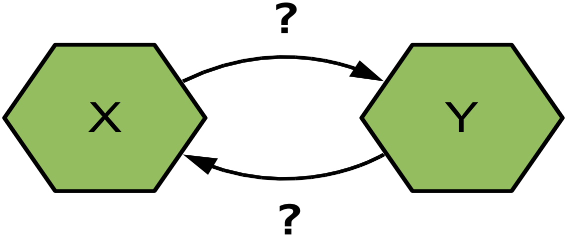
Cross-scale Information Transfer: Atmospheric Dynamics

Milan Paluš

Department of Nonlinear Dynamics and Complex Systems
Institute of Computer Science
Academy of Sciences of the Czech Republic
Prague, Czech Republic
E-Mail: mp@cs.cas.cz

CIDNET14, Dresden, June 17, 2014

Interactions in complex systems



- EVOLUTION OF (SUB)SYSTEMS: recorded time series
- INTERACTIONS:
  - COUPLING / DEPENDENCE $\rightarrow$ SYNCHRONIZATION ?
    - none, **unidirectional**, bidirectional

COUPLING / DEPENDENCE

RANDOM VARIABLES X, Y
Probability Distribution Functions $p(x)$, $p(y)$

INDEPENDENCE: $p(x, y) = p(x)p(y)$

digression from independence: $\log \frac{p(x,y)}{p(x)p(y)}$

a measure of dependence: MUTUAL INFORMATION

$$I(X; Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

Mutual information

- mutual information

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

- average amount of common information, contained in the variables X and Y
- measure of general statistical dependence
- $I(X;Y) \geq 0$
- $I(X;Y) = 0$ iff X and Y are independent

Conditional mutual information

- conditional mutual information $I(X;Y|Z)$ of variables X, Y given the variable Z

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

- Z independent of X and Y

$$I(X;Y|Z) = I(X;Y)$$

- $I(X;Y|Z) = I(X;Y;Z) - I(X;Z) - I(Y;Z)$
  here

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

- "net" dependence between X and Y without possible influence of Z

DIRECTED COUPLING – CAUSAL INFLUENCE

Sir Clive W. J. Granger, 2003 Nobel prize in economy
inspiration by the Wiener's work about causality:

1. The cause occurs before the effect; and
2. The cause contains information about the effect that is unique, and is in no other variable.

Granger causality

Sir Clive W. J. Granger, 2003 Nobel prize in economy

- causal variable can help to forecast the effect variable after other data has been first used
- restricted sense of causality, referred to as Granger causality (GC)
- *process* X_t *Granger causes another process* Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather then only past values of Y_t

Granger causality

- a linear regression model

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \xi_t, \qquad (1)$$

where ξ_t are uncorrelated random variables with zero mean and variance σ^2, L is the specified number of time lags, and $t = L + 1, \ldots, N$.

- The null hypothesis that X_t does not Granger cause Y_t is supported when $b_{2k} = 0$ for $k = 1, \ldots, L$:

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \tilde{\xi}_t. \qquad (2)$$

- process X_t Granger causes process Y_t iff $b_{2k} \neq 0$

Generalization of Granger causality

- time series $\{x(t)\}$ and $\{y(t)\}$: realizations of stationary and ergodic stochastic processes $\{X(t)\}$ and $\{Y(t)\}$
- we will mark $x(t)$ as x and $x(t+\tau)$ as x_τ, $\{y(t)\}$ ...
- mutual information $I(y; x_\tau)$ measures the average amount of information contained in the process $\{Y\}$ about the process $\{X\}$ in its future τ time units ahead (τ-future thereafter).
- $I(y; x_\tau)$ also contains an information about the τ-future of the process $\{X\}$ contained in this process itself if the processes $\{X\}$ and $\{Y\}$ are not independent, i.e., if $I(x; y) > 0$

