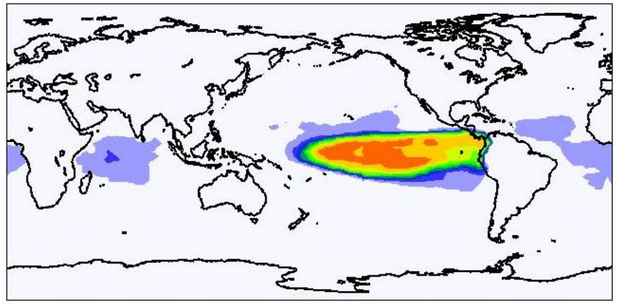


Inferring the connectivity of coupled dynamical units from time-series statistical similarity analysis

Cristina Masoller
Universitat Politecnica de Catalunya
Cristina.masoller@upc.edu
www.fisica.edu.uy/~cris

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

Workshop on Generalized Network Structures and Dynamics
The Mathematical Biosciences Institute
Ohio State University
March 2016

ICREA

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

Where are we?


Viernes, 25 de septiembre de 2009 **Diari de Terrassa**

El edificio Gaia centraliza grupos científicos consolidados y emergentes.

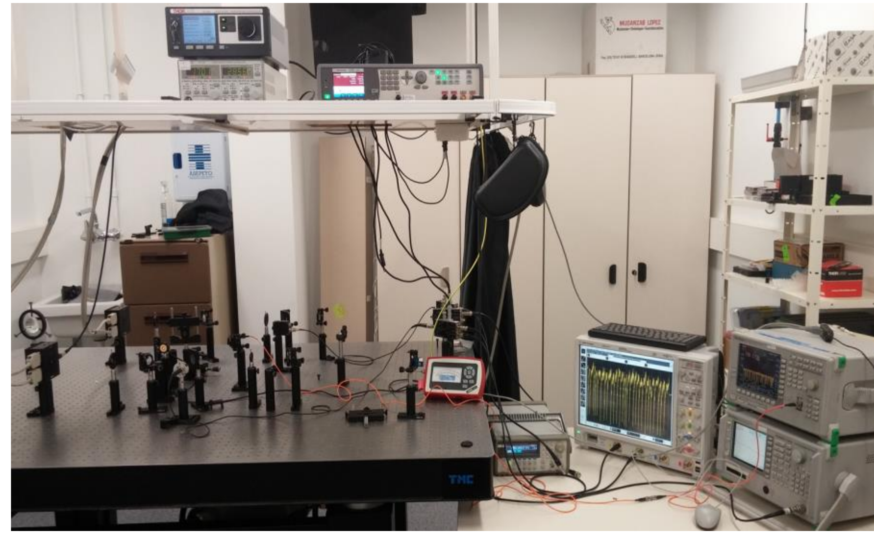
Who are we?

Research group on Dynamics, Nonlinear Optics and Lasers

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

What do we study?



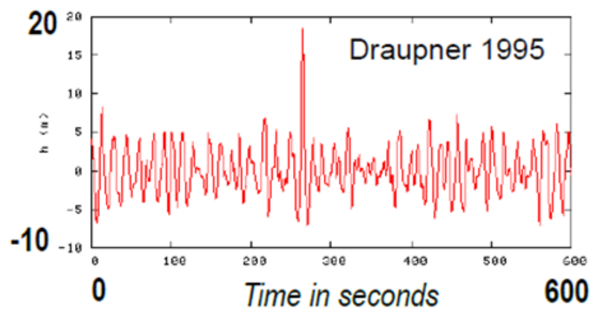
- Nonlinear phenomena in complex systems
  - Photonics (dynamics of lasers, nonlinear optics)
  - Biophysics (excitability, coupled oscillators)
  - Data analysis (climate time-series, biomedical images)

Climate and lasers??

- Optical systems allow recording long time-series under controlled conditions.
- This allows testing novel analysis tools (prediction, classification, etc.).

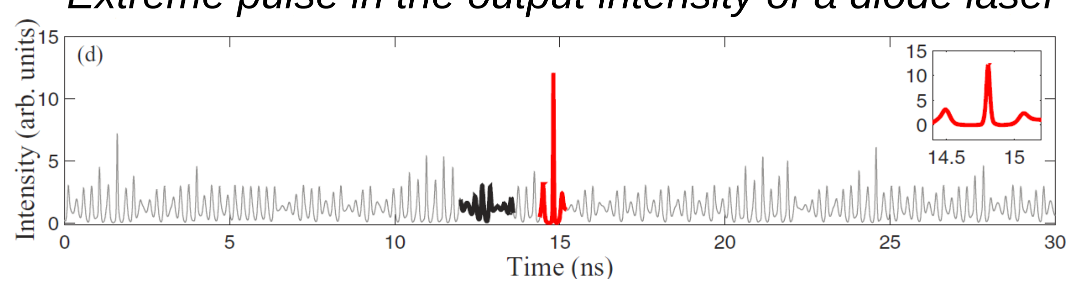
Elevation of the sea surface (in meters)

- Ocean rogue wave ⇒

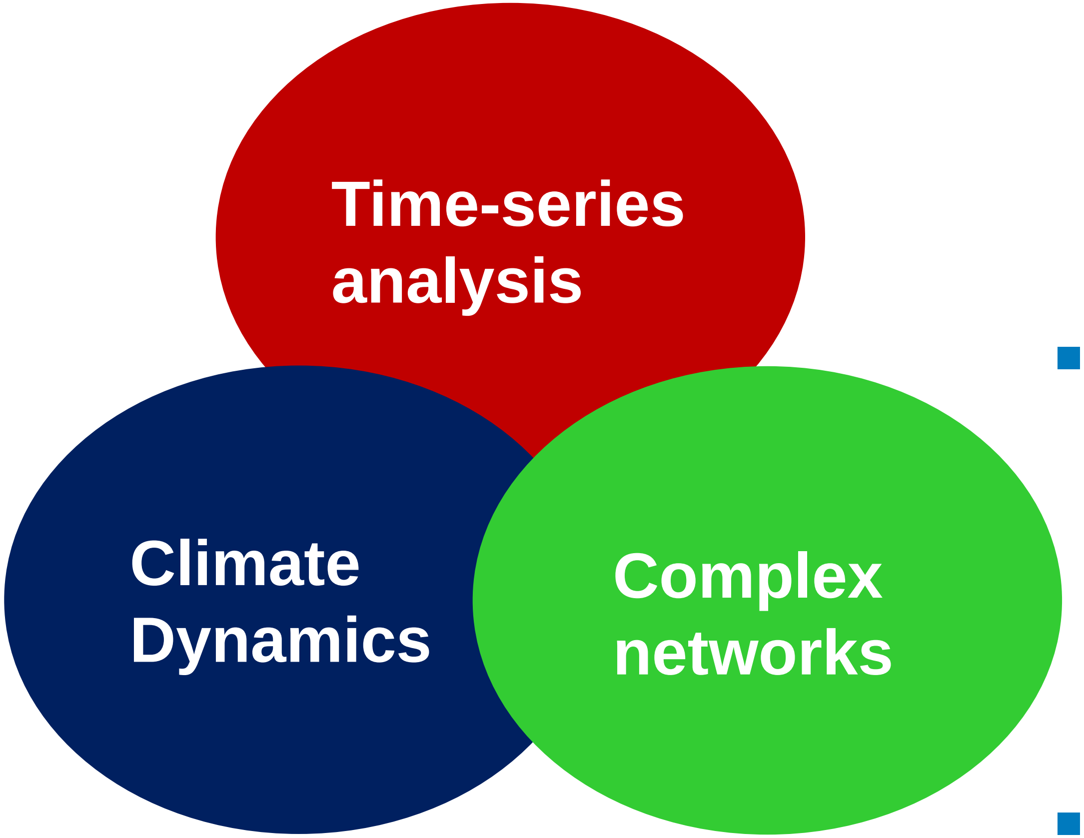


Extreme pulse in the output intensity of a diode laser

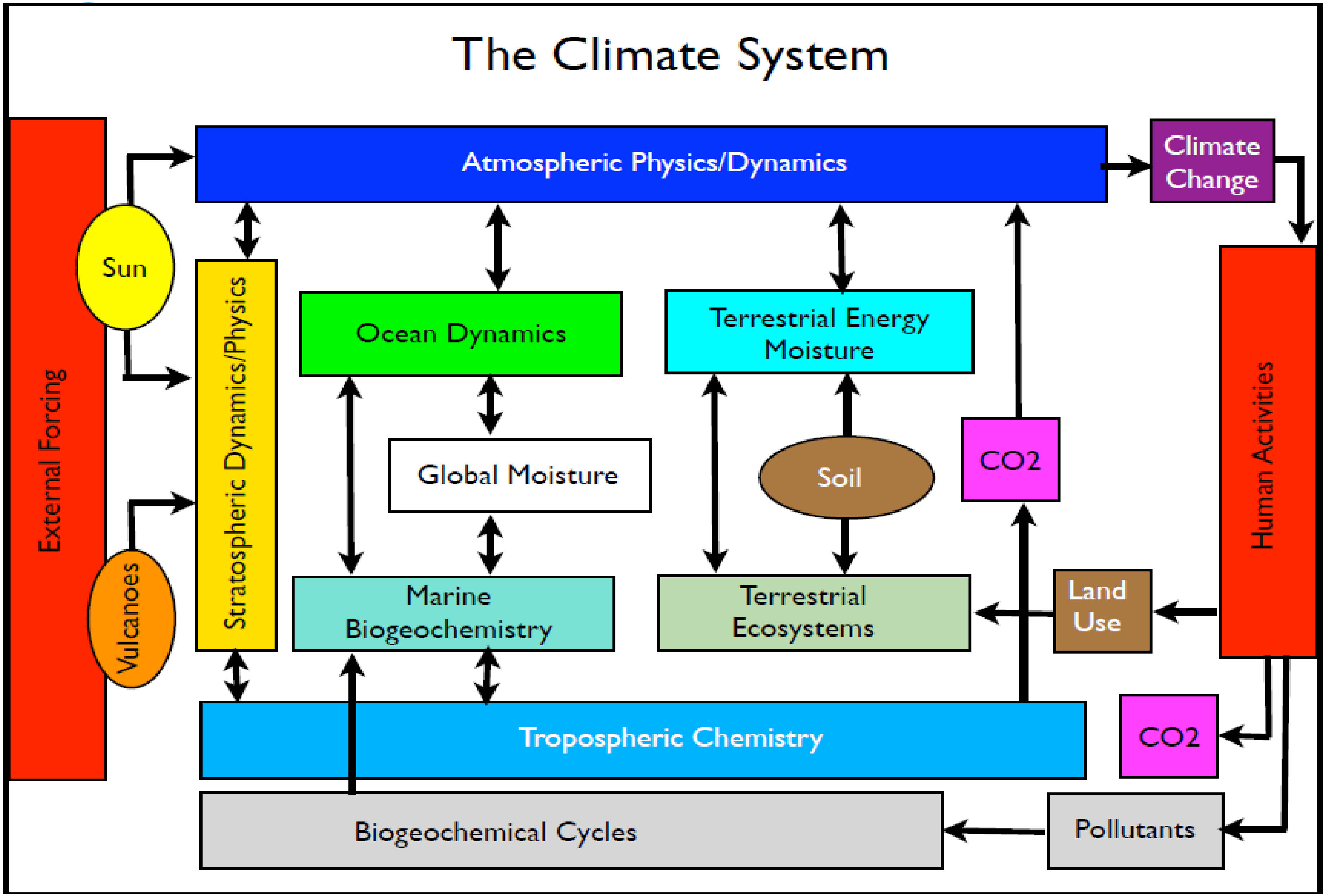
- Optical rogue wave ⇒



Outline



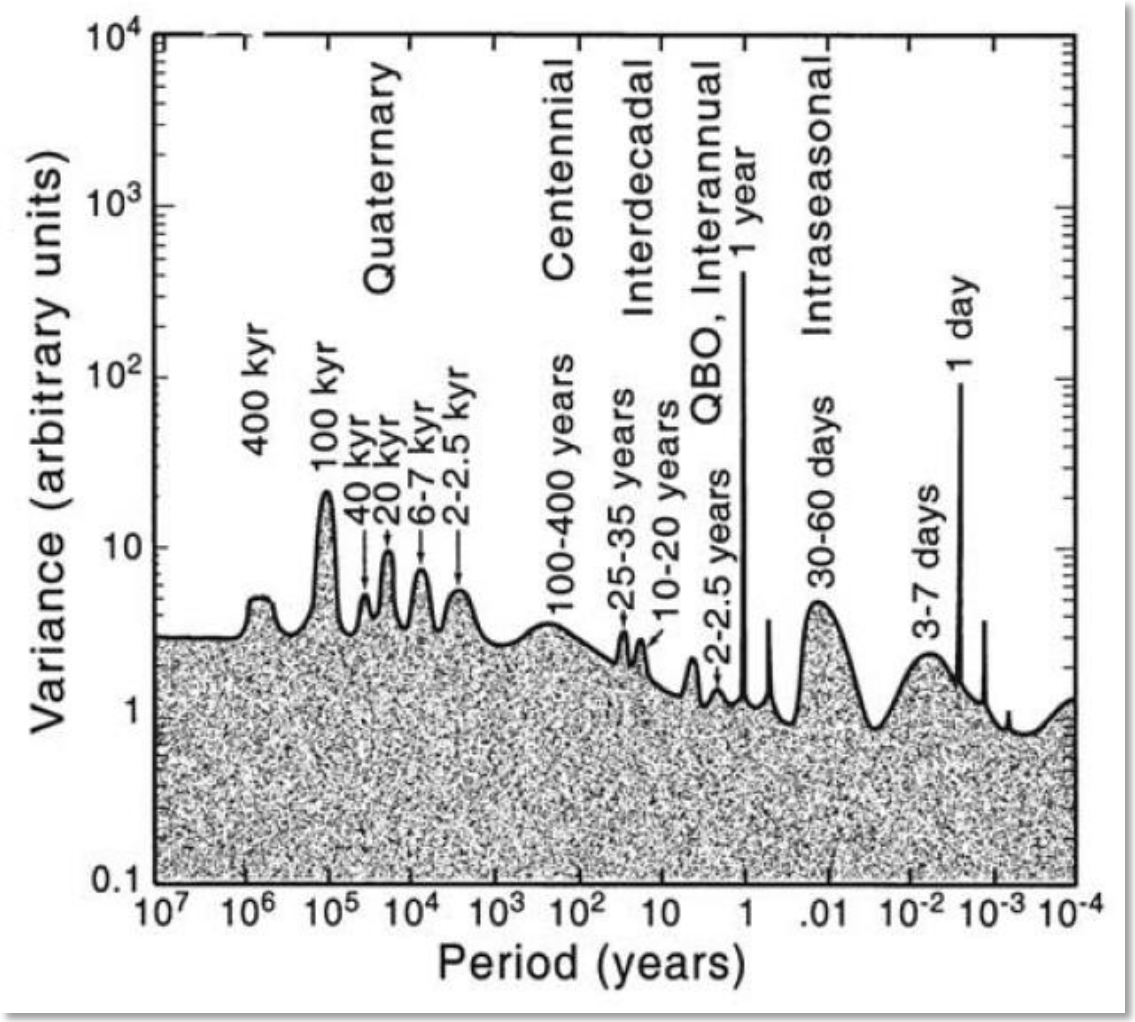
- Introduction
  - Climate dynamics
  - Symbolic method of time-series analysis
- Results: networks
  - Inferring the network connectivity
  - Inferring climate communities
- Summary



Courtesy of Henk Dijkstra (Ultrech University)

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

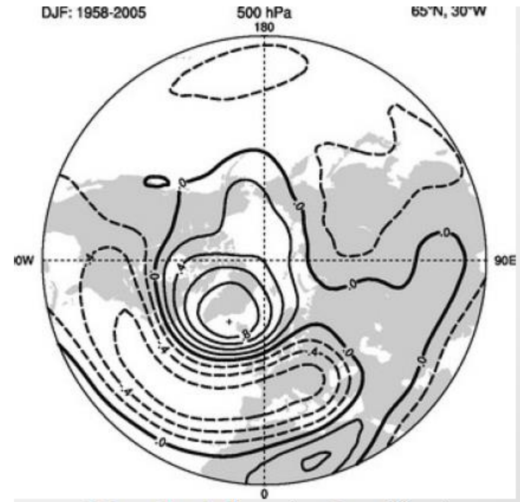
- hours to days,
- months to seasons,
- decades to centuries,
- and even longer...



