Inferring long memory processes in the climate network via nonlinear time series analysis

Cristina Masoller

Universitat Politecnica de Catalunya, Barcelona, Spain

Cristina.masoller@upc.edu

www.fisica.edu.uy/~cris

Marcelo Barreiro, Arturo Martí

Universidad de la República, Montevideo, Uruguay

SigmaPhi 2011, Larnaca, Cyprus, July 12

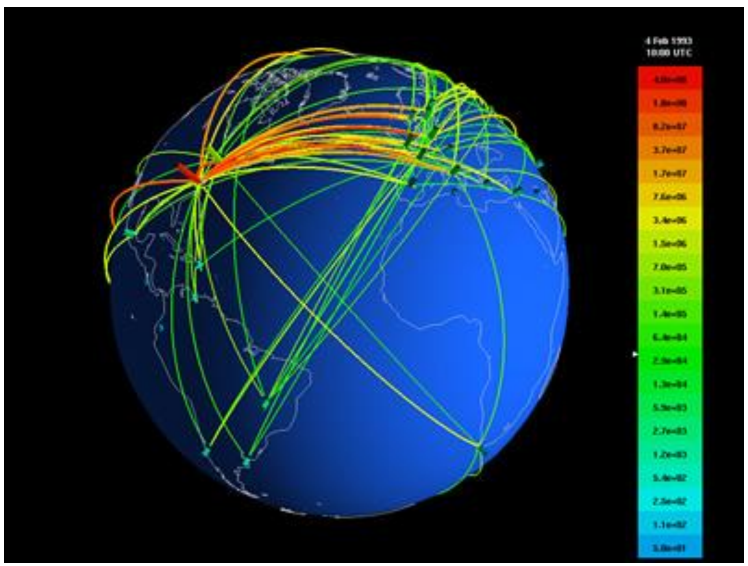
Outline of the talk

- Applying network concepts to the Earth's climate
- Ordinal patterns nonlinear time-series analysis
- Results
- Conclusions

Complex networks

UNIVERSIDAD DE LA REPÚBLICA
URUGUAY

- Scientific collaborations & social networks,
- Biological & ecological networks,
- Brain functional networks,
- Airport networks,
- Internet traffic, etc, etc



Stephen G. Eick, visualcomplexity.com

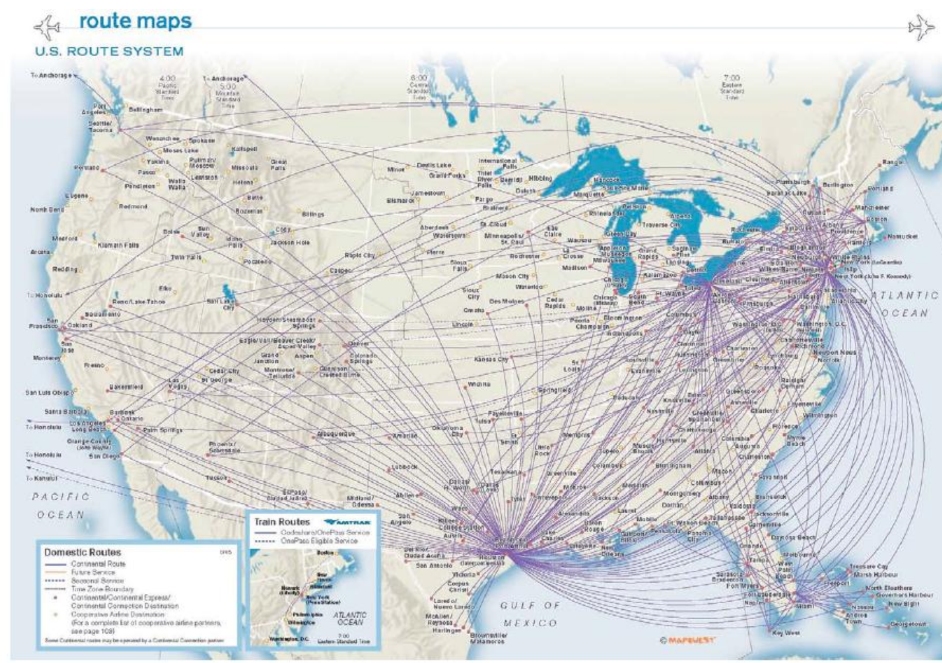
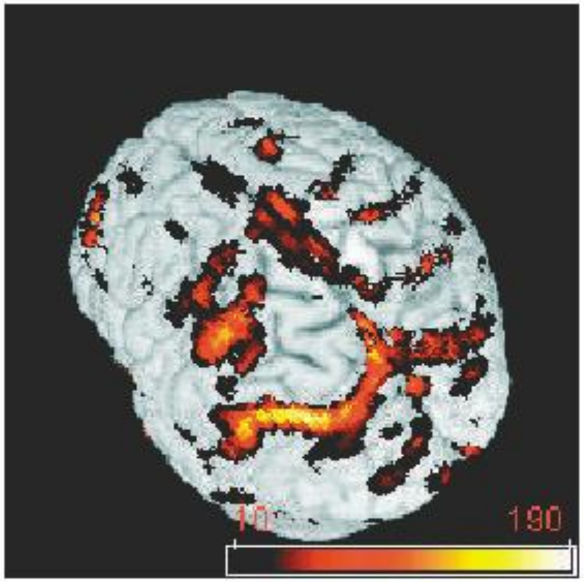


FIG. 2. Route map for Continental Airlines (courtesy of Continental Airlines).

Complex networks

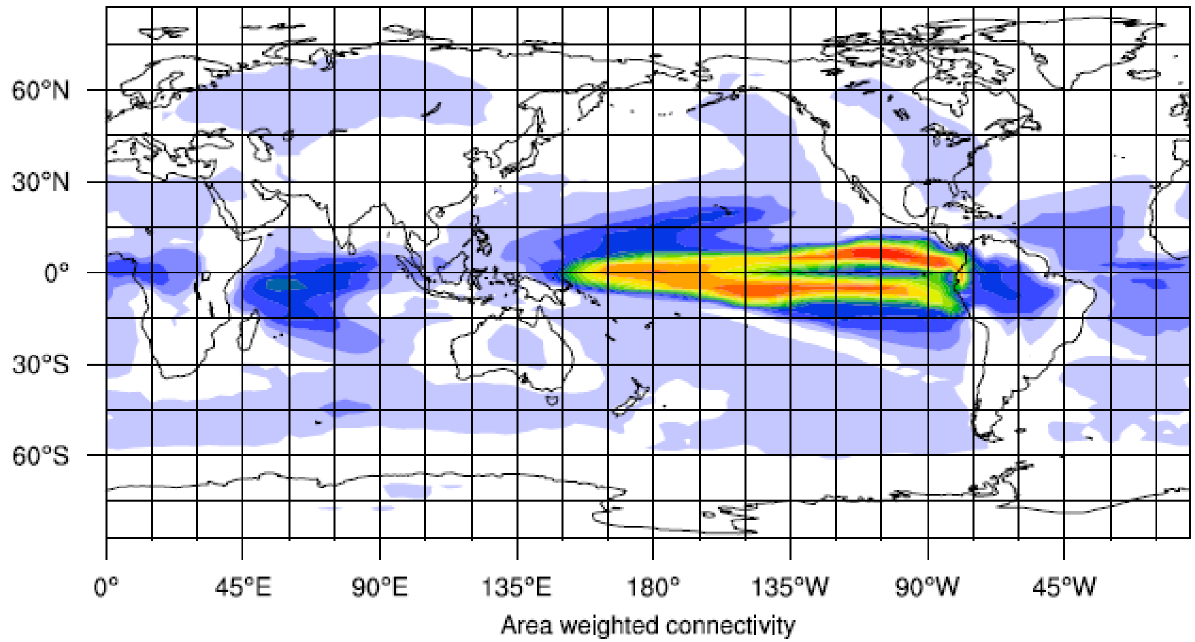
- network theory can improve our understanding of spatially extended complex systems such as the brain or the climate.
- The network approach considers such systems as being composed of dynamically interacting subsystems whose functional interdependencies are reflected as links.