Conditional mutual information

- In order to obtain the "net" information about the τ-future of the process $\{X\}$ contained in the process $\{Y\}$, use the **conditional mutual information** $I(y; x_\tau | x)$
- in time-series representation $I\big(\vec{Y}(t); \vec{X}(t+\tau) | \vec{X}(t)\big)$

$$= I\Big(\big(y(t), y(t-\rho), \ldots, y(t-(m-1)\rho)\big); x(t+\tau) |$$

$$\big(x(t), x(t-\eta), \ldots, x(t-(n-1)\eta)\big)\Big),$$

η, ρ: time lags, embedding of trajectories $\{\vec{X}(t)\}$, $\{\vec{Y}(t)\}$

- equivalent to transfer entropy (Schreiber, 2000)

Conditional mutual information

- typically

$$I\big(y(t); x(t+\tau)|x(t), x(t-\eta), \ldots, x(t-(n-1)\eta)\big)$$

  is sufficient to infer coupling direction between x and y
- dimensionality of condition matters

Estimation strategies

- mutual information $I(X_1; X_2; \ldots; X_n) =$

$$\sum_{x_1 \in \Xi_1} \sum_{x_2 \in \Xi_2} \cdots \sum_{x_n \in \Xi_n} p(x_1, x_2, \ldots, x_n) \log \frac{p(x_1, x_2, \ldots, x_n)}{p(x_1)p(x_2) \ldots p(x_n)}$$

- continuous variables
  - PDD – analytic solutions (Gaussian)
  - metric/distance based methods
- discrete variables
  - binning methods
  - symbolic/ranking methods
- parametric methods

Correlation coefficient

N observations x_i, y_i of two variables X and Y

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

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

the correlation between X and Y is

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

MI of Gaussian variables

Let variables X and Y have normal PDF $p(x, y)$, $p(x)$, $p(y)$
the correlation between X and Y is

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

then

$$I(X; Y) = -\frac{1}{2} \log \left(1 - c^2(X, Y)\right)$$

n random variables

- mutual information $I(X_1; X_2; \ldots; X_n)$
$= H(X_1) + H(X_2) + \cdots + H(X_n) - H(X_1, X_2, \ldots, X_n)$
- $X_1, \ldots, X_n$ an n-dimensional normally distributed random variable with zero mean and covariance matrix **C**

$$I_G(X_1; \ldots; X_n) = \frac{1}{2} \sum_{i=1}^{n} \log(c_{ii}) - \frac{1}{2} \sum_{i=1}^{n} \log(\sigma_i),$$

where c_{ii} are the diagonal elements (variances) and σ_i are the eigenvalues of the $n \times n$ covariance matrix **C**

n random variables

- using the correlation matrix instead of the covariance matrix, then particularly $c_{ii} = 1$ for every i, and we obtain

$$I_G(X_1; \ldots ; X_n) = -\frac{1}{2}\sum_{i=1}^{n}\log(\sigma_i)$$

Binning strategies

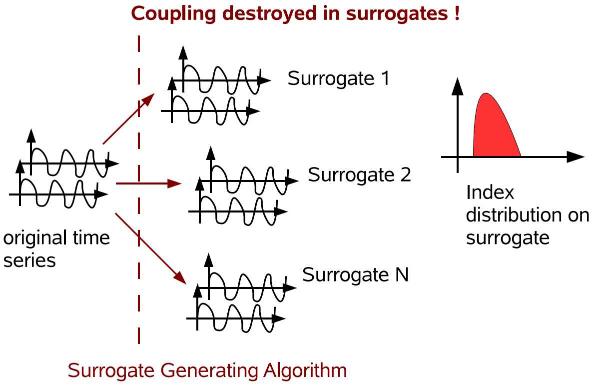
- EQUIDISTANT BINS
- ADAPTIVE BINS
  - adaptive separately for each variable
    marginally equiprobable bins
    marginal equiquantization
  - adaptive in 2-dim (n-dim) space
    Fraser-Swinney
    Darbellay-Vajda
- “FUZZY” BINS – B-splines

Direction of coupling

- CMI estimates
  - nonzero
  - different for $x \to y$ and $y \to x$
- how to discern unidirectional from bidirectional coupling?