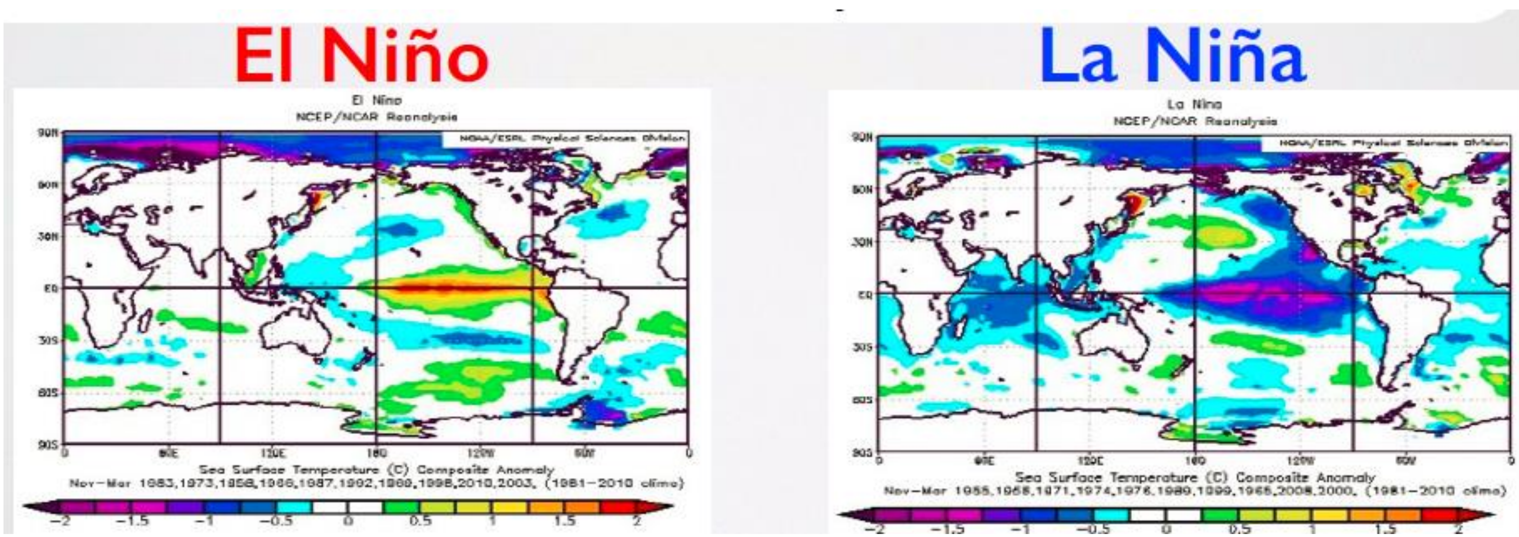
An "artist's representation" of the power spectrum of climate variability (Ghil 2002).

And a wide range of spatial modes of variability

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



North Atlantic Oscillation



WHAT DO NETWORKS HAVE TO DO WITH CLIMATE?

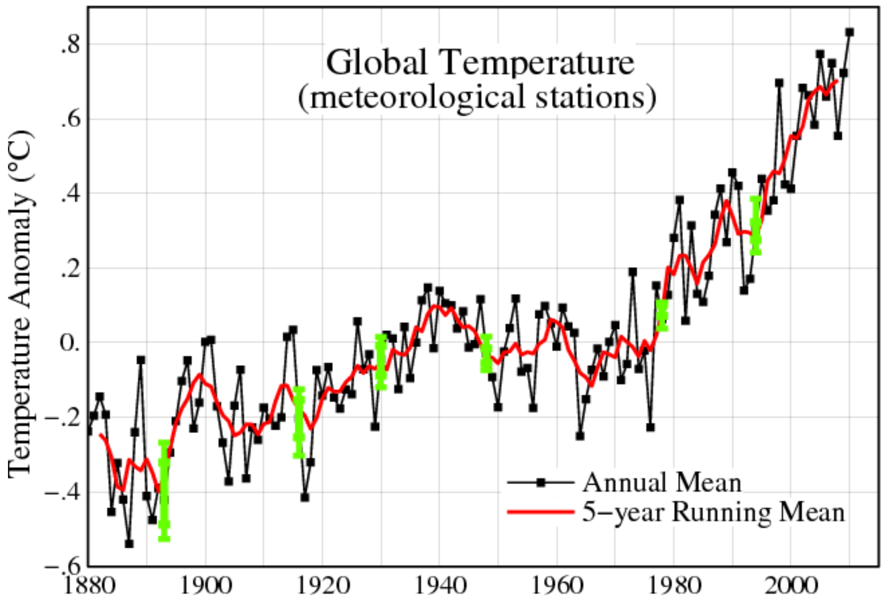
BY ANASTASIOS A. TSONIS, KYLE L. SWANSON, AND PAUL J. ROEBBER

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.

Time-scale of our analysis: weather vs. climate

- weather = short-term variability

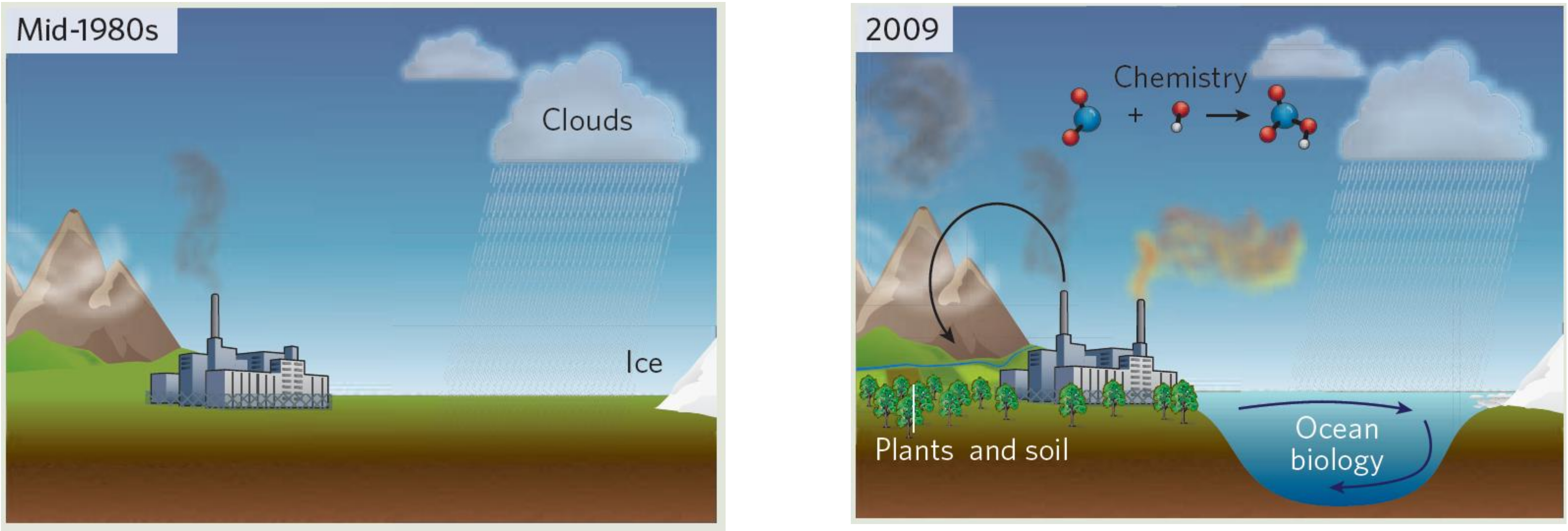
- climate = long-term trend



- Global warming

Monte Perdido (Spain)

Climate Modeling



- Nowadays climate models capture many physical and biophysical processes.
- BUT many "feedback loops" (e.g., due to human adaptation activity) are poorly understood and not represented in models.
- Clear need of "data driven" studies.
- Clear need of reliable high-resolution spatio-temporal data.

Nature, February 2010

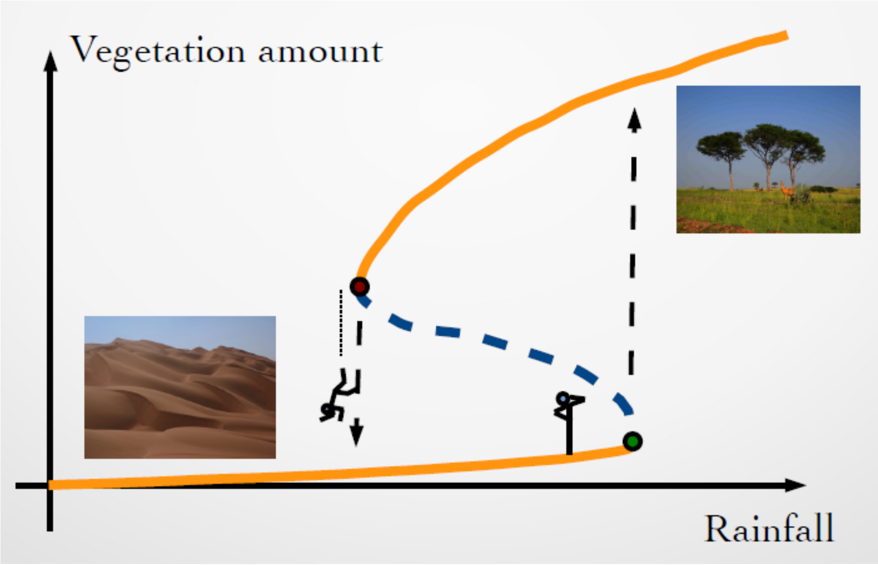
UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

The importance of being nonlinear

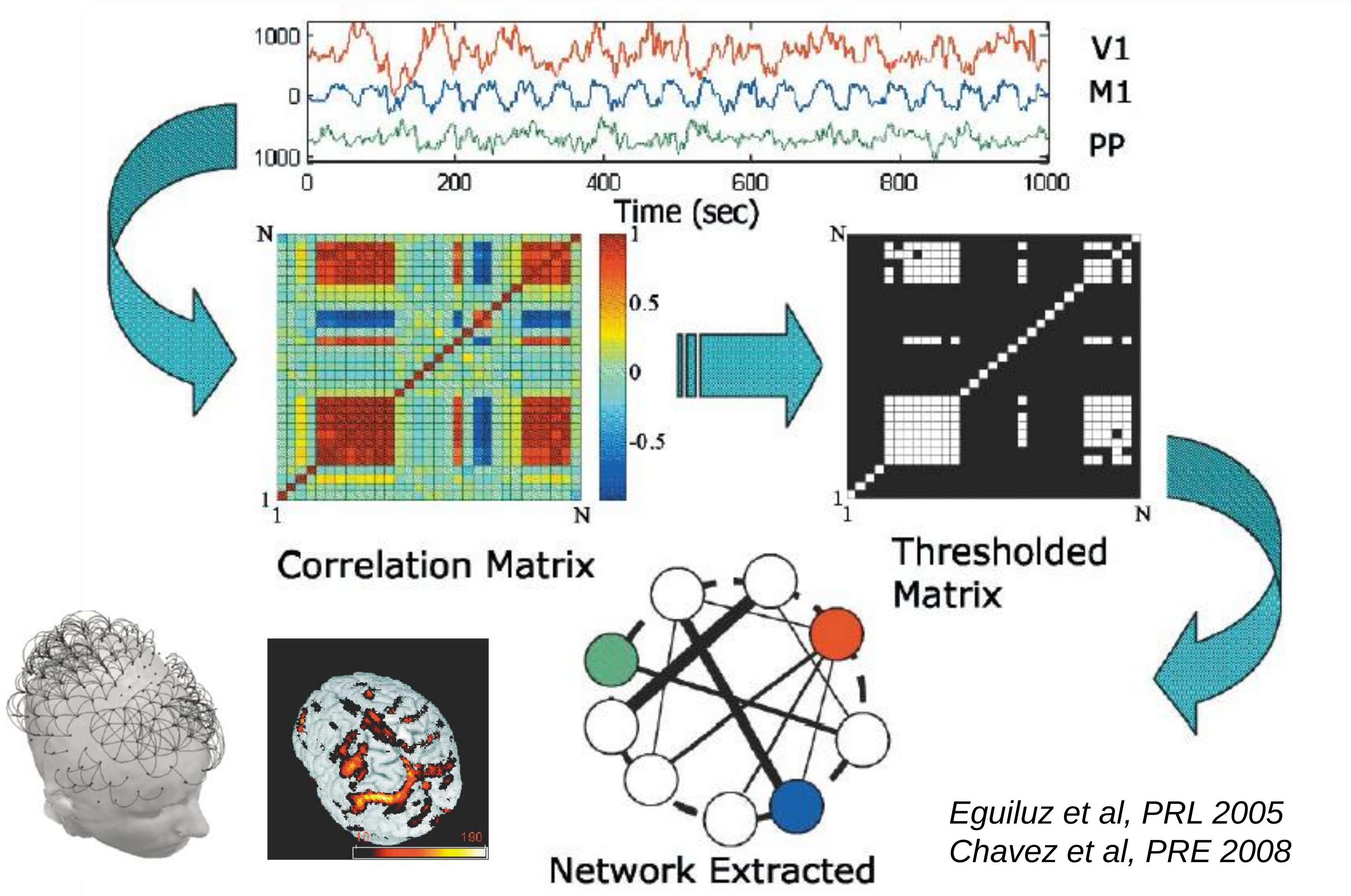
- Methods of data analysis are dominated by linear thinking (example: expectations of continuity; extrapolation of trends).

- BUT in complex systems nonlinear thinking is crucial!
- Examples: accurate forecasts of critical transitions & extremes.



