Brain connectivity analysis from neural data

SINC2 one-day workshop

Adrià Tauste-Campo
Andrea Insabato
Matthieu Gilson

Toy models:
http://github.com/cns-upf

Wifi: **event@upf.edu**
Password: **7fabrauni**

Outline

- Introduction
  - Define scientific question and context
  - Roadmap for connectivity measures
  - Value of model
- Linear analysis
  - Spectrum and statistical testing (null model)
  - Connectivity estimation in network (PC, MVAR)
- Non-linear analysis
  - Probabilistic modeling
  - Connectivity estimation using information-theoretic measures
- Machine learning
  - ML as an alternative to hypothesis testing
  - Classification
  - Evaluation of classifiers

Define a hypothesis that lead to a technical question

- Time will judge the validity of theories, models, etc.
- Replicability crisis

Reproducibility in Science

Improving the Standard for Basic and Preclinical Research

C. Glenn Begley, John P.A. Ioannidis

Abstract: **Medical and scientific advances are predicated on new knowledge that is robust and reliable and that serves as a solid foundation on which further advances can be built. In biomedical research, we are in the midst of a revolution with the generation of new data and scientific publications at a previously unprecedented rate. However, unfortunately, there is compelling evidence that the majority of these discoveries will not stand the test of time. To a large extent, this reproducibility crisis in basic and preclinical research may be as a result of failure to adhere to good scientific practice and the desperation to publish or perish. This is a multifaceted, multistakeholder problem. No single party is solely responsible, and no single solution will suffice. Here we review the reproducibility problems in basic and preclinical biomedical research, highlight some of the complexities, and discuss potential solutions that may help improve research quality and reproducibility. (*Circ Res*. 2015;116:116-126. DOI: 10.1161/CIRCRESAHA.114.303819.)**

Key Words: funding ■ journals ■ research integrity ■ universities

Define a hypothesis that lead to a technical question

- Time will judge the validity of theories, models, etc.
- Replicability crisis
- Measures on biological signals and interpretation:
  - Where is information?
  - Validation, Robustness
  - Simplicity

Define a hypothesis that lead to a technical question

- Time will judge the validity of theories, models, etc.
- Replicability crisis
- Measures on biological signals and interpretation:
  - Where is information? → underlying physics, neural correlates, variables of interest (single channel versus pairwise connection)
  - Validation, Robustness
  - Simplicity

Define a hypothesis that lead to a technical question

- Time will judge the validity of theories, models, etc.
- Replicability crisis
- Measures on biological signals and interpretation:
  - Where is information? → underlying physics, neural correlates, variables of interest (single channel versus pairwise connection)
  - Validation, Robustness → benchmarking with synthetic models, stability of measurement, statistics and null hypothesis (shuffling surrogates), etc.
  - Simplicity

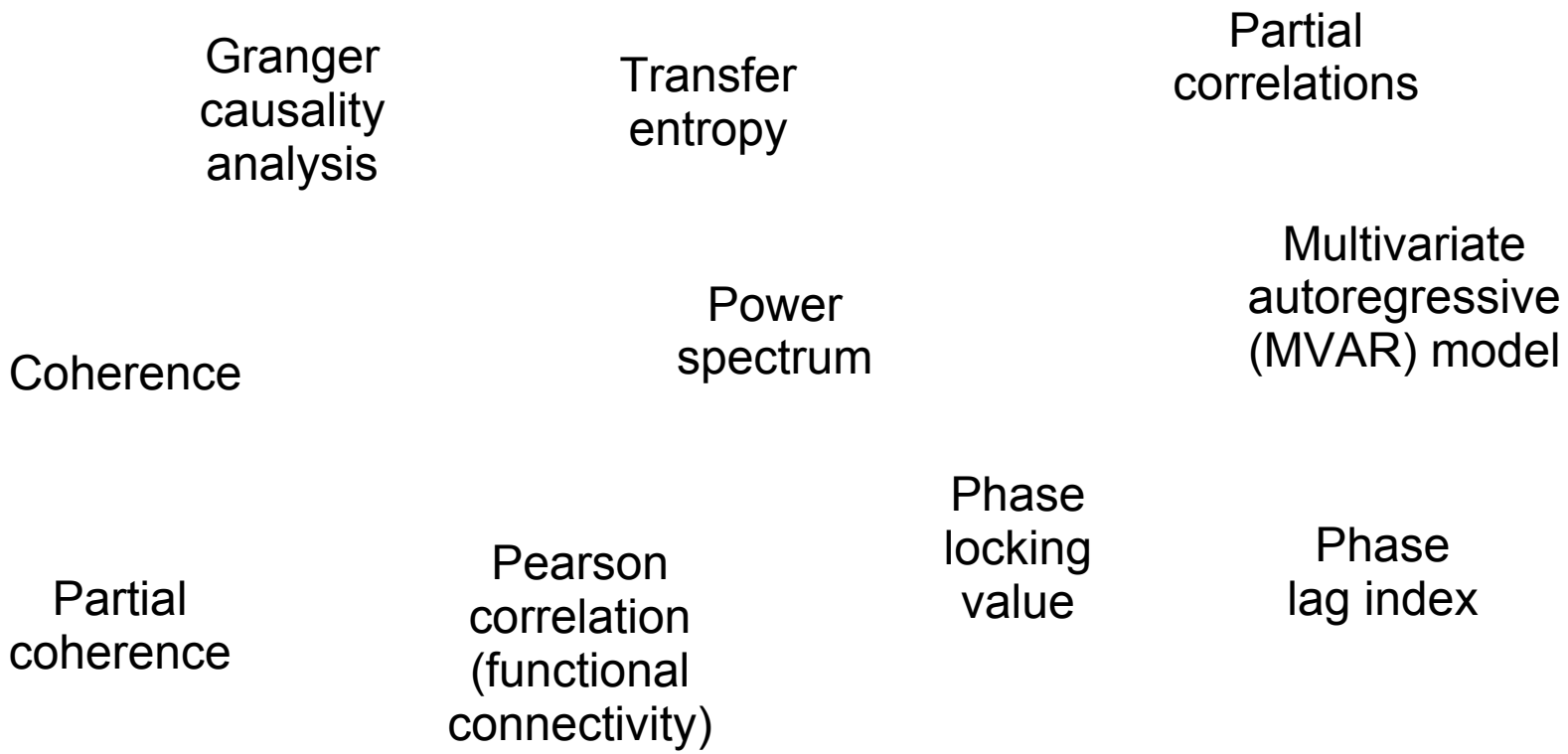
Define a hypothesis that lead to a technical question

- Time will judge the validity of theories, models, etc.
- Replicability crisis
- Measures on biological signals and interpretation:
  - Where is information? → underlying physics, neural correlates, variables of interest (single channel versus pairwise connection)
  - Validation, Robustness → benchmarking with synthetic models, stability of measurement, statistics and null hypothesis (shuffling surrogates), etc.
  - Simplicity → intuition provided by underlying model, adequate level of sophistication

Define a hypothesis that lead to a technical question

- Time will judge the validity of theories, models, etc.
- Replicability crisis
- Measures on biological signals and interpretation:
  - Where is information? → underlying physics, neural correlates, variables of interest (single channel versus pairwise connection)
  - Validation, Robustness → benchmarking with synthetic models, stability of measurement, statistics and null hypothesis (shuffling surrogates), etc.
  - Simplicity → intuition provided by underlying model, adequate level of sophistication
- Back to reality: limited data, uncontrolled (mixed) sources of noise, issues with stationarity of protocols/conditions

Roadmap / Typology



Roadmap / Typology

	Without time	Linear in time domain	Linear in frequency domain	Non-linear
Nodal measure	Variance	Auto-covariance	Power spectrum	
Connectivity measures and estimates	Pearson correlation Partial correlations	Cross-covariances MVAR Granger causality analysis	Coherence Partial coherence	Phase lag index Phase locking value Transfer entropy

