Complex networks of interacting stochastic dynamical systems: Discerning connectivity from dynamics

M. Paluš, D. Hartman, J. Hlinka & M. Vejmelka

Institute of Computer Science
Academy of Sciences of the Czech Republic
Prague, Czech Republic
E-Mail: mp@cs.cas.cz

Fluctuations & Coherence, Lancaster, July 15, 2011

INTRODUCTION

- Interacting dynamical systems
- Statistical physics
- Graph theory
- COMPLEX NETWORKS
- **Multivariate time series $\longrightarrow$ networks**
  - Nodes: measuring sites
  - Edges: dependence, **“connectivity” measures**
    - weighted graph
    - threshold $\rightarrow$ binary graph

INTRODUCTION

- **Multivariate time series $\longrightarrow$ networks**
  - Edges: dependence, “connectivity” measure
  - linear cross-correlation – the measure of first choice
- correlation – linearity – Gaussianity
- Nonlinearity? hidden connectivity patterns?
- Factors influencing connectivity measures
  - dynamics (serial correlations)
  - temporal and spatial sampling (time lags)
- Factors influencing network structure
  - uniform thresholding or individual statistical testing
  - thresholding Z-score, significance function

INTRODUCTION

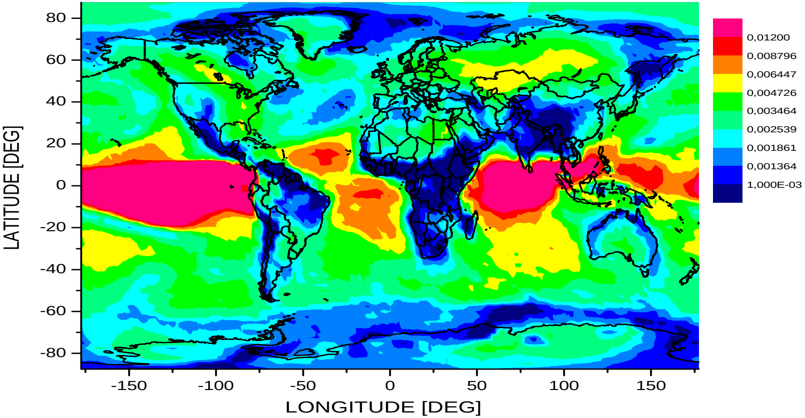
- **Multivariate time series $\longrightarrow$ networks**
  - Edges: dependence, “connectivity measure”
  - linear cross-correlation – the measure of first choice
- correlation – linearity – Gaussianity
- Nonlinearity hidden connectivity patterns?
- Factors influencing connectivity measures
  - **dynamics (serial correlations)**
  - temporal and spatial sampling (time lags)
- Factors influencing network structure
  - uniform thresholding or individual statistical testing
  - **thresholding Z-score**, significance function

CLIMATE NETWORKS

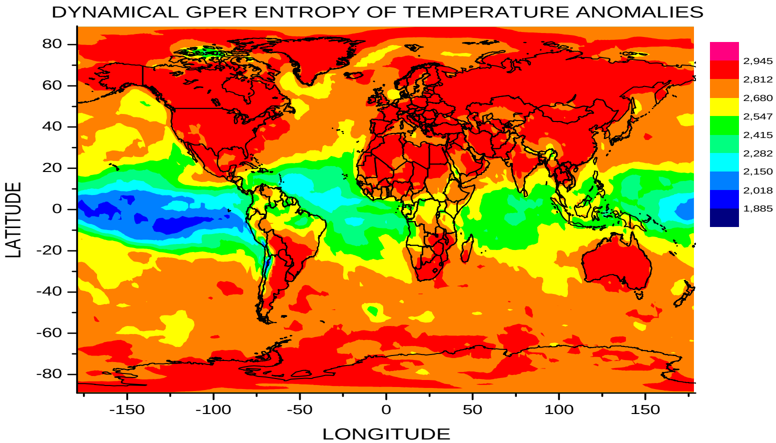
- **Multivariate time series**: gridded "reanalysis data" of atmospheric variables: air temperature, pressure, humidity, precipitation...
- Here: near-surface air temperature **anomalies**
  subtraction of seasonal means (mean Jan, mean Feb ...)
  removal of the annual cycle
  = **fluctuations** around seasonal means
- grid 2.5° x 2.5° $\longrightarrow 10^4$ nodes
- Pearson correlation $\longrightarrow$ weighted network
- thresholding $\longrightarrow$ binary network
- $\longrightarrow$ graph-theoretical analysis

Area Weighted Connectivity $\varrho = 0.005$ for

NCEP/NCAR SAT anomalies – absolute correlations



Connectivity vs. dynamics



Dynamical entropy (inverse to regularity) of temperature anomaly time series for each node.