Significance testing using surrogate data

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



PHYSICAL REVIEW E **75**, 056211 (2007)

Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections

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

(Received 14 December 2006; revised manuscript received 1 March 2007; published 18 May 2007)

We discuss some problems encountered in inference of directionality of coupling, or, in the case of two interacting systems, in inference of causality from bivariate time series. We identify factors and influences that can lead to either decreased test sensitivity or false detections and propose ways to cope with them in order to perform tests with high sensitivity and a low rate of false positive results.

DOI: 10.1103/PhysRevE.75.056211 PACS number(s): 05.45.Tp, 05.45.Xt, 02.50.Ey, 87.80.Tq

I. INTRODUCTION

Cooperative behavior of coupled complex systems has recently attracted considerable interest from theoreticians as well as experimentalists (see, e.g., the monograph [1]), since synchronization and related phenomena have been observed not only in physical but also in many biological systems

ries, is far from resolved. In this paper we identify some problems encountered in this task and give some practical advice for avoiding false detections of coupling asymmetry or causality. We will consider two interacting systems, possibly one of them driving the other. Then the coupling asymmetry, or, as it is called, the directionality of coupling, also

CMI – statistical evaluation

- estimator bias, variance
- proper surrogate data
- different dynamics $\rightarrow$ different bias in each direction:
  - different characteristic frequencies: I(slower $\rightarrow$ faster) $>$ I(faster $\rightarrow$ slower)
  - different complexity (entropy rate)
  - different noise content
- Do NOT use differences $I(X \rightarrow Y) - I(Y \rightarrow X)$
- test significance in each direction separately

PHYSICAL REVIEW E **77**, 026214 (2008)

Inferring the directionality of coupling with conditional mutual information

Martin Vejmelka*

Institute of Computer Science, Academy of Sciences of the Czech Republic, Praha, Czech Republic and Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Praha, Czech Republic

Milan Paluš

Institute of Computer Science, Academy of Sciences of the Czech Republic, Praha, Czech Republic

(Received 15 August 2007; published 21 February 2008)

Uncovering the directionality of coupling is a significant step in understanding drive-response relationships in complex systems. In this paper, we discuss a nonparametric method for detecting the directionality of coupling based on the estimation of information theoretic functionals. We consider several different methods for estimating conditional mutual information. The behavior of each estimator with respect to its free parameter is shown using a linear model where an analytical estimate of conditional mutual information is available. Numerical experiments in detecting coupling directionality are performed using chaotic oscillators, where the influence of the phase extraction method and relative frequency ratio is investigated.

DOI: 10.1103/PhysRevE.77.026214 PACS number(s): 05.45.Tp, 05.10.−a

Route to synchronization

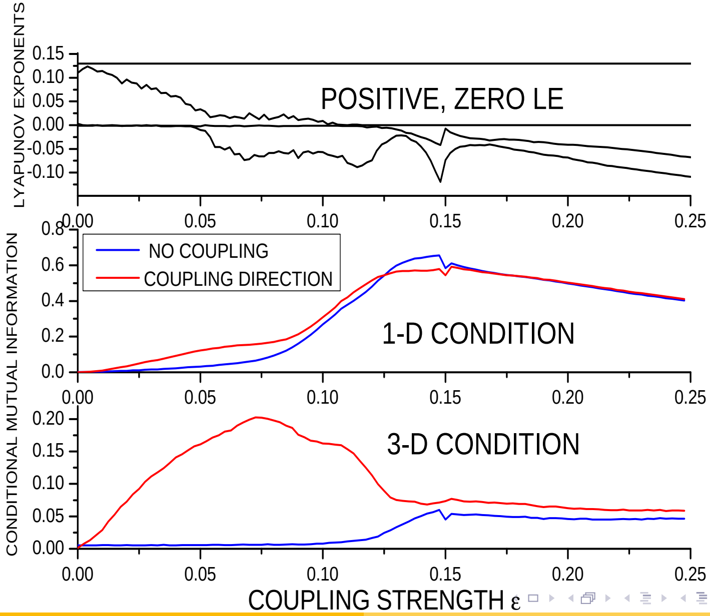
- unidirectionally coupled Rössler systems