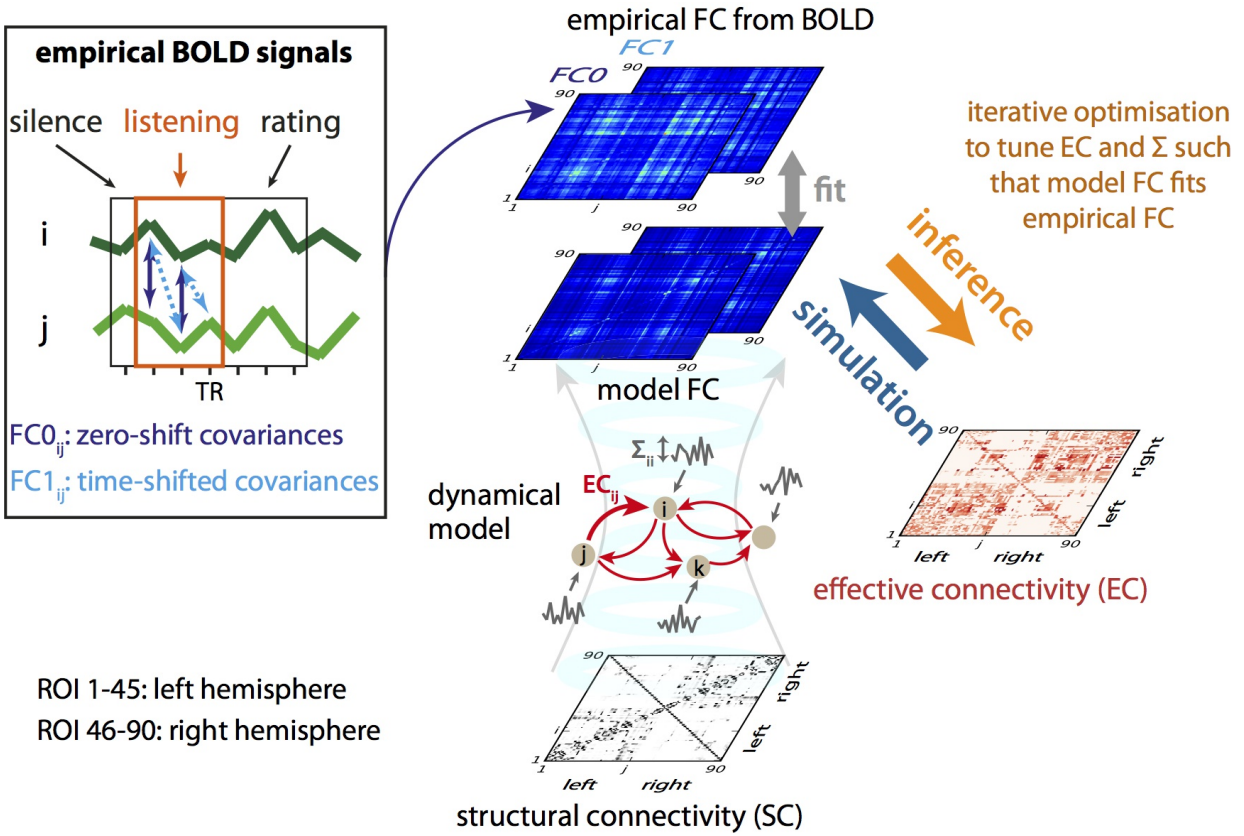
Roadmap / Typology

	Without time	Linear in time domain	Linear in frequency domain	Non-linear
Nodal measure	Variance	Auto-covariance	Power spectrum	
Connectivity measures and estimates	Pearson correlation Partial correlations	Cross-covariances MVAR Granger causality analysis	Coherence Partial coherence	Phase lag index Phase locking value Transfer entropy

There is always a model behind a measure (especially to interpret the values it takes), so better know the details about it!

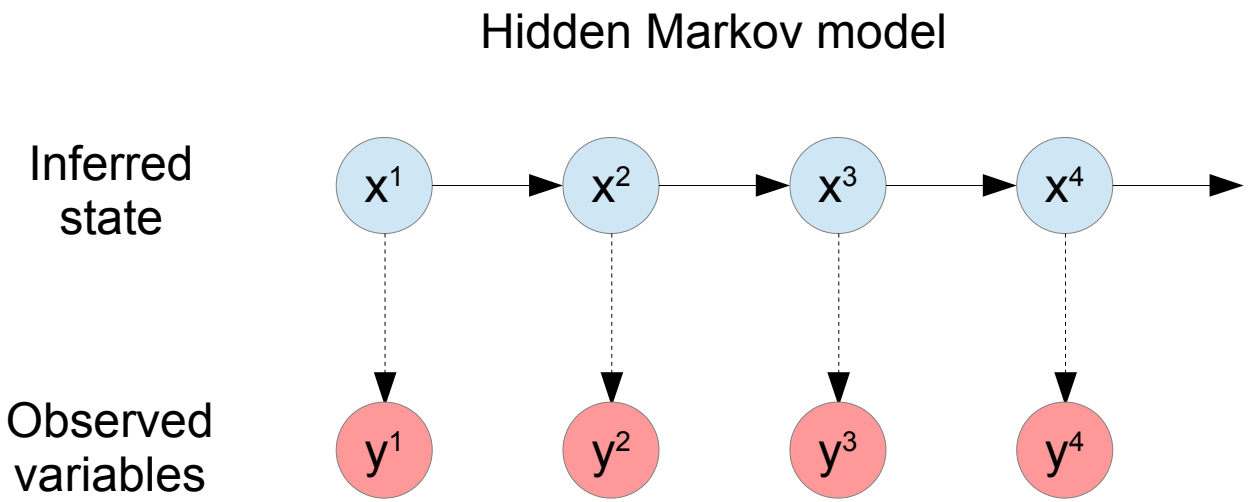
How to use a model?

- Exploration of concepts (qualitative approach)
- Fit and interpretation of data (quantitative approach)



How to use a model?

- Exploration of concepts (qualitative approach)
- Fit and interpretation of data (quantitative approach): investigate latent variables



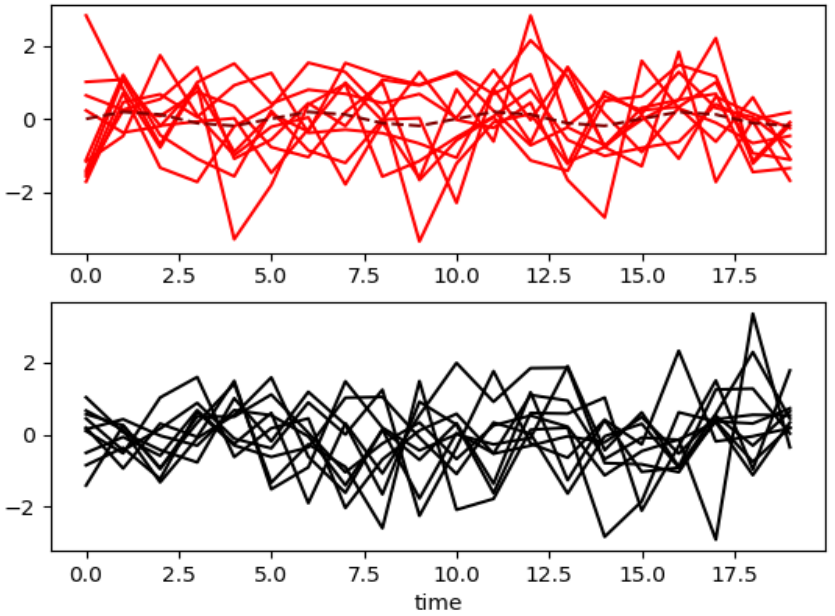
Outline

- Introduction
  - Define scientific question and context
  - Roadmap for connectivity measures
  - Value of model
- **Linear analysis**
  - **Spectrum and statistical testing (null model)**
  - Connectivity estimation in network (PC, MVAR)
- Non-linear analysis
  - Probabilistic modeling
  - Connectivity estimation using information-theoretic measures
- Machine learning
  - ML as an alternative to hypothesis testing
  - Classification
  - Evaluation of classifiers

Rhythm in time series

- http://github.com/cns-upf/test_modulation.py.ipynb

- Single time series for n subjects (behavioral data, electrophysiology, etc.)
- Is there a modulation in signal common to all subjects?



