MULTI-SCALE FACTOR ANALYSIS OF HIGH DIMENSIONAL TIME SERIES

Hernando Ombao

University of California at Irvine
http://ucispacetime.wix.com/spacetime

With Yuxiao Wang (UC Irvine) and Chee-Ming Ting (Univ Teknologi Malaysia)

February 1, 2016

Outline of Talk

Outline of Talk

1. Motivation

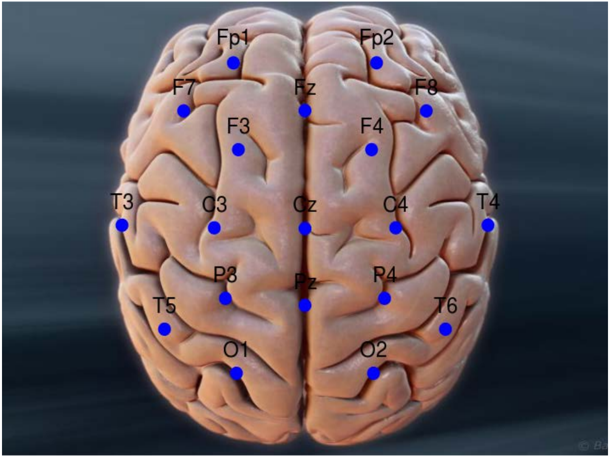
2. Multi-Scale Factor Analysis

Scientific Motivation

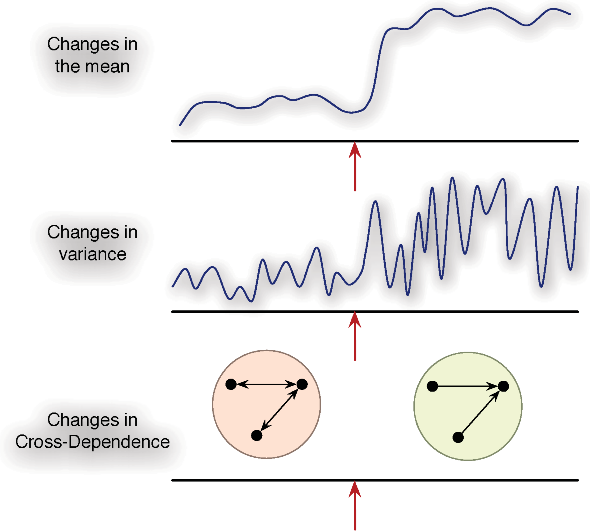
Electrophysiological Signals

- Indirect recordings: Electroencephalograms (EEGs) are recordings on the scalp that capture coordinated activity of cortical neurons
- Direct recordings: Local field potentials (LFPs) are direct recordings from a localized population of neurons
- These recordings consist of waveforms oscillating at different frequency bands

Scientific Motivation



From Stimulus to Neurons to Behavior
The Big Picture

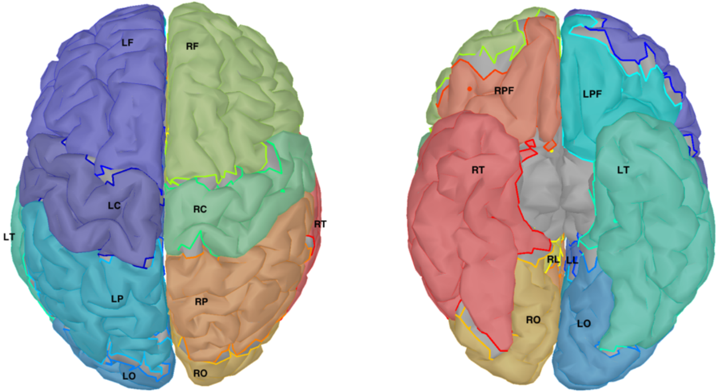


Scientific Questions on Connectivity

- Brain regions do not act in isolation. Rather brain regions act in a cooperative manner during a cognitive task.
- Does activity in one brain region excite or inhibit another?
- Are there experimental conditions or cognitive processes that require less or greater connectivity?
- Does connectivity evolve across time; vary across subjects?
- Important considerations
  - Scalability of statistical models
  - Interpretability of statistical output
  - Visualization!