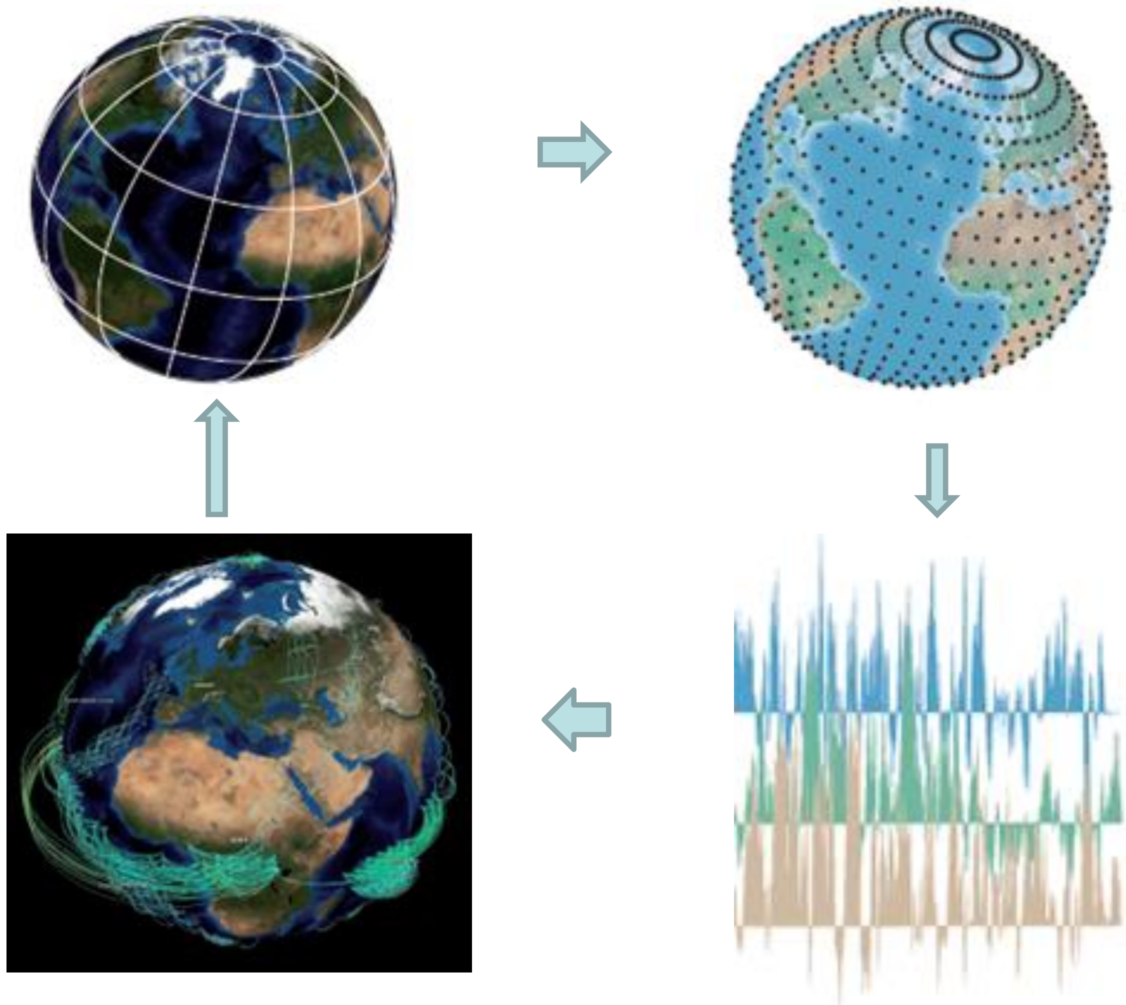
Bangladesh, Nature 2014

Brain functional network

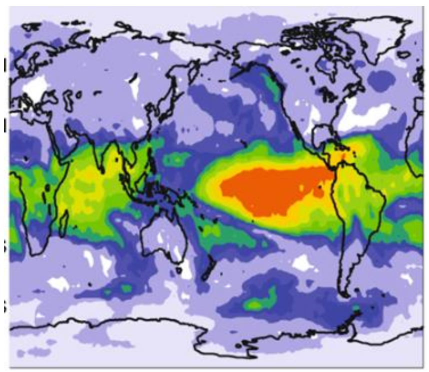


Eguiluz et al, PRL 2005
Chavez et al, PRE 2008

Climate networks



Area-weighted connectivity (weighted degree)

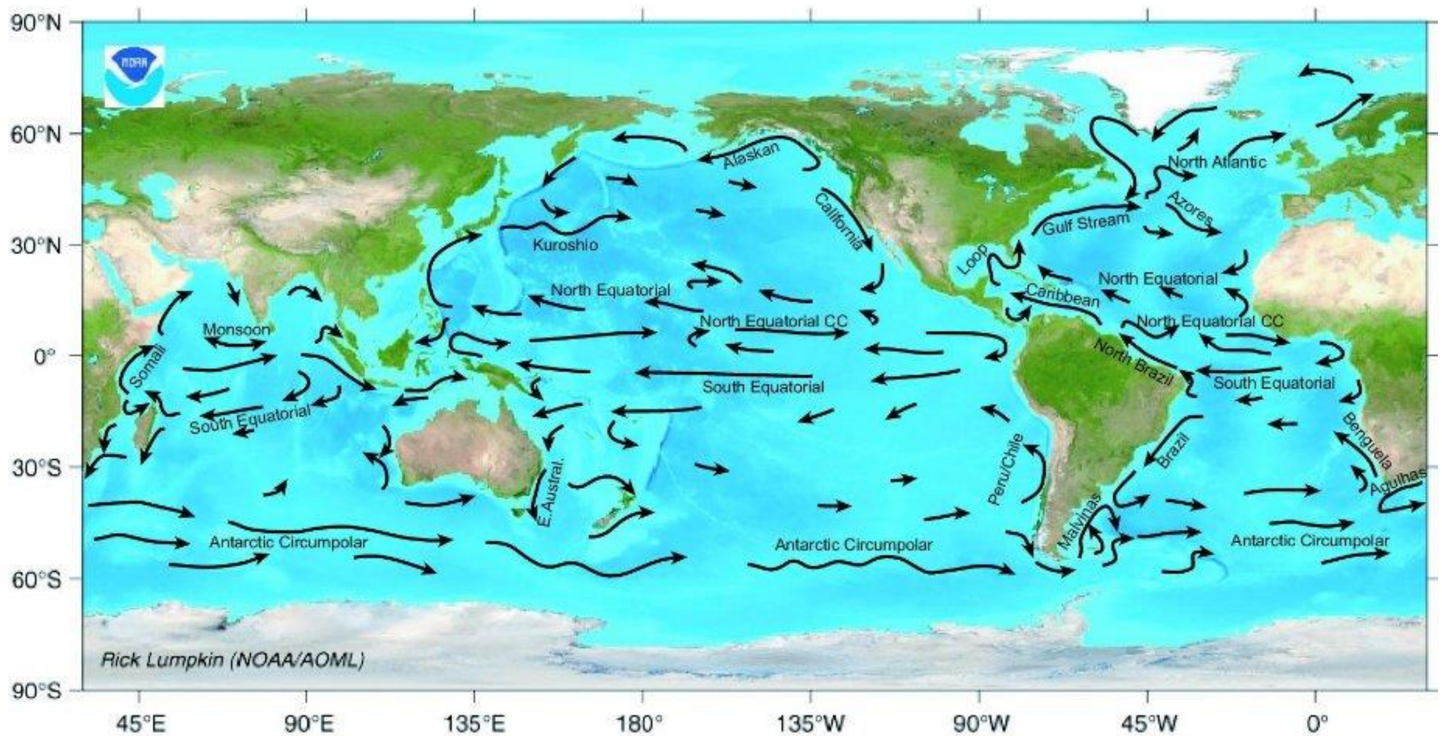


Deza et al, Chaos 2013

Donges et al, Chaos 2015

Physical mechanisms responsible for teleconnections

Winds, ocean currents and solar forcing.

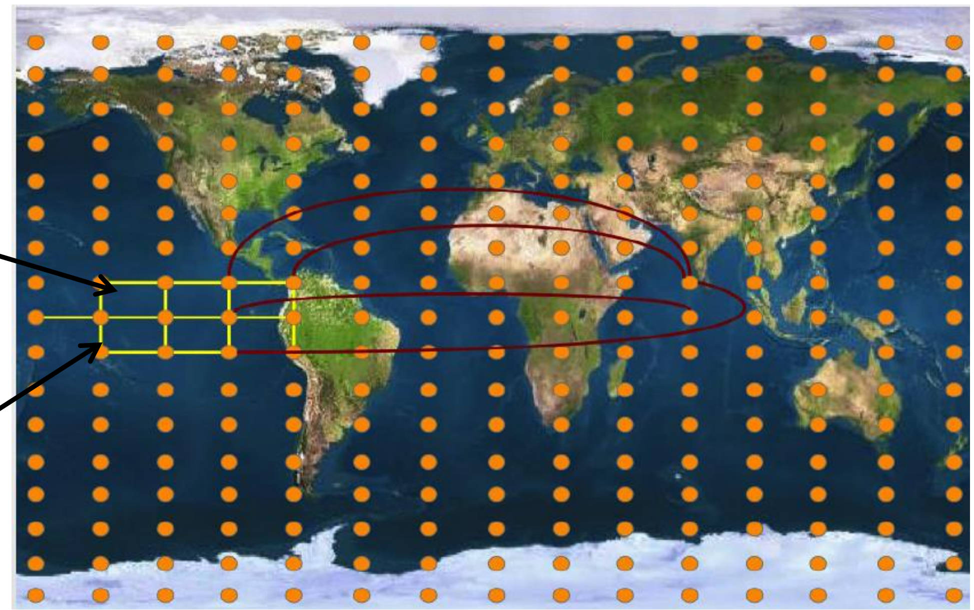


http://www.aoml.noaa.gov

Climate networks

Regular grid
2.5º x 2.5º
⇒ 10226 nodes

$X_i(t)$

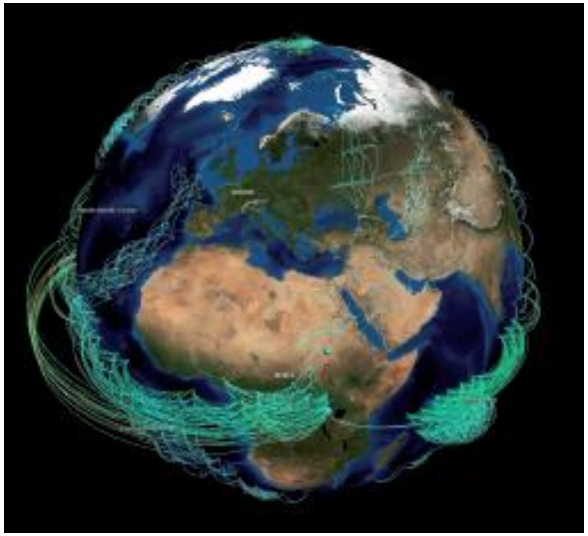


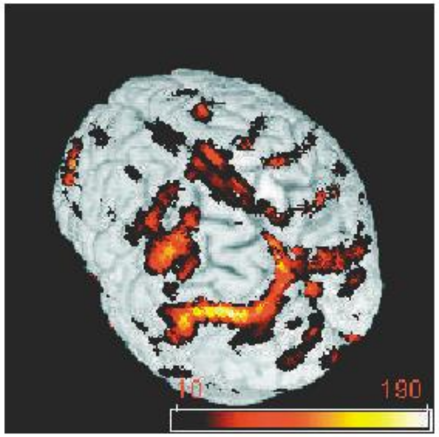
$X_j(t)$

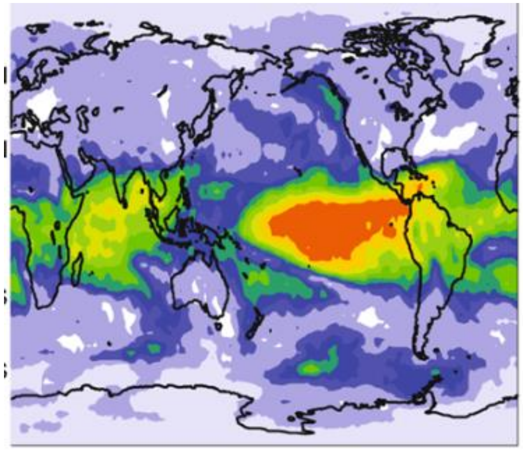
- CNs constructed from an **interdependency/causality** analysis of a climate variable.
- Which climate variable? **surface air temperature, surface sea temperature, wind velocity, precipitation, etc.**
- Interdependency measure: usually **cross-correlation or mutual information.**
- Causality measure: **conditional mutual information or Granger estimator**

Brain network

Climate network





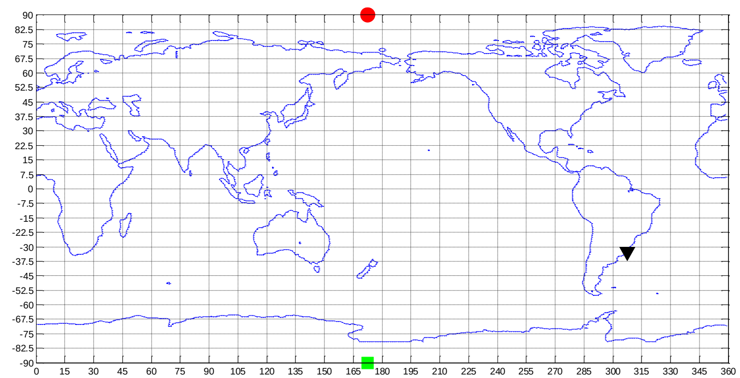


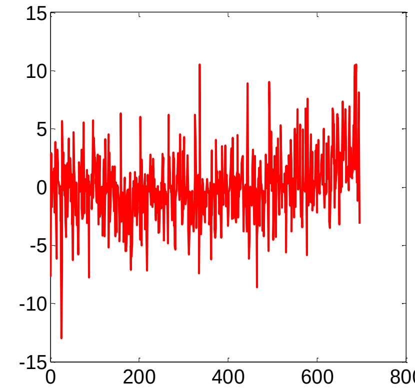
The data: monthly surface air temperature (SAT) 1949-2013

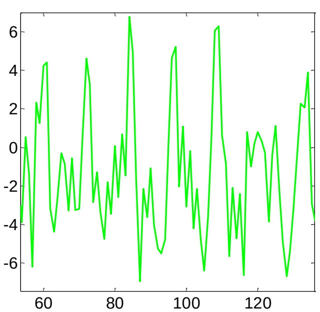
Anomalies = <u>annual solar cycle removed</u>

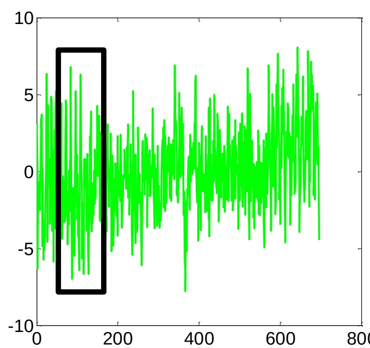
In each of the 10226 nodes ≈700 data points (60 years x 12 months)

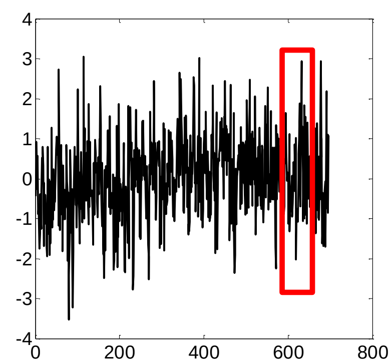
How does the data look like?

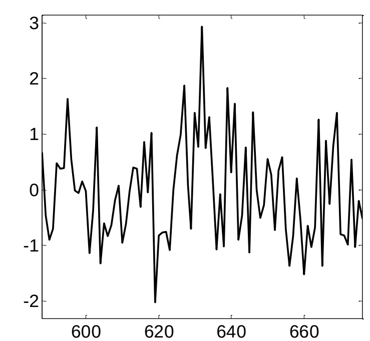












Where does the data come from?

- National Center for Environmental Prediction, National Center for Atmospheric Research (NCEP-NCAR).

- Freely available.

- Reanalysis = run a sophisticated model of general atmospheric circulation and feed the model (data assimilation) with empirical data, where and when available.

- This process restricts the solution of the model to one as close to reality as possible in regions/times where there are data available, and to a solution physically "plausible" in regions/times where no data is available.

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Our analysis: nonlinear in three aspects

- We use a **nonlinear measure** to quantify 'statistical interdependency' between the climate in different regions.

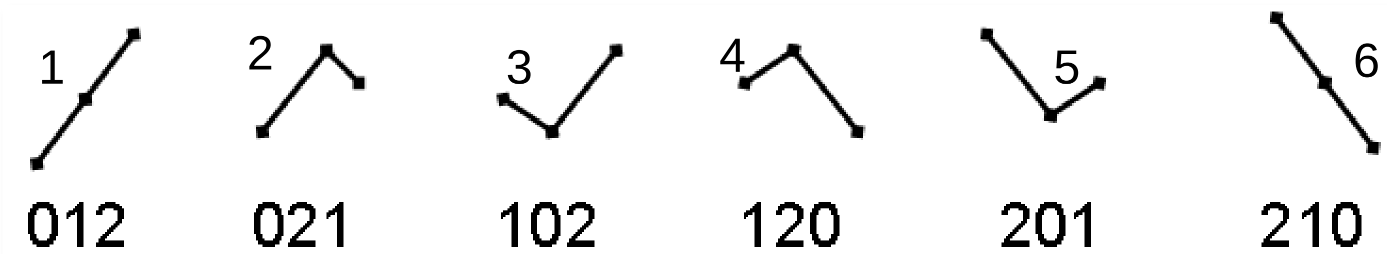
$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

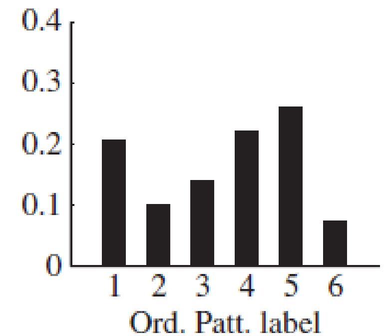
- We use a **threshold** to select the significant M_{ij} values (contrasting M_{ij} values obtained from original time-series with M_{ij} values obtained from surrogates).
- We use **symbolic** time-series analysis (ordinal patterns) to compute the probabilities.