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

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

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

Route to synchronization



MULTISCALE DYNAMICS

DECOMPOSITION OF BROAD-BAND SIGNALS

- DIGITAL FILTERING
- WAVELET DECOMPOSITION
- EMPIRICAL MODE DECOMPOSITION
- SINGULAR SPECTRUM ANALYSIS

IN-SCALE OR ACROSS SCALES INTERACTIONS

- SCALE-SPECIFIC SYNCHRONIZATION
- SCALE-SPECIFIC GRANGER CAUSALITY
- CROSS-SCALE INTERACTIONS
- CROSS-FREQUENCY COUPLING

Phase dynamics approach

ANALYTIC SIGNAL

$$\psi(t) = s(t) + j\hat{s}(t) = A(t)e^{j\phi(t)} \tag{3}$$

INSTANTANEOUS PHASE

$$\phi(t) = \arctan \frac{\hat{s}(t)}{s(t)} \tag{4}$$

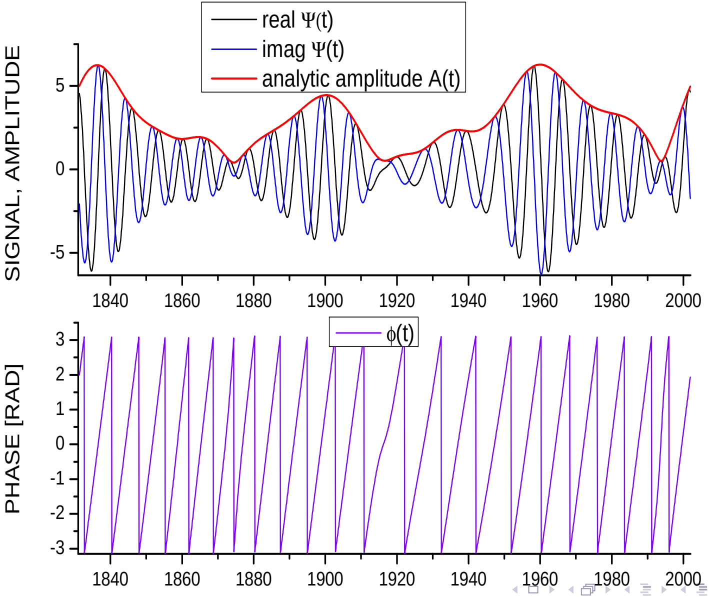
INSTANTANEOUS AMPLITUDE

$$A(t) = \sqrt{\hat{s}(t)^2 + s(t)^2} \tag{5}$$

FILTERING $\longrightarrow$ HILBERT TRANSFORM

COMPLEX CONTINUOUS WAVELET TRANSFORM

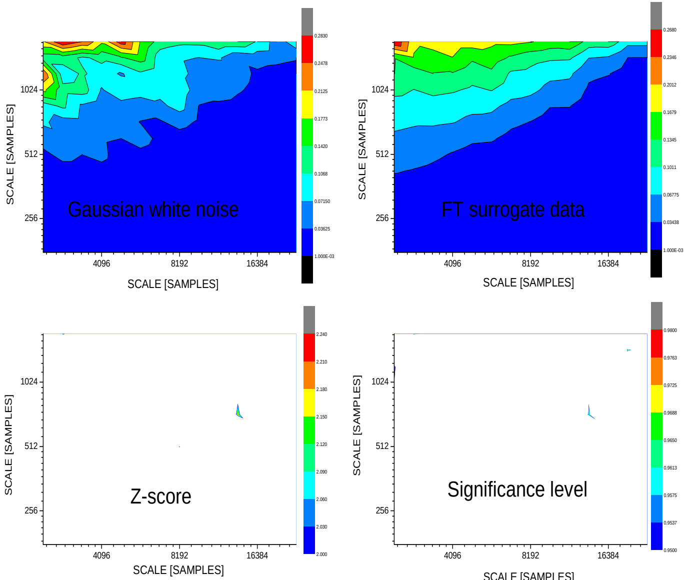
Phase dynamics in each scale (frequency)



Cross-frequency interactions

- phase–phase
- amplitude–amplitude
- **phase–amplitude**

Wavelet phase-phase method error



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)