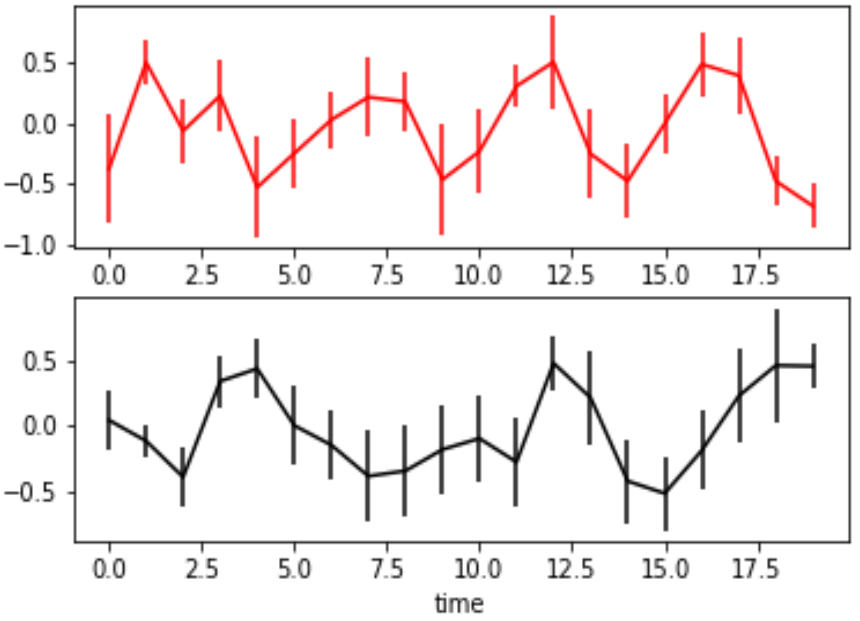
Here:
- n_sub = 10 subjects
- Weak amplitude a = 0.2

Random (white noise)

Rhythm in time series

- http://github.com/cns-upf/test_modulation.py.ipynb

- Single time series for n subjects (behavioral data, electrophysiology, etc.)
- Is there a modulation in signal common to all subjects?



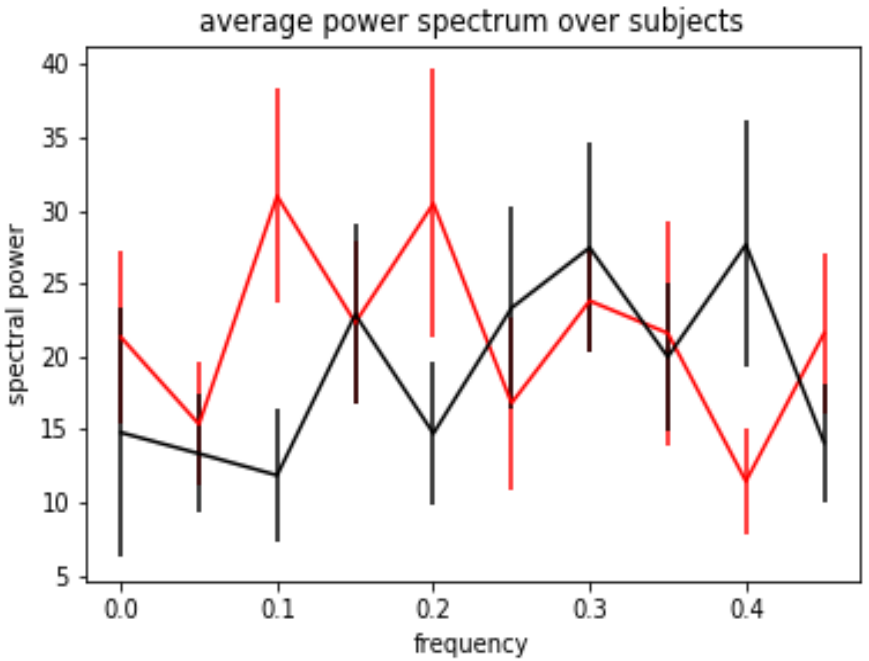
Here:
- n_sub = 10 subjects
- Weak amplitude a = 0.2

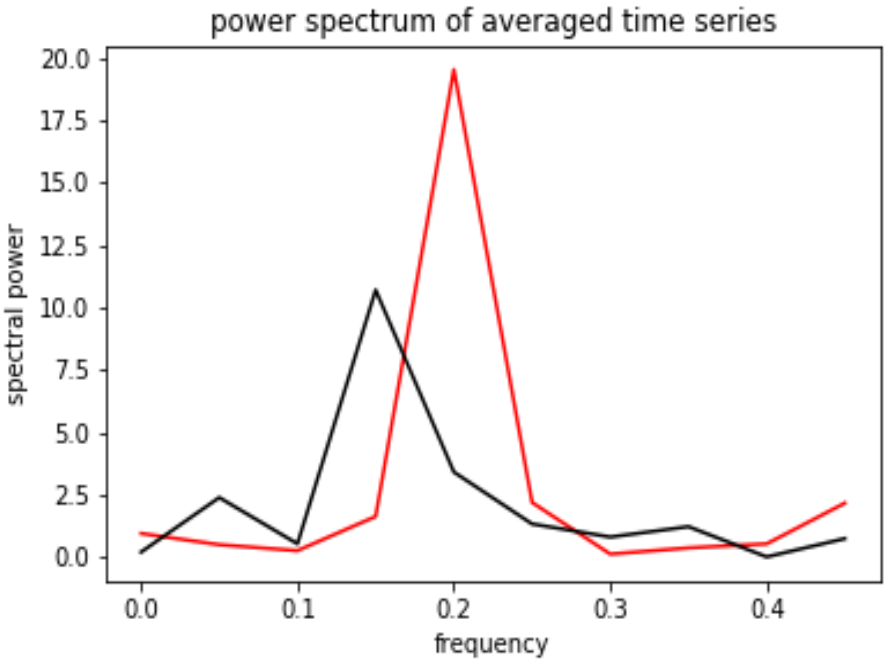
Random (white noise)

Rhythm in time series

- http://github.com/cns-upf/test_modulation.py.ipynb

- Power spectrum to investigate rhythms
- Average before or after Fourier transform?

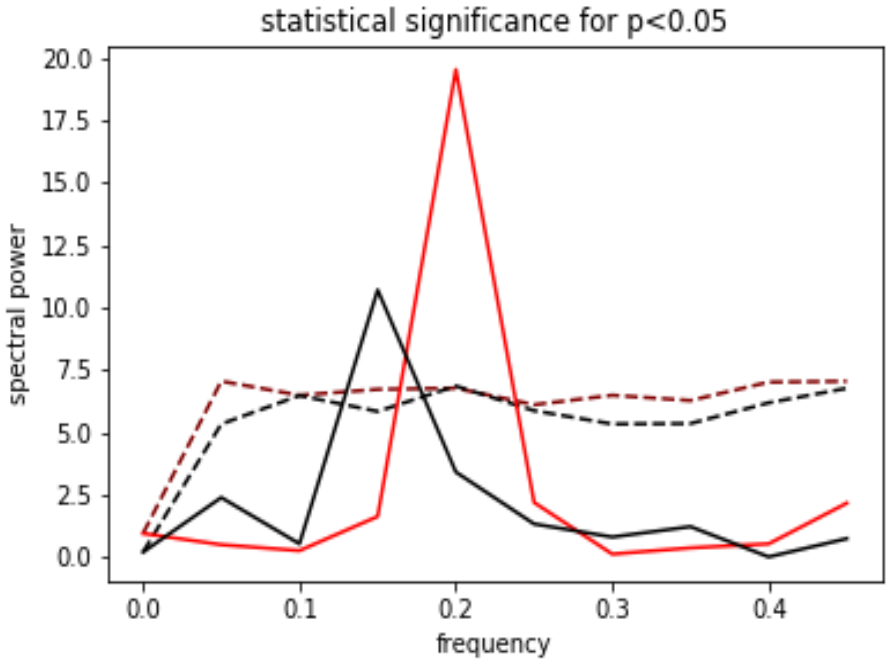


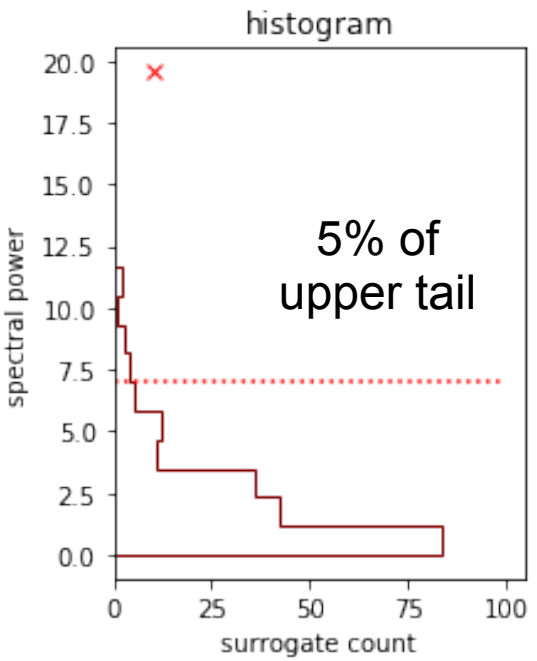


Rhythm in time series

- http://github.com/cns-upf/test_modulation.py.ipynb

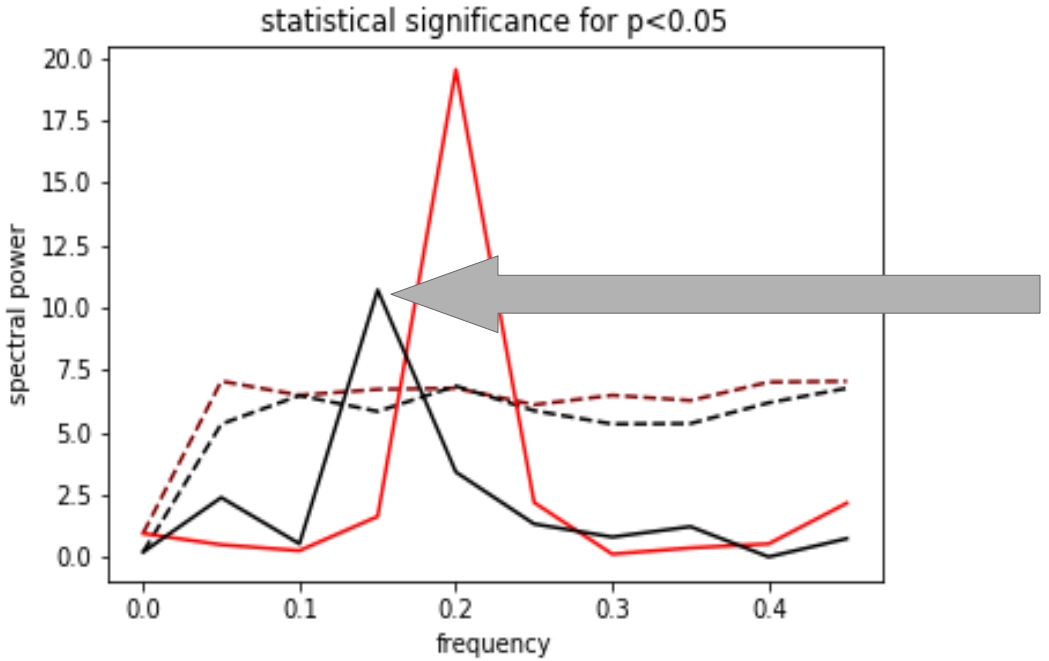
- Significance test for spectrum of average time series
- Generate surrogates by shuffling data for each subject (then averaging...)





Rhythm in time series

- http://github.com/cns-upf/test_modulation.py.ipynb
- Define threshold for multiple comparisons?



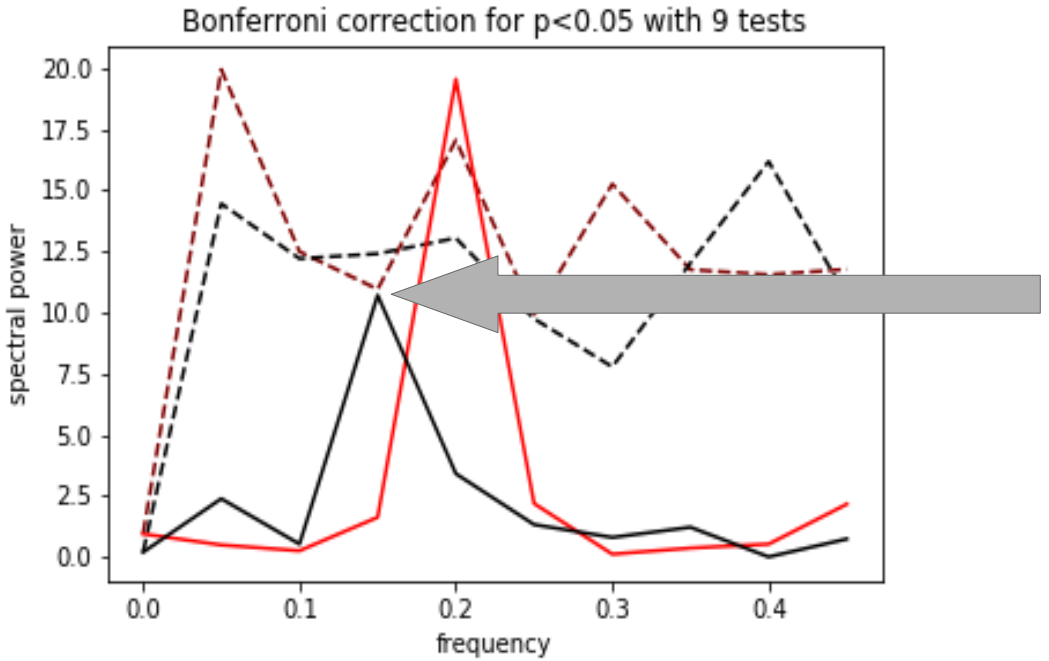
False positive (5% allowed here)

Here 9 tests for
all frequencies

Rhythm in time series

- http://github.com/cns-upf/test_modulation.py.ipynb

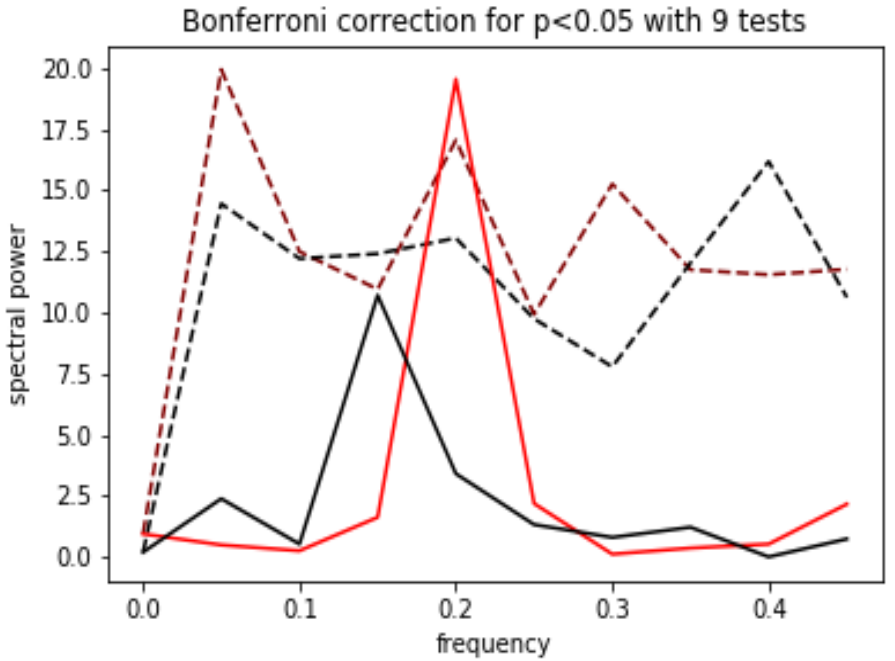
- Bonferroni correction (conservative threshold)
- Assuming independent tests for all 9 frequencies



p-value = 0.05 / 9 = 0.0056

Rhythm in time series

- http://github.com/cns-upf/test_modulation.py.ipynb
- What is your question?
  - Is there a modulation at any frequency?
  - Do you want to test for a specific frequency?



Play with:

- signal duration ***T***
- modulation amplitude ***a***
- number of subjects ***n_sub***
- number of surrogate ***n_shuf***
- surrogate type
- desired p-value ***pval***

Outline

- Introduction
  - Define scientific question and context
  - Roadmap for connectivity measures
  - Value of model
- **Linear analysis**
  - Spectrum and statistical testing (null model)
  - **Connectivity estimation in network (PC, MVAR)**
- Non-linear analysis
  - Probabilistic modeling
  - Connectivity estimation using information-theoretic measures
- Machine learning
  - ML as an alternative to hypothesis testing
  - Classification
  - Evaluation of classifiers

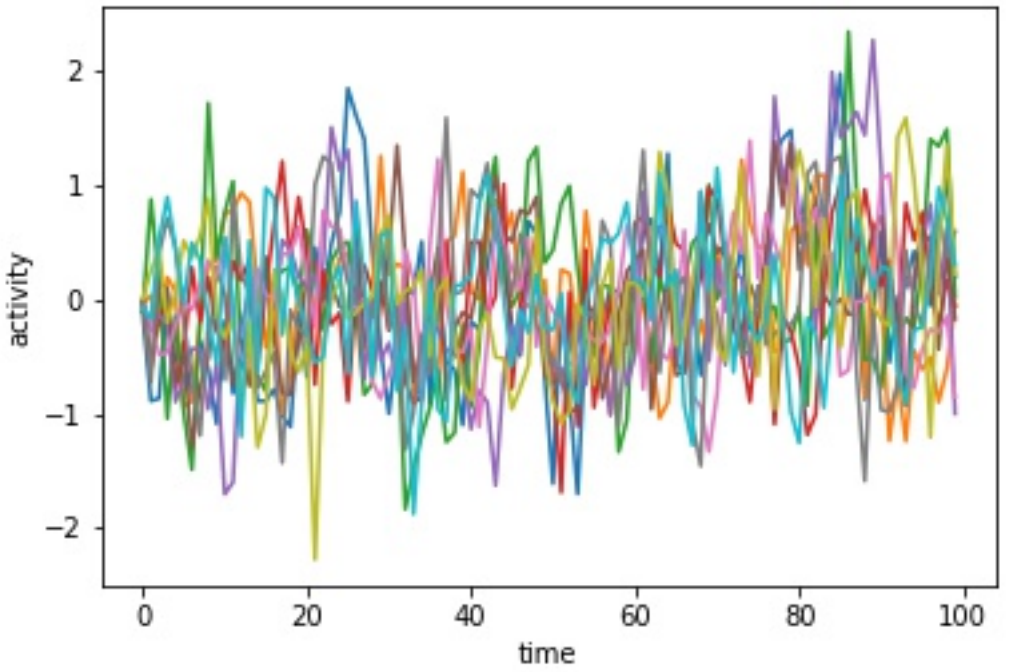
Connectivity estimation in network

- Is there temporal information in the data?
- Stationarity? On which time scale?
- Does directionality matter?

	Without time	Linear in time domain	Linear in frequency domain	Non-linear
Nodal measure	Variance	Auto-covariance	Power spectrum	
Connectivity measures and estimates	Pearson correlation Partial correlations	Cross-covariances MVAR Granger causality analysis	Coherence Partial coherence	Phase lag index Phase locking value Transfer entropy

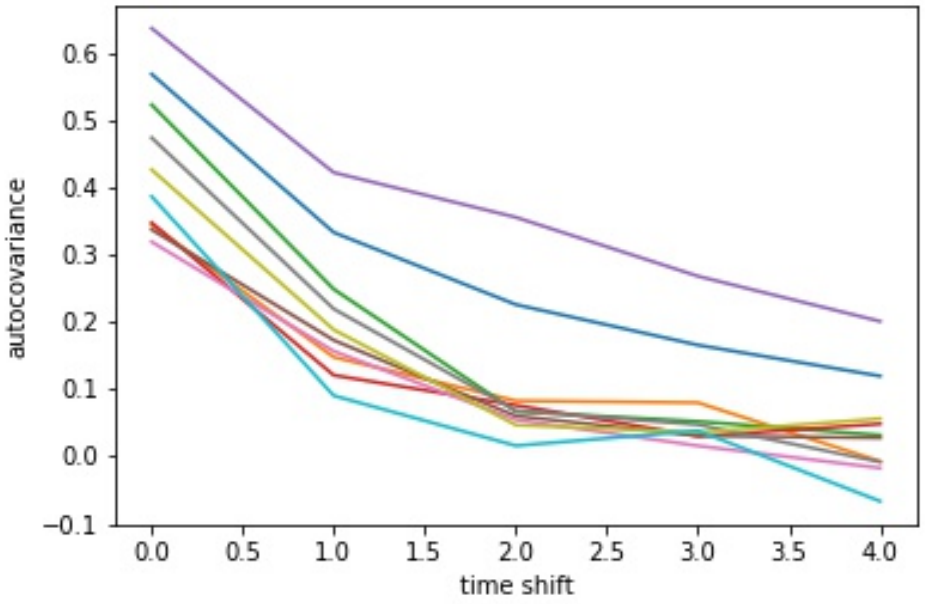
Connectivity estimation in network

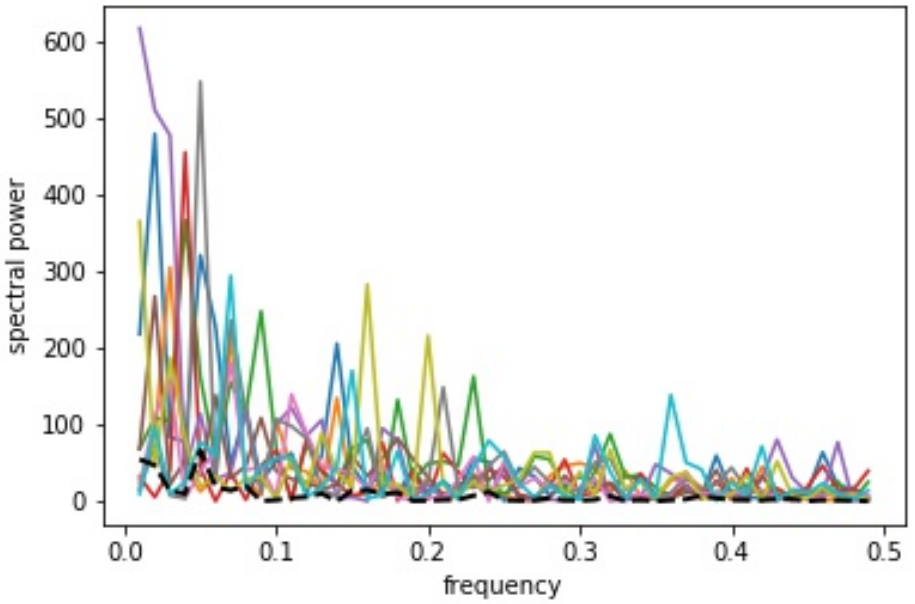
- Is there temporal information in the data?
- Stationarity? On which time scale?
- Does directionality matter?



Connectivity estimation in network

- Is there temporal information in the data?
- Stationarity? On which time scale?
- Does directionality matter?





Connectivity estimation in network

- Is there temporal information in the data?
- Stationarity? On which time scale?
- Does directionality matter?

	Without time	Linear in time domain	Linear in frequency domain	Non-linear
Nodal measure	Variance	Auto-covariance	Power spectrum	
Connectivity measures and estimates	Pearson correlation Partial correlations	Cross-covariances MVAR Granger causality analysis	Coherence Partial coherence	Phase lag index Phase locking value Transfer entropy

Connectivity estimation in network

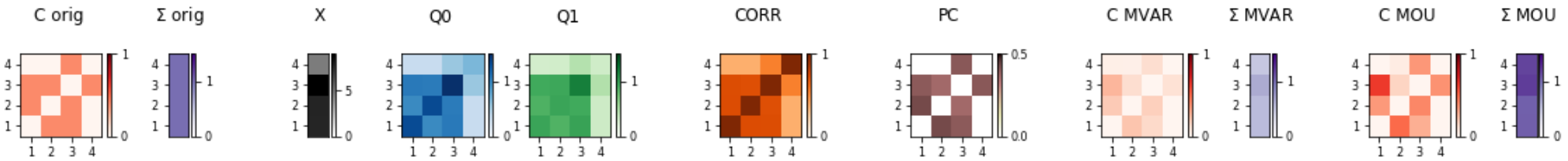
- Is there temporal information in the data?
- Stationarity? On which time scale?
- Does directionality matter?

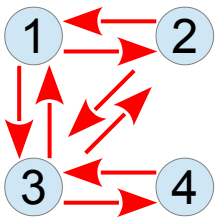
	Without time	Linear in time domain	Linear in frequency domain	Non-linear
Nodal measure	Variance	Auto-covariance	Power spectrum	
Connectivity measures and estimates	Pearson correlation Partial correlations	Cross-covariances MVAR Granger causality analysis	Coherence Partial coherence	Phase lag index Phase locking value Transfer entropy

Connectivity estimation in network: Linear methods in time domain

- http://github.com/cns-upf/estimate_directed_connectivity.ipynb

- Pearson correlation
- Partial correlation (related to graphical model)
- Multivariate autoregressive model (MVAR)
- Multivariate Ornstein-Uhlenbeck (MOU)





Signal generated using the MOU dynamics (continuous time), observations involve down-sampling

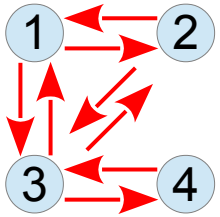
MVAR $x^{t+1} = A\,x^t + \zeta$

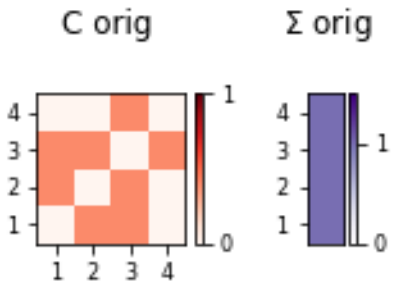
MOU $dx^t = J\,x^t\,dt + dW^t$

Connectivity estimation in network: Linear methods in time domain

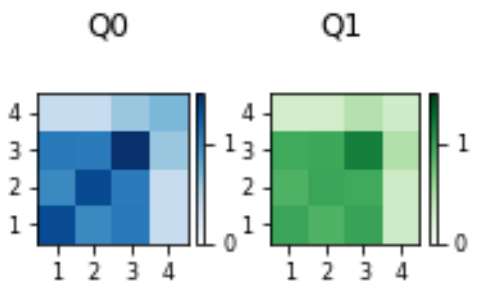
- http://github.com/cns-upf/estimate_directed_connectivity.ipynb

