From simple measures of dependence to complex network topologies

Milan Paluš

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

Dynamics Days Europe, Gothenburg, September 6, 2012

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

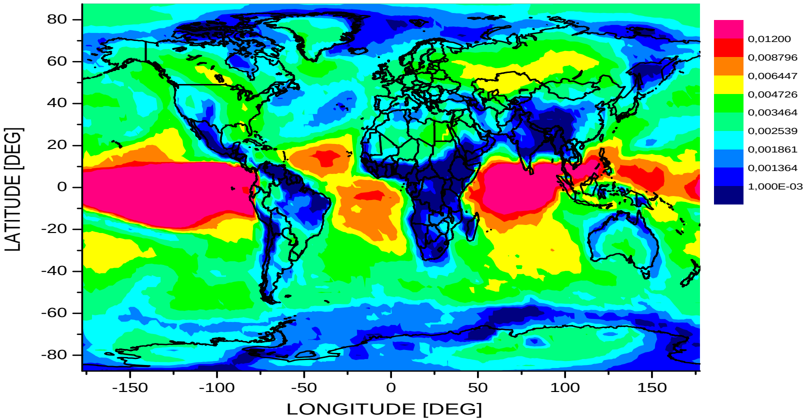
- **Multivariate time series $\longrightarrow$ networks**
  - Edges: dependence, “connectivity measure”
  - linear cross-correlation – the measure of first choice
- correlation – linearity – Gaussianity
- Nonlinearity hidden connectivity patterns?
- Factors influencing connectivity measures
  - **dynamics (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 $2.5^\circ \longrightarrow 10^4$ 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

- n discrete random variables $X_1, \ldots, X_n$
  values $(x_1, \ldots, x_n) \in \Xi_1 \times \cdots \times \Xi_n$
- PDF for an individual X_i is $p(x_i) = \Pr\{X_i = x_i\}$, $x_i \in \Xi_i$
- joint distribution for the n variables $X_1, \ldots, X_n$ is
  $p(x_1, \ldots, x_n) = \Pr\{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\}$
- the joint entropy of the n variables $X_1, \ldots, X_n$ with the joint distribution $p(x_1, \ldots, x_n)$:

$$H(X_1, \ldots, X_n) = -\sum_{x_1 \in \Xi_1} \cdots \sum_{x_n \in \Xi_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)$$

- stochastic process $\{X_i\}$:
  indexed sequence of random variables, characterized by $p(x_1, \ldots, x_n)$
- **entropy rate** of $\{X_i\}$ is defined as

$$h = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$$

- dynamical systems: *Kolmogorov-Sinai entropy*
- for a Gaussian process with spectral density function $f(\omega)$

$$h_G = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega$$

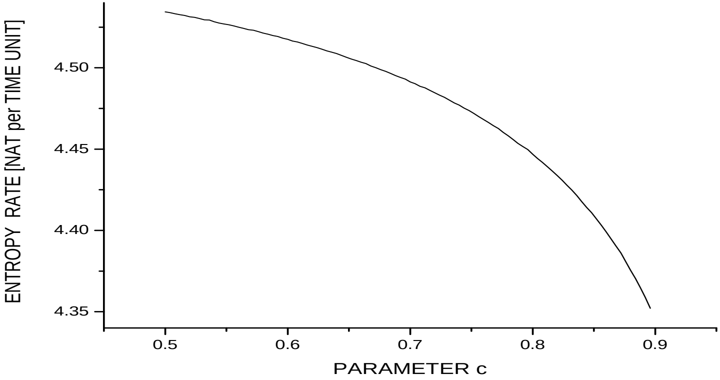
Connectivity vs. dynamics

- autoregressive process

$$y_t = c \sum_{k=1}^{10} a_k y_{t-k} + \sigma e_t, \qquad (1)$$

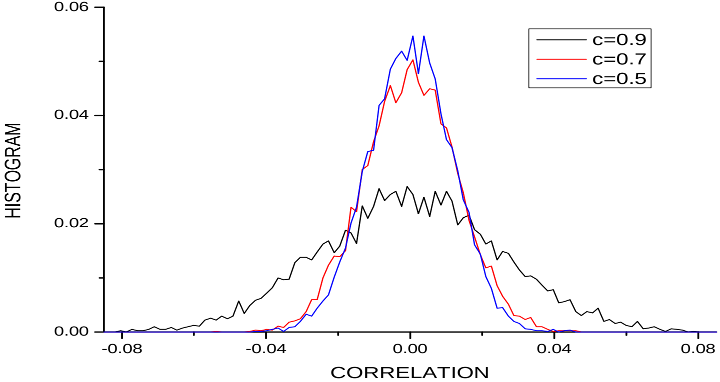
where $a_{k=1,..,10} = 0, 0, 0, 0, 0, .19, .2, .2, .2, .2$, $\sigma = 0.01$ and e_t are Gaussian deviates with zero mean and unit variance autoregressive process