Multi-scale Factor Analysis

Brain Parcellation



Multi-scale Factor Analysis
Factor Model within a region

Why multi-scale?

- Local: connectivity within a region
  - High correlation between voxels/channels in a region
  - Suggests a lower-dimensional representation of activity in each region
- Global: connectivity between regions in a network

Multi-scale Factor Analysis
Factor Model within a region

- Here, a "region" ($r = 1, 2, \ldots, R$) is either
  - An actual region on the cortex with anatomically defined boundaries
  - A set of channels that project directly from a cortical patch
- Activity at region r
  - $\mathbf{Z}_r(t) \in \mathbb{R}^{n_r}$
  - Channels or voxels within region r are typically highly correlated
  - $\mathrm{Dim}[\mathbf{Z}_r(t)] = n_r$ (number of channels/voxels)
- Factors driving activity at region r
  - $\mathbf{f}_r(t) \in \mathbb{R}^{m_r}$
  - $\mathrm{Dim}[\mathbf{f}_r(t)] = m_r$
- Dimension reduction: $m_r < n_r$

Multi-scale Factor Analysis

Factor Model within a region

Activity in a region r, denoted $\mathbf{Z}_r(t)$, is either

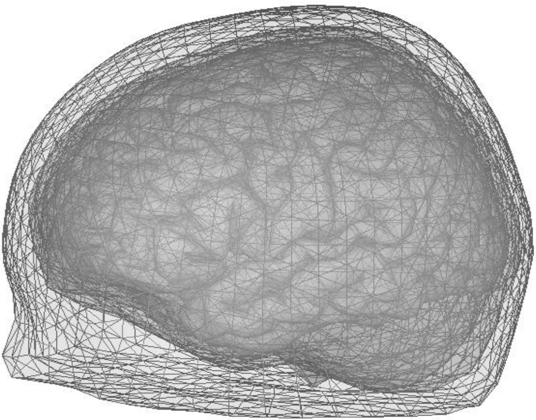
- Directly observed: EEG from different channels – (today's example)
- Estimated via source localization methods [Wang, Ting and Ombao (2016)]
  - standardized Low Resolution Brain Electromagnetic Tomography (sLORETA) by Pascual et al (2002)
  - Given: lead field matrix $\mathbb{A}$, EEG $\mathbf{X}$ and penalty parameter α
  - $\mathbf{Z}(t)$ minimizes the penalized criterion

$$F = \sum_{t=1}^{T} (\|\mathbf{X}(t) - \mathbb{A}\mathbf{S}(t) - c\mathbf{1}\|^2 + \alpha\|\mathbf{S}(t)\|^2)$$

  - $\mathbf{Z}(t) = \mathbf{G}\mathbf{X}(t)$, where
    - $\mathbf{G} = \mathbb{A}'\mathbf{H}(\mathbf{H}\mathbb{A}\mathbb{A}'\mathbf{H} + \alpha\mathbf{H})^{-}$
    - $\mathbf{H} = \mathbf{I} - \mathbf{1}\mathbf{1}'/\mathbf{1}'\mathbf{1}$.