Connectivity vs. dynamics: significance of dependence

SURROGATE DATA / BOOTSTRAP

- generated by a model
- obtained by manipulation (randomization) of the original data (surrogate data)

- IID (scrambled) surrogate data
- FT (AAFT, IAAFT ...) surrogate data
- wavelet
- recurrence
- constrained randomization ...

FT surrogates: preserve magnitudes of Fourier coefficients (spectra), randomize Fourier phases

Wavelet randomization preserving multifractality

PRL **101,** 134101 (2008) PHYSICAL REVIEW LETTERS week ending 26 SEPTEMBER 2008

Bootstrapping Multifractals: Surrogate Data from Random Cascades on Wavelet Dyadic Trees

Milan Paluš*

Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic

(Received 30 March 2007; revised manuscript received 21 June 2008; published 25 September 2008)

A method for random resampling of time series from multiscale processes is proposed. Bootstrapped series—realizations of surrogate data obtained from random cascades on wavelet dyadic trees—preserve the multifractal properties of input data, namely, interactions among scales and nonlinear dependence structures. The proposed approach opens the possibility for rigorous Monte Carlo testing of nonlinear dependence within, with, between, or among time series from multifractal processes.

DOI: 10.1103/PhysRevLett.101.134101 PACS numbers: 05.45.Tp, 05.45.Df, 89.75.Da

The estimation of any quantity from experimental data, with the aim to characterize an underlying process or its change, is incomplete without assessing the confidence of the obtained values or significance of their difference from natural variability. With the increasing performance and availability of powerful computers, Efron [1] proposed to replace (not always possible) analytical derivations based on (not always realistic) narrow assumptions by computational estimation of empirical distributions of quantities under interest using so-called Monte Carlo randomization procedures. In statistics, the term “bootstrap” [2] is coined for random resampling of experimental data, usually with the aim to estimate confidence intervals (“error bars”)

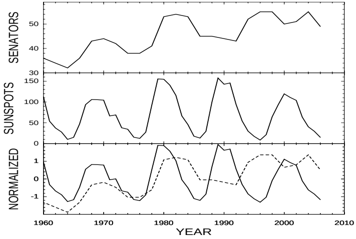
data in combinations with some constraints. Possible nonlinear dependence between a signal $s(t)$ and its history $s(t - \eta)$ is destroyed, as well as interactions among various scales in a potentially hierarchical, multiscale process. Multiscale processes that exhibit hierarchical information flow or energy transfer from large to small scales, successfully described by using the multifractal concepts (see [7] and references therein) have been observed in diverse fields from turbulence to finance [8], through cardiovascular physiology [9] or hydrology, meteorology, and climatology [10]. Angelini *et al.* [11] express the need for resampling techniques in evaluating data from atmospheric turbulence and other hierarchical processes. They apply a sophisti-

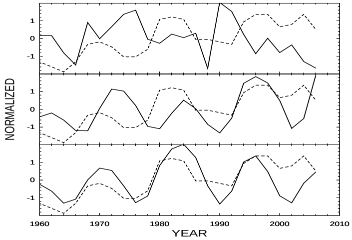
Connectivity (correlation) vs. dynamics

Correlation of sunspot numbers and the number of Republican senators

For the part of the record: 1960–1986, correlation c=0.52

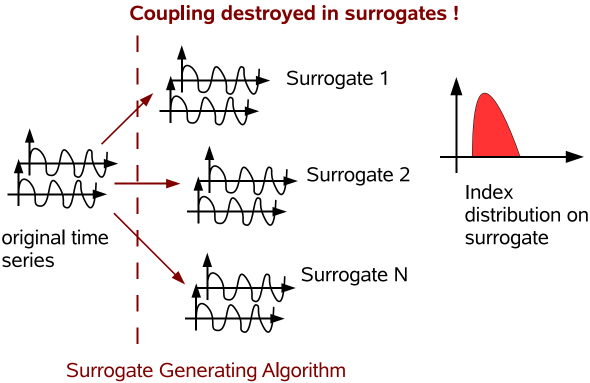
Table critical c=0.458 (for IID), but 0.73 for surrogate data test