$$y_t = c \sum_{k=1}^{10} a_k y_{t-k} + \sigma e_t$$



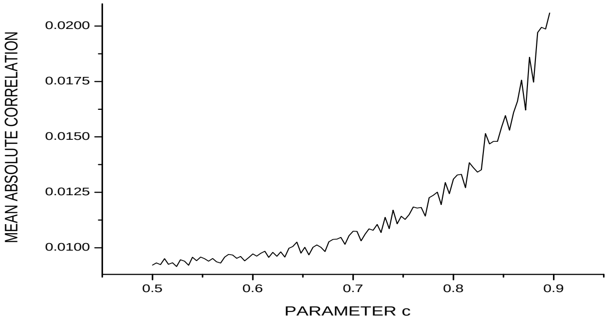
correlations of INDEPENDENT realizations of

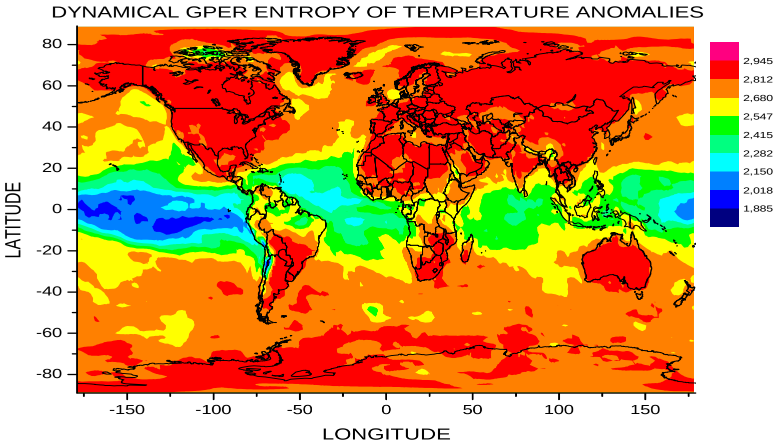
$$y_t = c \sum_{k=1}^{10} a_k y_{t-k} + \sigma e_t$$



mean ABSOLUTE correlations of INDEPENDENT realizations of

$$y_t = c \sum_{k=1}^{10} a_k y_{t-k} + \sigma e_t$$





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

Connectivity vs. dynamics: significance of dependence

SURROGATE DATA / BOOTSTRAP

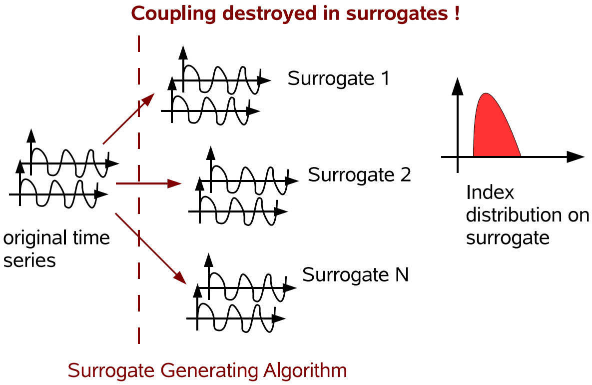
- generated by a model
- obtained by manipulation (randomization) of the original data (surrogate data)

