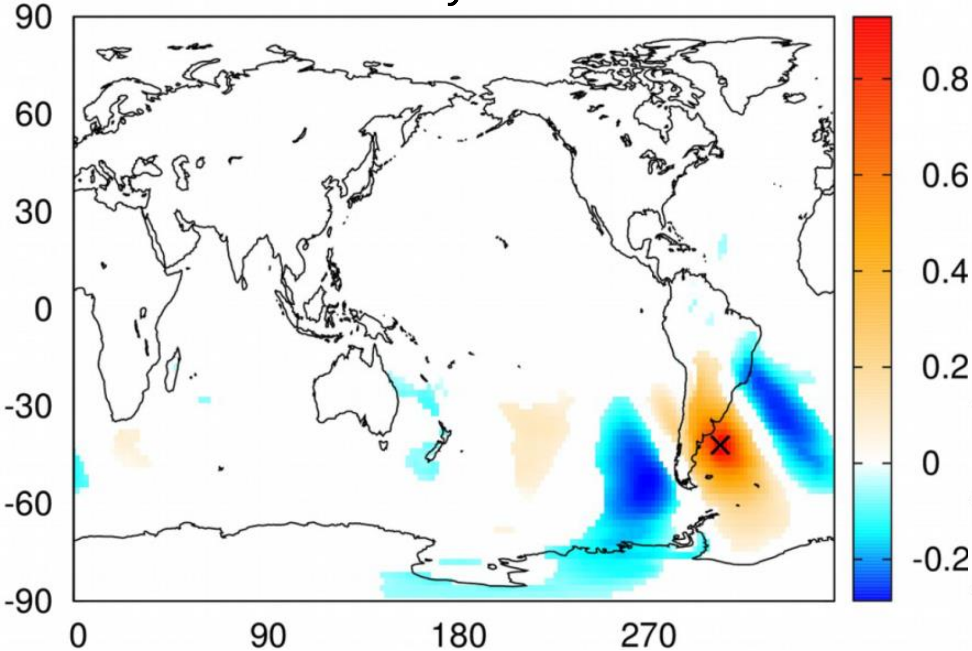
Climate network built from zero-lag CC analysis



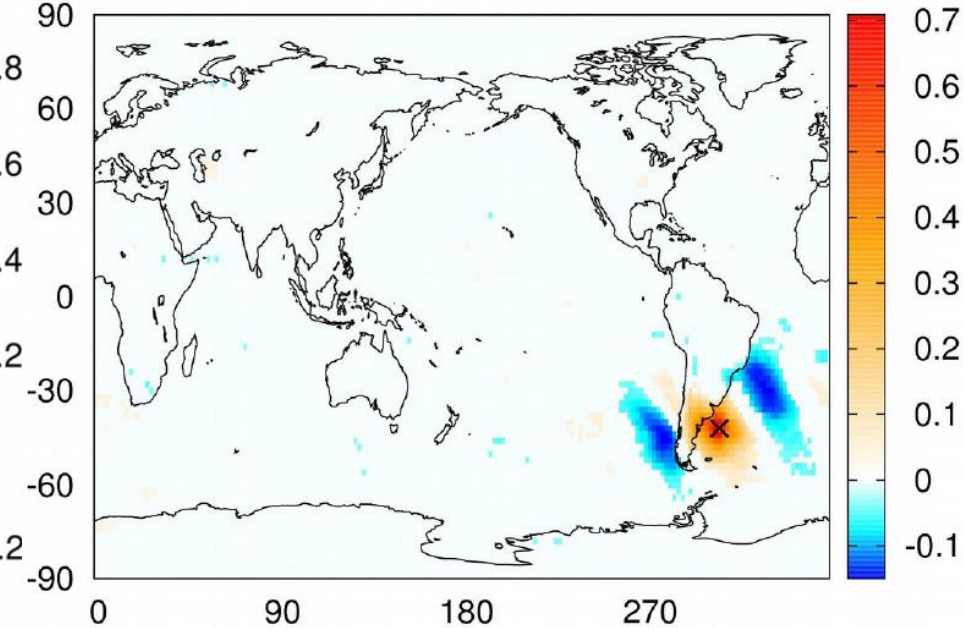
D. Zappala et al, in preparation.