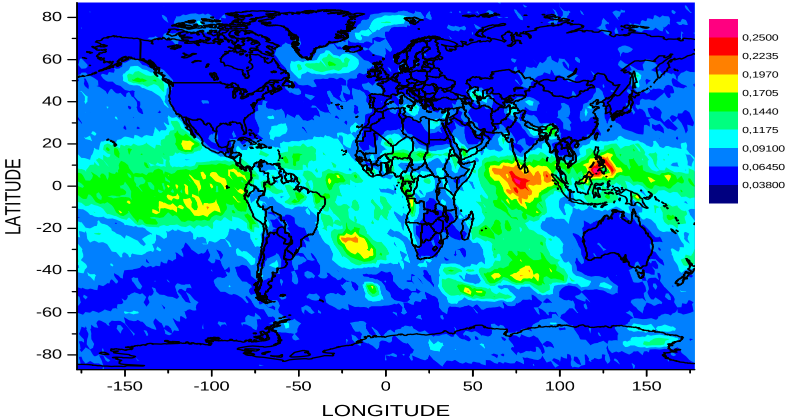
Significance testing using surrogate data

- Use of bootstrap-like strategy (surrogate time series)
- Ideally preserve all properties except tested (coupling)



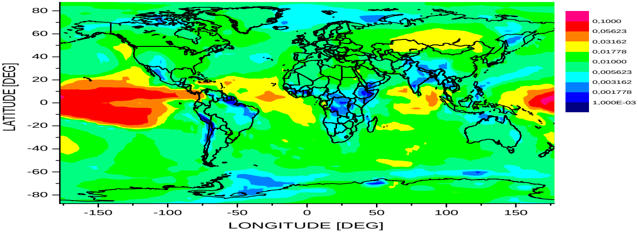
Mean absolute correlation of NCEP/NCAR SAT anomalies

with FT surrogate data

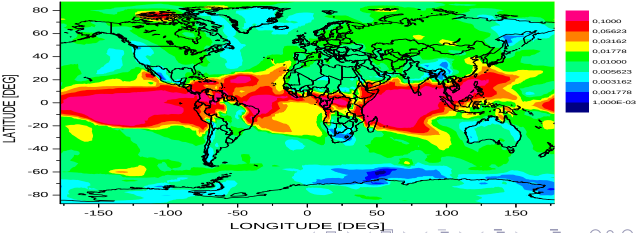


Area Weighted Connectivity absolute correlations > 0.5
(Tsonis & Swanson, PRL 100, 228502, 2008)

ENSO-

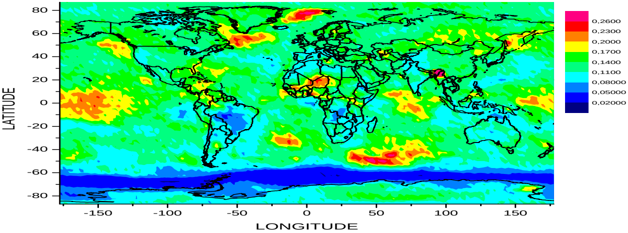


ENSO+

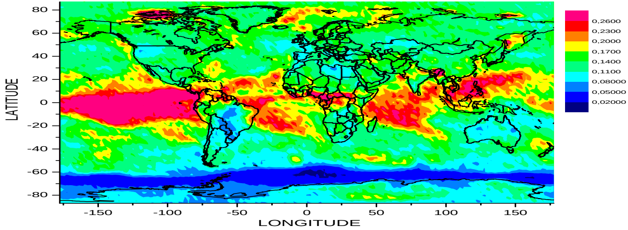


Mean absolute correlations SATA w/ FT surrogates

ENSO-



ENSO+



Connectivity vs. dynamics

Correct for dynamics (serial correlations):