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

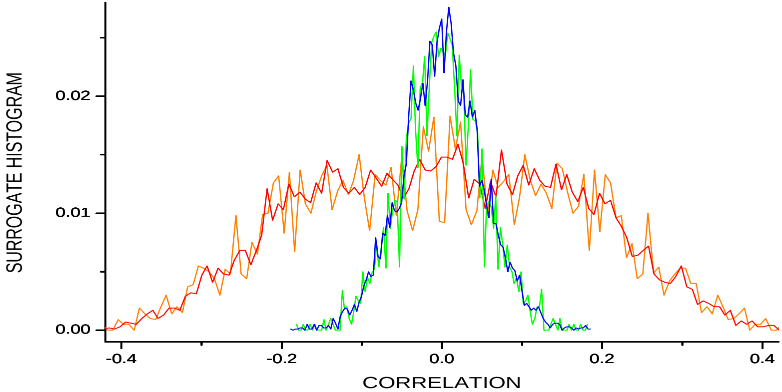
FT surrogates: preserve magnitudes of Fourier coefficients (spectra), randomize Fourier phases

Significance testing using surrogate data

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



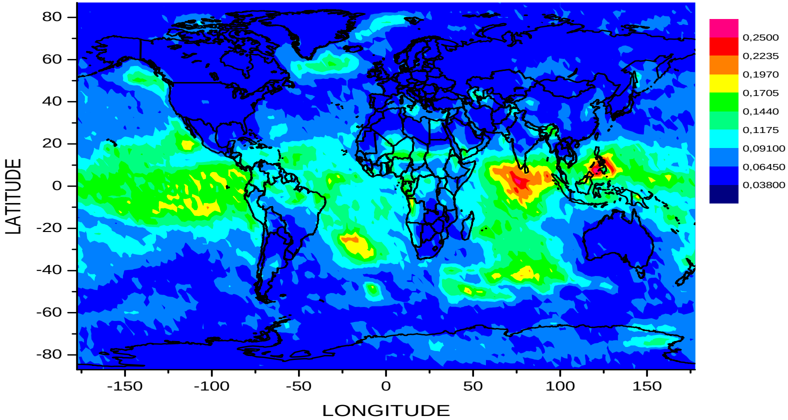
Connectivity vs. dynamics



Surrogate cross-correlation for high-ER (green, blue) and low-ER (orange, red) NCEP/NCAR grid-points. FT (green, orange), AAFT (blue, red).

Mean absolute correlation of NCEP/NCAR SAT anomalies

with FT surrogate data



Connectivity vs. dynamics

Correct for dynamics (serial correlations):

For each link a statistical test with FT surrogate data

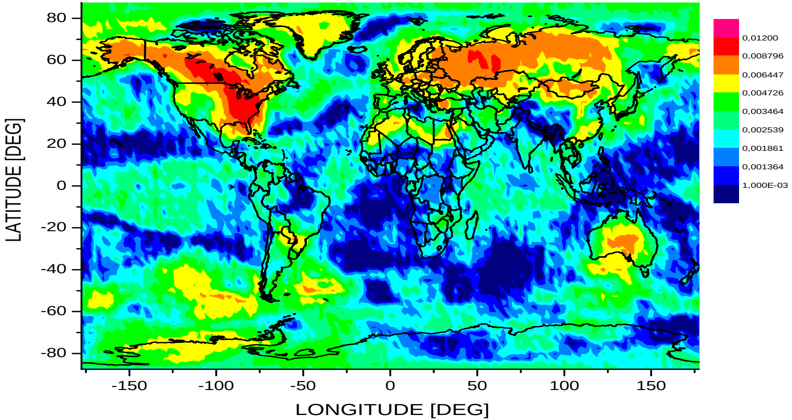
evaluated by using **Z-score**

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

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



Simple dependence measures

- two variables X and Y:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$$\widetilde{x}_i = \frac{x_i - \bar{x}}{\sigma}$$

- correlation between x and y is

$$c(x, y) = \frac{1}{N}\sum_{i=1}^{N} \widetilde{x}_i \widetilde{y}_i$$

Simple dependence measures

- two variables X and Y:
- p(x), H(X), p(y), H(Y), joint PDF p(x,y), joint entropy H(X,Y)
- mutual information

$$I(X;Y) = H(X) + X(Y) - H(X,Y)$$

- static p(x) – entropy H(X)
- characterization of dynamics – entropy rate

- static joint p(x,y) – mutual information I(X;Y) (correlation)
- similarity of dynamics – mutual information rate

Mutual information rate

- stochastic processes $\{X_i\}$, $\{Y_i\}$, characterized by $p(x_1, \ldots, x_n)$ and $p(y_1, \ldots, y_n)$
- **mutual information rate**

$$i(X_i; Y_i) = \lim_{n\to\infty} \frac{1}{n} I(X_1, \ldots, X_n; Y_1, \ldots, Y_n)$$