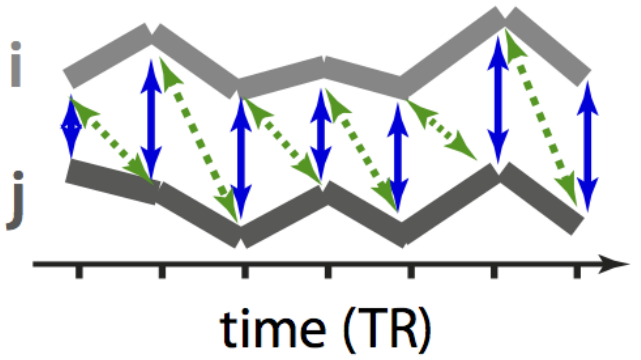
Original network





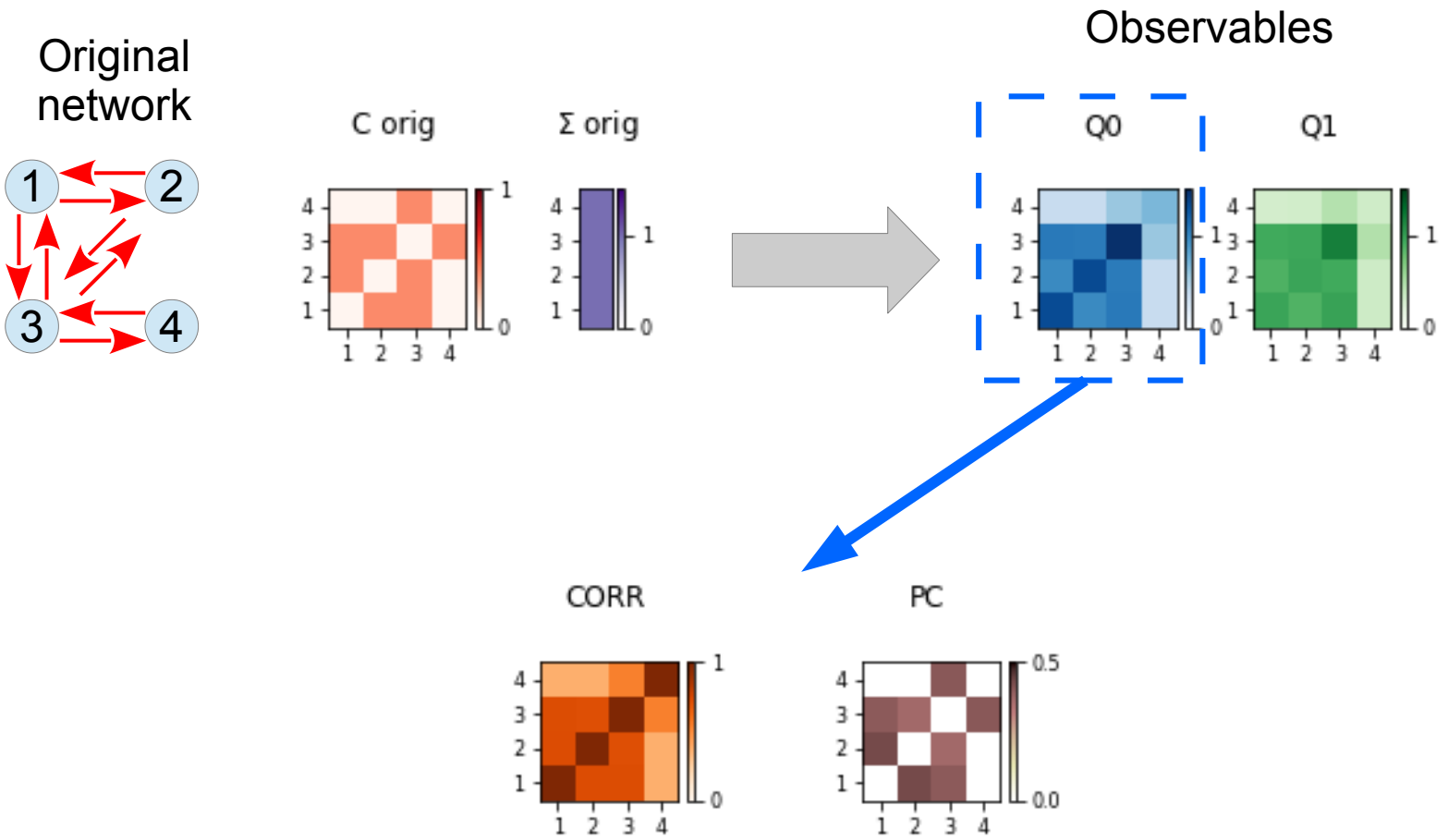
Observables





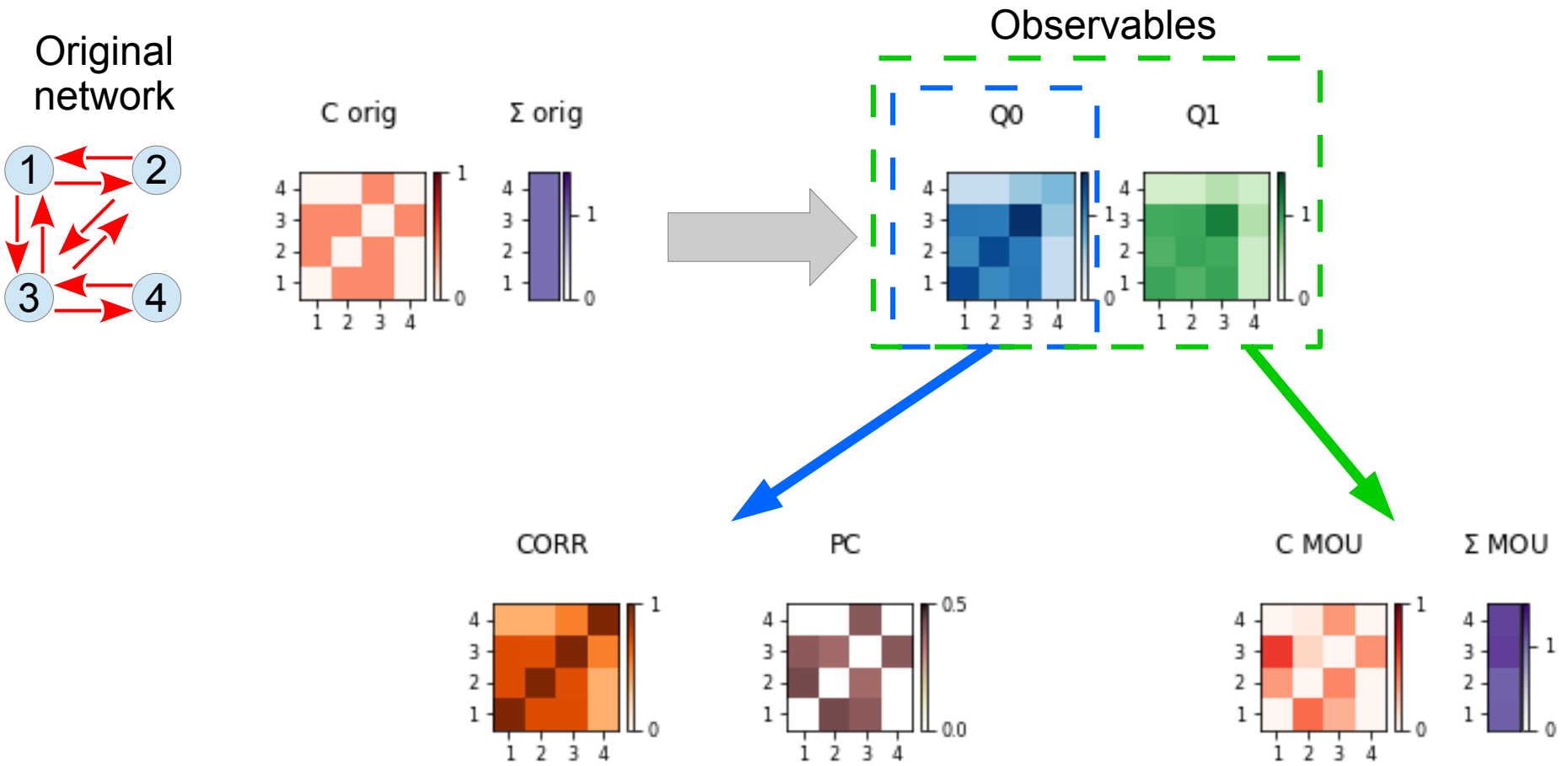
Connectivity estimation in network: Linear methods in time domain

- http://github.com/cns-upf/estimate_directed_connectivity.ipynb



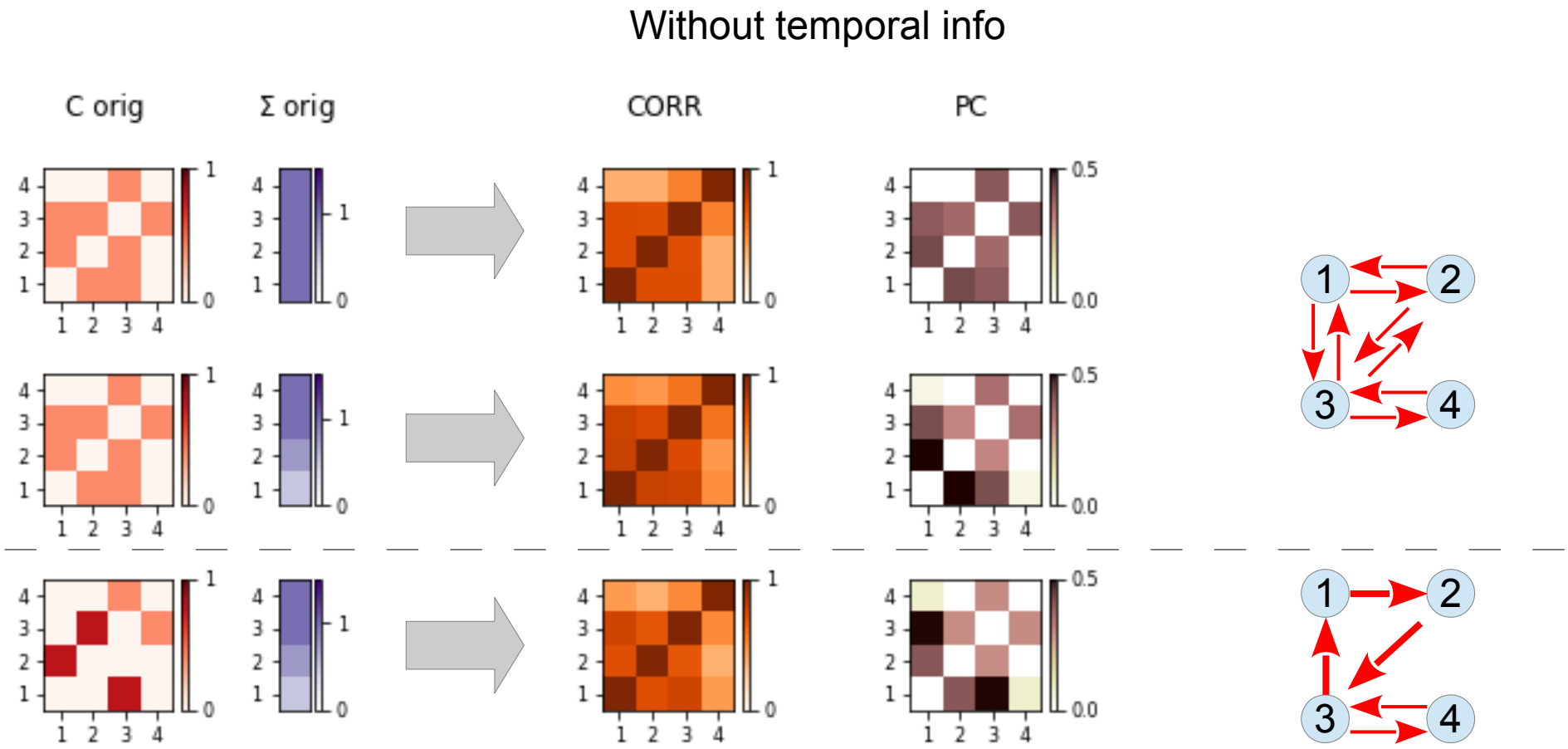
Connectivity estimation in network: Linear methods in time domain