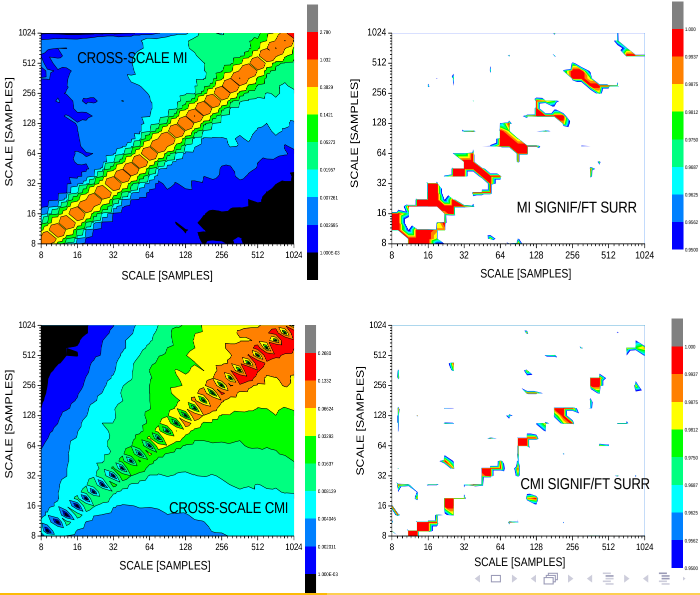
A method for random resampling of time series from multiscale processes is proposed. Bootstrapped series—realizations of surrogate data obtained from random cascades on wavelet dyadic trees—preserve the multifractal properties of input data, namely, interactions among scales and nonlinear dependence 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

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, success-

Multifractal process



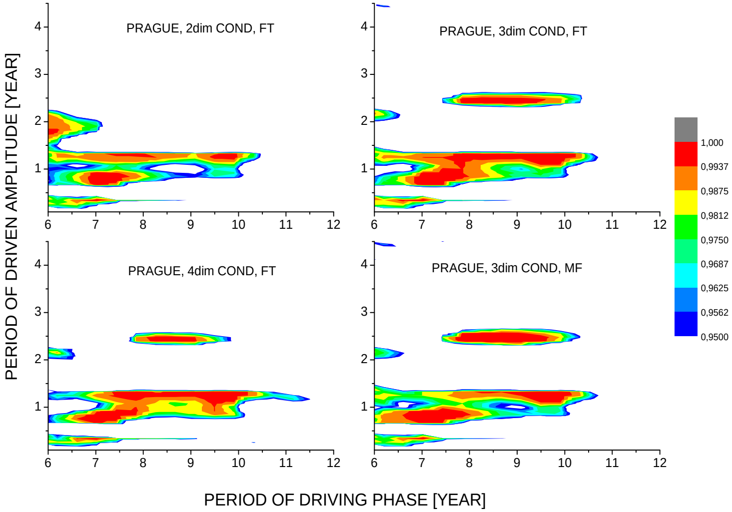
CAUSAL PHASE –> AMPLITUDE INTERACTIONS
in about a century long records of daily near-surface air temperature records from European stations

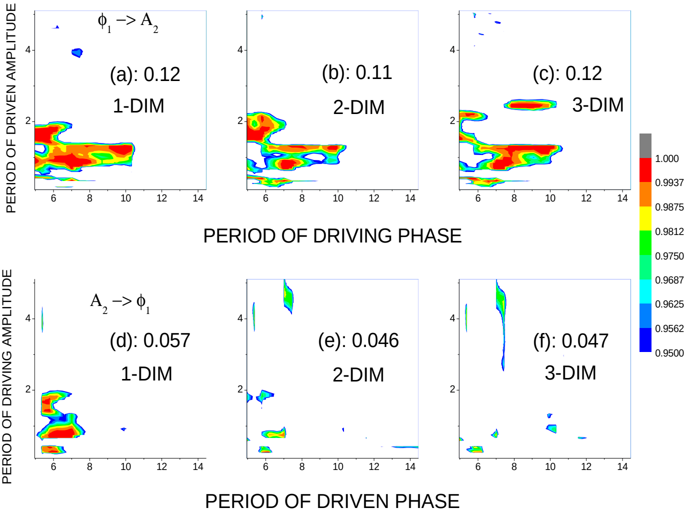
- phase ϕ_1 of slow oscillations (around 10 year period)
- amplitude A_2 of higher-frequency variability (periods 5 years and less)
- $I(\phi_1(t); A_2(t+\tau)|A_2(t), A_2(t-\eta), \ldots, A_2(t-m\eta))$
- testing using surrogate data approach
    - Fourier transform (FT) surrogate data (Theiler et al.)
    - multifractal (MF) surrogate data (Paluš)

CAUSAL PHASE –> AMPLITUDE INTERACTIONS

- $I(\phi_1(t); A_2(t+\tau)|A_2(t), A_2(t-\eta), \ldots, A_2(t-m\eta))$
- series length 32768
- forward lags $\tau = 1 - 750$ days
- backward condition lags η = 1/4 of the slow period
- Gaussian process estimator
- conditioning dimension: stable results from 3
- raw data include annual cycle
- seasonal mean and variance removed before surrogate randomization
- seasonal mean and variance added back to surrogate realizations

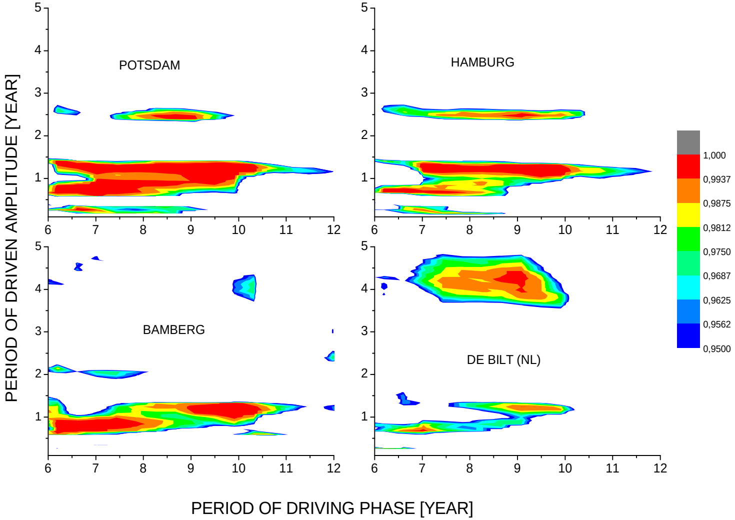
CAUSAL PHASE –> AMPLITUDE INTERACTIONS



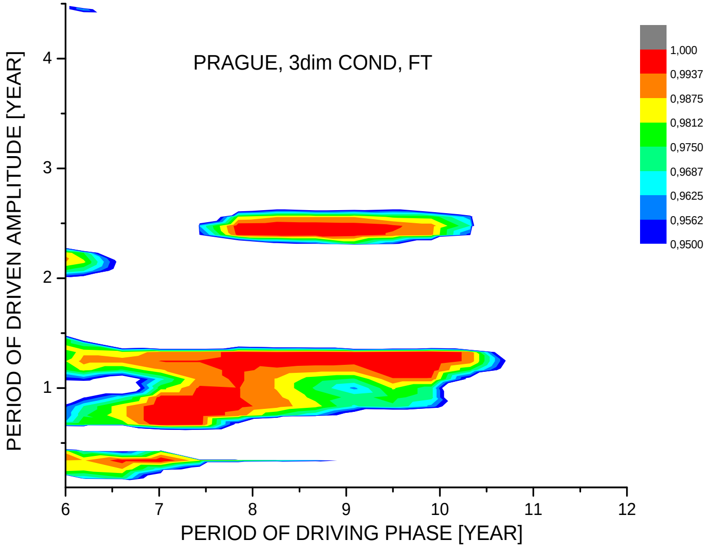


ϕ1 -> A2
(a): 0.12
1-DIM
(b): 0.11
2-DIM
(c): 0.12
3-DIM
PERIOD OF DRIVEN AMPLITUDE
PERIOD OF DRIVING PHASE
A2 -> ϕ1
(d): 0.057
1-DIM
(e): 0.046
2-DIM
(f): 0.047
3-DIM
PERIOD OF DRIVING AMPLITUDE
PERIOD OF DRIVEN PHASE
2
4
6
8
10
12
14
1.000
0.9937
0.9875
0.9812
0.9750
0.9687
0.9625
0.9562
0.9500