Mutual information rate

- for Gaussian stochastic processes $\{X_i\}$, $\{Y_i\}$, characterized by power spectral densities (PSD) $\Phi_X(\omega)$, $\Phi_Y(\omega)$ and cross PSD $\Phi_{X,Y}(\omega)$
- **mutual information rate**

$$i(X_i; Y_i) = -\frac{1}{4\pi} \int_0^{2\pi} \log(1 - |\gamma_{X,Y}(\omega)|^2) d\omega$$

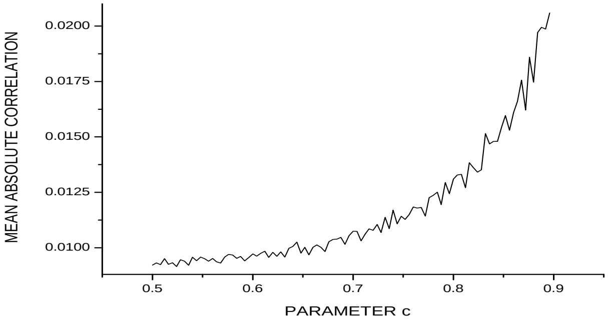
- magnitude-squared coherence

$$|\gamma_{X,Y}(\omega)|^2 = \frac{|\Phi_{X,Y}(\omega)|^2}{\Phi_X(\omega)\Phi_Y(\omega)}$$

AR process - remainder

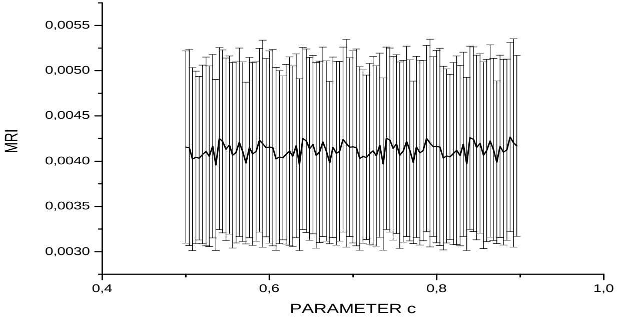
mean ABSOLUTE correlations of INDEPENDENT realizations of

$$y_t = c \sum_{k=1}^{10} a_k y_{t-k} + \sigma e_t$$



mean (Gaussian) MRI of 1000 INDEPENDENT realizations of

$$y_t = c \sum_{k=1}^{10} a_k y_{t-k} + \sigma e_t$$



Gaussian ER and nonlinear DS

ELSEVIER

24 March 1997

PHYSICS LETTERS A

Physics Letters A 227 (1997) 301–308

On entropy rates of dynamical systems and Gaussian processes

Milan Paluš

School of Mathematics, Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia
Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou Věží 2, 182 07 Prague 8, Czech Republic [1]

Received 9 September 1996; revised manuscript received 12 January 1997; accepted for publication 13 January 1997
Communicated by C.R. Doering

Abstract

The possibility of a relation between the Kolmogorov–Sinai entropy of a dynamical system and the entropy rate of a Gaussian process isospectral to time series generated by the dynamical system is numerically investigated using discrete and continuous chaotic dynamical systems. The results suggest that such a relation as a nonlinear one-to-one function may exist when the Kolmogorov–Sinai entropy varies smoothly with variations of the system parameters, but is broken in critical states near bifurcation points.

1. Entropy rates

Entropy rates will be considered as a tool for quantitative characterization of dynamic processes evolving in time. Let $\{x_i\}$ be a time series, i.e., a series of mea-

$$H(X_1, \ldots, X_n) = -\sum_{x_1} \ldots \sum_{x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n). \quad (2)$$

Gaussian ER and nonlinear DS

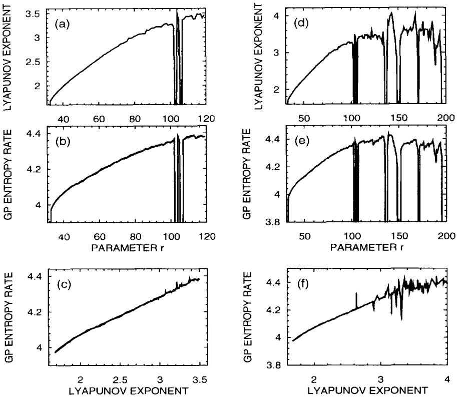


Fig. 3. Further results for the Lorenz system: (a) The positive Lyapunov exponents computed from the Lorenz equations for the parameter r varying from 33 to 120 in steps of 1. (b) The GP entropy rates estimated from 15 realizations of 16k time series (mean: thick line; mean $\pm$ SD: thin lines, coinciding with the mean) for different values of the parameter r varying as in (a). (c) Plot of GPER (the same line codes as before) versus LE. (d), (e), (f) The same as (a), (b), (c), respectively, except for the parameter r varying from 33 to 200 in steps of 1.