Multi-scale Factor Analysis

Head model for source estimation



The BEM head model

Multi-scale Factor Analysis

Factor model within a region

- Model for activity in region r

$$\mathbf{Z}_r(t) \approx \mathbf{Q}_r \mathbf{f}_r(t)$$

- Loading matrix $\mathbf{Q}_r$ - captures dependence between channels/voxels in a region
- Identifiability constraints
  - $\mathbf{Q}_r' \mathbf{Q}_r = \mathbf{I}_{m_r}$
  - $\mathbb{C}\mathrm{ov}(\mathbf{f}_r(t))$ is diagonal matrix with distinct elements

Multi-scale Factor Analysis
Factor model for all regions

$$\begin{bmatrix} \mathbf{Z}_1(t) \\ \mathbf{Z}_2(t) \\ \dots \\ \mathbf{Z}_R(t) \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1 & 0 & \dots & 0 \\ 0 & \mathbf{Q}_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \mathbf{Q}_R \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(t) \\ \mathbf{f}_2(t) \\ \dots \\ \mathbf{f}_R(t) \end{bmatrix}$$

$$\mathbf{Z}(t) = \mathbf{Q}\mathbf{f}(t)$$

NOTE: Factor components within a region are uncorrelated; but correlated with other regions.

Multi-scale Factor Analysis
VAR Model for the Factors

$$\mathbf{Z}(t) = \mathbf{Q}\mathbf{f}(t)$$
$$\mathbf{f}(t) = \sum_{\ell=1}^{P} \Phi^{\mathrm{f}}(\ell)\mathbf{f}(t-\ell) + \eta(t)$$

- $\Phi^{\mathrm{f}}(\ell)$ is the VAR coefficient matrix at lag ℓ
- $\eta(t) \sim N(\underline{0}, \Sigma_\eta)$.
- Dependence between cortical sources $\mathbf{Z}_r(t)$ and $\mathbf{Z}_s(t)$
  - Characterized by the dependence between the factors $\mathbf{f}_r(t)$ and $\mathbf{f}_s(t)$
  - Dimension reduction: $(n_r \times n_r)$ vs. $P(m_r \times m_r)$
    - EEG: $n_r = 256$ vs $m_r = 3$
    - Source estimates: $n_r = 10K$ vs $m_r \approx 10$

Multi-scale Factor Analysis

VAR Model for all regions

$$\mathbf{Z}(t) = \mathbf{Q}\mathbf{f}(t) = \sum_{\ell=1}^{P} \mathbf{Q}\Phi^{\mathrm{f}}(\ell)\mathbf{f}(t-\ell) + \mathbf{Q}\eta(t)$$

$$\mathbf{Z}(t) = \sum_{\ell=1}^{P} \Phi^{\mathbf{Z}}(\ell)\mathbf{Z}(t-\ell) + \mathbf{E}(t)$$

where

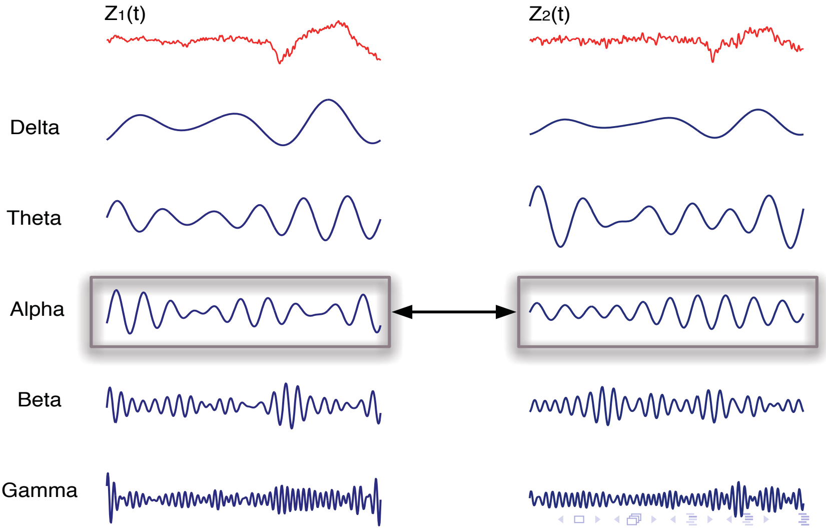
- $\Phi^{\mathbf{Z}}(\ell) = \mathbf{Q}\Phi^{\mathrm{f}}(\ell)\mathbf{Q}'$
- $\mathbf{E}(t) = \mathbf{Q}\eta(t)$