Connectivity maps

Anomaly network



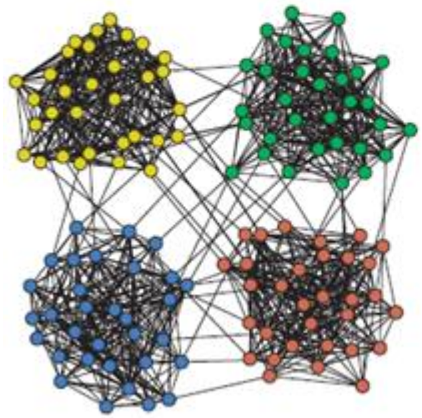
Hilbert frequency network



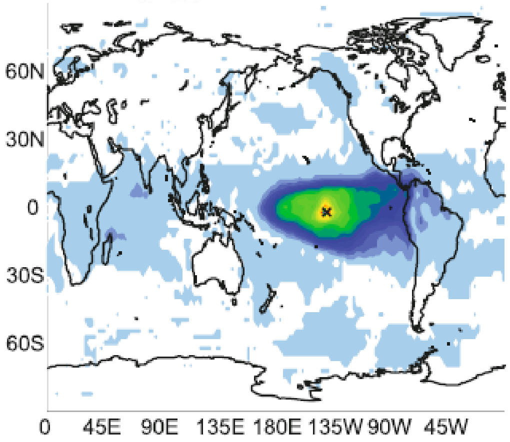
Climate "communities"

How to identify regions with similar climate?

- Goal: to construct a network in which regions with similar climate (e.g., continental) are in the same "community".



- Problem: not possible with the "usual" method to construct the network because NH and SH are only indirectly connected.



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Campus d'Excel·lència Internacional

Network construction based on similar symbolic dynamics

- Step 1: transform SAT anomalies in each node in a sequence of symbols (we use ordinal patterns)

  s_i = {012, 102, 210, 012…} s_j = {201, 210, 210, 012, …}

- Step 2: in each node compute the <u>transition probabilities</u>

  $TP^i_{\alpha\beta} = \#(\alpha \rightarrow \beta)/N$

- Step 3: define the weights $$w_{ij} = \frac{1}{\sum_{\alpha\beta}\left(TP^i_{\alpha\beta} - TP^j_{\alpha\beta}\right)^2}$$

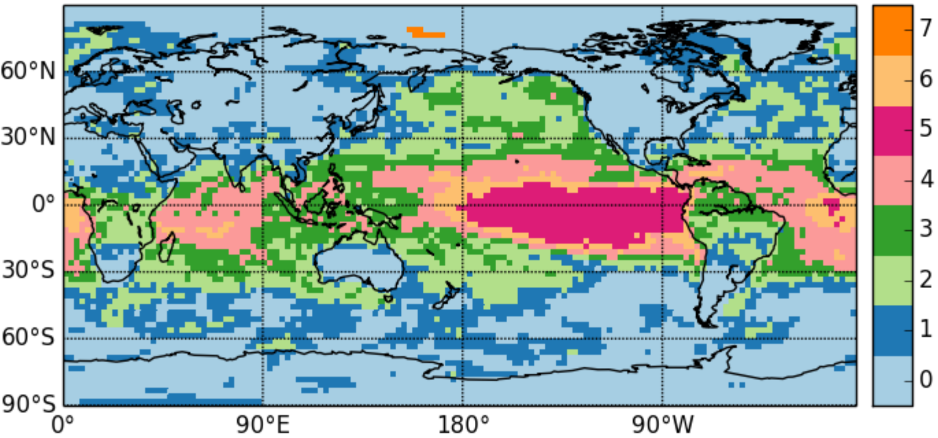
  High weight if similar symbolic "language"