$r > 65$ enters the bifurcation region (Figs. 3a, 3b and plot was obtained by increasing the parameter a from

Route to synchronization

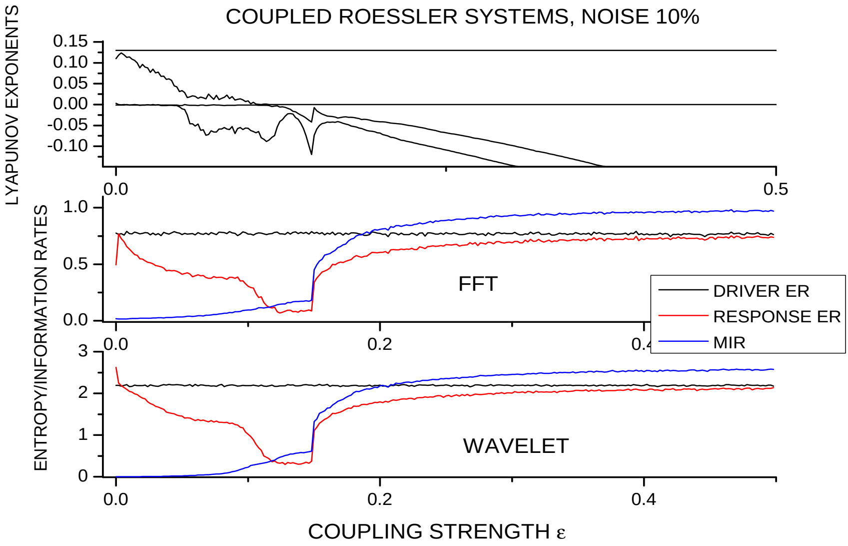
- unidirectionally coupled Rössler systems

$$\dot{x}_1 = -\omega_1 x_2 - x_3$$
$$\dot{x}_2 = \omega_1 x_1 + a_1\, x_2$$
$$\dot{x}_3 = b_1 + x_3(x_1 - c_1)$$

$$\dot{y}_1 = -\omega_2 y_2 - y_3 + \epsilon(x_1 - y_1)$$
$$\dot{y}_2 = \omega_2 y_1 + a_2\, y_2$$
$$\dot{y}_3 = b_2 + y_3(y_1 - c_2)$$

$a_1 = a_2 = 0.15$, $b_1 = b_2 = 0.2$, $c_1 = c_2 = 10.0$
frequencies $\omega_1 = 1.015$, $\omega_2 = 0.985$.

Route to synchronization and MIR, ER



Route to synchronization and MIR, ER

PHYSICAL REVIEW E, VOLUME 63, 046211

Synchronization as adjustment of information rates: Detection from bivariate time series

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

Vladimír Komárek, Zbyněk Hrnčíř, and Katalin Štěrbová
Clinic of Paediatric Neurology, 2nd Medical Faculty of Charles University, V úvalu 84, 150 06 Prague 5–Motol, Czech Republic

(Received 5 July 2000; revised manuscript received 4 December 2000; published 28 March 2001)

An information-theoretic approach for studying synchronization phenomena in experimental bivariate time series is presented. ''Coarse-grained'' information rates are introduced and their ability to indicate generalized synchronization as well as to establish a ''direction of information flow'' between coupled systems, i.e., to discern the driving from the driven (response) system, is demonstrated using numerically generated time series from unidirectionally coupled chaotic systems. The method introduced is then applied in a case study of electroencephalogram recordings of an epileptic patient. Synchronization events leading to seizures have been found on two levels of organization of brain tissues and ''directions of information flow'' among brain areas have been identified. This allows localization of the primary epileptogenic areas, also confirmed by magnetic resonance imaging and pasitron emission tomography scans.