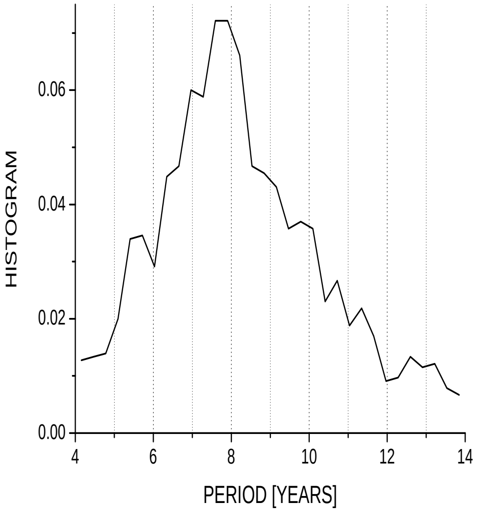
CAUSAL PHASE –> AMPLITUDE INTERACTIONS



CAUSAL PHASE –> AMPLITUDE INTERACTIONS



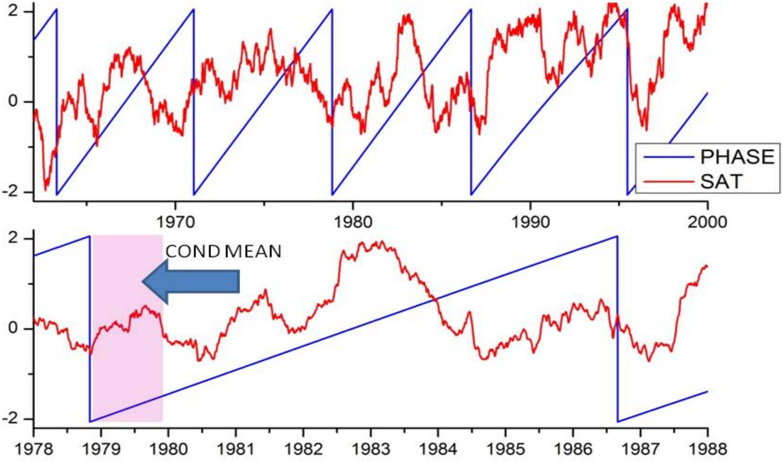
SSA-extracted "7–8 yr cycle", Prague SAT



EFFECT PHASE –> AMPLITUDE COUPLING

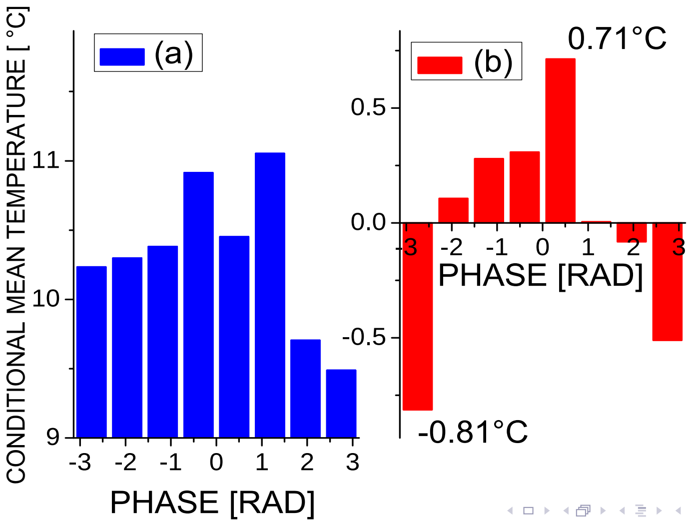
- HOW TO QUANTIFY THE EFFECT OF PHASE –> AMPLITUDE COUPLING ?
- EXTRACT THE CYCLE WITH PERIOD AROUND 8 YEARS
- EXTRACT ITS PHASE
- DIVIDE THE PHASE INTO 8 BINS
- COMPUTE CONDITIONAL TEMPERATURE MEANS $< T|\phi \in (\phi_1, \phi_2) >$