Method of symbolic time-series analysis: ordinal patterns

- X= $\{\ldots x_i, x_{i+1}, x_{i+2}, \ldots\}$

Brandt & Pompe, PRL 88, 174102 (2002)





The OP probabilities allow to identify frequent patterns in the *ordering* of the data points

Random data $\Rightarrow$ OPs are equally probable

– Advantage: the probabilities uncover temporal correlations.

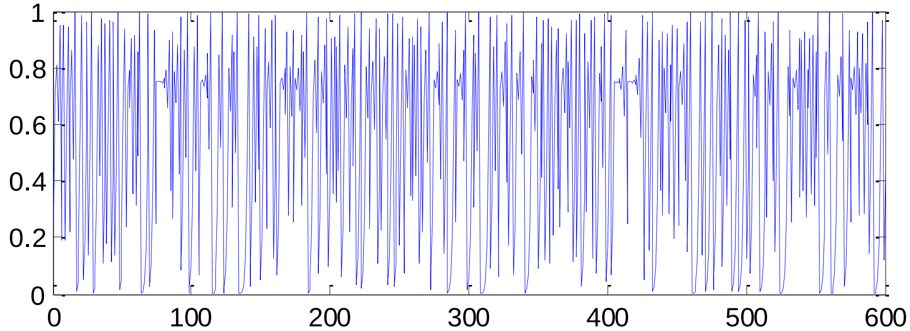
– Drawback: we lose information about the actual values.

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

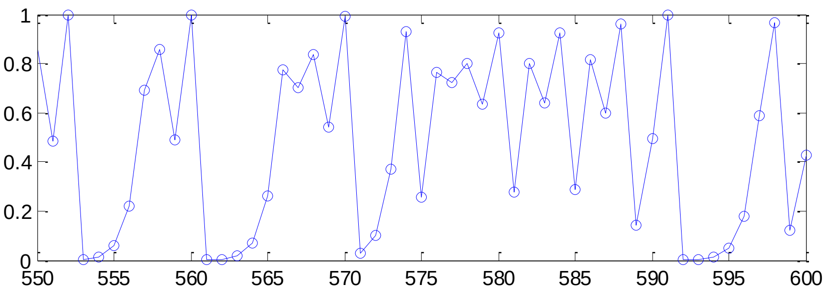
Campus d'Excel·lència Internacional

Example: the logistic map $x(i+1)=4x(i)[1-x(i)]$

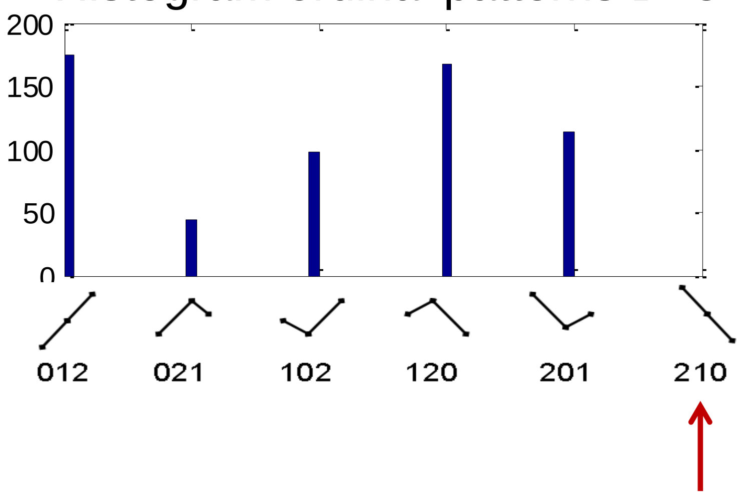
Time series



Detail

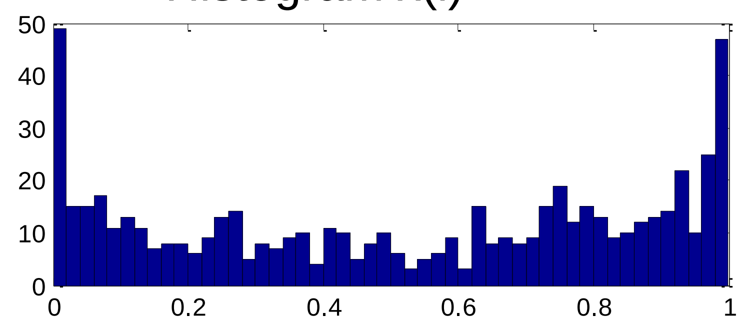


Histogram ordinal patterns D=3



Forbidden pattern

Histogram x(i)



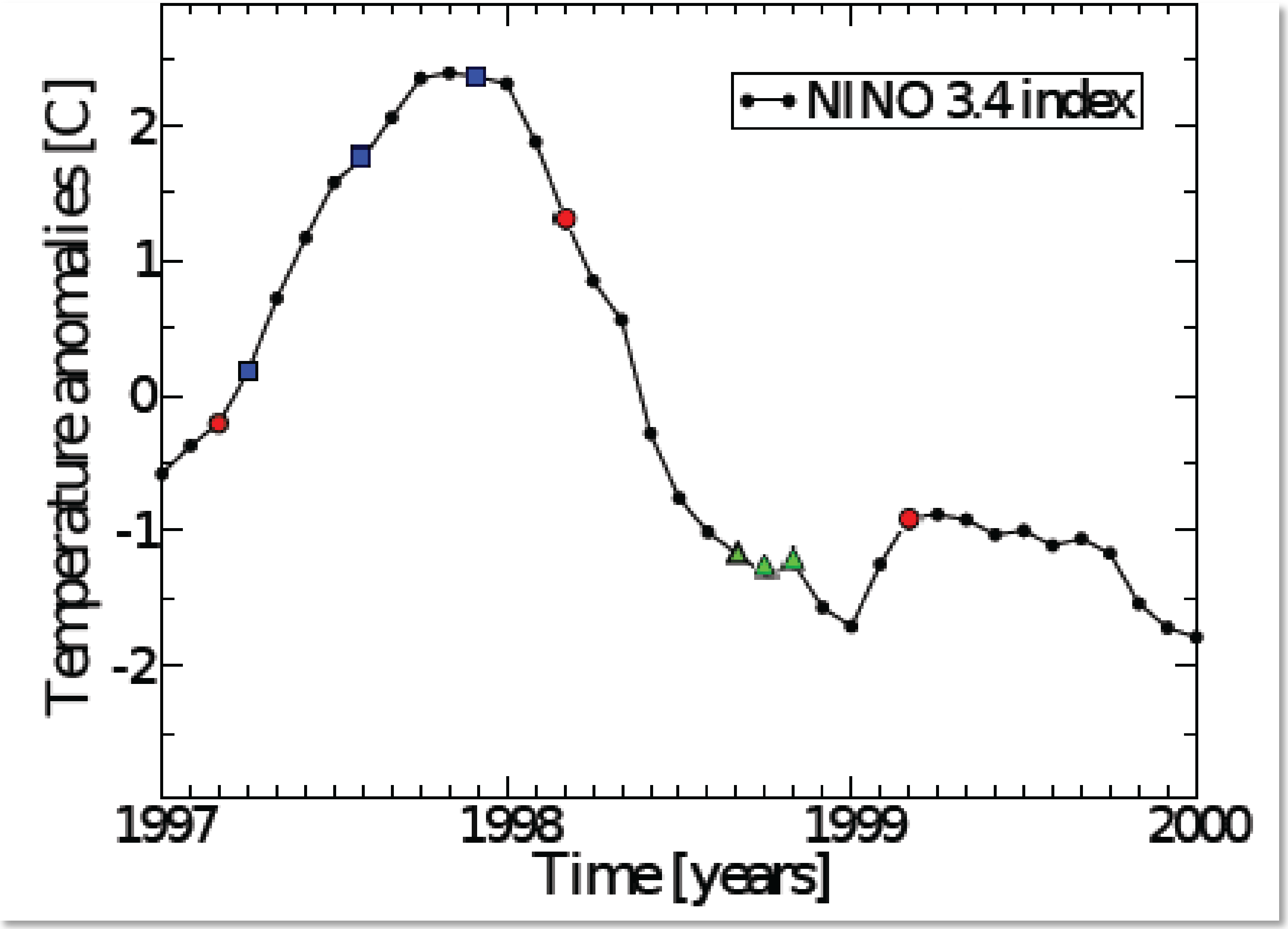
Ordinal analysis provides complementary information.

Ordinal analysis allows selecting the time scale of the analysis

Intra-season 102

Intra-annual 012

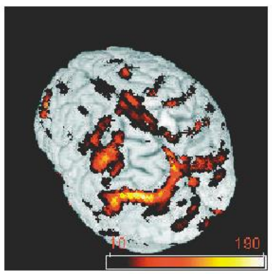
Inter-annual 120

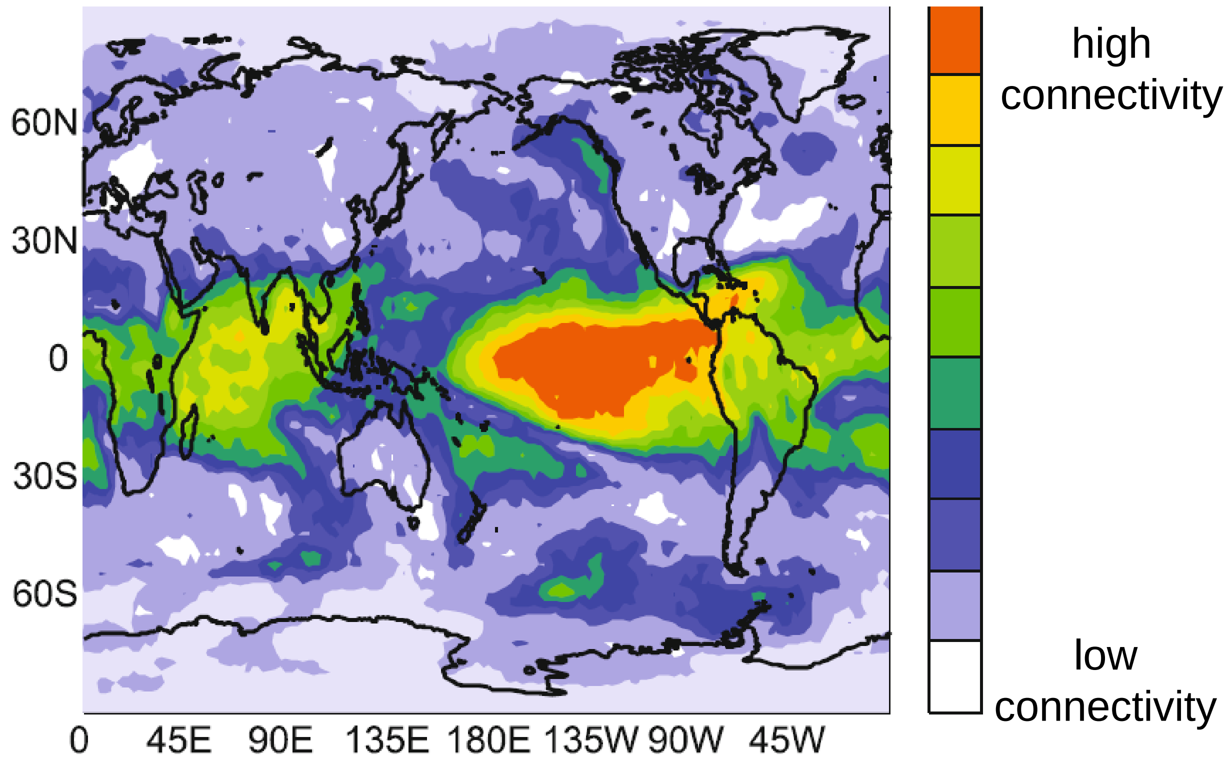


Graphical representation of the climate network

$$\mathrm{AWC}_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$$

Network obtained with ordinal analysis using <u>inter-annual</u> time-scale (3 consecutive years). The color-code indicates the Area Weighted Connectivity (weighted degree)





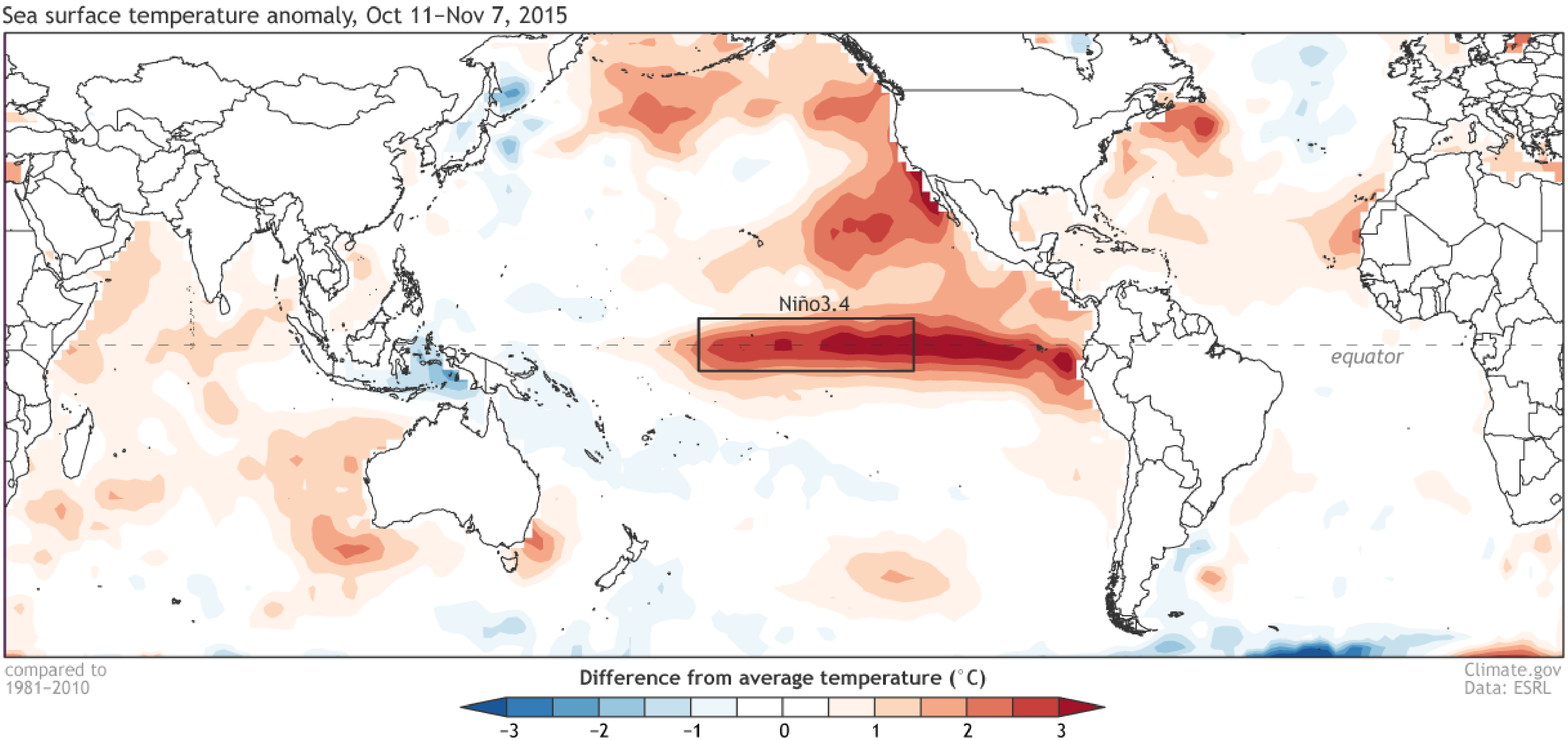
J. I. Deza, M. Barreiro and C. Masoller, Eur. Phys. J. Special Topics 222, 511 (2013)

El Niño/La Niña-Southern Oscillation (ENSO)

Is the most important climate phenomena on the planet

- Occurs across the tropical Pacific Ocean with ≈ 3-6 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).
- These two variations are coupled:
  - **El Niño** (ocean warming) -- high air surface pressure,
  - **La Niña** (ocean cooling) -- low air surface pressure.