DOI: 10.1103/PhysRevE.63.046211 PACS number(s): 05.45.Tp, 05.45.Xt, 89.70.+c

I. INTRODUCTION

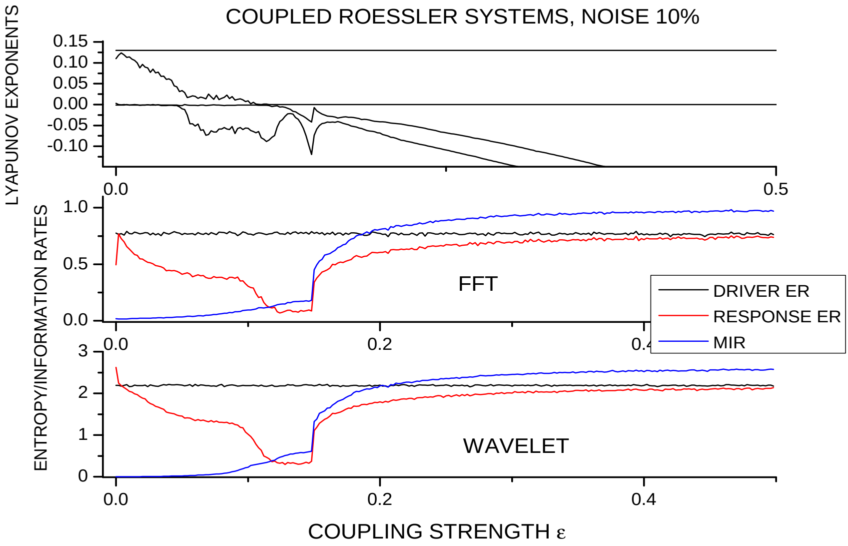
During the last decade there has been considerable interest in the study of the cooperative behavior of coupled chaotic systems [1]. Synchronization phenomena have been observed in many physical and biological systems, even in

electroencephalogram (EEG) recordings of an epileptic patient. A conclusion is given in Sec. V.

II. COARSE-GRAINED INFORMATION RATES

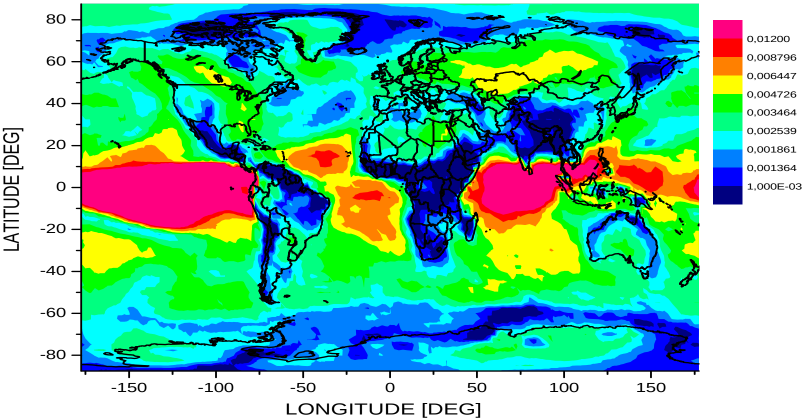
Consider discrete random variables X and Y with sets of

Route to synchronization and MIR, ER



Area Weighted Connectivity $\varrho = 0.005$ for

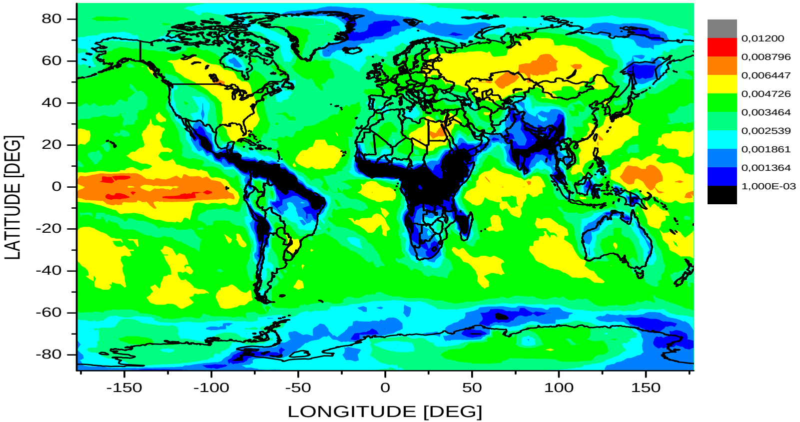
NCEP/NCAR SAT anomalies – absolute correlations