Computing SAT conditional means



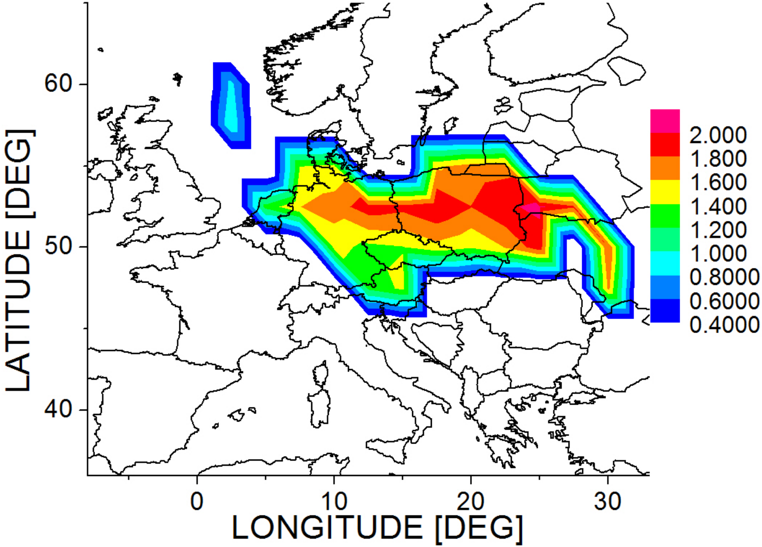
EFFECT of PHASE –> AMPLITUDE COUPLING

a) Prague SAT conditional means (CM)
b) Prague SAT anomalies CM



EFFECT of PHASE –> AMPLITUDE COUPLING

ERA SATA (ERA-40 + ERA-Interim reanalysis data)



Cross-scale information transfer

PRL **112,** 078702 (2014) PHYSICAL REVIEW LETTERS week ending 21 FEBRUARY 2014

Multiscale Atmospheric Dynamics: Cross-Frequency Phase-Amplitude Coupling in the Air Temperature

Milan Paluš*

Department of Nonlinear Dynamics and Complex Systems, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic

(Received 29 November 2012; published 21 February 2014)

Interactions between dynamics on different temporal scales of about a century long record of data of the daily mean surface air temperature from various European locations have been detected using a form of the conditional mutual information, statistically tested using the Fourier-transform and multifractal surrogate data methods. An information transfer from larger to smaller time scales has been observed as the influence of the phase of slow oscillatory phenomena with the periods around 6–11 yr on the amplitudes of the variability characterized by the smaller temporal scales from a few months to 4–5 yr. The overall effect of the slow oscillations on the interannual temperature variability within the range 1–2 °C has been observed in large areas of Europe.

DOI: 10.1103/PhysRevLett.112.078702 PACS numbers: 92.60.Ry, 05.45.Tp, 89.75.Da, 92.70.Gt

Open problems

- Conditioning dimension: objective, automatic, ..., data limitations
- Estimator
  - Gaussian: "sees" only interactions of ϕ_1 and A_2 in the same time scale (A_2 related to fast variability given by ϕ_2, but A_2 varies slowly, in time scale of ϕ_1)
  - binning: low sensitivity (32768 daily samples, only 11 cycles with 8yr period)
  - k-nn: can see more interactions? do they exist?
- time scale estimation
  - WT: uncertainty in frequency and time localization
  - natural cycles - varying frequency
  - other methods? SSA, EMD

Open problems

- Statistical significance
  - sensitivity & specificity depend on number of cycles
  - on the estimator type
  - multiple testing, but dependent tests
  - tests of the method: artificial multiscale data
- Measurable effects / conditional means, variance
  - nonstationarity, segmentation
  - WT redundacy - non-redundant decomposition?
  - other variables – precipitation

Thank you for your attention

**Support: Czech Science Foundation, Projects
No. P103/11/J068 and No. P303/14-02634S.**

NONLINEAR DYNAMICS WORKGROUP
http://ndw.cs.cas.cz
SW for interaction analysis
SW for network analysis
Electronic preprints at http://www.cs.cas.cz/mp/