Oct.-Nov. 2015: how ocean surface temperature differed from average



El Niño has huge impact world-wide

A few examples:

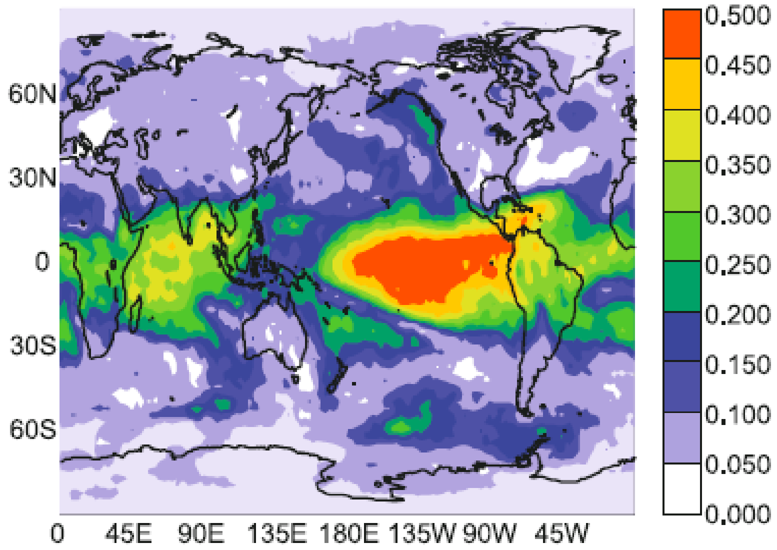
- Extra rainfall in South America: malaria outbreaks.
- Devastating forest fires in Indonesia.
- Dry conditions in South Africa: stress in water availability.
- Enhanced hurricane season in the Pacific.
- etc. etc. etc.

A lot of work to forecast El Niño evolution and to design mitigation/adaptation strategies.

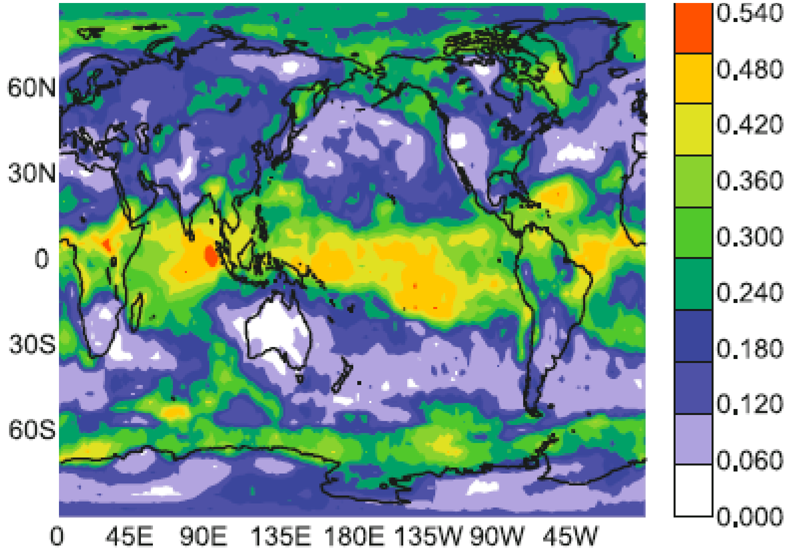
Contrasting two methods for inferring the climate network

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

Network when the probabilities are computed with ordinal analysis

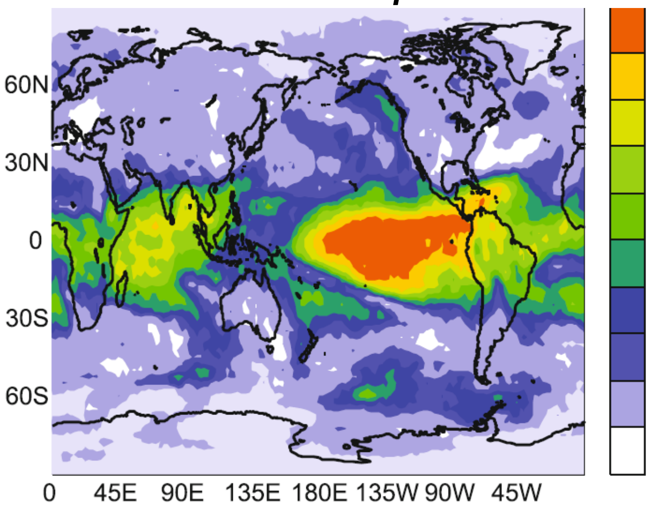


Network when the probabilities are computed with histogram of values

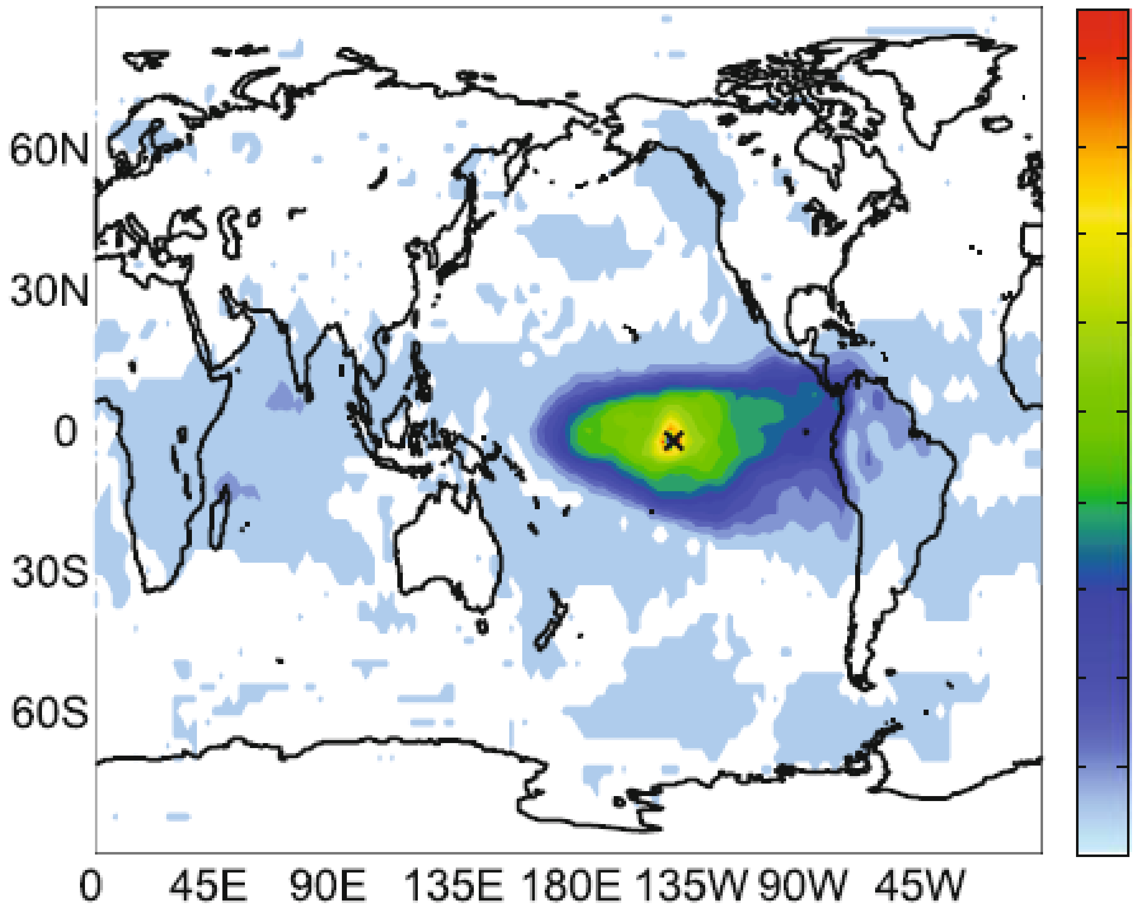


Who is connected to who?

AWC map



color-code indicates the MI values (only significant values)

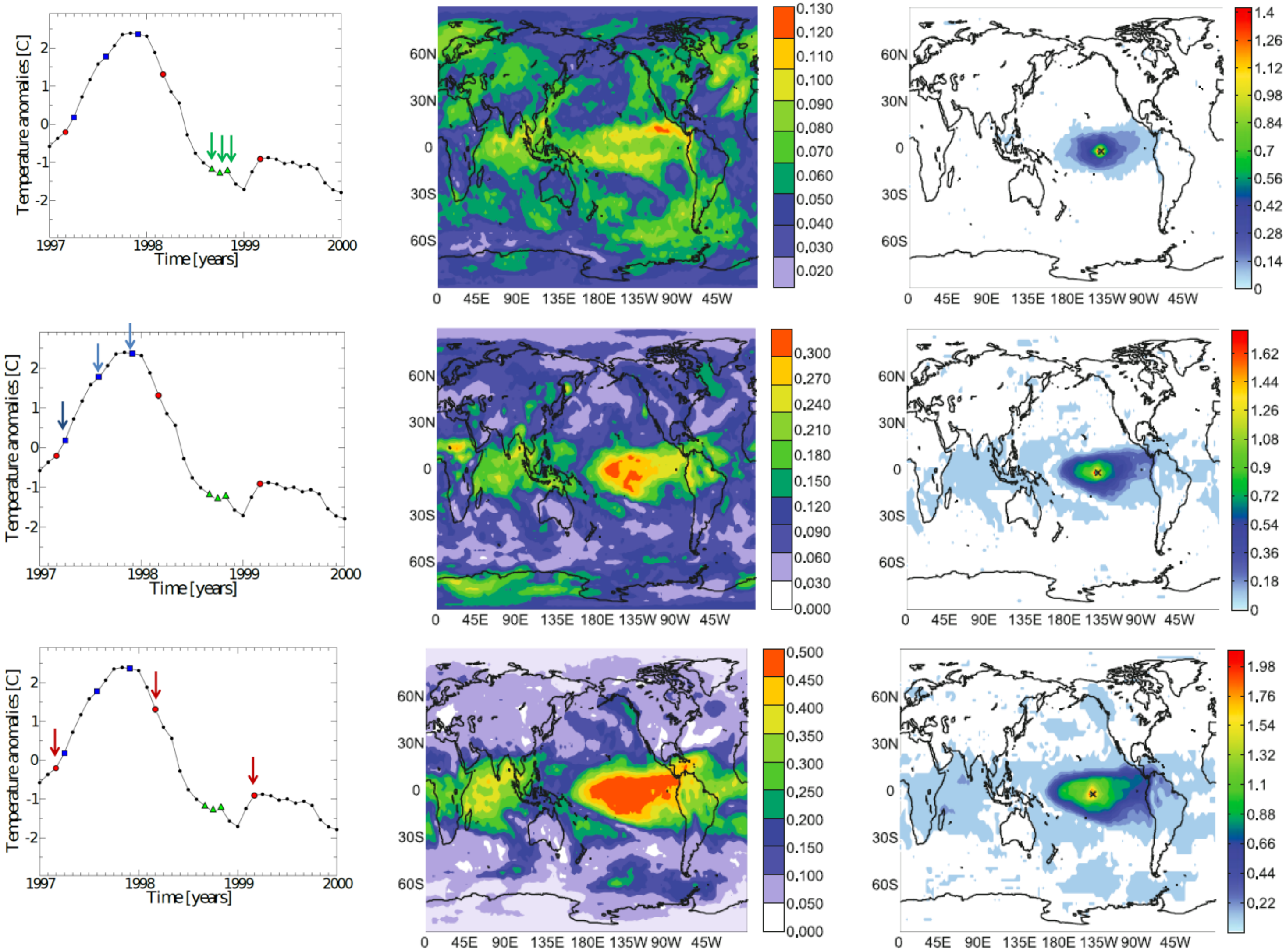


J. I. Deza, M. Barreiro, and C. Masoller, Eur. Phys. J. Special Topics 222, 511 (2013)

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

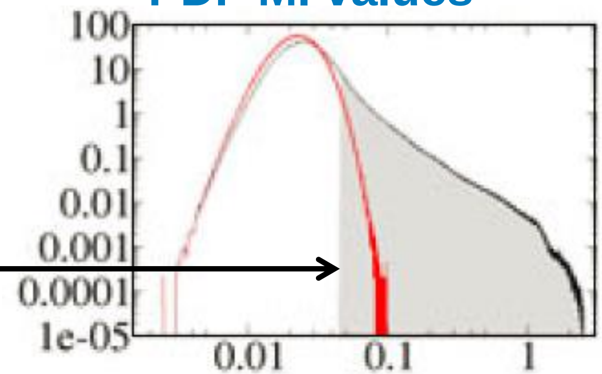
How do we assess the significance of the links?

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

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

PDF MI values

$\tau = \mu + 3\sigma$ →



Original data
Surrogated

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

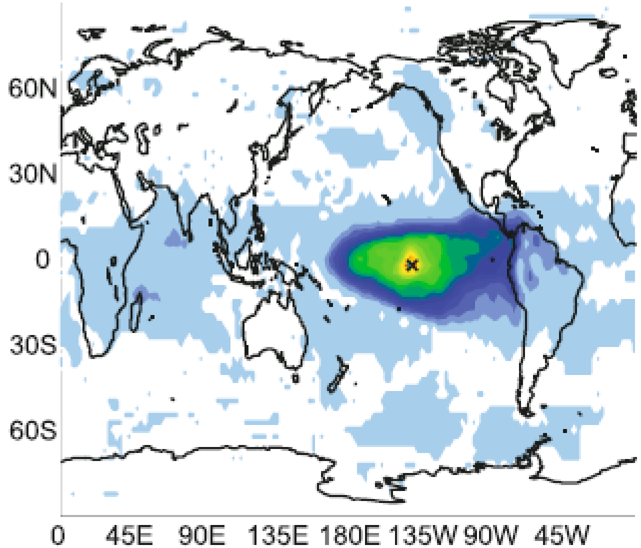
J. I. Deza, M. Barreiro, and C. Masoller, Eur. Phys. J. Special Topics 222, 511 (2013)

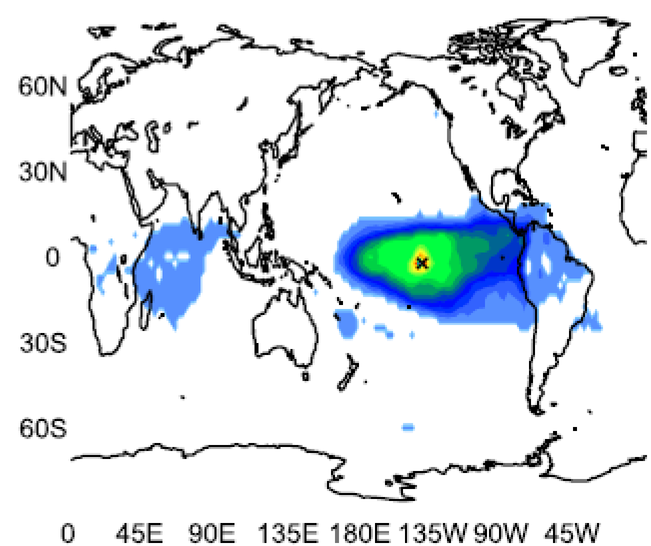
Are the links significant? Influence of the threshold

Low threshold *(11% link density)*

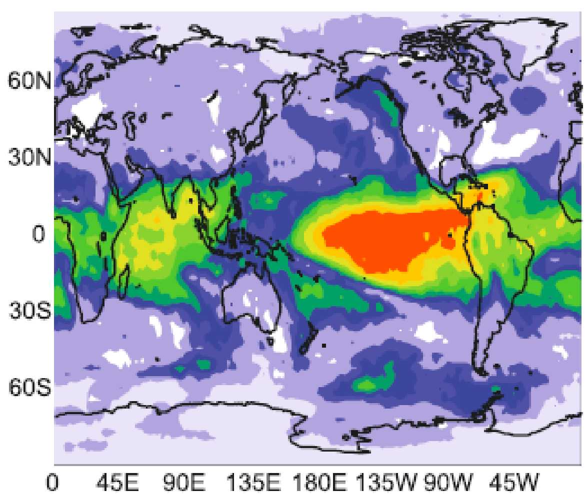
High threshold *(3% link density)*

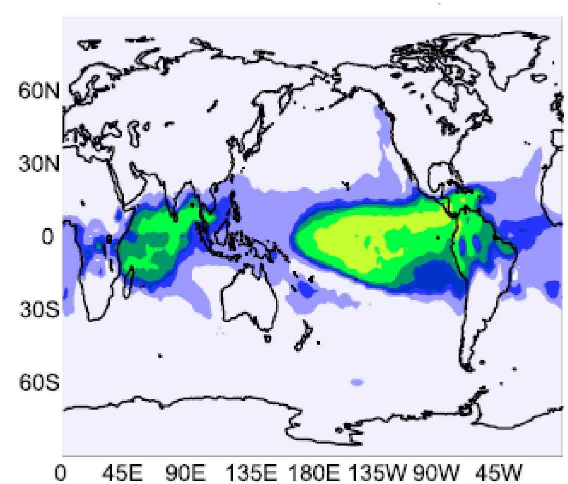
Color code: MI





Color code: AWC





How to improve climate predictability?