Connectivity vs. dynamics in climate network

Area Weighted Connectivity $\varrho = 0.005$ for

NCEP/NCAR SAT anomalies – mutual information rate



Spurious small-world networks

CHAOS **22**, 033107 (2012)

Small-world topology of functional connectivity in randomly connected dynamical systems

J. Hlinka, D. Hartman, and M. Paluš
Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 2, 18207 Prague, Czech Republic

(Received 15 March 2012; accepted 18 June 2012; published online 11 July 2012)

Characterization of real-world complex systems increasingly involves the study of their topological structure using graph theory. Among global network properties, small-world property, consisting in existence of relatively short paths together with high clustering of the network, is one of the most discussed and studied. When dealing with coupled dynamical systems, links among units of the system are commonly quantified by a measure of pairwise statistical dependence of observed time series (functional connectivity). We argue that the functional connectivity approach leads to upwardly biased estimates of small-world characteristics (with respect to commonly used random graph models) due to partial transitivity of the accepted functional connectivity measures such as the correlation coefficient. In particular, this may lead to observation of small-world characteristics in connectivity graphs estimated from generic randomly connected dynamical systems. The ubiquity and robustness of the phenomenon are documented by an extensive parameter study of its manifestation in a multivariate linear autoregressive process, with discussion of the potential relevance for nonlinear processes and measures.
[http://dx.doi.org/10.1063/1.4732541]

In the field of complex systems study, new measurement and computational resources have lead to increased inter-

dynamical system with linear dynamics and random coupling matrix, the functional connectivity approach gener-

Spurious small-world networks

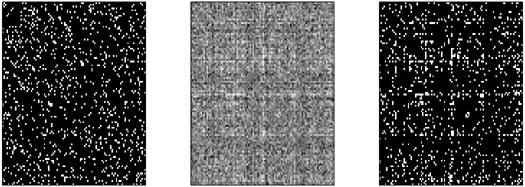


FIG. 1. An example of binary functional connectivity matrix (right) generated from random structural connectivity matrix (left) by thresholding the correlation matrix of AR-model generated time series (center, light shades of gray indicate higher correlation values). Network with $N = 100$ nodes shown. Note that the functional connectivity matrix shows a specific structure although the entries of the generating structural connectivity matrix were chosen randomly. See text for further details.

$\{0.2, 0.5, 0.75, 0.9, 0.99\}$, $\alpha \in \{0, 1\}$. We further varied p_{SC} and p_{FC} logarithmically in 24 steps within the $(0, 1)$ interval – more exactly both variables are defined as 2^n where n is an arithmetic progression from 0 to -6.9 with step -0.3. The lowest density was therefore smaller than 0.01.

For robustness of evidence, for each parameter setting we compute 20 independent realizations of the coupling

tical significance in most cases (p-values $< 10^{-5}$, sign test of hypothesis of median equal to 1, no correction for multiple comparisons; similar results obtained for t-test). The only exceptions were observed for the case of exactly equal densities of structural and functional connectivity matrix, when this commong density was very low (e.g. only for $p_{FC} = p_{SC} \lesssim 0.03$ for the specific settings in Figure 2), where the σ values were relatively close to 1;

NETWORKS FROM MULTIVARIATE TIME SERIES

- interesting, useful, dangerous
- (partial) transitivity connectivity measure $\rightarrow$
  $\rightarrow$ spurious small-world network topology
- biased connectivity measure $\rightarrow$
  $\rightarrow$ spurious highly connected hubs
- stability of connectivity, network structure
- significance of changes in time and space
- (climate) network variability vs. external influence

Dependence and directional coupling in time series

- bias due to different level of dynamical complexity in different nodes
  - symmetric measures: relate dynamics not static PDF use MIR rather than MI, corr
- directional measures: the same problem, only partial solution
  - surrogate data
  - test individually each direction
  - existence of directional coupling
  - but not its strength !

CONCLUSION

Thank you for your attention

Support: Czech Science Foundation,
Project No. P103/11/J068.

NONLINEAR DYNAMICS WORKGROUP
http://ndw.cs.cas.cz
SW for interaction analysis
SW for network analysis
Preprints