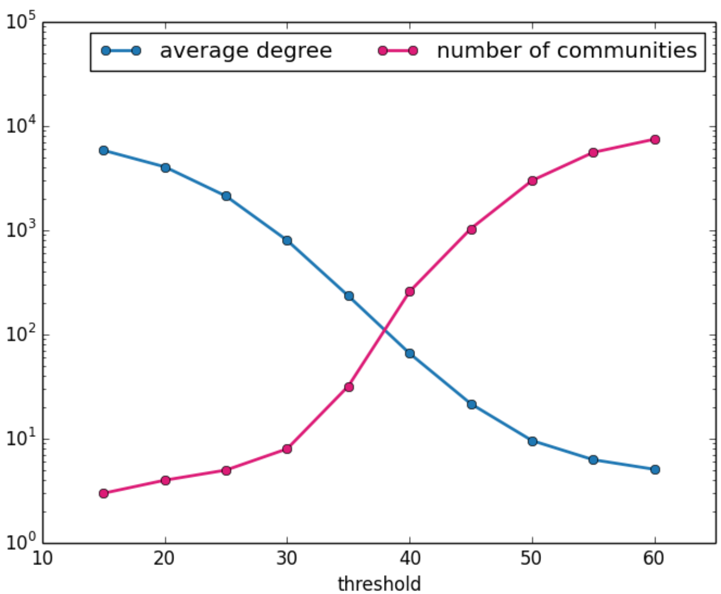
- Step 4: threshold w_{ij} to obtain the adjacency matrix.

- Step 5: run a community detection algorithm (Infomap).