Eguiluz et al, PRL 2005

Climate networks

- constructed over a regular grid (**nodes**) covering the Earth's surface.
- interdependencies are reflected as **links**
- graphical representation of the **area-weighted connectivity**



Donges et al, Eur. Phys. J. Spe. Top. 2009:
Understanding the Earth as a Complex System

Climate system: wide range of time scales

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

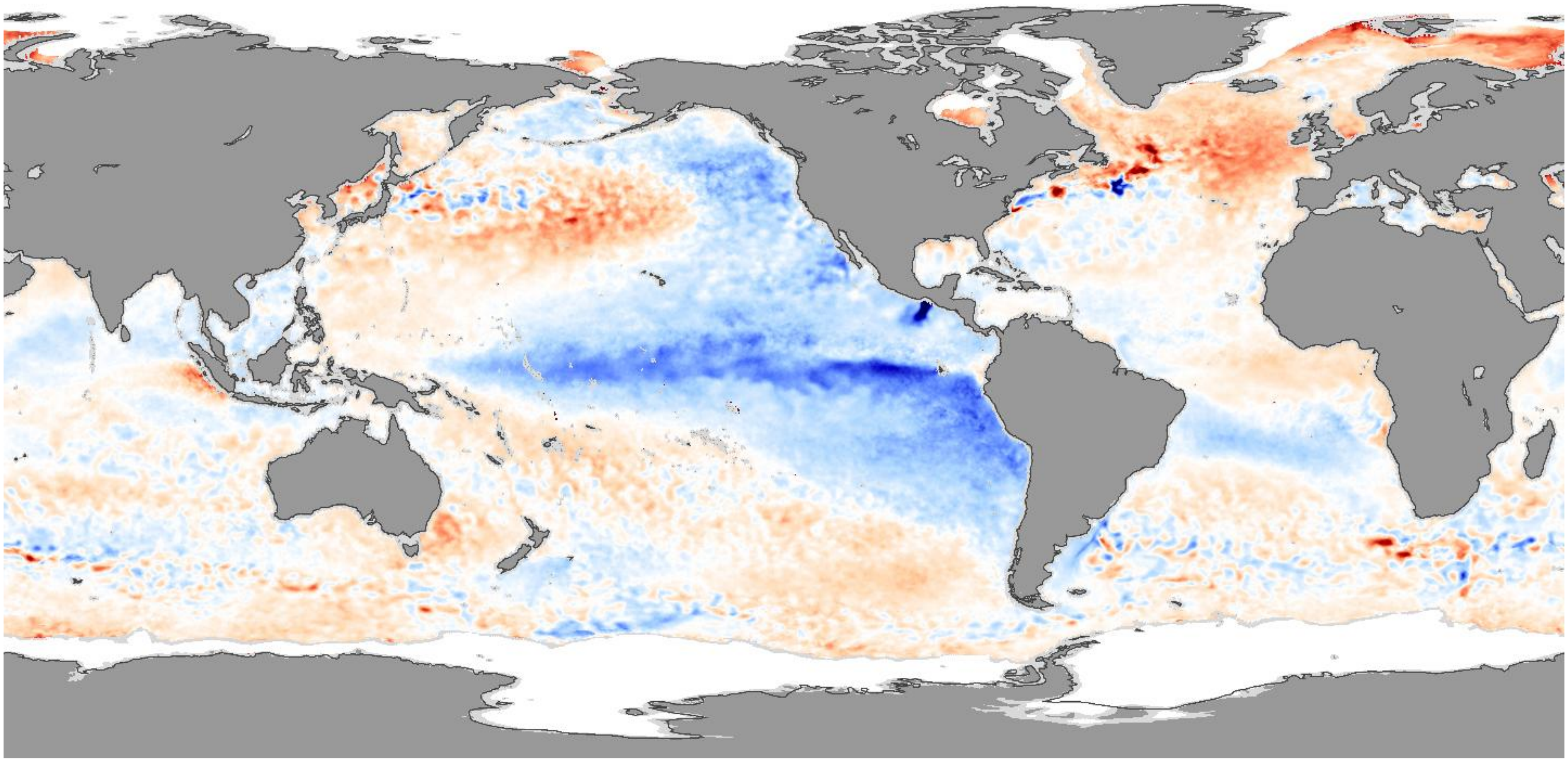
Example: El Niño/La Niña-Southern Oscillation (**ENSO**)

- Occurs across the tropical Pacific Ocean with ≈ 5 years periodicity.
- variations in the surface temperature of the tropical eastern Pacific Ocean (warming: El Niño, cooling: La Niña)
- variations in the air surface pressure in the tropical western Pacific (the Southern Oscillation).
- The two variations are coupled:

  El Niño (ocean warming) -- high air surface pressure,

  La Niña (ocean cooling) -- low air surface pressure.

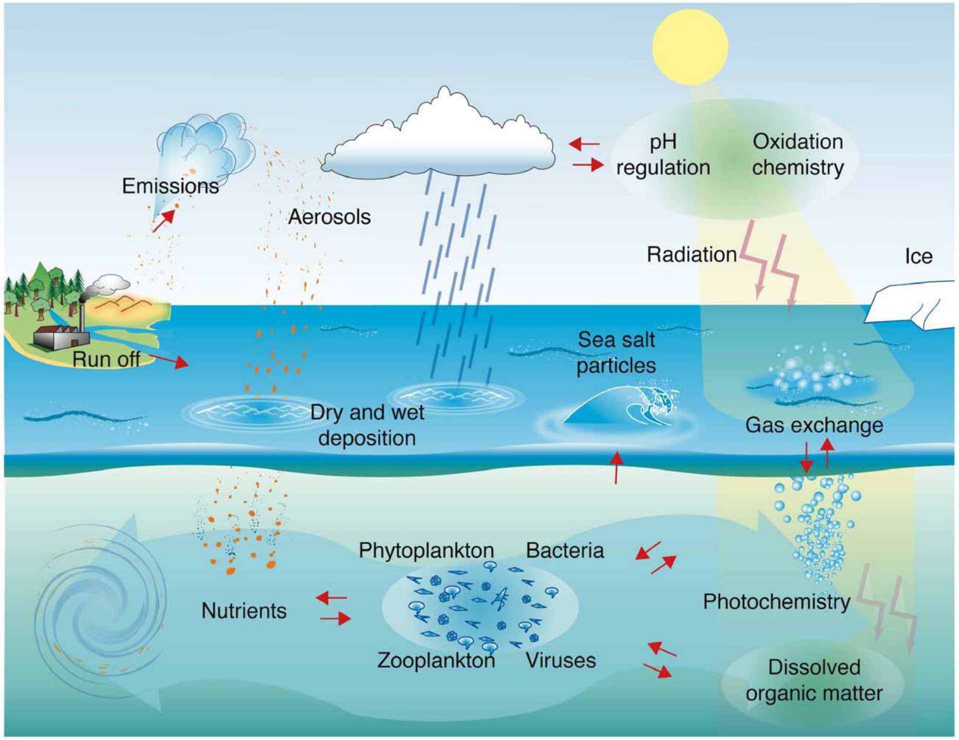
Surface Sea Temperature anomalies during La Niña (November 2007)