For each link a statistical test with FT surrogate data

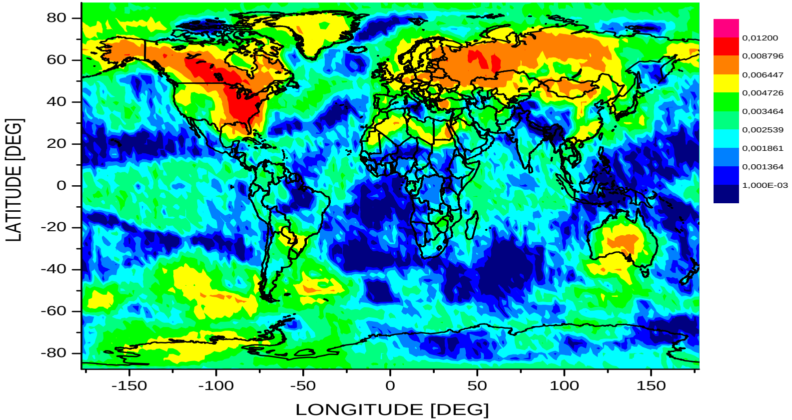
evaluated by using **Z-score**

$$Z_{i,j} = \frac{c_{i,j} - mean[c_{i,j}(surr)]}{SD[c_{i,j}(surr)]}$$

Z-score $Z_{i,j}$ used instead of $c_{i,j}$ for the link weights

Area Weighted Connectivity, NCEP/NCAR SATA, $\varrho = 0.005$

Z-score for absolute correlations + FT surrogate data

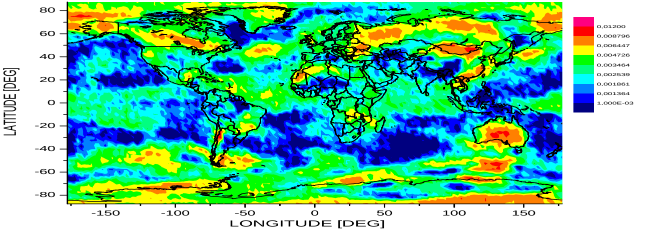


Connectivity vs. dynamics

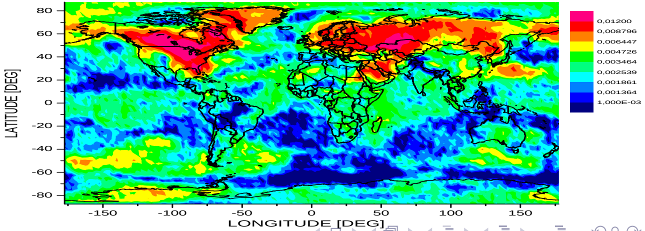
Z-score Area Weighted Connectivity, $\varrho = 0.005$

North Atlantic Oscillation influence

NAO-



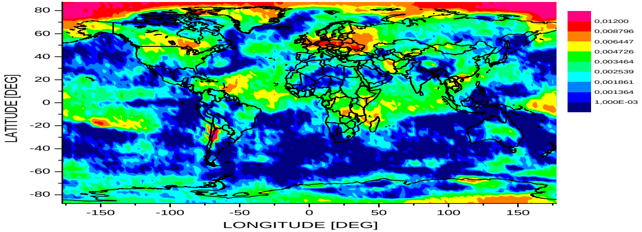
NAO+



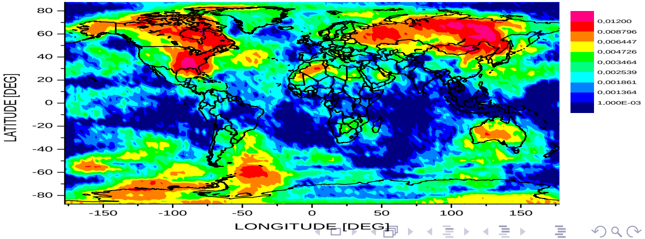
Z-score Area Weighted Connectivity, $\varrho = 0.005$

Solar influence: radio flux at 2800 MHz 10.7 cm

F10.7-



F10.7+



CONCLUSION: problems to be solved

- connectivity vs. dynamics
- connectivity vs. spatial/temporal scale
- stability of connectivity, network structure
- significance of changes in time and space
- (climate) network variability vs. external influence

Software package for complex network analysis:

`http://ndw.cs.cas.cz/software/ndw-graph`

Thank you for your attention

Preprints:
`http://ndw.cs.cas.cz`
`http://www.cs.cas.cz/mp`

Acknowledgement
This study was supported by the Czech Science Foundation project No. P103/11/J068.