Results

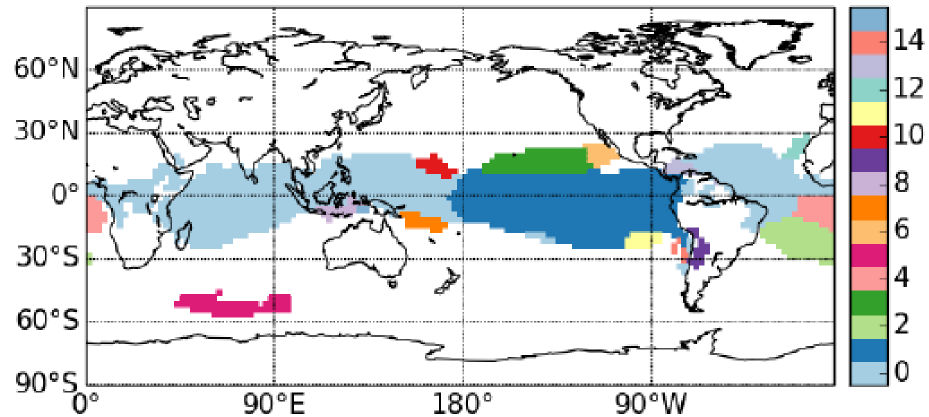
TP Network

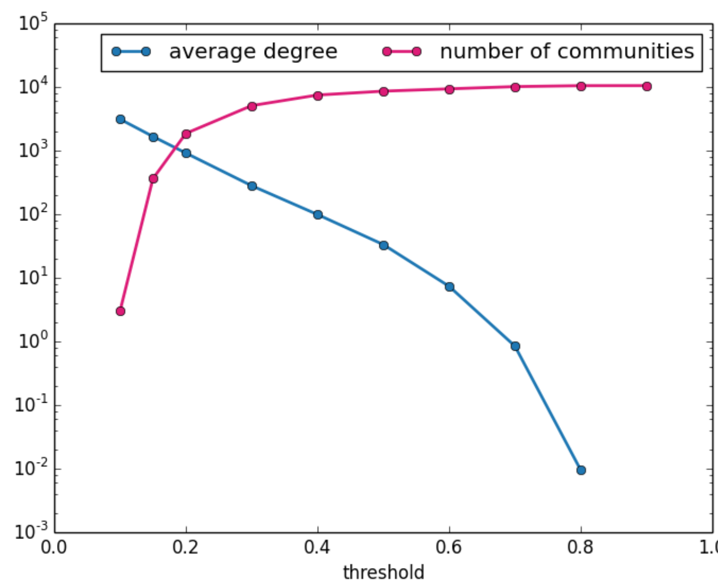




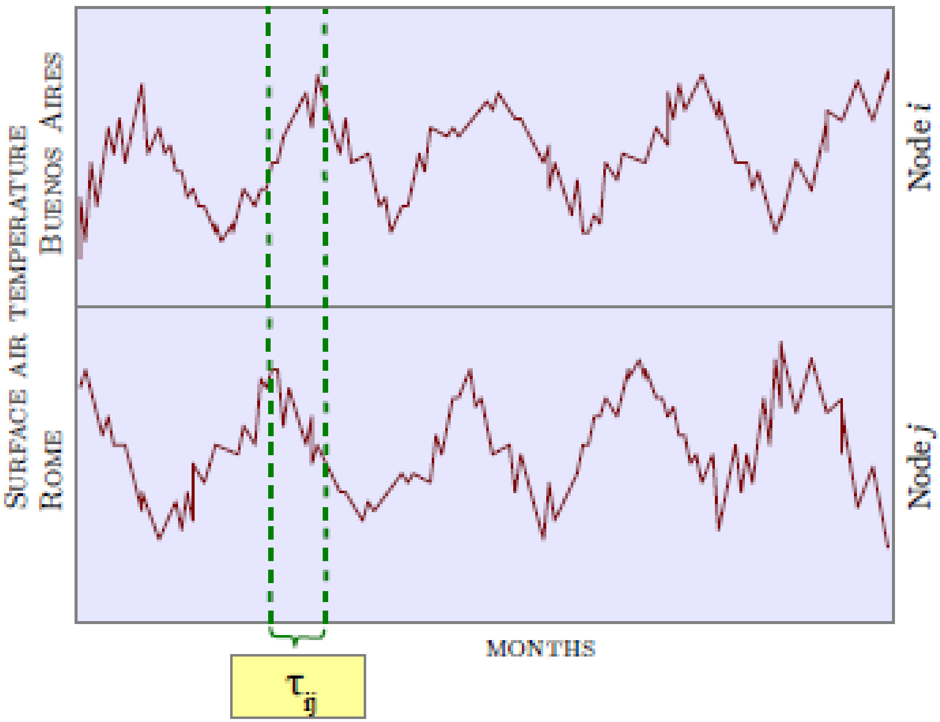
CC Network

(only the largest 16)

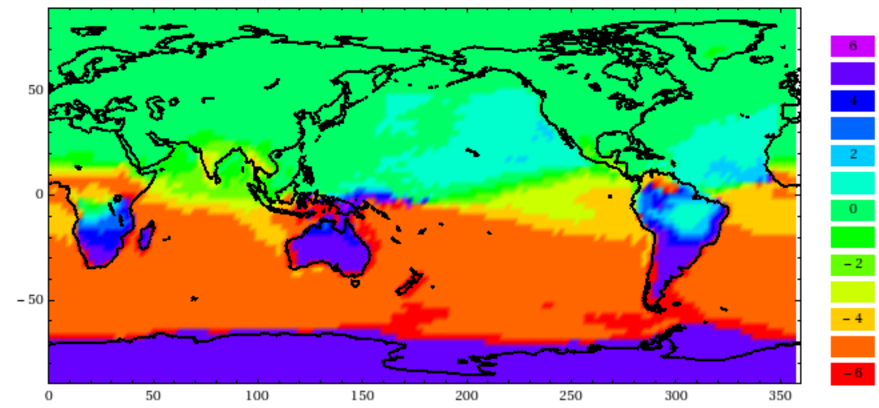




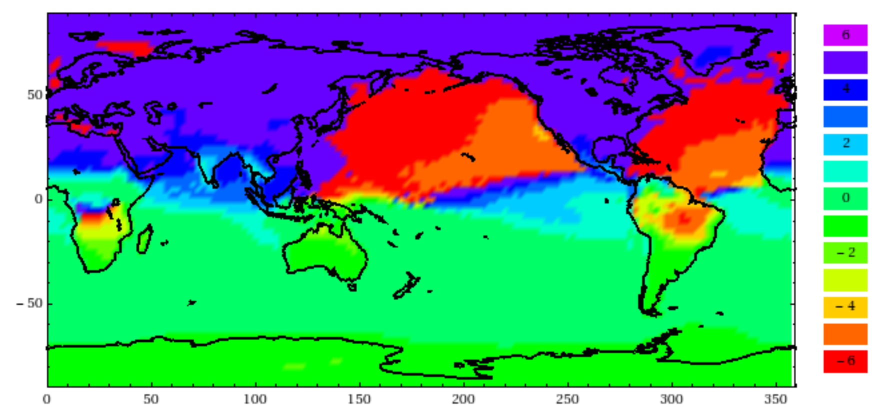
Another way to identify communities: lag-times between seasonal cycles