Source: Wikipedia

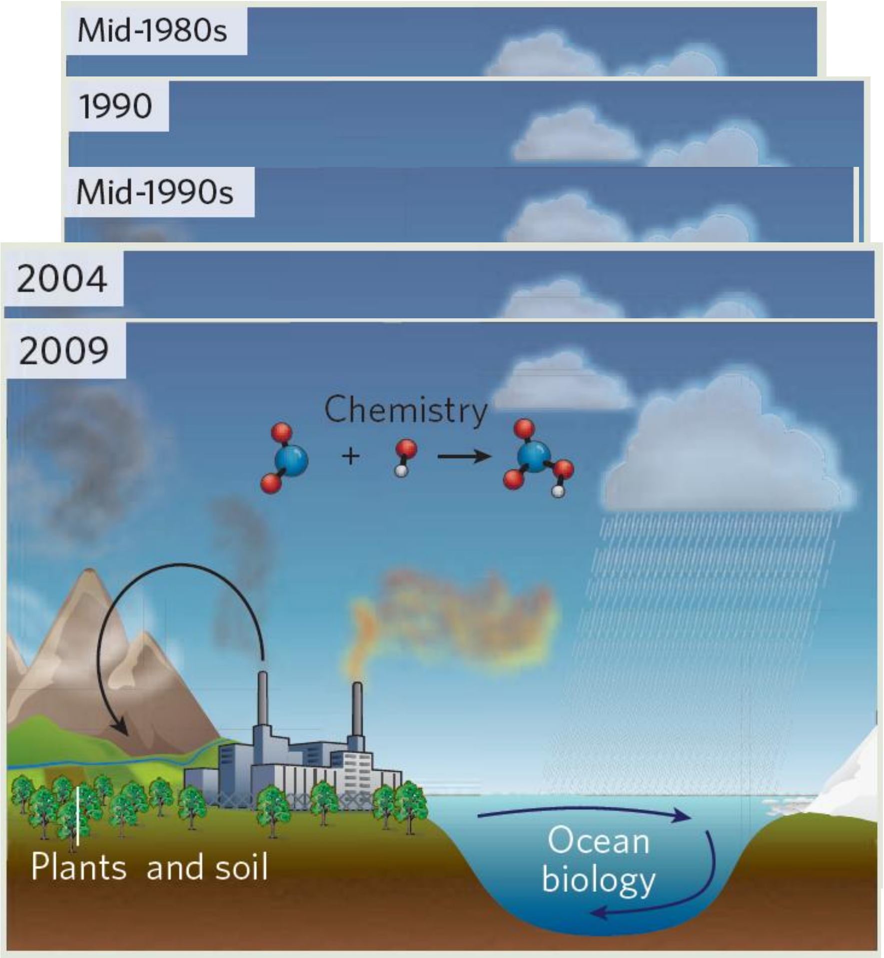
State of the art: climate modeling

Our climate: a (very!) Nonlinear Complex System



Adapted from Elliott and Maltrud, Los Alamos Nat. Lab.

MODEL EVOLUTION



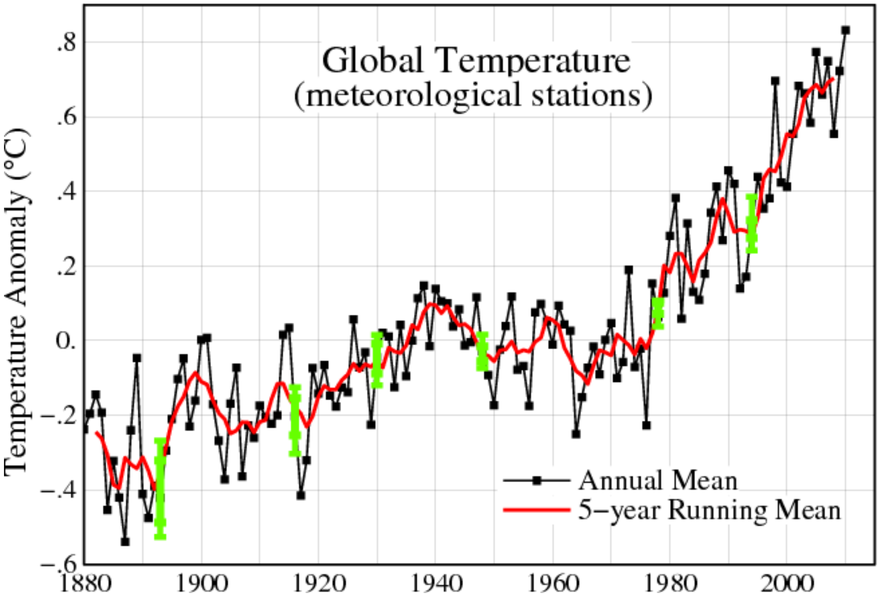
Nature, February 2010

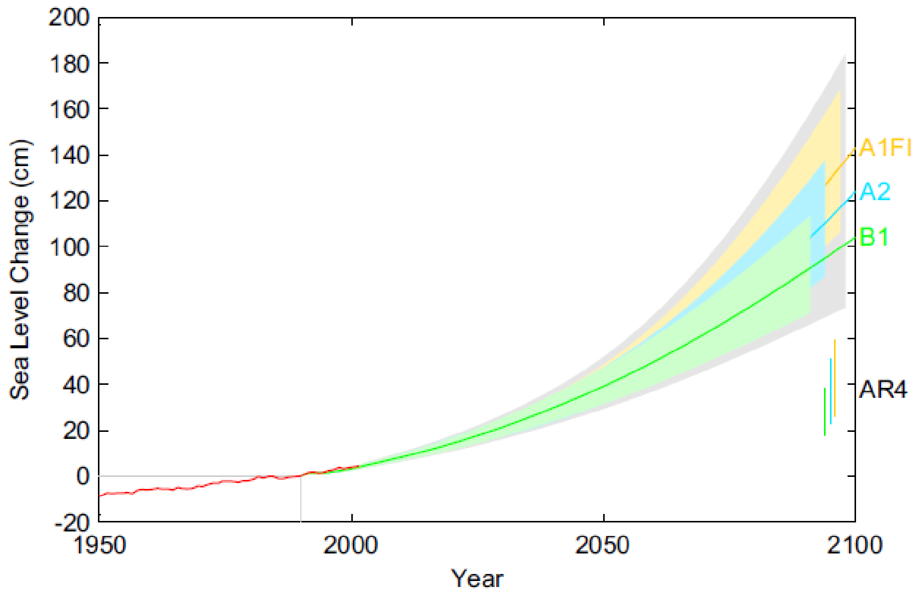
Climate data analysis

- Models are increasingly sophisticated but... methods for data analysis remain dominated by linear thinking (e.g., expectations of continuity and extrapolation of trends).

- Nonlinear thinking is particularly important when dealing with Climate Change, as adaptation strategies strongly depend on the accuracy and reliability of the forecasts.

Climate change





Global Warming

Global Annual Mean Surface Air Temperature (SAT) Anomaly

Source:
http://data.giss.nasa.gov/gistemp/graphs/(12/1/2011)

Sea Level Rise

Source:
Vermeer and Rahmstorf, PNAS 2009

Climate nonlinear analysis

- Our complex-systems approach to climate data analysis is nonlinear in three aspects:

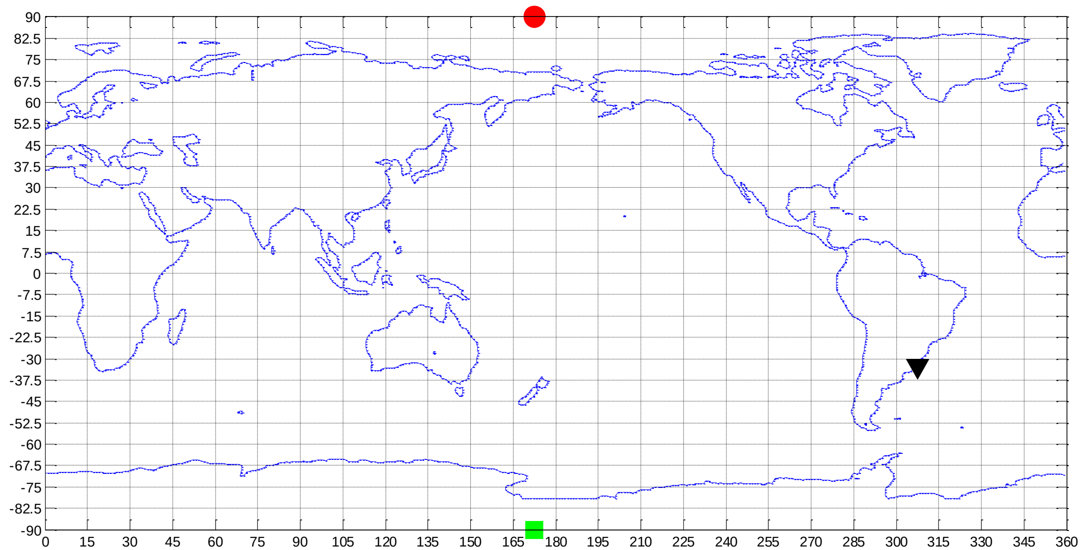
1. We use a nonlinear measure to quantify the degree of 'statistical interdependency' between the climate in two nodes "i" and "j" :

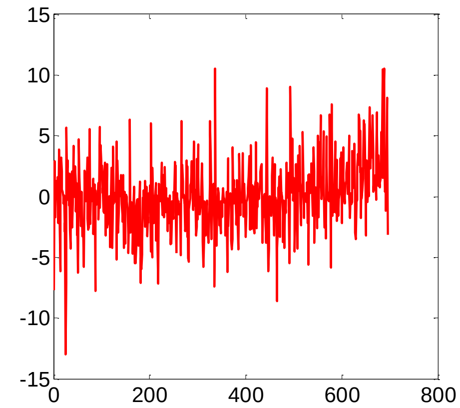
$x_i(t)$ = monthly-averaged SAT anomaly in node "i",
$x_j(t)$ = monthly-averaged SAT anomaly in node "j".

SAT = surface air temperature
Anomaly = annual cycle removed

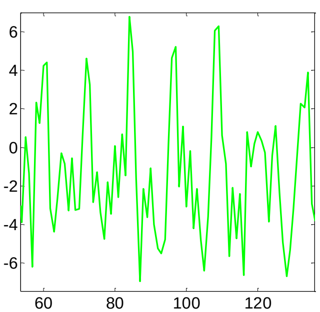
The data: SAT Anomalies
January 1949 -- December 2006

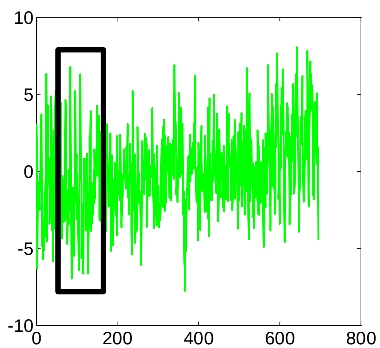
In each 'node' 696 data points (58 years x 12 months)



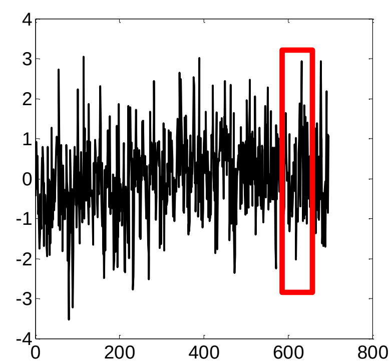


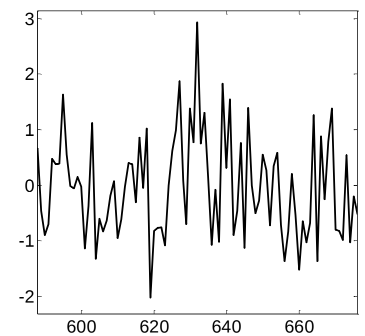
Time (months)





Time (months)





Time (months)

Where does the data comes from?

- Reanalysis of National Center for Environmental Prediction, National Center for Atmospheric Research (NCEP-NCAR, USA).
- Reanalysis = run a sophisticated model of general atmospheric circulation and feed it with the available experimental data, in the different points of the earth, at their corresponding times.
- This process restricts the solution of the model to one as close to reality as possible in regions where there are data available, and to a solution physically “plausible” in regions where no data is available.

Measures of statistical interdependency

- Linear: |Cross-correlation coefficient|

$$C_{ij} = \frac{\sum_{t=1}^{N} (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)}{\sigma(x_i)\sigma(x_j)}$$

- We use a Nonlinear Measure:
  the Mutual Information

$$M_{ij} = \sum_{m,n=1}^{N_{bin}} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)\, p_j(n)}$$

- $M_{ij} = 0 \Leftrightarrow \{x_i\}$ and $\{x_j\}$ are independent: $p_{ij}(m,n) = p_i(m)\, p_j(n)$

  - Our methodology is nonlinear also because:

2. We use a threshold to define the links: "i" $\leftrightarrow$ "j" only if $M_{ij} > \tau$.

Ordinal Pattern time-series analysis

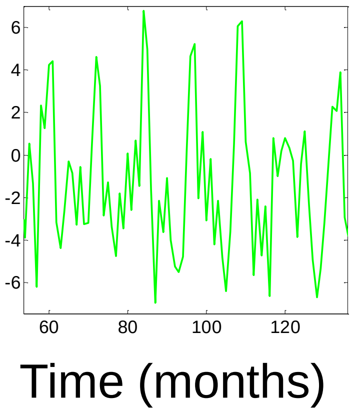
$$M_{ij} = \sum_{m,n=1}^{N_{bin}} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)\,p_j(n)}$$

- PDFs can be calculated from SAT anomalies with histogram method.
  - But our methodology is nonlinear also because:

3. We use nonlinear time-series analysis (ordinal patterns) to compute the PDFs

- In each node we transform the SAT time-series into a sequence of "Ordinal Patterns (OPs)" and compute the PDF of the various OPs.

- The central paradigm is that in climatological data there are patterns of oscillations that repeat from time to time.

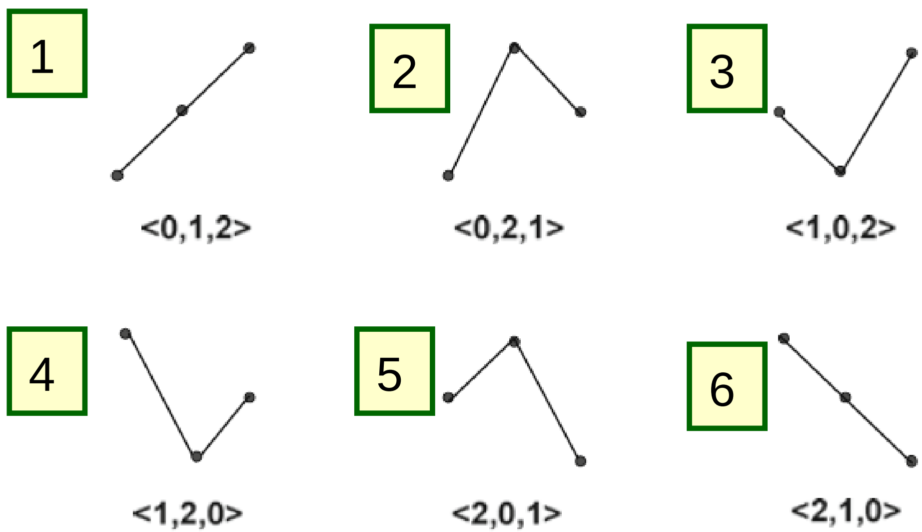


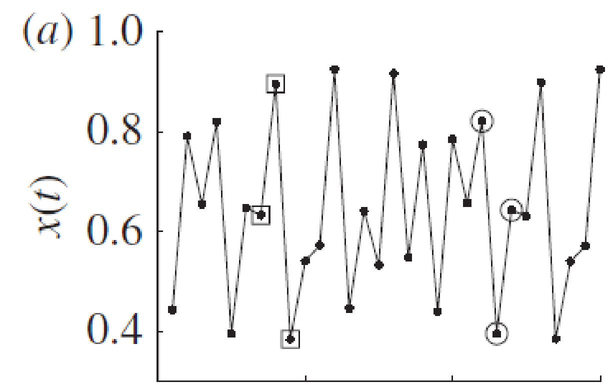
Ordinal Patterns

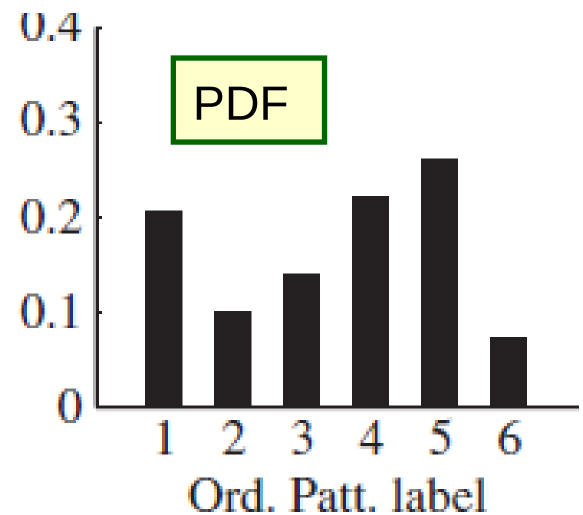
OPs take into account the **order relations** of values in a sequence of values:

$$\{\ x(1), x(2), \dots, \underbrace{x(t+0), x(t+1), x(t+2)}, \dots, x(N-1), x(N)\ \}$$

Geometrical representation of 6 OPs of length 3:







- Good statistics if: N >> D!

Brandt & Pompe, PRL 2002

Ordinal pattern analysis of *climatological* data

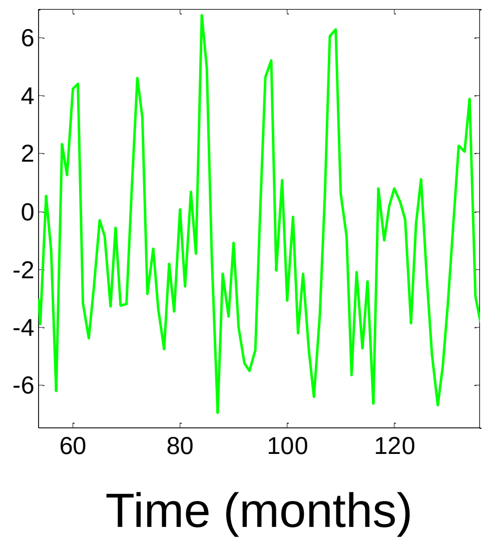
One can construct the OPs comparing monthly-averaged SAT anomalies on:

consecutive years or consecutive months

$[x_i(t), x_i(t+12), x_i(t+24)]$ (inter-annual time-scale)

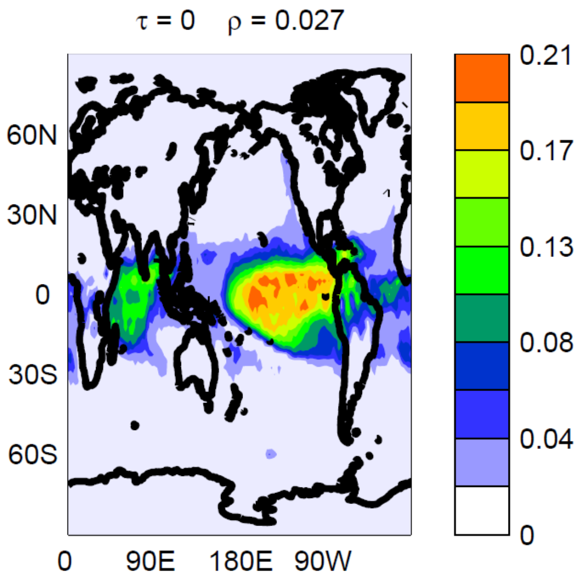
$[x_i(t), x_i(t+1), x_i(t+2)]$ (intra-season time-scale)

- Good statistics if: N=696 >> D! $\Rightarrow$ D=3,4,5
  3! = 6, 4! = 24, 5! = 120

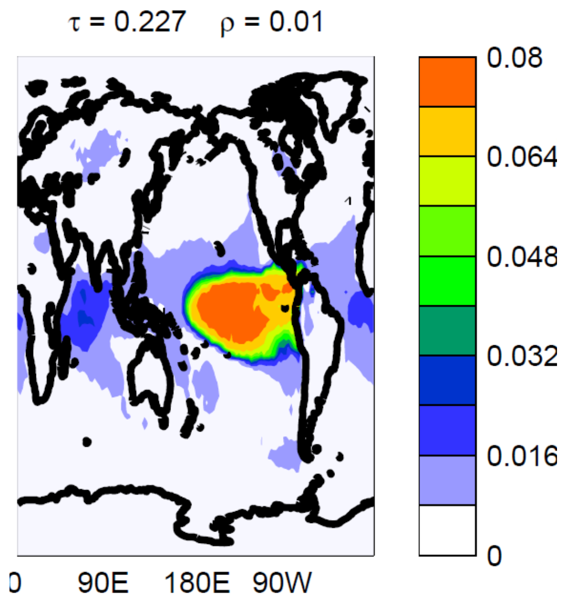


PDF defined from ordinal patterns, concatenating four consecutive years

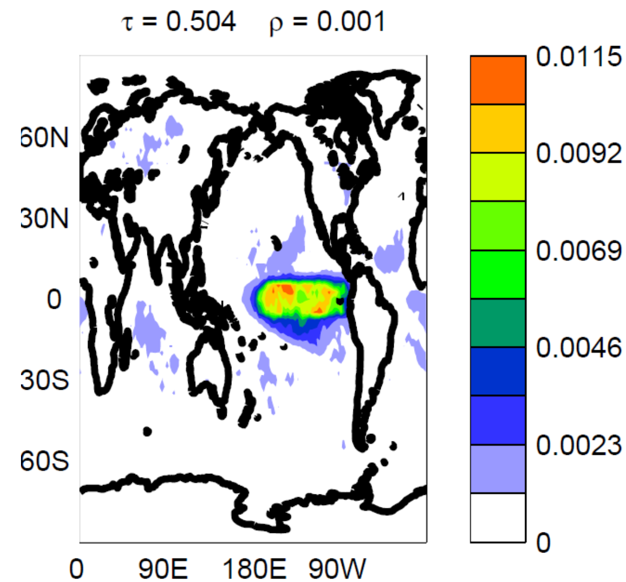
D=4 $[x_i(t), x_i(t+12), x_i(t+24), x_i(t+36)]$



No threshold (all the ***significant links***)



With a threshold such that the network has 1% of the total links



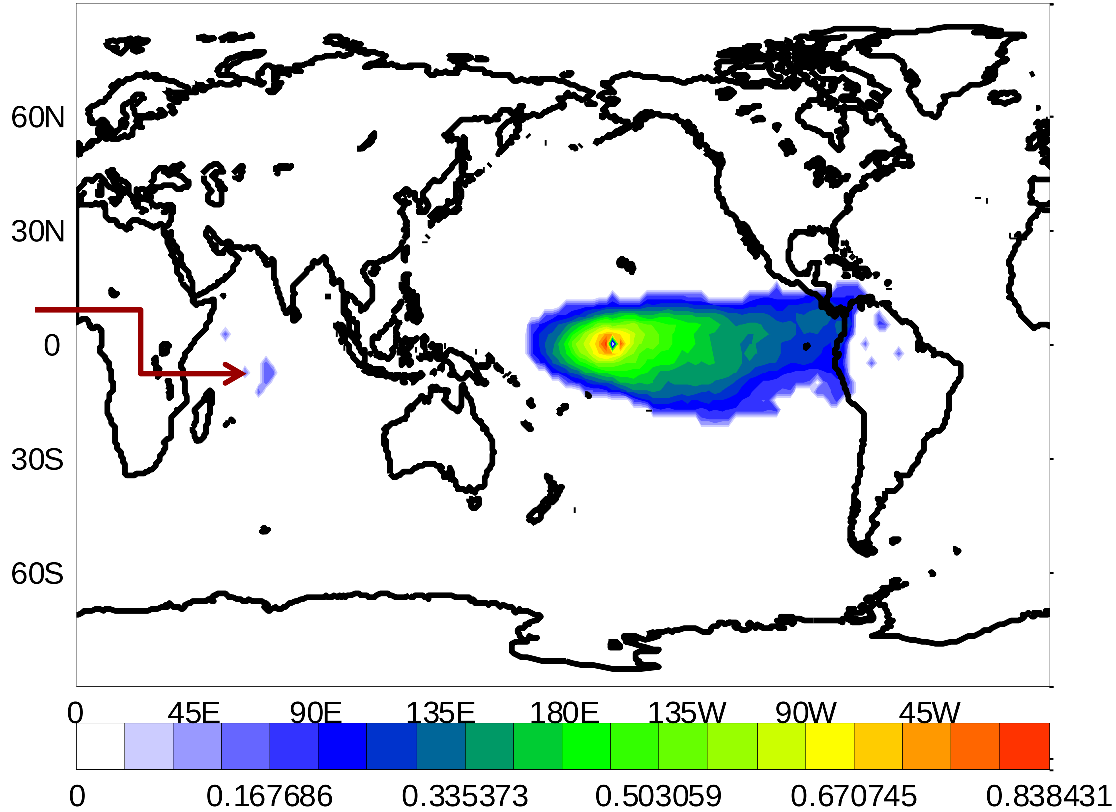
Higher threshold (only the strongest links)

Colors code the Area Weighted Connectivity

Barreiro, Martí and Masoller, Chaos 21, 013101 (2011)

Area to which the "hub" node is connected

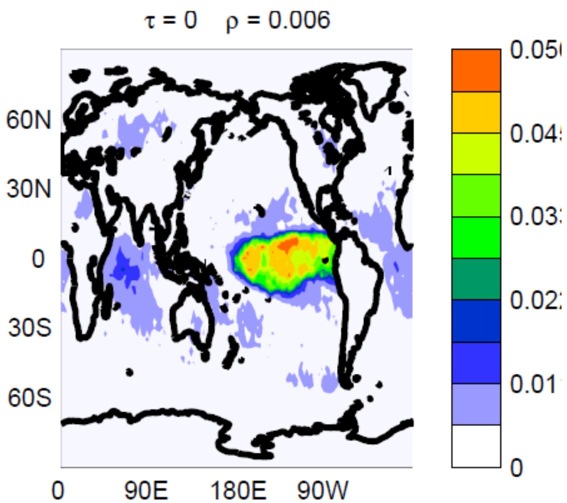
Tele-connection



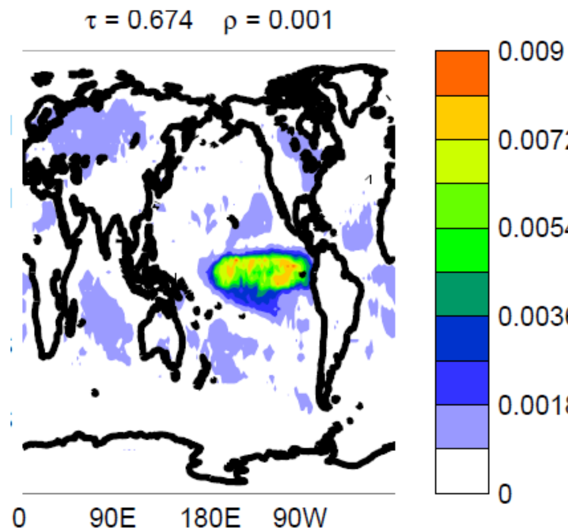
Colors code the mutual information

Longer ordinal pattern

D=5 $[x_i(t), x_i(t+12), x_i(t+24), x_i(t+36), x_i(t+48)]$



No threshold (all the significant links)



With a threshold such that the network has 1% of the total links

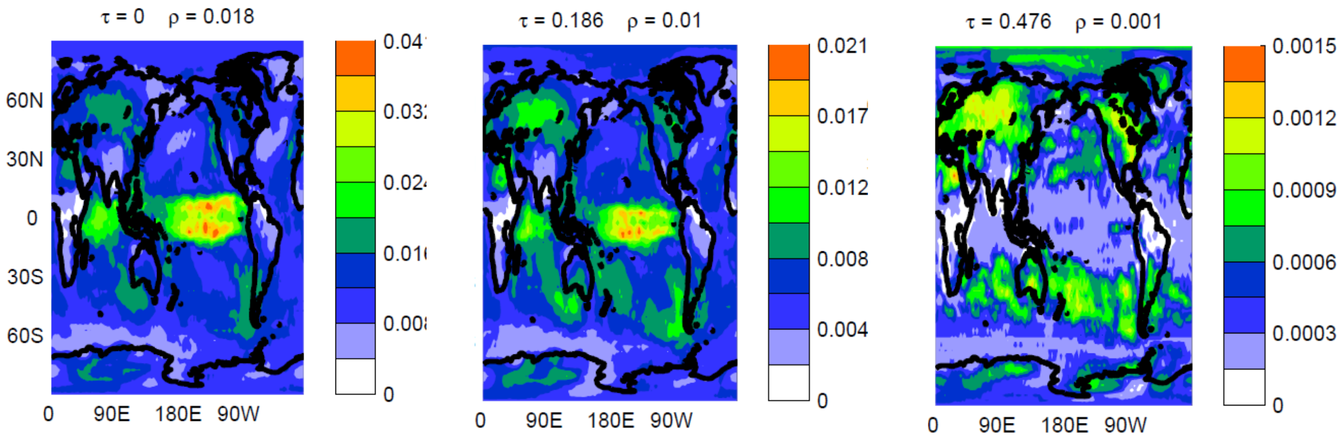
Most of the links that exist for D=4 remain for D=5.

Colors code the area weighted connectivity

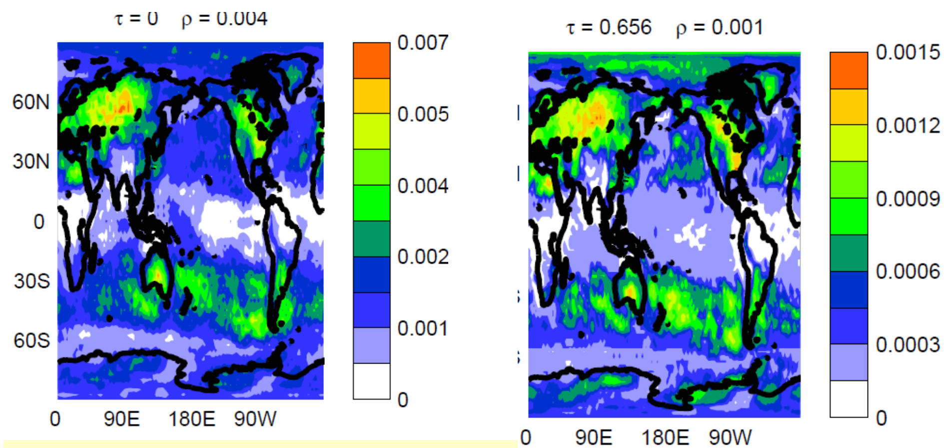
UNIVERSIDAD DE LA REPÚBLICA URUGUAY

OPs constructed by concatenating consecutive months

D=4 $[x_i(t), x_i(t+1), x_i(t+2), x_i(t+3)]$



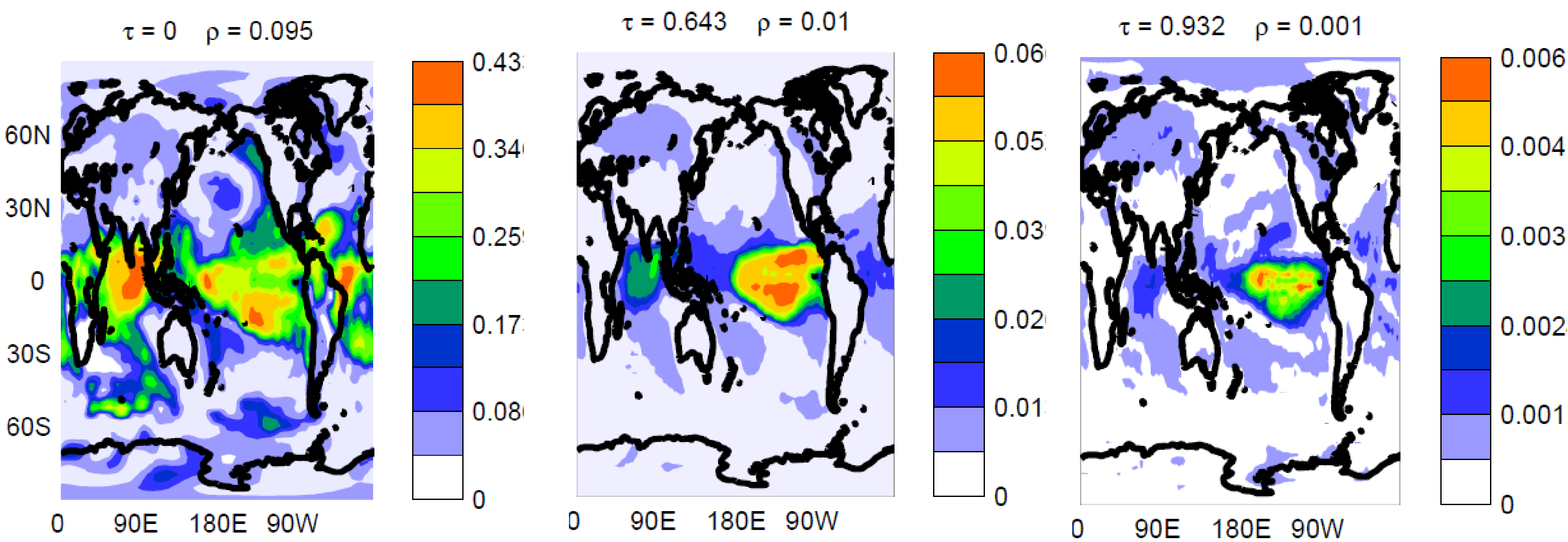
D=5 $[x_i(t), x_i(t+1), x_i(t+2), x_i(t+3), x_i(t+4)]$



- 1% and 0.1% connectivity: very different networks.
- Stronger links (0.1%): the network is almost the same for D=4 and D=5.

Colors code the area weighted connectivity

When using the |cross-correlation| as a measure of statistical interdependency



$$C_{i\,j} = \frac{\sum_{t=1}^{N}(x_i(t)-\bar{x}_i)(x_j(t)-\bar{x}_j)}{\sigma(x_i)\sigma(x_j)}$$

Similar results are obtained when using the Mutual Information, with PDFs defined in terms of histograms of SAT anomalies.

Longer memory effects?

D≥6 ?

Problem: the number of possible ordinal patterns for D=6 is 6! = 720.
the length of the time series is N=696 (58 years)

Not enough data!

A possible solution: Binary representations

$s = 1$ if $x > x_o$, else $s=0$

SAT anomalies: $x_0 = 0$

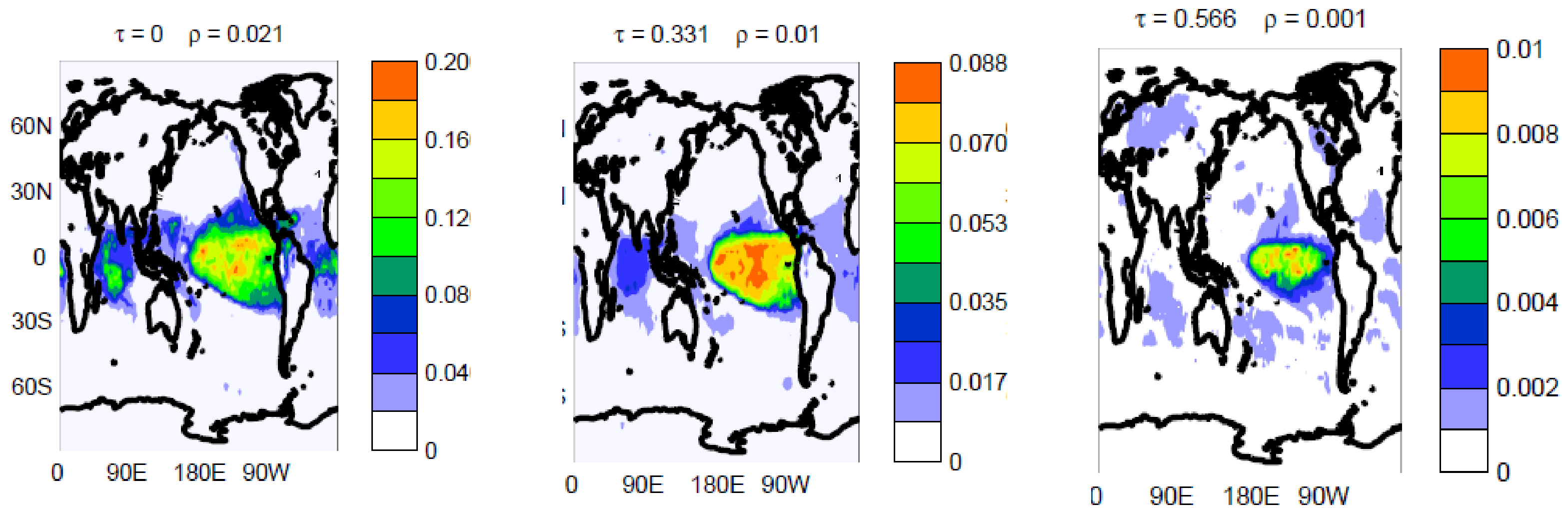
For binary patterns of length *D*, the # of possible patterns is 2^D

For $D=6$, $2^6 = 64 \ll 720$

Binary representation, concatenating consecutive years

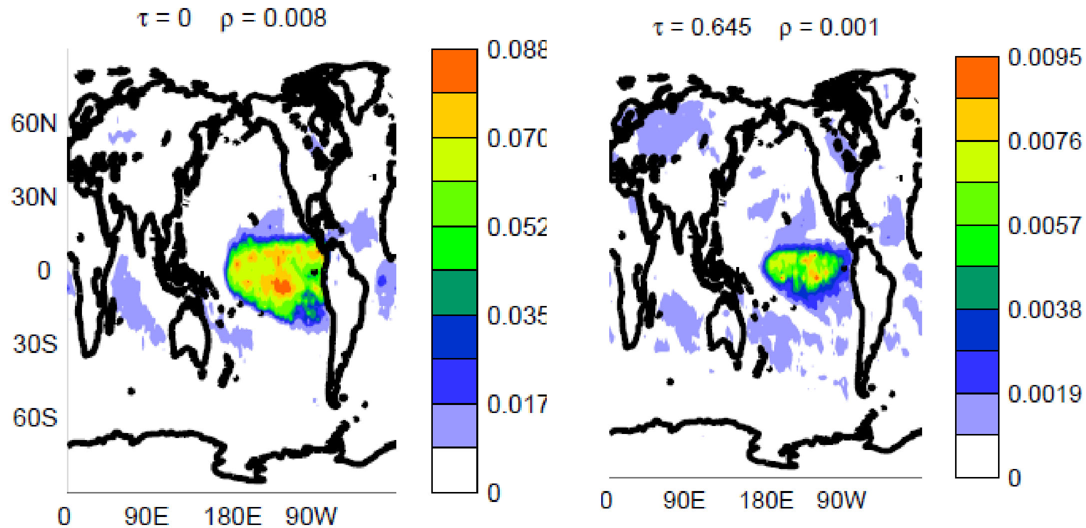
D=5,

2^5=32



D=6,

2^6=64

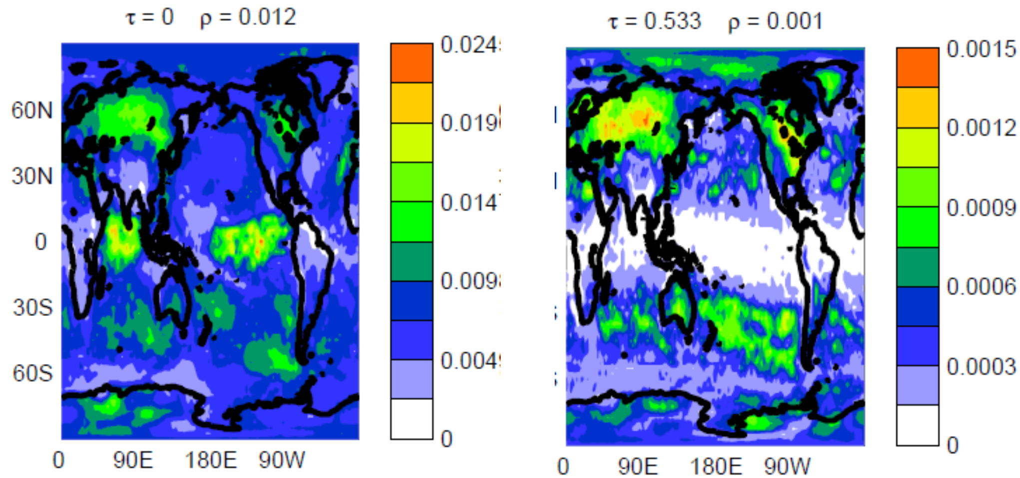


Most of the links that exist for D=5 remain for D=6.

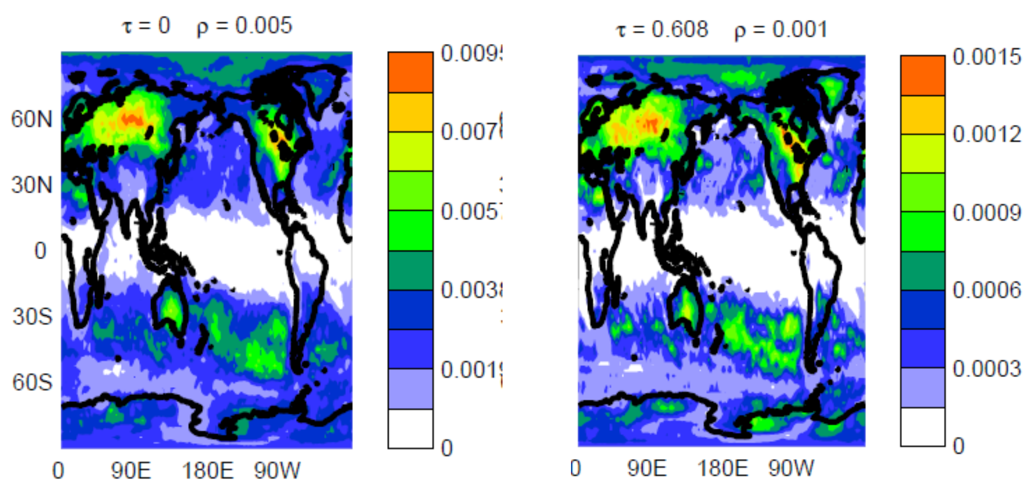
Colors code the area weighted connectivity

Binary representation, concatenating consecutive months

D=5

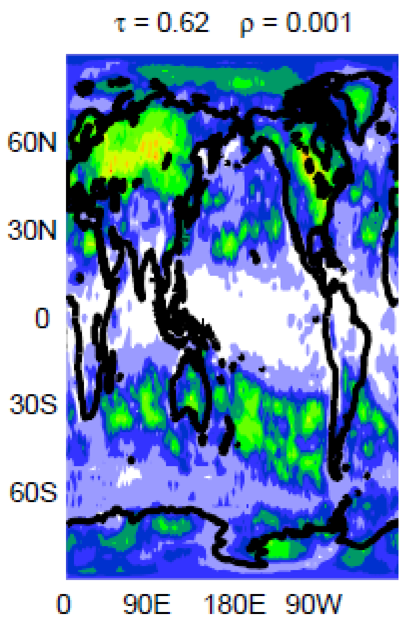


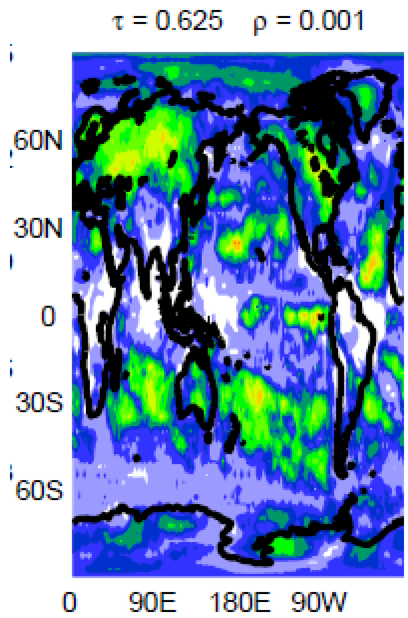
D=6

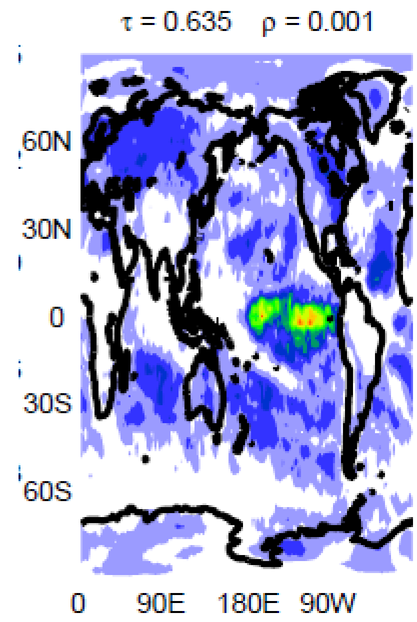


- 1% and 0.1% connectivity: very different networks.
- Stronger links (0.1%): the network is almost the same for D=5 and D=6.

Influence of the pattern time interval keeping fixed the pattern size (D=6) and the network density (0.1%)







6 months

$$[x_i(t), x_i(t+1), x_i(t+2), x_i(t+3), x_i(t+4), x_i(t+5)]$$

1 year

$$[x_i(t), x_i(t+2), x_i(t+4), x_i(t+6), x_i(t+8), x_i(t+10)]$$

2 years

$$[x_i(t), x_i(t+4), x_i(t+8), x_i(t+12), x_i(t+16), x_i(t+22)]$$

Summary and future work

- We have shown that ordinal patterns and symbolic analysis are powerful tools for the analysis of the large-scale topology of the climate network.
- The success of the method is based on an appropriate partition of the phase space that results in **a probability distribution function (PDF) that fully characterizes the diversity of patterns** present in the climate.
- We applied the method to the analysis of monthly-averaged Surface Air Temperature anomalies.
- Ordinal and Binary Patterns covering different time intervals (intra-season and inter-annual) reveal **long memory processes**.
- Future work: detection of directionality and causal relations.

THANK YOU FOR YOUR ATTENTION