Assessing the directionality of the links

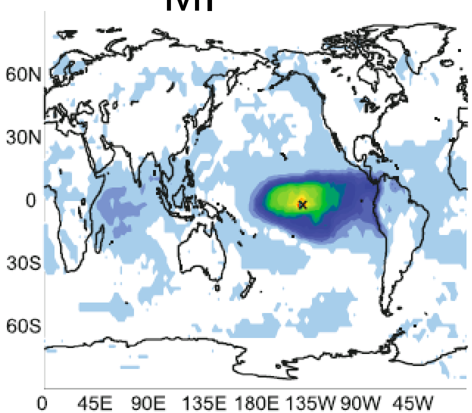
- $I_{xy}(\tau)$: <u>*conditional*</u> mutual information
- τ: *time-scale* of information transfer
- *D*: *<u>net</u> direction* of information transfer

$$D_{XY}(\tau) = \frac{I_{XY}(\tau) - I_{YX}(\tau)}{I_{XY}(\tau) + I_{YX}(\tau)}$$

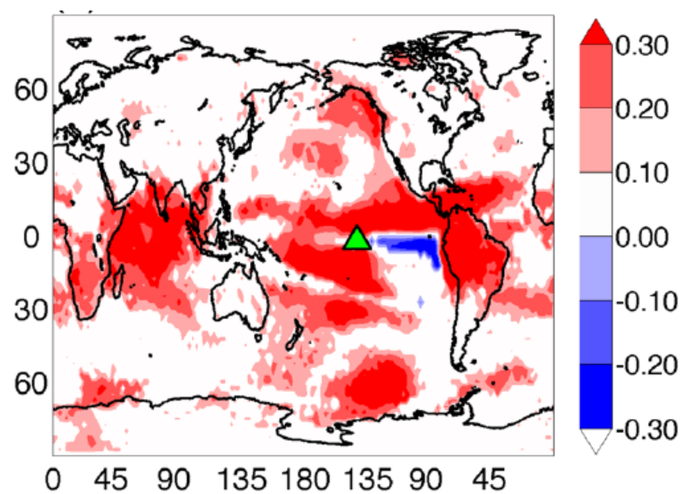
x → y
x → z

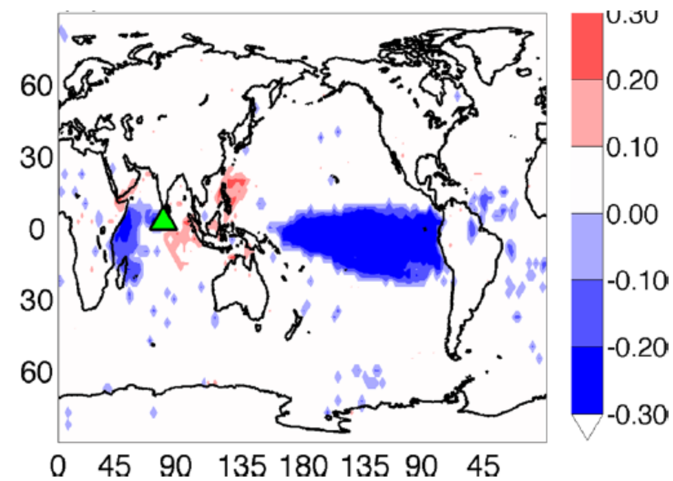
y ↔ z ??

MI



DI





Computed from daily SAT anomalies, PDFs estimated from histograms of values.
MI and DI are <u>*both significant*</u> (>3σ, bootstrap surrogates), τ=30 days.

Deza, Barreiro and Masoller, Chaos 25, 033105 (2015)

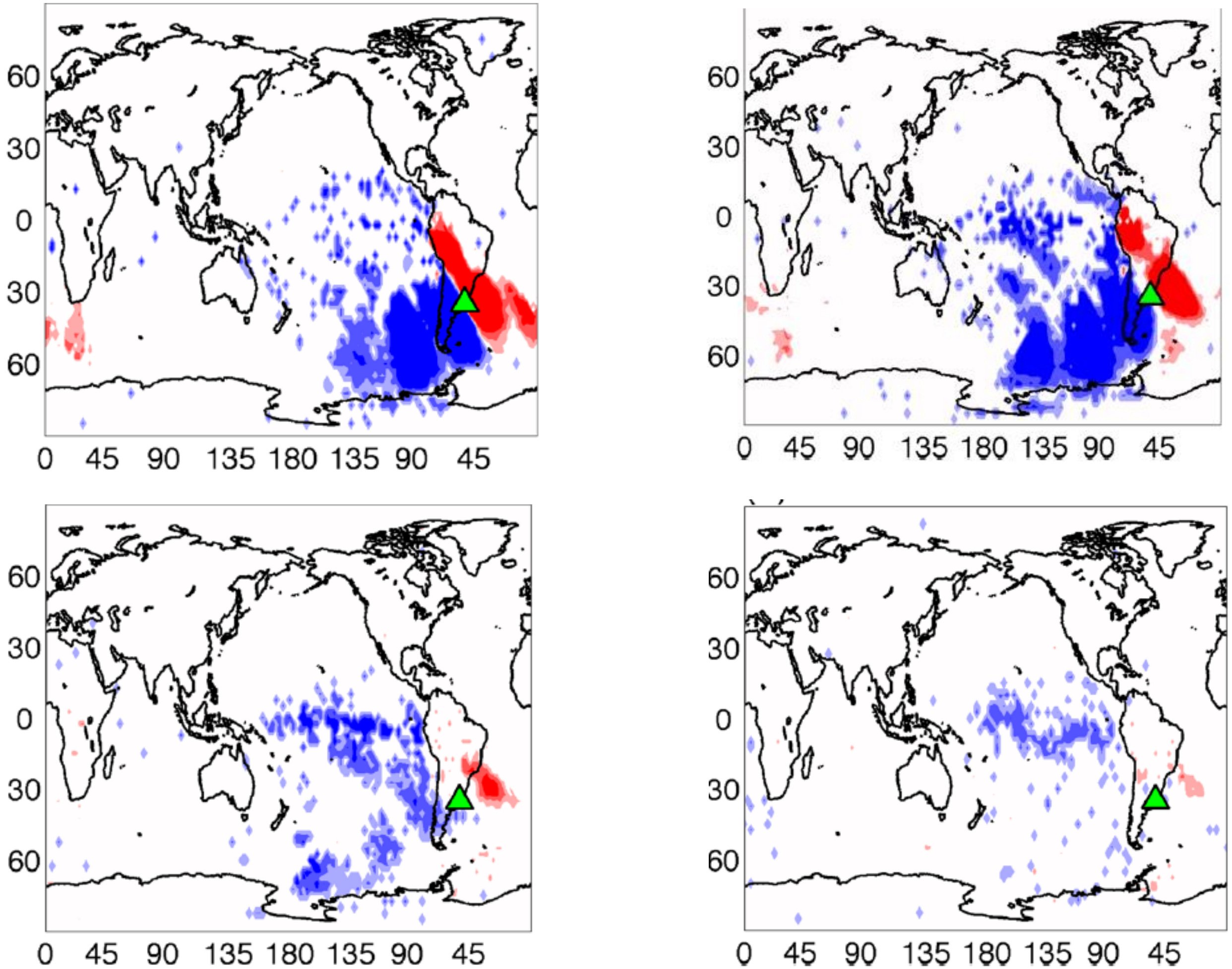
Time-scale of interactions

τ=1 day

τ=3 days

τ=7 days

τ=30 days



Link directionality reveals wave trains propagating from west to east

Deza, Barreiro and Masoller, Chaos 25, 033105 (2015)

Can we test the method used to built the climate network?

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Kuramoto oscillators in a random network

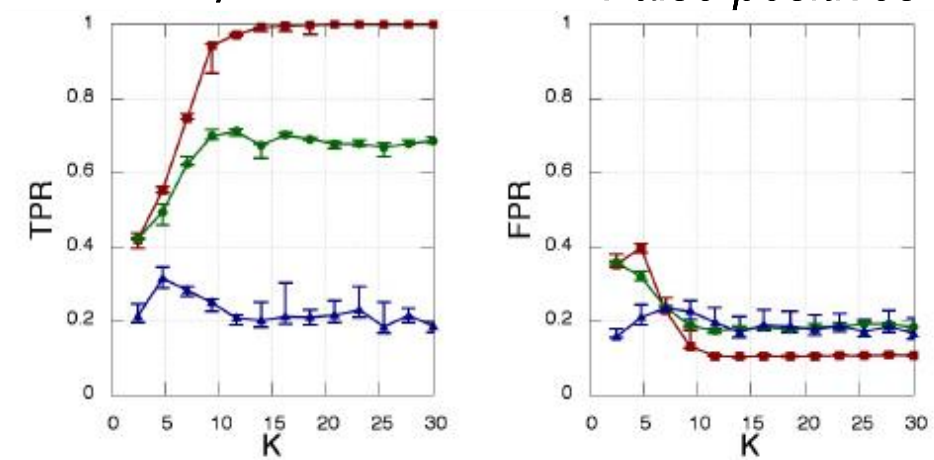
$$d\theta_i = \omega_i dt + \frac{K}{N}\sum_{j=1}^{N} A_{ij}\sin(\theta_j - \theta_i)dt + D\, dW_t^i$$

A$_{ij}$ is a known symmetric random matrix;
N=12 time-series, each with 10^4 data points.

CC MI MIOP

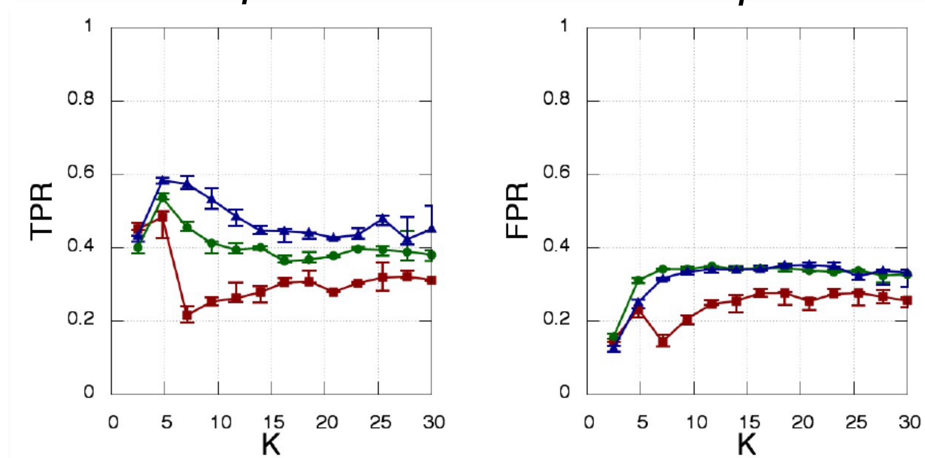
Phases (θ)

True positives *False positives*



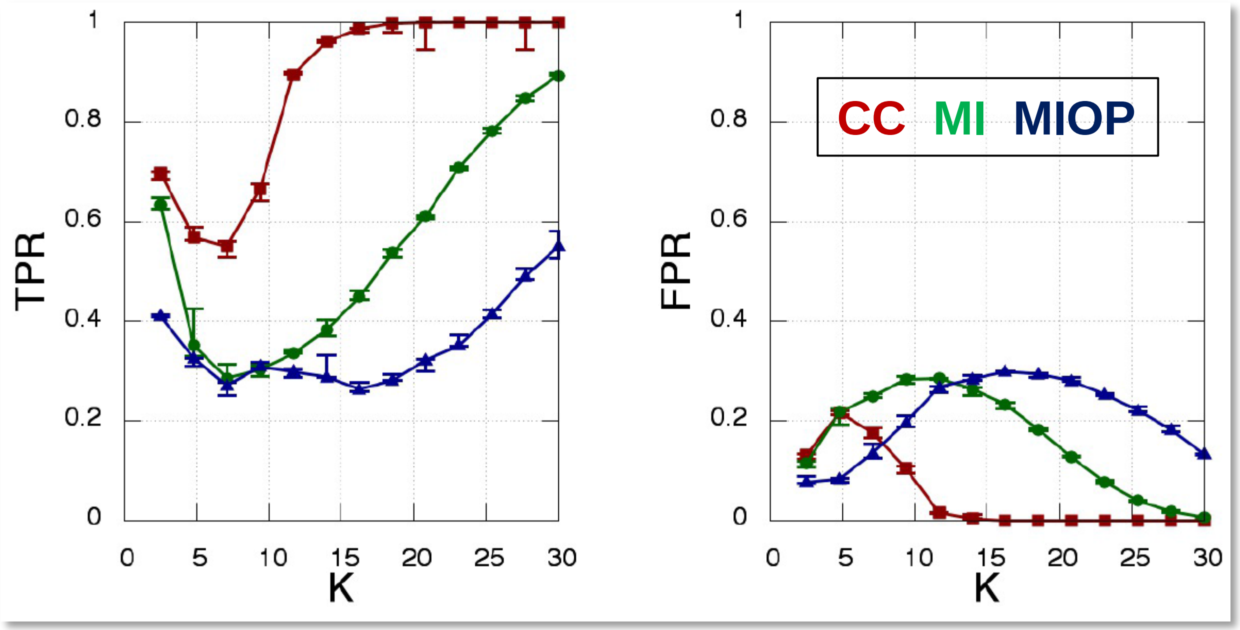
"Observable" Y=sin(θ)

True positives *False positives*



Results of a 100 simulations with different oscillators' frequencies, random matrices, noise realizations and initial conditions.
For each *K*, the threshold was varied to obtain optimal reconstruction.

Instantaneous frequencies (dθ/dt)



Perfect network inference is possible!

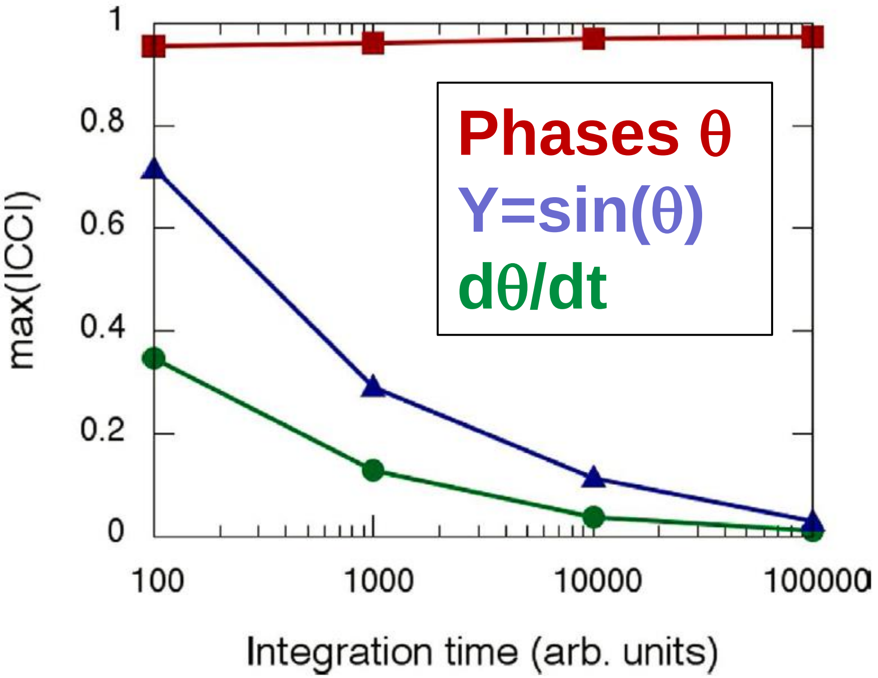
BUT

- 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)

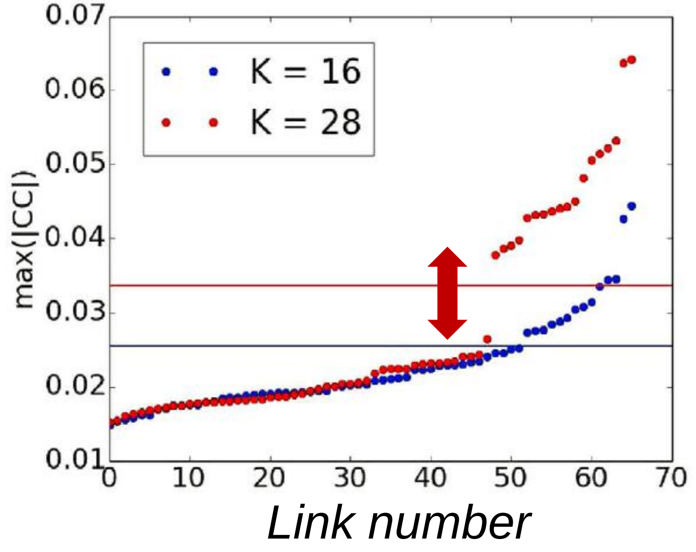
G. Tirabassi et al, Sci. Rep. ***5*** *10829 (2015)*

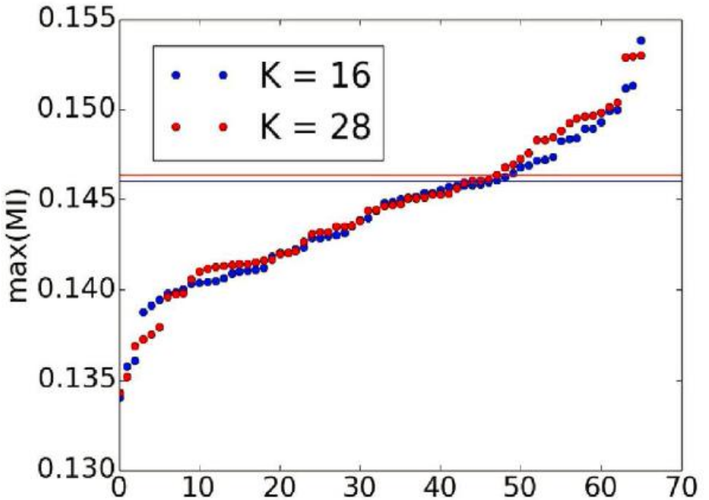
Why instantaneous frequencies are better than phases and "observables"?

Correlation analysis of two UNCOUPLED oscillators (K=0)



Why does CC outperforms MI?

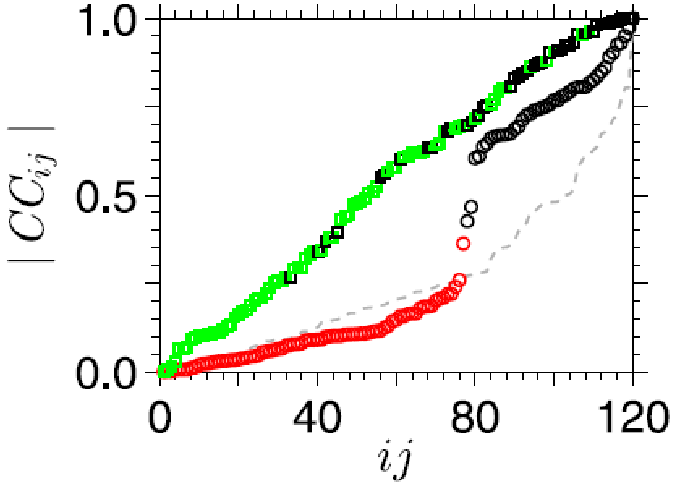




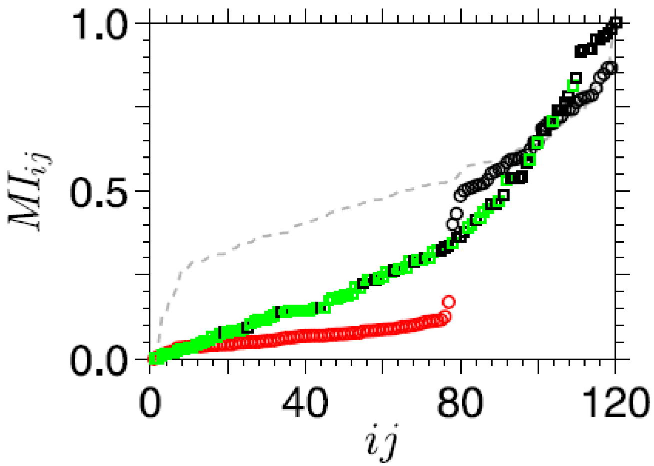
Gap in agreement with previous analysis with chaotic maps (e.g. logistic map)

$$x_{n+1}^{(i)} = (1-\epsilon) f\left(r_i,\ x_n^{(i)}\right) + \epsilon \sum_{j=1}^{N} \frac{W_{ij}}{\mathrm{d}_i} f\left(r_j,\ x_n^{(j)}\right)$$

Correlation analysis



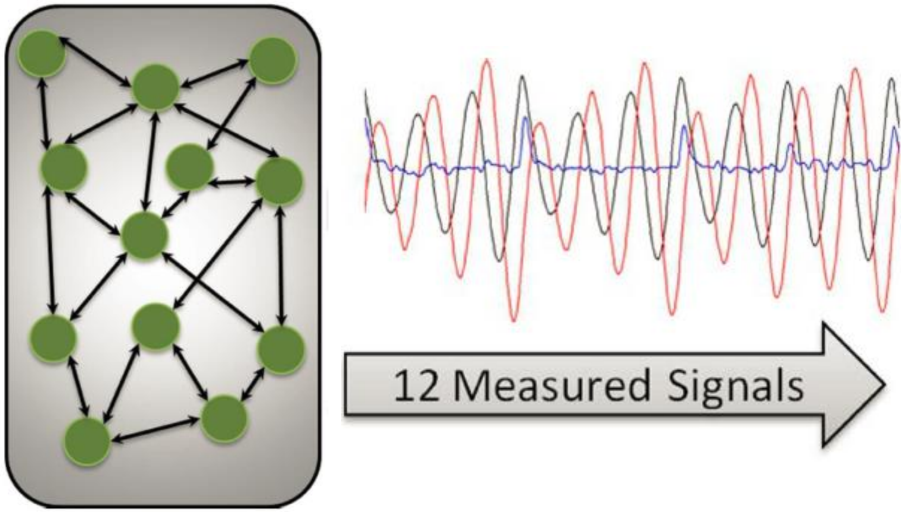
Mutual information analysis



- Dashed: ε=0
- Squares (green/black): ε=0.5
- Circles (red/ black): ε=0.06
- Perfect reconstruction possible for ε=0.06, but **wider gap with MI**

N. Rubido et al, New J. Phys. 16 (2014) 093010

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



The Hilbert Transform was used to obtain phases from experimental data

Experiments by J. Buldu & R. Sevilla-Escoboza.

- Kuramoto Oscillators' Network

$$\theta_i$$

$$f_i = \dot{\theta}_i$$

$$Y_i = \sin(\theta_i)$$

- Rössler Oscillators' Network

$$\varphi_i = HT(x_i)$$

$$f_i = \dot{\varphi}_i$$

$$x_i$$

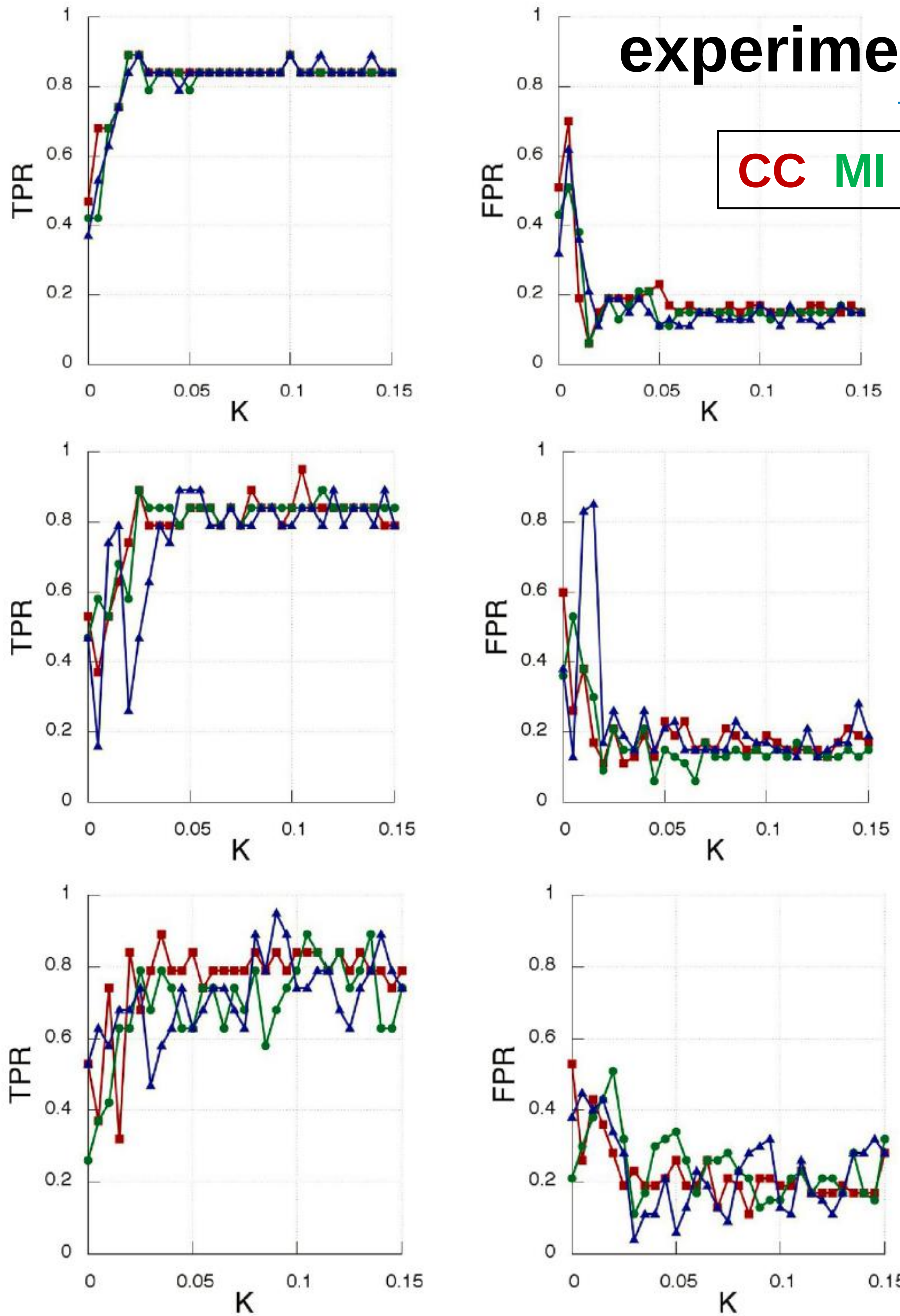
Results obtained with experimental data

CC MI MIOP

Observed variable (x)

Hilbert phase

Hilbert frequency

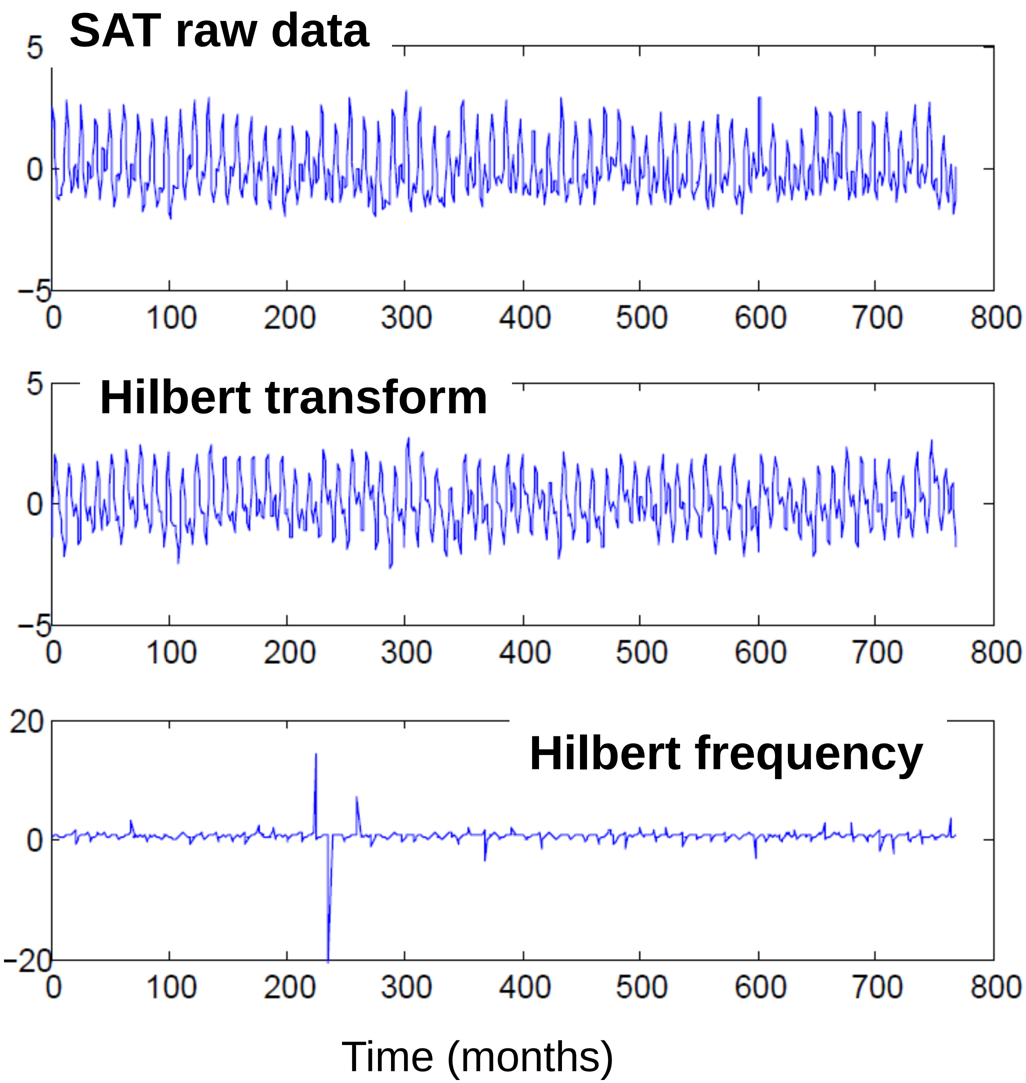


- No perfect reconstruction
- No important difference found among the 3 methods & 3 variables

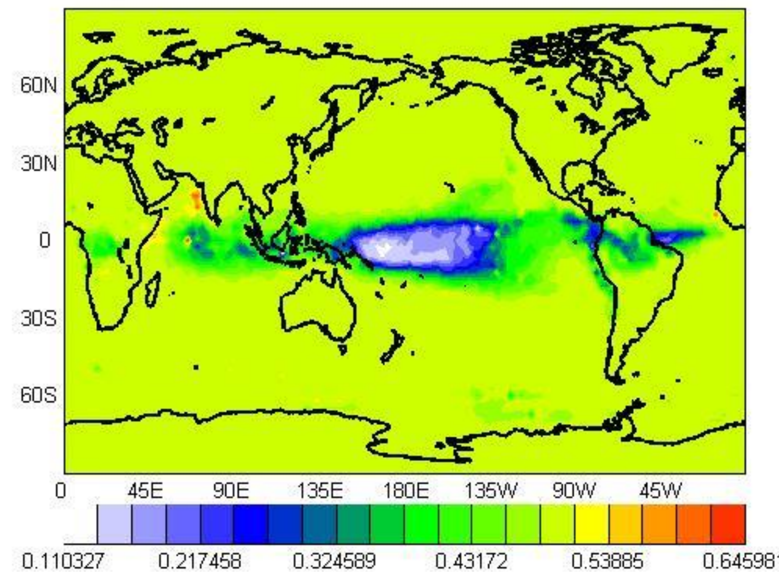
UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

Ongoing work: application of Hilbert transform to climate data

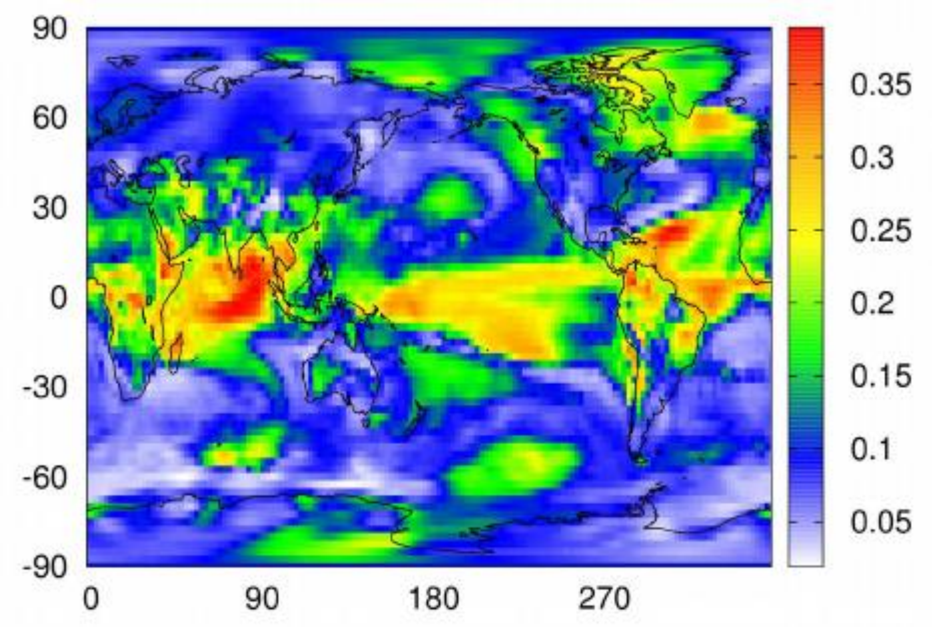


Time-averaged Hilbert frequency

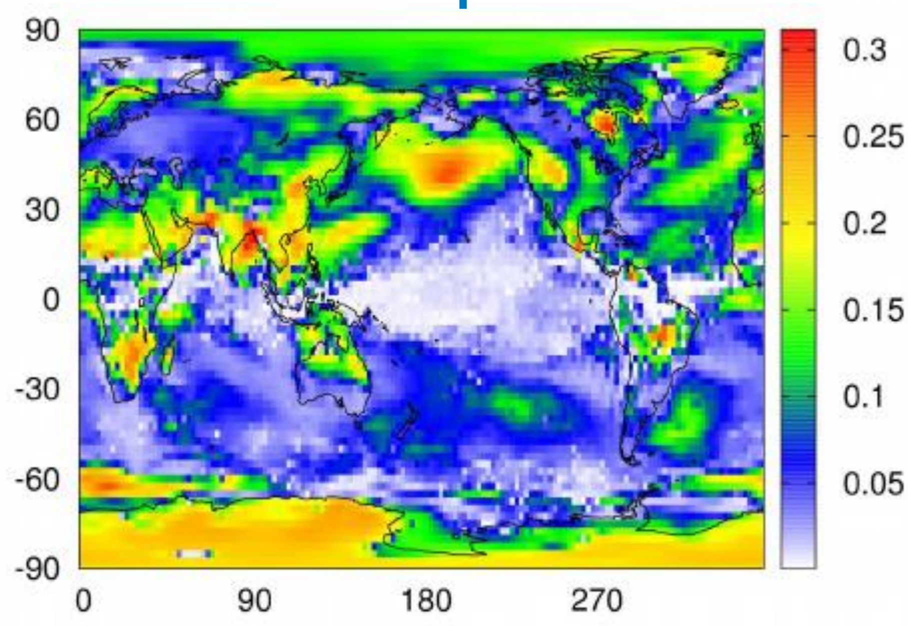


Contrasting (again) two methods for inferring the climate network

Network constructed from correlation analysis of SAT anomalies



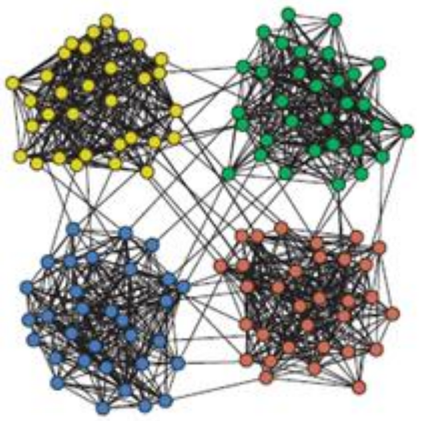
Network constructed from correlation analysis of Hilbert frequencies



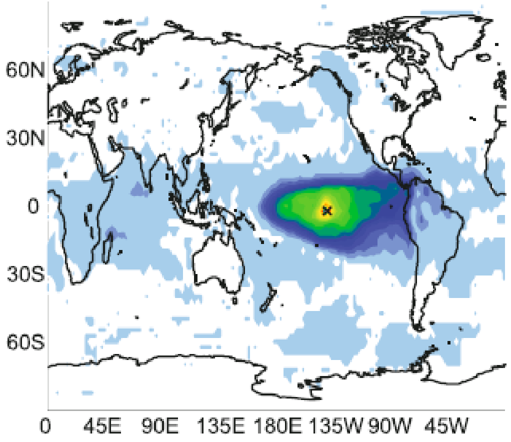
Climate "communities"

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.



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 transition probabilities

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

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

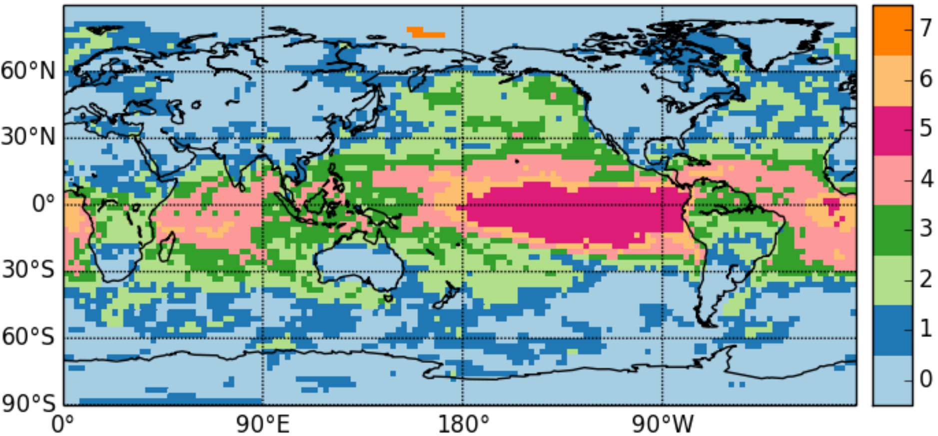
  High weight if similar symbolic "language"

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

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

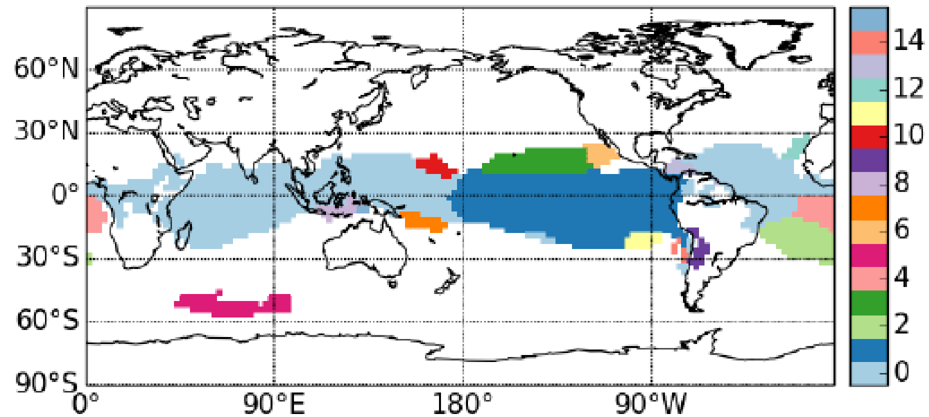
Results

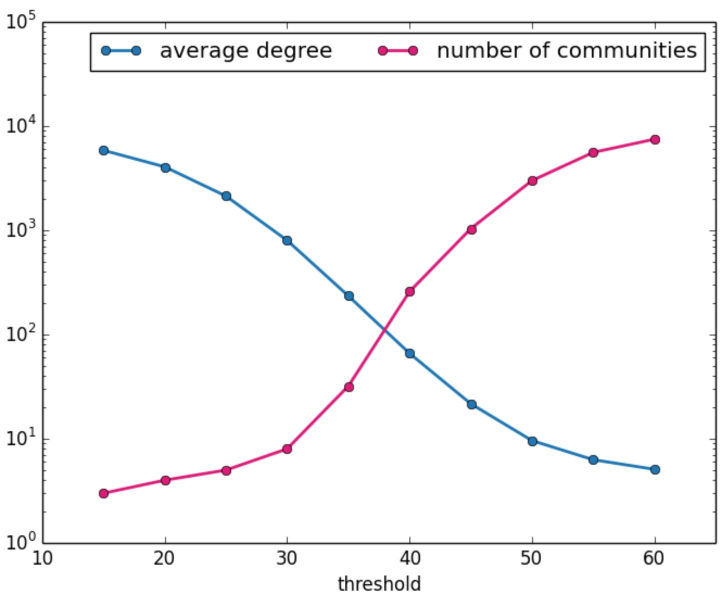
TP Network

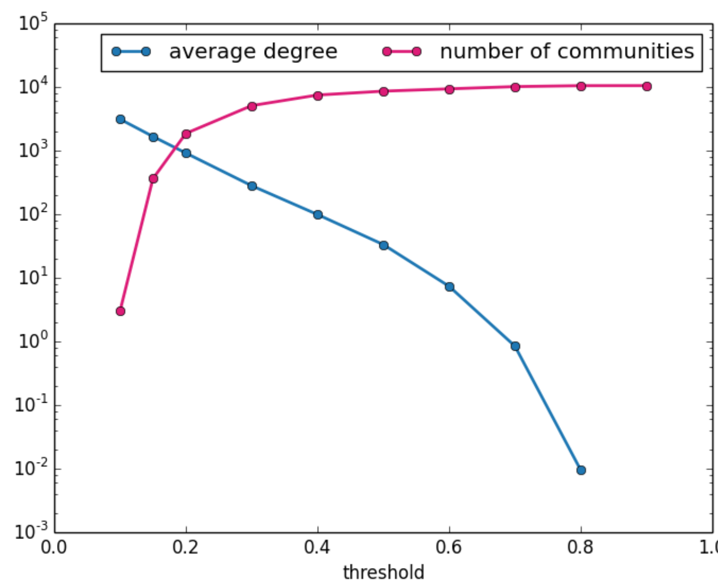


CC Network

(only the largest 16)







Outline

- Introduction
- Results
- **Summary**

What did we learn?

- Take home message:

  The network approach provides an opportunity for improving our understanding of climate phenomena.

  The Challenge: can we use networks to improve climate predictability?

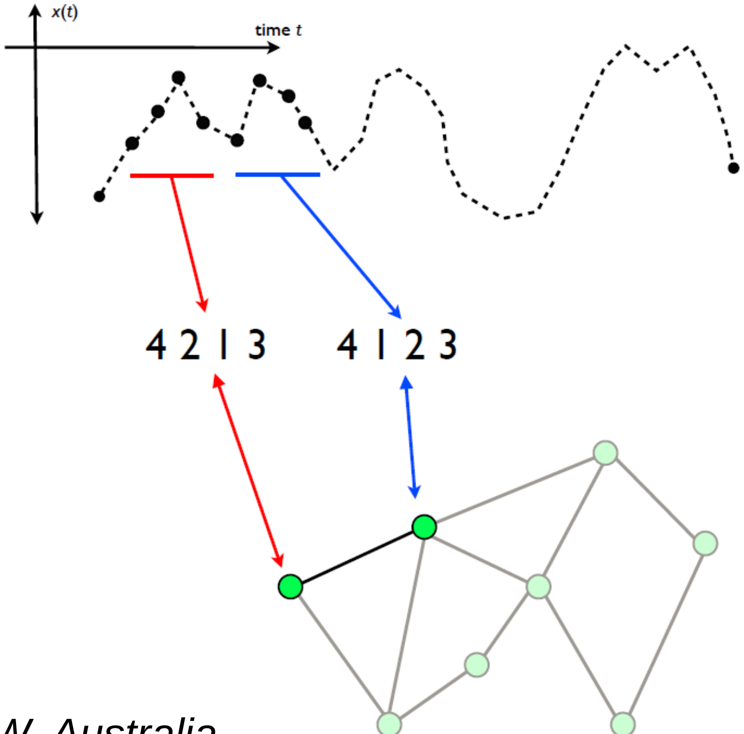
- A few specific conclusions:
  - Ordinal analysis allows identifying climate communities and time-scales of climate interactions.
  - Conditional mutual information allows identifying net direction of climate interactions.
  - In small synthetic networks, under appropriate conditions, perfect network inference is possible.
  - The similarity method to be used (CC or MI) and the variable to be analyzed, for optimal network reconstruction can depend on the system (not the same for Kuramotos, logistic maps, or electronic circuits).

Ongoing and future work

- In climate data, is there relevant information in Hilbert phases and frequencies?
- Are there favored / infrequent symbolic patterns in the climate dynamics?
- Potential for advancing sub-seasonal predictability?

- Time-series → network,

Potential for predicting *El Niño* "symbolic" dynamics?



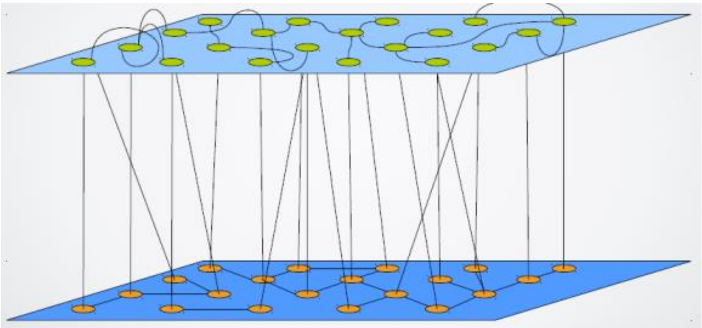
M. Small, Univ. W. Australia

Ongoing and future work

- Dissimilarity measure to quantify time-evolution of climate network: potential for uncovering climate regime transitions?

- Multilayer networks (Granger causality analysis of air-ocean interactions in the South America Convergence Zone – SACZ)

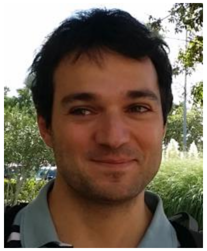
SST, pressure, wind, precipitation, etc.



SAT

Collaborators & funding





- Ignacio Deza
- Giulio Tirabassi
- Dario Zappala
- Marcelo Barreiro, Nicolas Rubido, and Arturo Martí (Universidad de la República, Uruguay)
- Experiments with chaotic electronic circuits: Javier Buldu (Technical University of Madrid), Ricardo Sevilla-Escoboza (Universidad de Guadalajara, Mexico)
- Coupled maps: Celso Grebogi and Murilo Baptista (University of Aberdeen)

ITN LINC
FP7-289447

ICREA

THANK YOU FOR YOUR ATTENTION !

<cristina.masoller@upc.edu>

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

- M. Barreiro et al, Chaos 21, 013101 (2011).
- J. I. Deza et al, Eur. Phys. J. Special Topics 222, 511 (2013).
- N. Rubido et al, New J. Phys. 16, 093010 (2014).
- G. Tirabassi et al, Sci. Rep. 5, 10829 (2015).