Rome

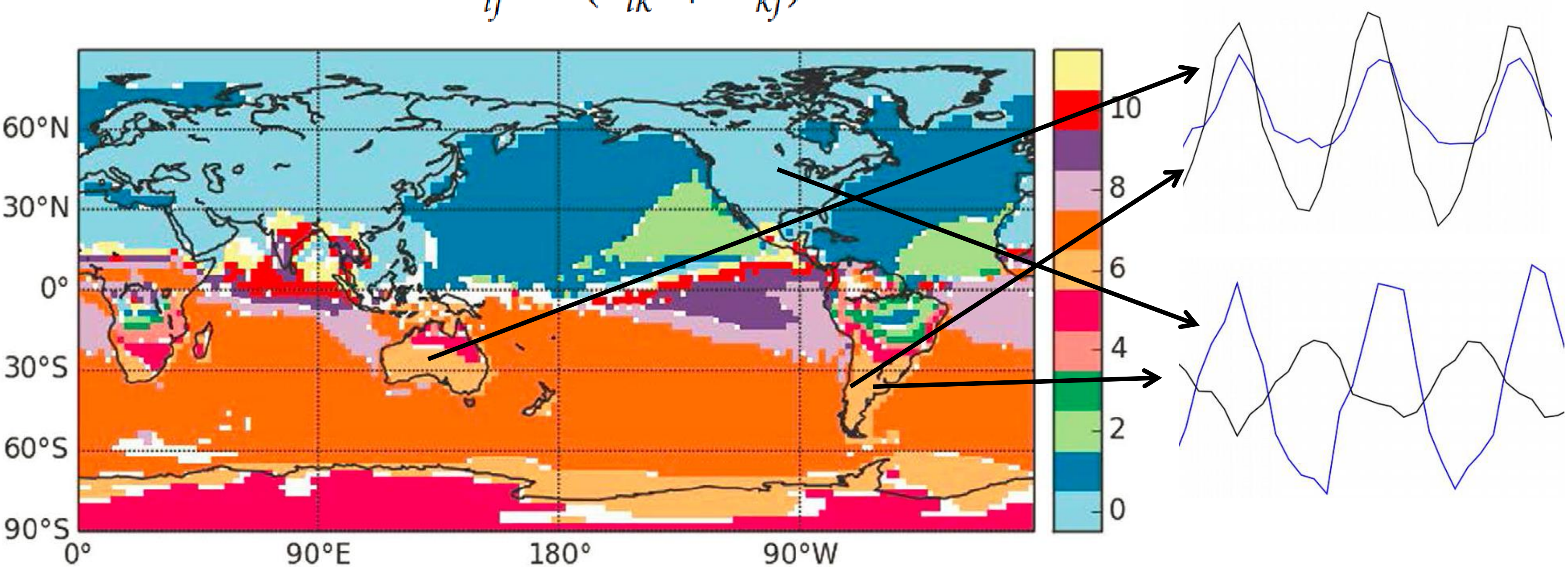


Buenos Aires



Climate communities: regions with inphase seasonal cycles

$$\ell_{ij} = (\ell_{ik} + \ell_{kj}) \bmod 12$$



- Oceans tend to have a one-month delay with respect to the landmasses.
- Well defined long-range spatial patterns.

G. Tirabassi and C. Masoller, Sci. Rep. 6:29804 (2016)

Summary

- Take home message: network tools and symbolic ordinal analysis provide an opportunity for advancing understanding and predictability of our climate.

- Ordinal analysis allows identifying time-scales of climate interactions and climate communities.

- In small synthetic networks, under appropriate conditions, perfect network inference is possible.

Papers at **http://www.fisica.edu.uy/~cris/**

- J. I. Deza et al, Eur. Phys. J. Special Topics 222, 511 (2013).
- G. Tirabassi et al, Sci. Rep. 5, 10829 (2015).
- G. Tirabassi and C. Masoller, Sci. Rep. 6, 29804 (2016).

60 two-year postdoctoral research positions in the Catalan science and technology system (Beatriu de Pinós programme 2016)

Agència de Gestió d'Ajuts Universitaris i de Recerca
AGAUR

agaur.gencat.cat

Thank you for your attention!

ITN LINC
FP7-289447

ICREA

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN