Synergetic and Redundant Information Flow Detected by Causal Analysis of Dynamical Networks

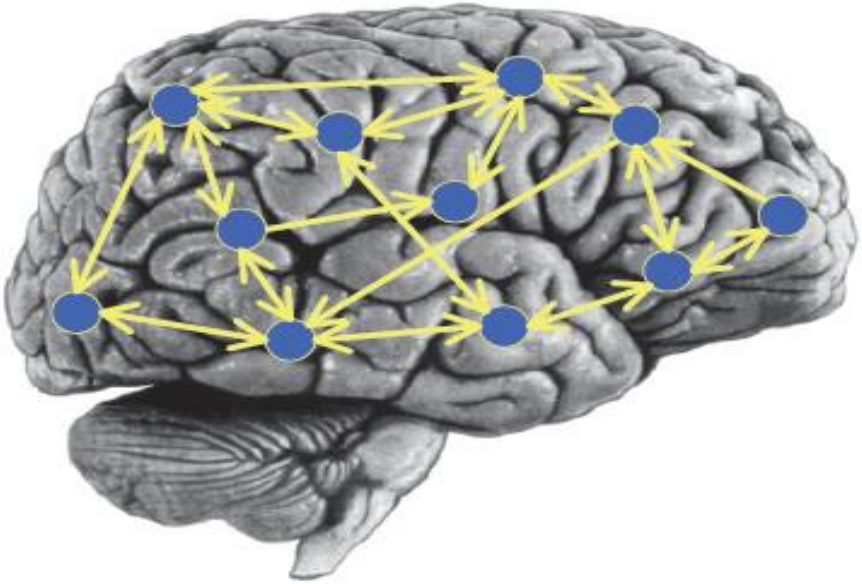
Sebastiano Stramaglia

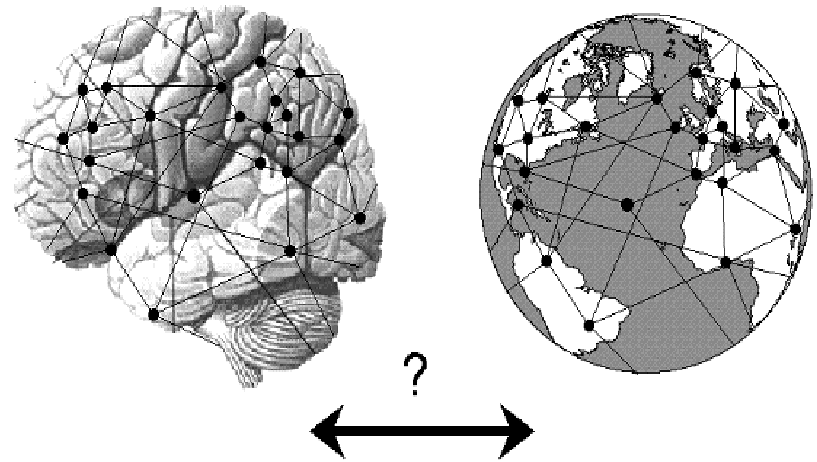
Physics Dept. University of Bari & INFN, Italy

Coworkers:
Luca Faes (Trento, Italy) and Daniele Marinazzo (Ghent, Belgium)

Plan of the talk

- Systems Neuroscience: brain connectivity
- Granger causality-Transfer Entropy
- Synergy and Redundancy
- Conclusions





Segregation-Integration

Systems neuroscience description of brain: Networks

BRAIN CONNECTIVITY:

- **Anatomical (structural) connectivity**
- **Functional connectivity**
- **Effective connectivity**

Structural connectivity

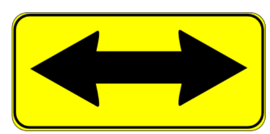
Levels of structural connectivity

- Microscale (micrometer)
- Mesoscale (0.1 millimeter)
- Macroscale (>= 1 millimeter)

Connectivity is a scale dependent notion

Functional Connectivity

- Statistical dependency between neuronal units (also distant ones)
- Highly dynamic (unlike structural conn.)
- Symmetric

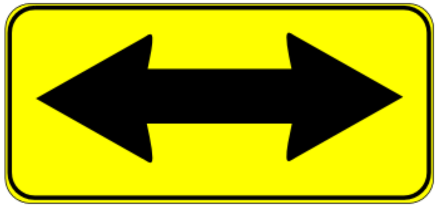


FC is a superposition of the interaction effects coming from structural connectivity (real physical connections) plus the effects coming from having a common functionality.

Effective connectivity

It is important not only to detect functional relations, but also to identify cause-effect (drive-response) relationships between neuronal units.

Functional:

A  B

Effective:

A  B

The relation structure-function

- Which are the properties of structural networks that allow them to support a huge number of different functional patterns?
- Dynamical state is critical? Should be weakly dependent on details of the underlying structural connectivity
- Many others

functional connectivity: Synchronization

Synchronization is the coordination of events to operate a system in unison.

Hyper-synchronization of EEG in migraine under visual stimulation

VOLUME 93, NUMBER 3 PHYSICAL REVIEW LETTERS week ending 16 JULY 2004

Steady-State Visual Evoked Potentials and Phase Synchronization in Migraine Patients

L. Angelini,[1,2,3] M. De Tommaso,[1,4] M. Guido,[1] K. Hu,[5] P. Ch. Ivanov,[5,6] D. Marinazzo,[1] G. Nardulli,[1,2,3] L. Nitti,[1,7,3] M. Pellicoro,[1,2,3] C. Pierro,[1] and S. Stramaglia[1,2,3]

[1]*TIRES: Center of Innovative Technologies for Signal Detection and Processing, University of Bari, Italy*
[2]*Physics Department, University of Bari, Italy*
[3]*Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy*
[4]*Department of Neurological and Psychiatric Sciences, University of Bari, Italy*
[5]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts, USA*
[6]*Institute of Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia, Bulgaria*
[7]*D.E.T.O., University of Bari, Italy*

(Received 8 October 2003; revised manuscript received 5 April 2004; published 15 July 2004)

We investigate phase synchronization in EEG recordings from migraine patients. We use the analytic signal technique, based on the Hilbert transform, and find that migraine brains are characterized by enhanced alpha band phase synchronization in the presence of visual stimuli. Our findings show that migraine patients have an overactive regulatory mechanism that renders them more sensitive to external stimuli.

DOI: 10.1103/PhysRevLett.93.038103 PACS numbers: 87.19.La, 05.45.Xt, 05.45.Tp, 42.66.Lc

Effective connectivity

X and Y two (vectorial) time series

x, the future values of X

1969: Granger causality

Definition: Y is cause of X if the knowledge of Y allows to make more precise prediction about x

This definition is meaningful only for irreversible processes, the direction of time is needed.

Absence of causality: Generalized Markov property

$$P(x \mid X) = P(x \mid X, Y)$$

Transfer entropy (Schreiber 2000)

Measuring the violation of the generalized Markov property:

$$T(Y \rightarrow X) = \int P(x, X, Y) \log\left(\frac{P(x \mid X, Y)}{P(x \mid X)}\right) dx\,dX\,dY$$

T measures the information flowing from one series to the other.

T is connected but not equivalent to coupling

Transfer entropy

$$S_X = -\int dxdXp(x,X)\log[p(x\,|\,X)]$$

$$S_{XY} = -\int dxdXdYp(x,X,Y)\log[p(x\,|\,X,Y)]$$

Regression

$$E_X = \int dxdX\,p(x,X)\left(x - \int dx'\,p(x'\,|\,X)x'\right)^2$$

$$E_{XY} = \int dxdXdY\,p(x,X,Y)\left(x - \int dx'\,p(x'\,|\,X,Y)x'\right)^2$$

For Gaussian variables (Barnett et al., PRL 2009)

$$Granger\ causality = 2\ Transfer\ Entropy$$

For linear systems, there is complete equivalence of the notions of Granger causality and transfer entropy.

Unifies information-theoretic and autoregressive approaches, GC measures the flow of information

Analytical expression for transfer entropy-GC

Nonlinearity: kernelization

- Using the theory of Reproducing kernel Hilbert spaces, the new formulation can be generalized to the nonlinear case.
- The inner product is to be replaced by a suitable kernel function with spectral representation:

$$K(x,x') = \sum_i \lambda_i \psi(x)\psi(x') = \sum_i \phi_i(x)\phi_i(x')$$

$$K_{ij} = K(X_i, X_j) \qquad K'_{ij} = K(Z_i, Z_j)$$

Equivalent to perform linear granger causality in the space of the eigenfunctions of K

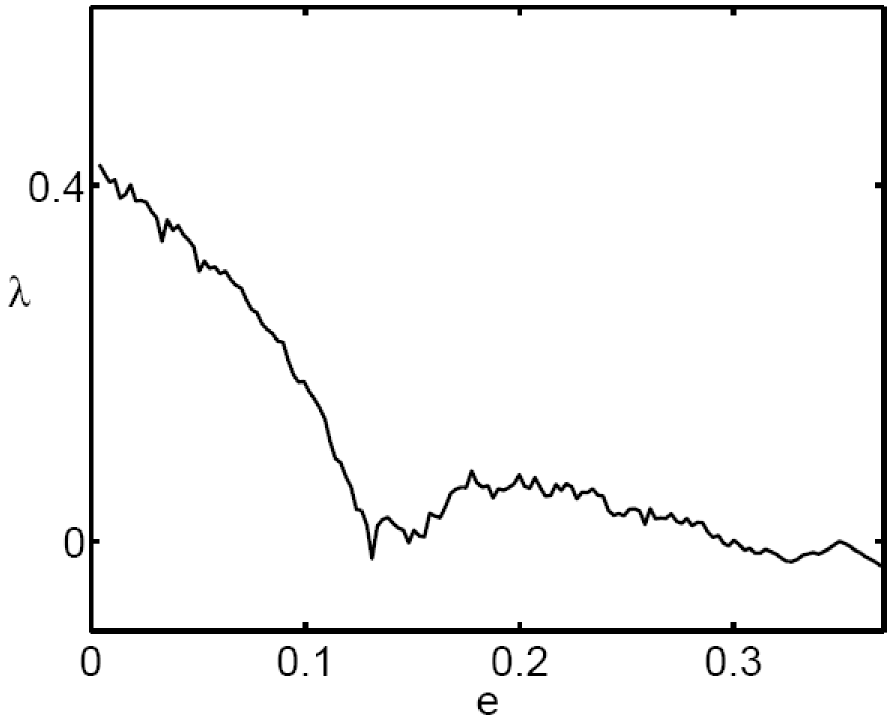
Marinazzo, Pellicoro, Stramaglia, Physical Review Letters 2008

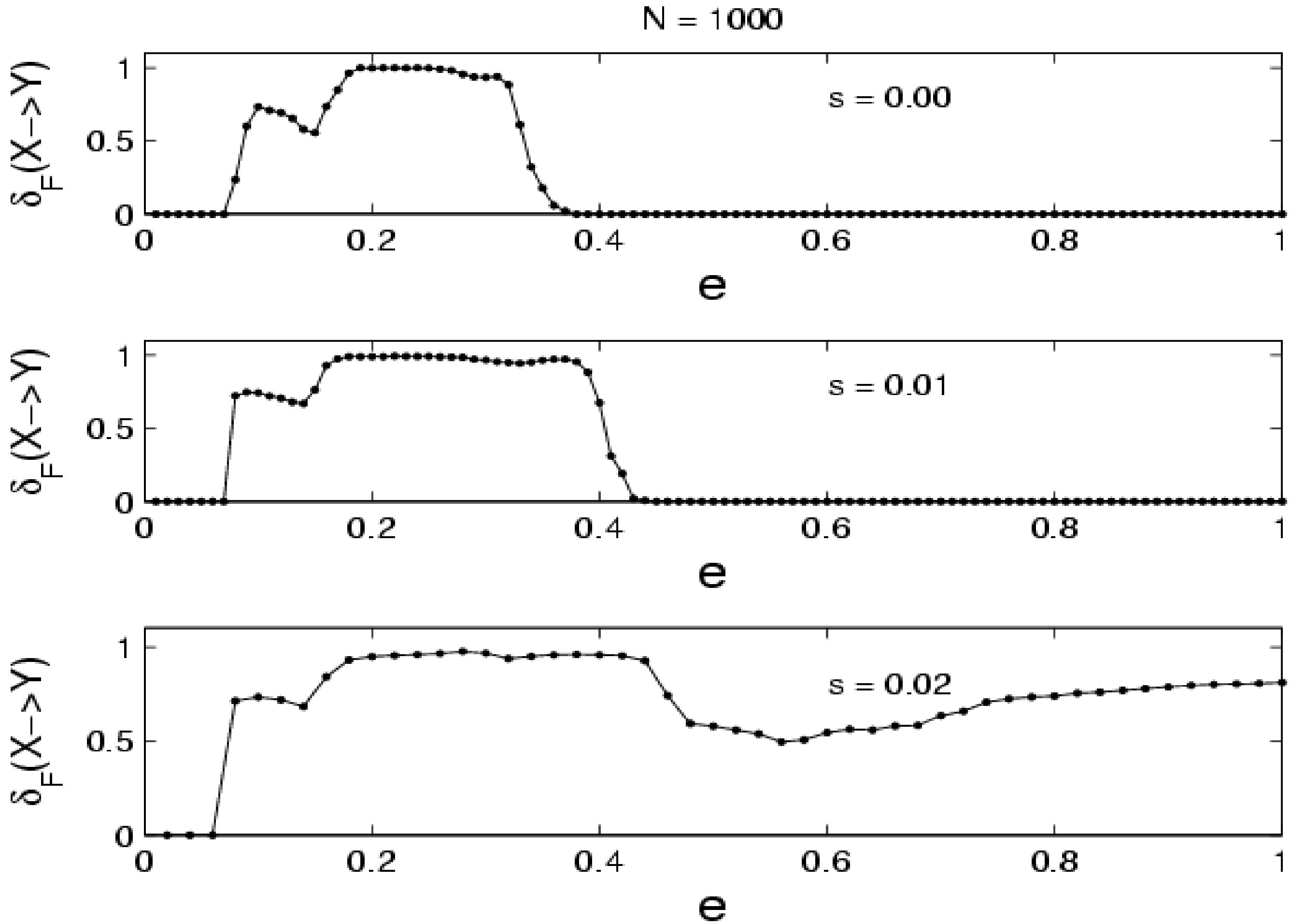
Nonlinear approach: Symbolic Transfer Entropy (Lehnertz et al.)

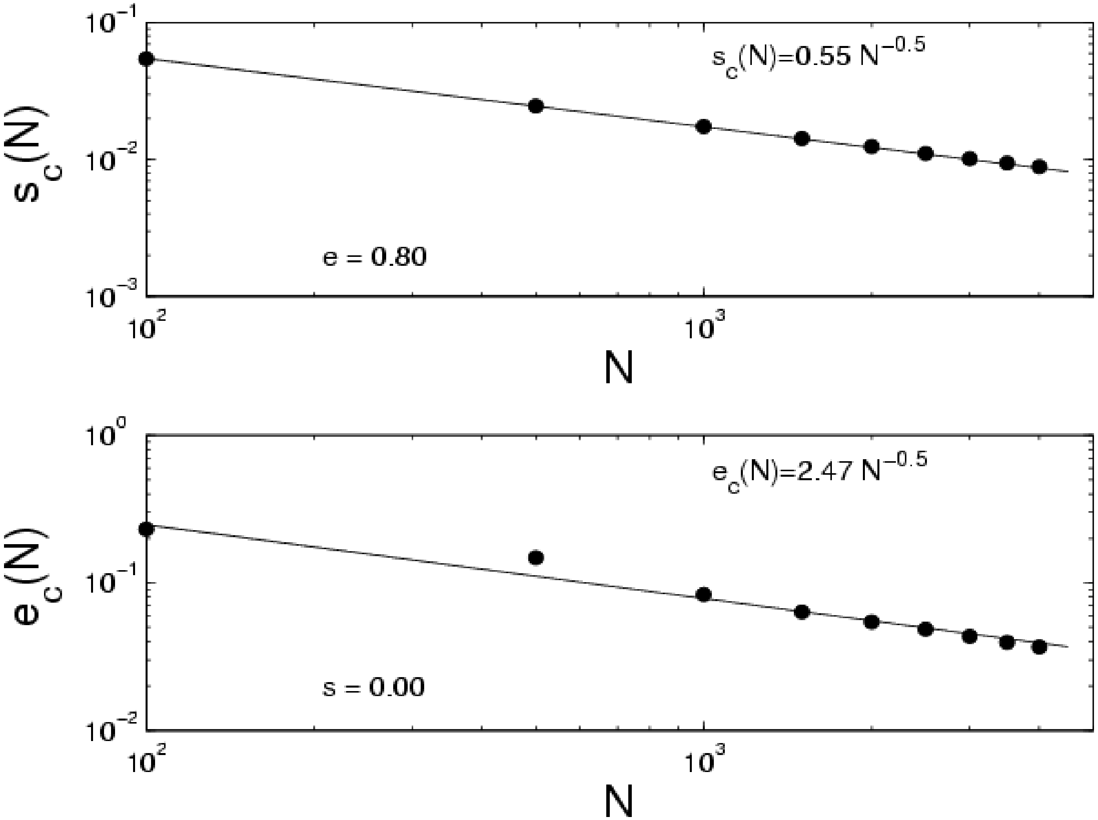
Example of application of the method

Pair of noisy logistic maps:

$$x_{n+1} = a\left(1 - x_n^2\right) + s\eta_{n+1},$$
$$y_{n+1} = (1-e)a\left(1 - y_n^2\right) + ea\left(1 - x_n^2\right) + s\xi_{n+1};$$




N = 1000
s = 0.00
s = 0.01
s = 0.02
δ_F(X–>Y)
e
0
0.5
1
0.2
0.4
0.6
0.8



$s_c(N)=0.55\ N^{-0.5}$
$s_c(N)$
e = 0.80
N
$e_c(N)=2.47\ N^{-0.5}$
$e_c(N)$
s = 0.00
N

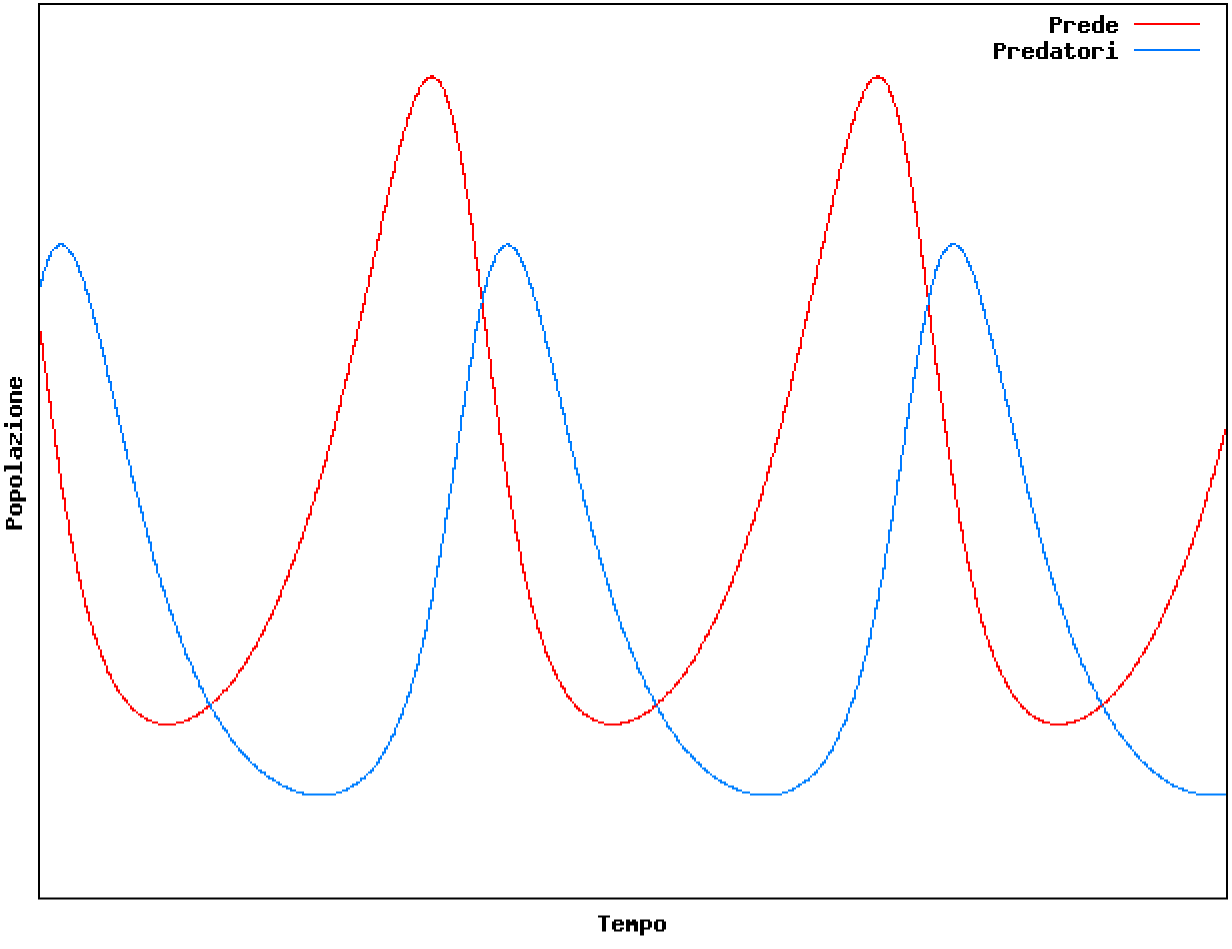
Oceanography

Detecting Causality in Complex Ecosystems

George Sugihara,[1]* Robert May,[2] Hao Ye,[1] Chih-hao Hsieh,[3]* Ethan Deyle,[1] Michael Fogarty,[4] Stephan Munch[5]

Lotka Volterra

$$dx/dt = (A - By)x$$
$$dy/dt = (Cx - D)y$$



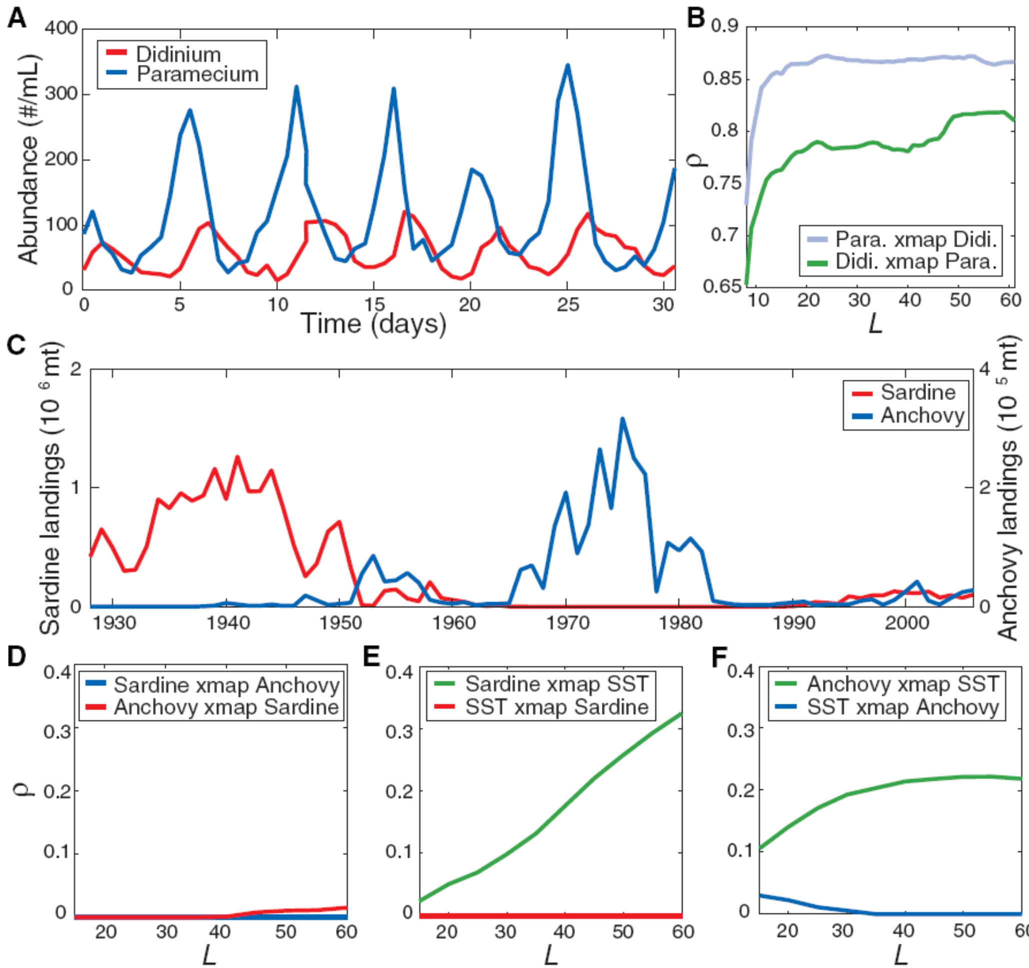
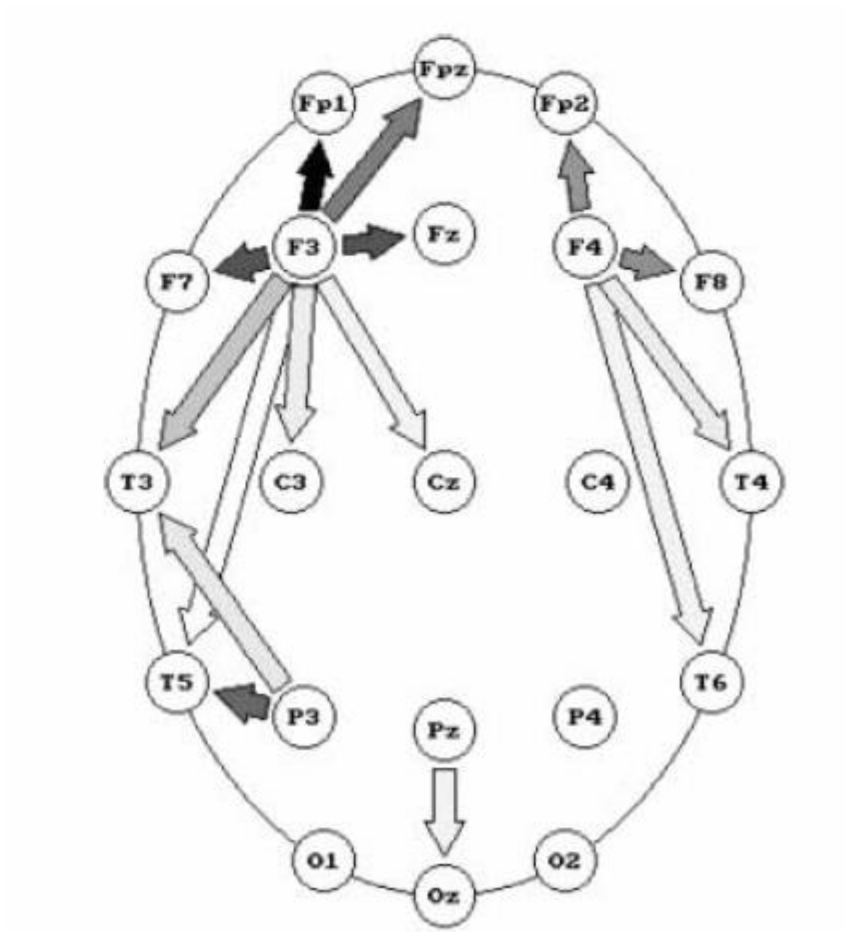


Fig. 5. Detecting causation in real time series. (**A**) Abundance time series of *Paramecium aurelia* and *Didinium nasutum* as reported in (*28*). (**B**) CCM of *Paramecium* and *Didinium* with increasing time-series length *L*. The pattern suggests top-down predator control. (**C**) California landings of Pacific sardine (*Sardinos sagax*) and northern anchovy (*Engraulis mordax*). (**D** to **F**) CCM (or lack thereof) of sardine versus anchovy, sardine versus SST (Scripps Pier), and anchovy versus SST (Newport Pier), respectively. This shows that sardines and anchovies do not interact with each other and that both are forced by temperature.

Information flow pattern (NxN matrix)



Networks Motifs (Alon, 2003)

- Characteristic network building blocks
- Small connected subgraphs that occur significantly more frequently than in randomized networks
- Transcription networks, signal transduction networks
- Brain networks: small set of structural motifs, large number of functional motifs (Sporns,Bullmore)

Mining informational motifs from data

- Possible strategy: Transfer Entropy -> Graph -> Motifs of the graph
- Problem: the presence of redundant variables renders the performance of multivariate transfer entropy poor (Angelini,Pellicoro,Stramaglia, PRE 2010)
- Multiplets of variables, constituting informational circuits, must be sought for directly from time series data

Expanding the transfer entropy

- Formal expansion of the transfer entropy to put in evidence irreducible sets of variables which provide information for the future state of each assigned target
- Multiplets characterized by a large contribution to the expansion are associated to informational circuits present in the system, with an informational character (synergetic or redundant) which can be associated to the sign of the contribution.
- For the sake of computational complexity, we adopt the assumption of Gaussianity and use the corresponding exact formula for the conditional mutual information.

INTERACTION INFORMATION: informational character of circuits of three variables

$$R(x, y, z) = I(x; z) + I(y; z) - I(\{x, y\}; z)$$

Example: s stimulus, r1 and r2 the response from two brain areas

R=0 Information Independence

$$I(\{r_1, r_2\}; s) = I(r_1; s) + I(r_2; s)$$

The two brain areas are sensitive to completely different features of the stimulus

E. Schneidman, W. Bialek, M.J. Berry, J. Neuroscience 23,11539 (2003).

R<0 Synergy

$$I\left(\left\{r_1, r_2\right\}; s\right) > I\left(r_1; s\right) + I\left(r_2; s\right)$$

The joint response from the two brain areas conveys more information than treating them separately

S is a function of both r1 and r2

R>0 Redundancy

$$I\left(\{r_1, r_2\}; s\right) < I\left(r_1; s\right) + I\left(r_2; s\right)$$

The two brain areas are sensitive to the same features of the stimulus

The two responses r1 and r2 share a certain amount of common information about the stimulus

Flow of information

$$S\left(x_0 | \{Y_k\}_{k=1}^{n}\right) - S(x_0) = -I\left(x_0; \{Y_k\}_{k=1}^{n}\right)$$

$$S\left(x_0 | \{Y_k\}_{k=1}^{n}\right) - S(x_0) =$$
$$\sum_i \frac{\Delta S(x_0)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(x_0)}{\Delta Y_i \Delta Y_j} + \cdots + \frac{\Delta^n S(x_0)}{\Delta Y_i \cdots \Delta Y_n}.$$

Conditioning on the past of the target

$$\mathcal{C}_{Y_0} S(X) = S(X|Y_0)$$

$$S\left(x_0|\{Y_k\}_{k=1}^{n}, Y_0\right) - S(x_0|Y_0) = -I\left(x_0; \{Y\}_{k=1}^{n}|Y_0\right) =$$
$$\sum_i \frac{\Delta S(x_0|Y_0)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(x_0|Y_0)}{\Delta Y_i \Delta Y_j} + \cdots + \frac{\Delta^n S(x_0|Y_0)}{\Delta Y_i \cdots \Delta Y_n}.$$

First terms in the expansion

$$A_i^0 = \frac{\Delta S(x_0|Y_0)}{\Delta Y_i} = -I\left(x_0; Y_i|Y_0\right)$$

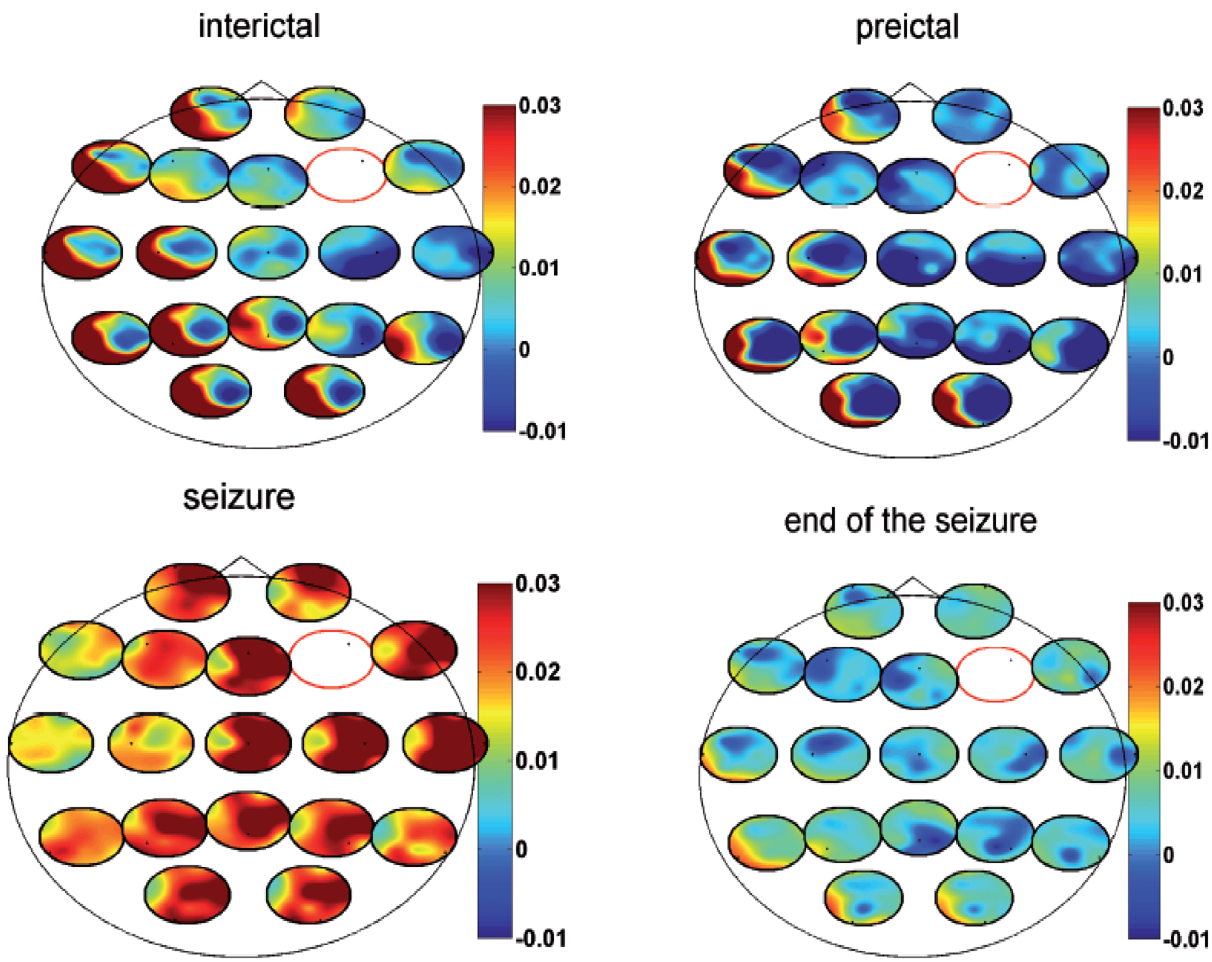
$$B_{ij}^0 = I\left(x_0; Y_i|Y_0\right) - I\left(x_0; Y_i|Y_j, Y_0\right)$$

$$\begin{aligned} C_{ijk}^0 = \; & I\left(x_0; Y_i|Y_j, Y_0\right) + I\left(x_0; Y_i|Y_k, Y_0\right) \\ & -I\left(x_0; Y_i|Y_0\right) - I\left(x_0; Y_i|Y_j, Y_k, Y_0\right) \end{aligned}$$

$$\frac{\Delta^n S\left(x_0 \middle| Y_0\right)}{\Delta Y_{i_1} \Delta Y_{i_2} \cdots \Delta Y_{i_n}}$$

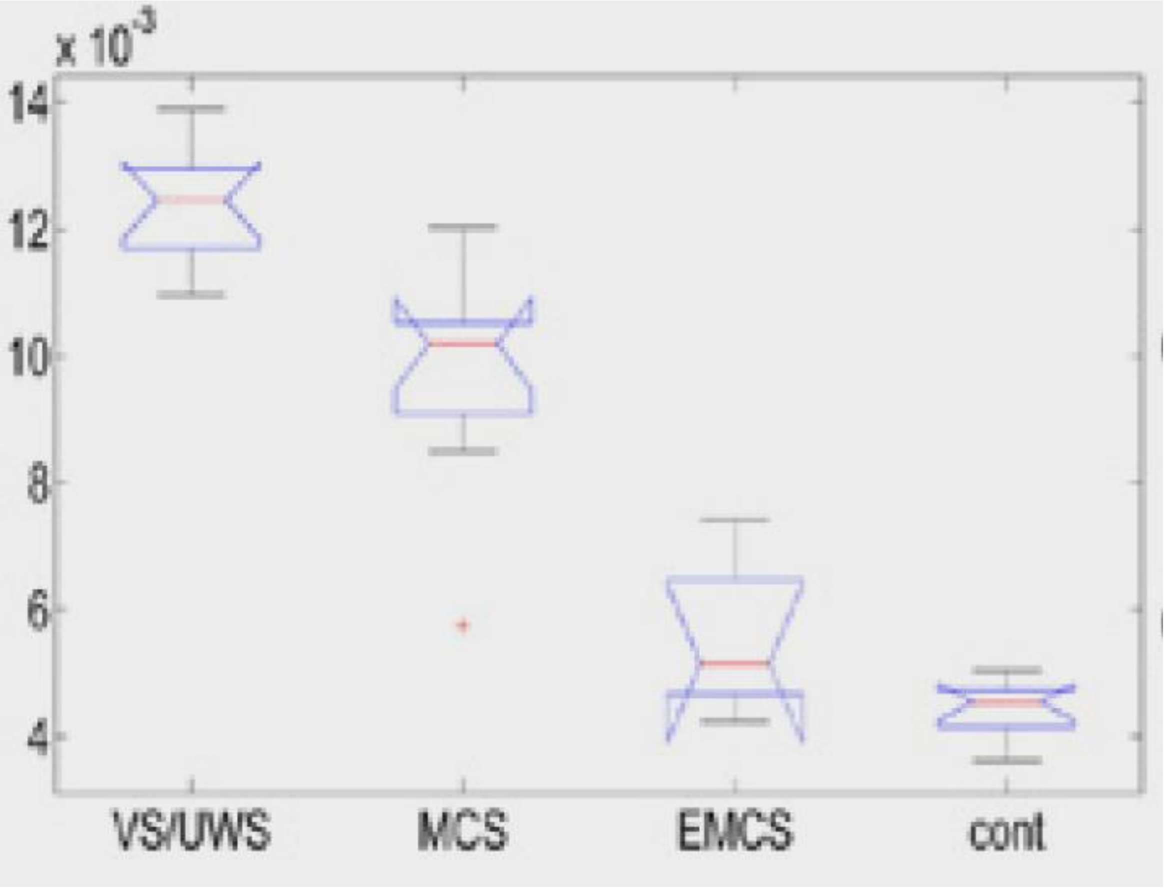
- Symmetrical under permutations of variables
- Independence among any of the Y results in vanishing contribution
- Each nonvanishing term provides an irreducible set of variables which send information to the target variable
- The sign of the contribution is related to the informational character: positive for redundancy, negative for synergy

Epilepsy: scalp EEG



TBI analysis: healthy controls are characterized by a greater amount of synergetic contributions from duplets of variables, w.r.t vegetative state, minimally conscious state, and emergence of the minimally conscious state

Red - syn



D. Marinazzo et al, Clinical EEG and neuroscience 2014

Partial Information Decomposition

Is it possible to separate a redundant and a synergetic contribution?

Is it possible to estimate an unique information coming from each variable?

Two variables are drivers, one is the target

Information Transfer Decomposition

$$\mathcal{T}_{i \to j} = I(Y_{j,n}; Y_{i,n}^{-} | Y_{j,n}^{-})$$

$$\mathcal{T}_{ik \to j} = I(Y_{j,n}; Y_{i,n}^{-}, Y_{k,n}^{-} | Y_{j,n}^{-}).$$

Joint Transfer Entropy

$$\mathcal{T}_{ik \to j} = \mathcal{T}_{i \to j} + \mathcal{T}_{k \to j} + \mathcal{I}_{ik \to j}.$$

Interaction Information Transfer

Partial Information Decomposition

$$\mathcal{T}_{ik\to j} = \mathcal{U}_{i\to j} + \mathcal{U}_{k\to j} + \mathcal{R}_{ik\to j} + \mathcal{S}_{ik\to j},$$

$$\mathcal{T}_{i\to j} = \mathcal{U}_{i\to j} + \mathcal{R}_{ik\to j},$$

$$\mathcal{T}_{k\to j} = \mathcal{U}_{k\to j} + \mathcal{R}_{ik\to j}.$$

$$\mathcal{I}_{ik\to j} = \mathcal{S}_{ik\to j} - \mathcal{R}_{ik\to j}$$

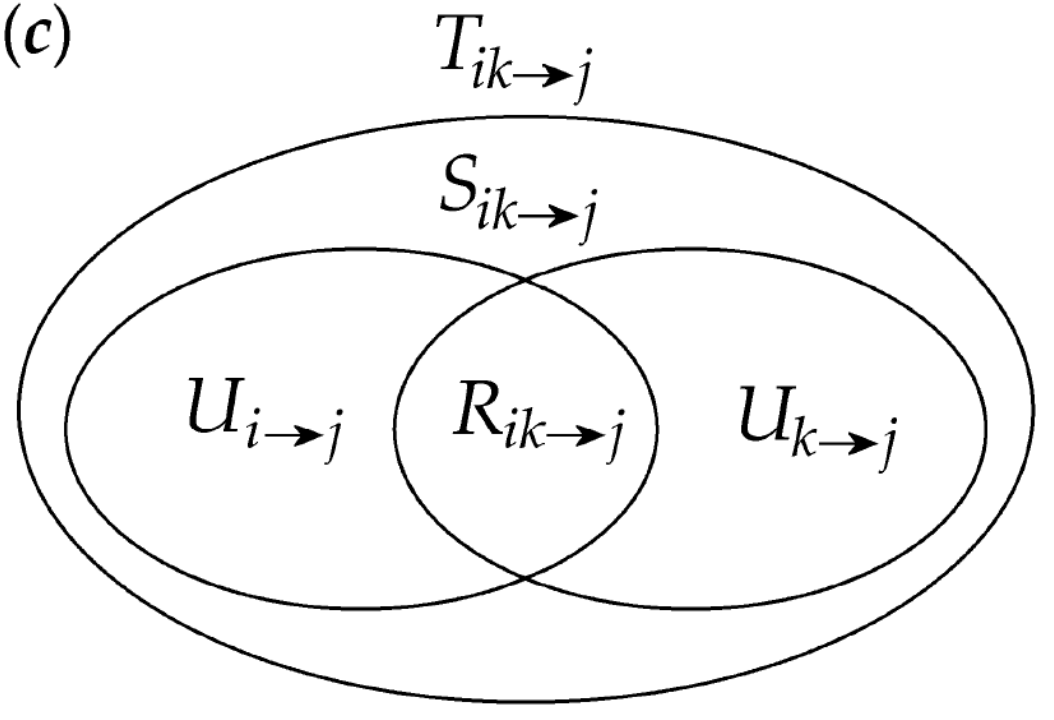
One more relation is needed to solve all the quantities. Shannon information theory does not univocally determine this decomposition

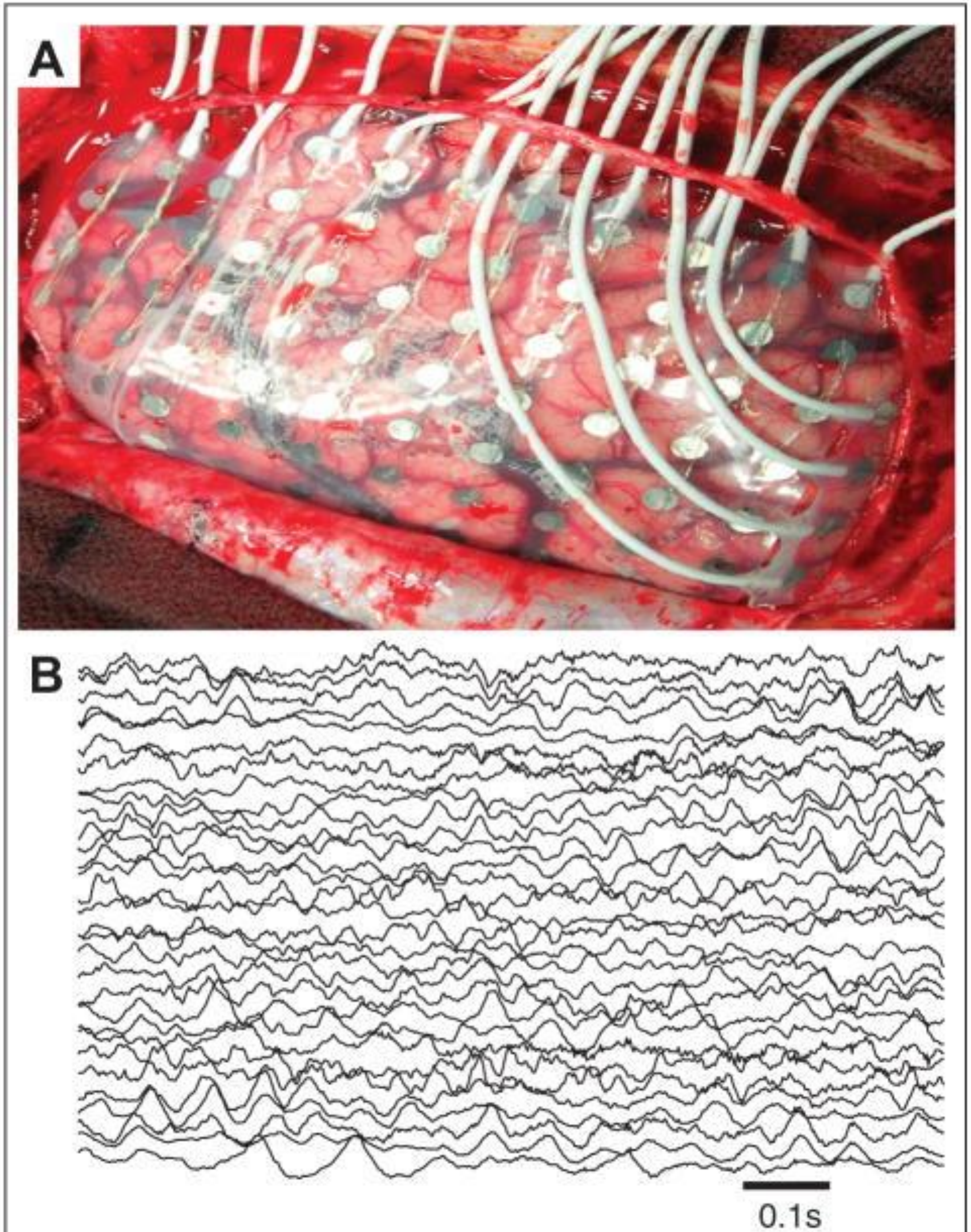
Minimum MI

$$\mathcal{R}_{ik\to j} = \min\{\mathcal{T}_{i\to j}, \mathcal{T}_{k\to j}\}$$

“Nonnegative Decomposition of Multivariate Information”
Paul L. Williams, Randall D. Beer
-M. Wibral, J. Lizier, ….

Partial Information Decomposition

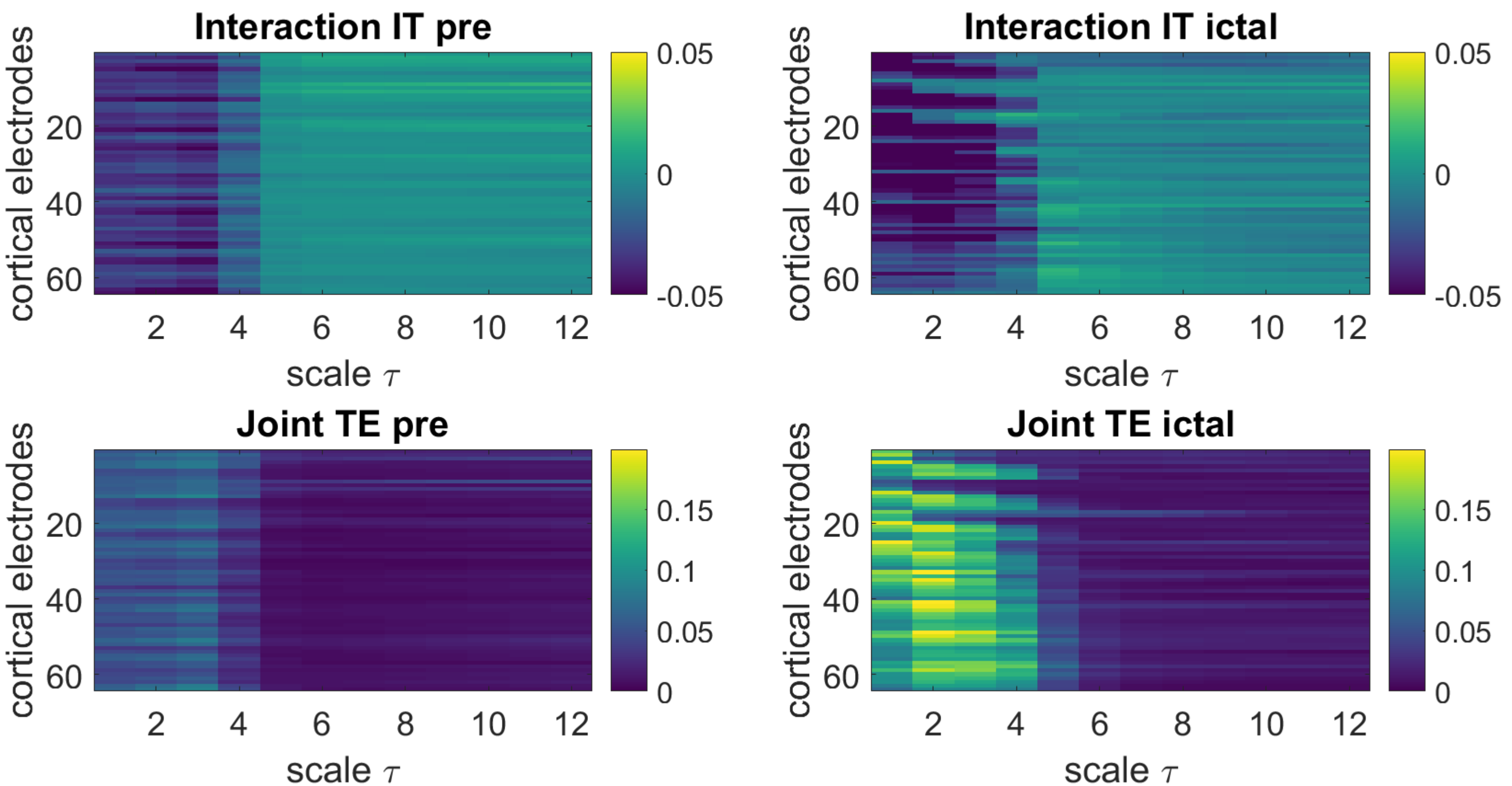


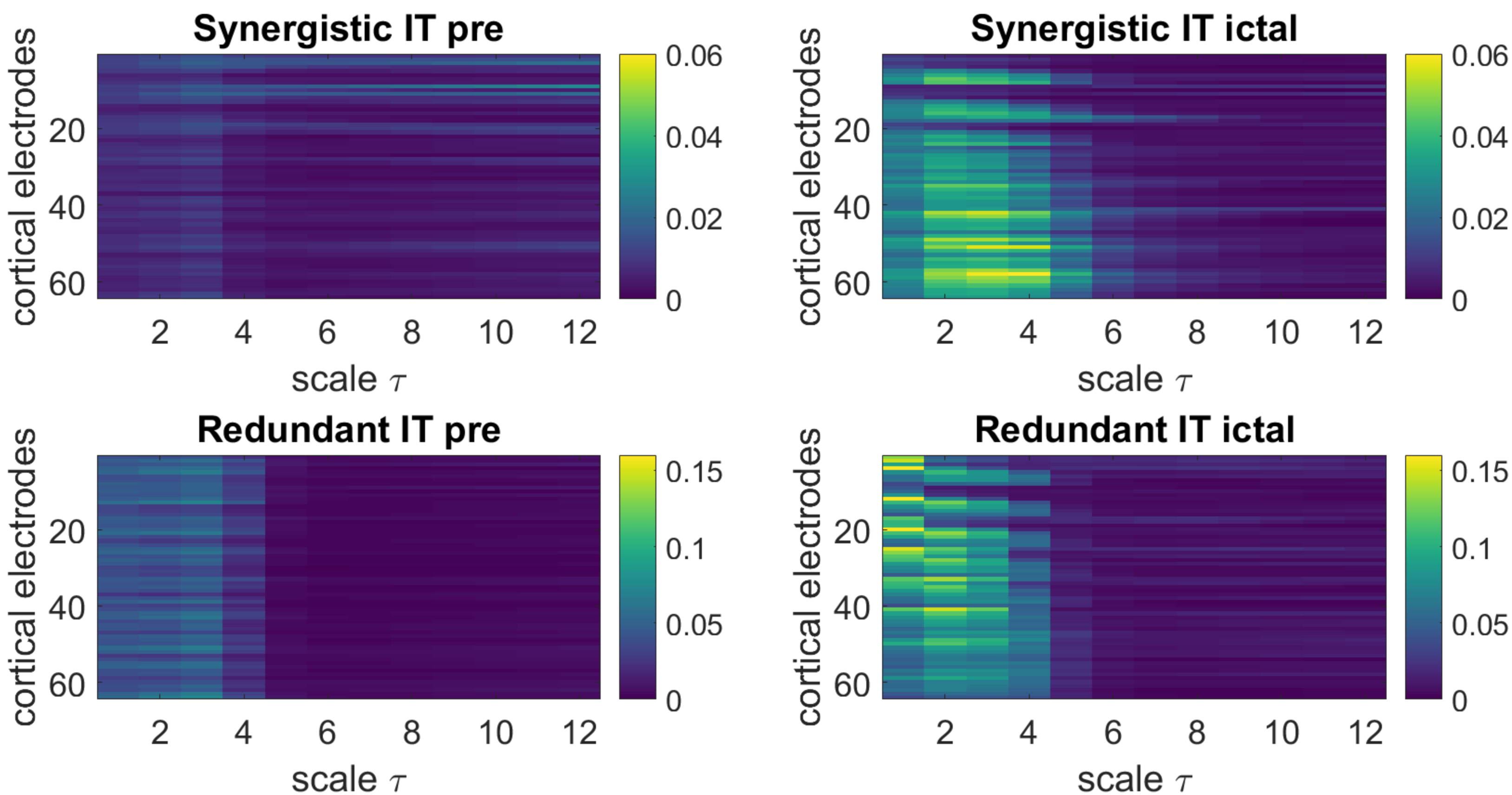


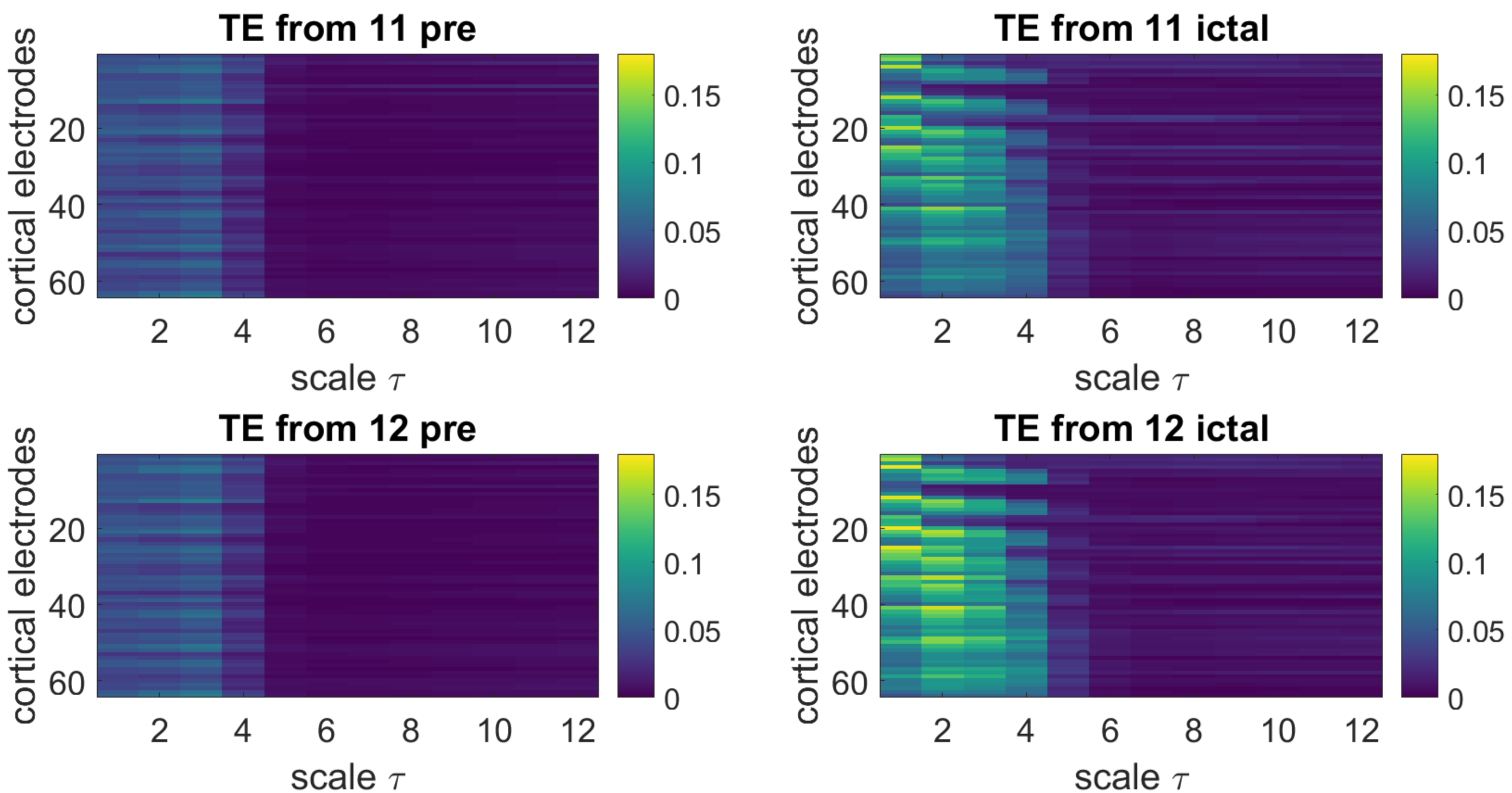
ECoG recording. An 8 × 8 electrode grid is placed directly on the cortical surface and recordings are made usually for several days or even weeks.

Kramer, M.A., Kolaczyk, E.D., and Kirsch, H.E. (2008). Emergent network topology at seizure onset in humans. Epilepsy Research,79, 173-186.

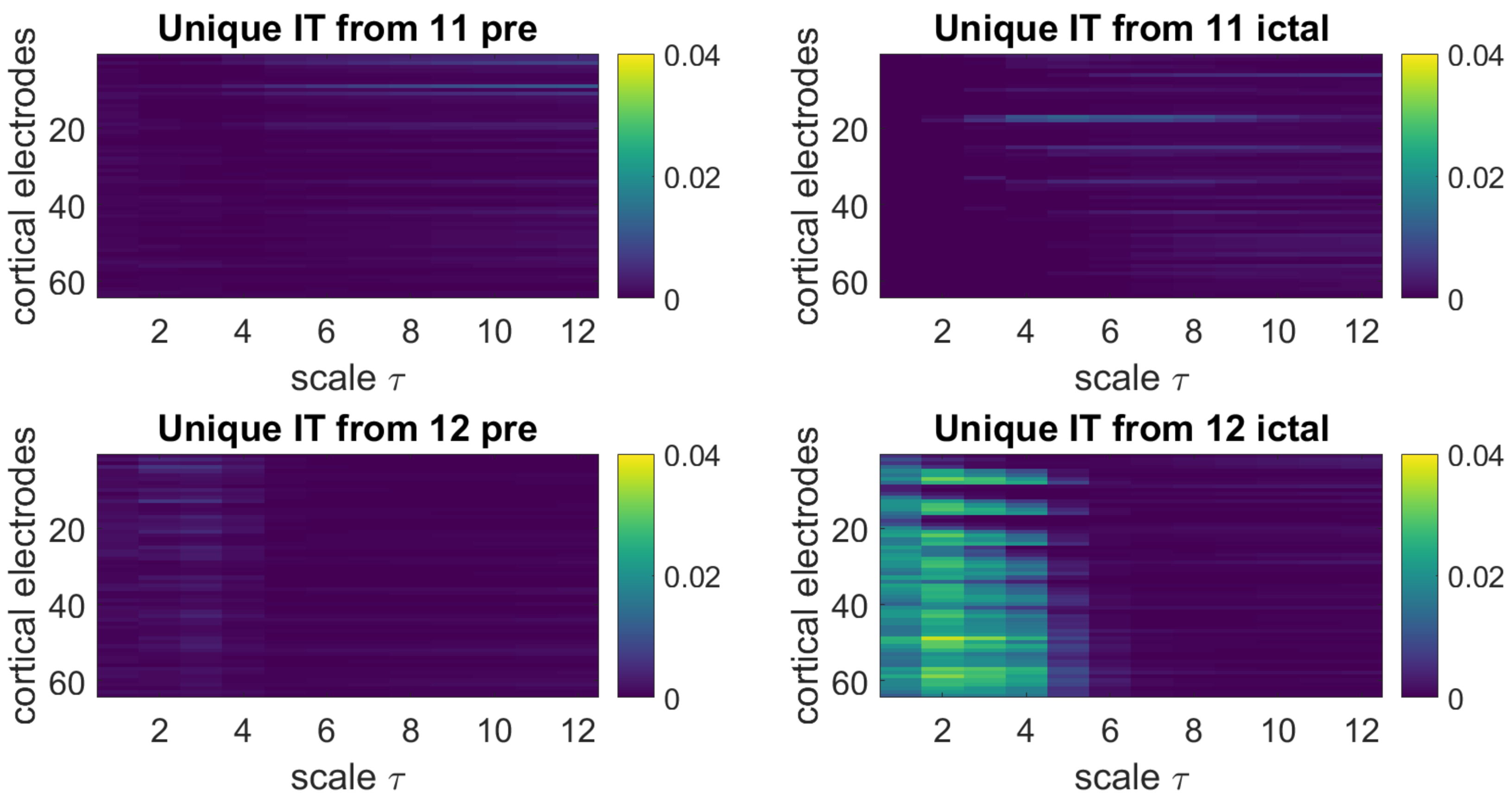
Drivers: depth electrodes 11 and 12. Target: Cortical electrodes




Synergistic IT pre
Synergistic IT ictal
Redundant IT pre
Redundant IT ictal
cortical electrodes
scale τ
20
40
60
2
4
6
8
10
12
0.06
0.04
0.02
0
0.15
0.1
0.05



TE from 11 pre
TE from 11 ictal
TE from 12 pre
TE from 12 ictal
cortical electrodes
scale τ
20
40
60
2
4
6
8
10
12
0
0.05
0.1
0.15


Unique IT from 11 pre
Unique IT from 11 ictal
Unique IT from 12 pre
Unique IT from 12 ictal
cortical electrodes
20
40
60
2
4
6
8
10
12
scale τ
0.04
0.02
0

Conclusions

Causal interactions occur at multiple scales and involve informational circuits of multiplets of variables.
New approaches are under development to identify redundant, synergetic and unique contributions to the total information flow

- Barrett 2015: Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems (PRE) https://arxiv.org/abs/1411.2832
- Faes et al. 2016: Predictability decomposition detects the impairment of brainheart dynamical networks during sleep disorders and their recovery with treatment (PTRS A) http://rsta.royalsocietypublishing.org/content/374/2067/20150177
- Stramaglia et al. 2014: Synergy and redundancy in the Granger causal analysis of dynamical networks (NJP) http://iopscience.iop.org/article/10.1088/1367-2630/16/10/105003
- Stramaglia et al. 2016: Synergetic and Redundant Information Flow Detected by Unnormalized Granger Causality (IEEE TBME) https://arxiv.org/abs/1504.03584
- Faes et al. 2017: Multiscale Information Decomposition: Exact Computation for Multivariate Gaussian Processes https://arxiv.org/abs/1706.07136