- http://github.com/cns-upf/estimate_directed_connectivity.ipynb



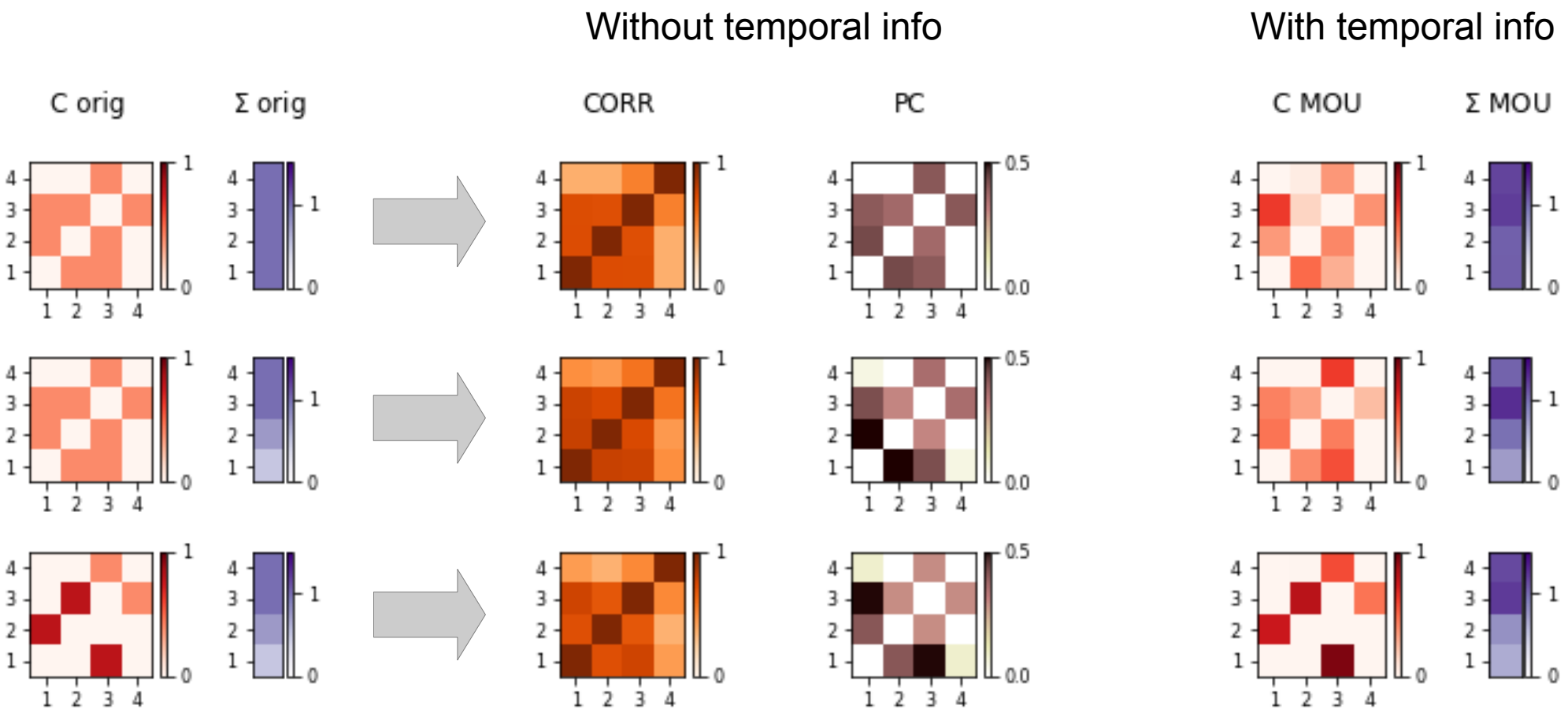
Connectivity estimation in network: Linear methods in time domain

- http://github.com/cns-upf/estimate_directed_connectivity.ipynb



Connectivity estimation in network: Linear methods in time domain

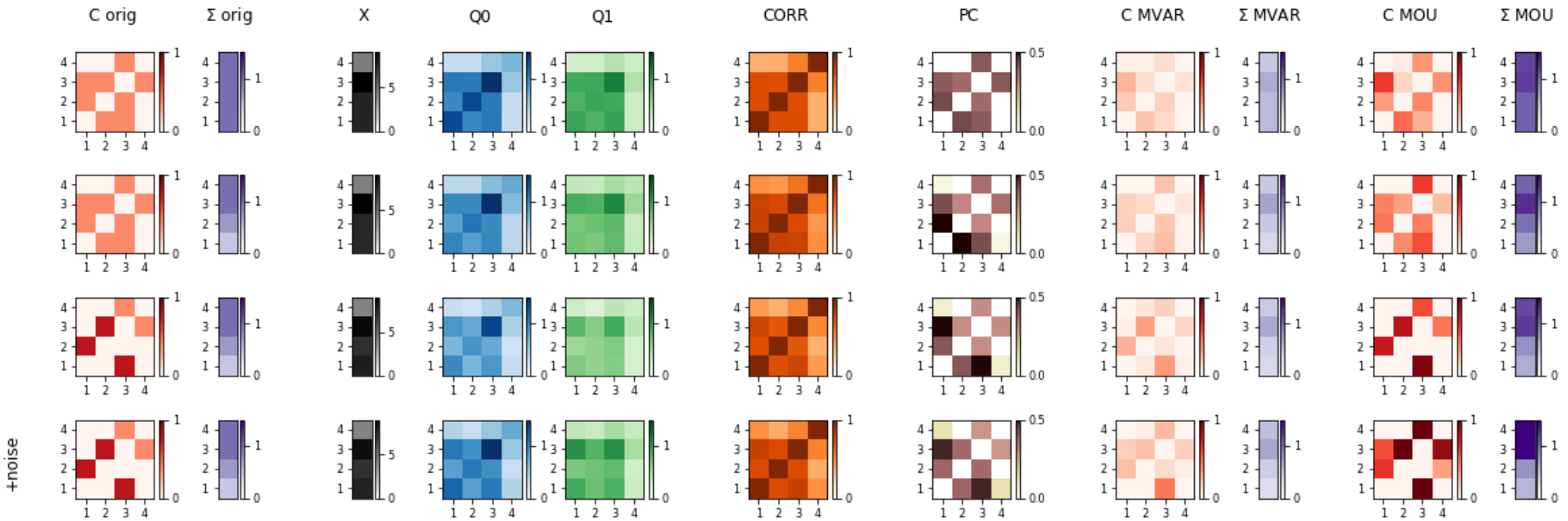
- http://github.com/cns-upf/estimate_directed_connectivity.ipynb



Connectivity estimation in network: Linear methods in time domain

- http://github.com/cns-upf/estimate_directed_connectivity.ipynb

- Play with parameters of original network, observation noise, etc.



Connectivity estimation in large networks

- Toy models: (http://github.com/cns-upf)
  - test_modulation
  - estimate_directed_connectivity
  - fMRI data: http://github.com/MatthieuGilson/WBLEC_toolbox
- Linear formalism is often the only computationally-feasible approach for large networks (>20 nodes)
  - Spectral domain
  - All nodes have same type of signals

Connectivity estimation in large networks

- Toy models: (http://github.com/cns-upf)
  - test_modulation
  - estimate_directed_connectivity
  - fMRI data: http://github.com/MatthieuGilson/WBLEC_toolbox
- Linear formalism is often the only computationally-feasible approach for large networks (>20 nodes)
  - Spectral domain
  - All nodes have same type of signals
- Non-parametric testing to avoid assumption that observed signals are Gaussian
  - Statistical testing for MVAR using surrogates (Gilson, Tauste Campo et al. Net Neurosci 2017)
- Heterogeneous signals
  - behavioral data and neural signals (with distinct timescales)
- Multiple comparisons (N nodes → N^2 connections): What type of correction to apply? Independent tests?
  - Detection of effect for connection clusters instead of individual connections
  - Machine learning to measure effect size