Dimension in reduction

- $\mathrm{Dim}[\mathbf{Z}(t)] = n$; $\mathrm{Dim}[\mathbf{f}(t)] = m \quad m \ll n$
- $\mathrm{Dim}[\Phi^{\mathbf{Z}}(\ell)] = n^2$; $\mathrm{Dim}[\Phi^{\mathrm{f}}(\ell)] = m^2$ for each lag ℓ
- $P \times (m^2) \times R \quad \ll \quad P \times (n^2)$

Multi-scale Factor Analysis

Spectral measures of connectivity



Multi-scale Factor Analysis
Spectral measures of connectivity

Between a pair of chanels or voxels v_1 and v_2

- Undirected coherence
- Undirected partial coherence
- Partial directed coherence (PDC)

Between a pair of regions r and r'

- Summrized connectivity between all pairs of channels/voxels (v_1, v_2) where v_1 is in region r and v_2 is in region r'

MULTI-SCALE FACTOR ANALYSIS
MEASURES OF CONNECTIVITY

Partial Directed Coherence between channels/voxels v_1 and v_2

- Recall Φ_ℓ^Z is the VAR coefficient matrix at lag ℓ;
- Define $\Phi^Z(\omega) = \mathbf{I} - \sum_{\ell=1}^{P} \Phi_\ell^Z \exp(-i2\pi\omega\ell/\Omega_s)$
    - Fourier transform of Φ^Z at frequency ω, where Ω_s is the sampling frequency.

Partial directed coherence (PDC) between voxels v_1 and v_2 is

$$\pi_{12}^2(\omega) = \frac{|\Phi_{12}^Z(\omega)|^2}{\sum_{k=1}^{N} \Phi_{k2}^Z(\omega)(\Phi_{k2}^Z(\omega))^*}$$

Multi-scale Factor Analysis

Measures of Connectivity

Partial Directed Coherence between
Remarks.

- $\pi_{12}^2(\omega)$ provides a measure of the linear influence of Z_2 on Z_1 at frequency ω
- Interpretation: direct impact of a change in the amplitude of ω-oscillatory activity in voxel v_2 on the amplitude of the ω-oscillation on v_1

Multi-scale Factor Analysis
Measures of Connectivity

Undirected Coherence between channels/voxels v_1 and v_2

- Cross-spectral matrix of the global factor $\mathbf{f}(t)$ at frequency ω

$$S^{\text{f}}(\omega) = H^{\text{f}}(\omega)\Sigma_{\eta}(H^{\text{f}}(\omega))^{*} \text{ where}$$

  - $H^{\text{f}}(\omega) = (\Phi^{\text{f}}(\omega))^{-1}$
  - $\Phi^{\text{f}}(\omega) = \mathbf{I} - \sum_{\ell=1}^{P} \Phi^{\text{f}}(\ell) \exp(-i2\pi\omega\ell/\Omega_s)$.

- Cross-spectral matrix of $\mathbf{Z}(t)$ at frequency ω

$$S^{Z}(\omega) = \mathbf{Q}S^{\text{f}}(\omega)\mathbf{Q}'$$

- (Undirected) coherence between channel/voxel v_1 and v_2 is

$$\rho_{12}^{2} = \frac{|S_{12}^{Z}(\omega)|^{2}}{S_{11}^{Z}(\omega)S_{22}^{Z}(\omega)}$$

Multi-scale Factor Analysis
Measures of Connectivity

Undirected **partial** Coherence between channels/voxels v_1 and v_2

- $P(\omega) = \left[S^{Z}(\omega)\right]^{-1}$
- $D(\omega) = \text{diag}[P_{11}^{-\frac{1}{2}}(\omega), \ldots, P_{nn}^{-\frac{1}{2}}(\omega)]$
- $\Lambda(\omega) = -D(\omega)P(\omega)D(\omega)$
- Partial coherence between channel/voxel v_1 and v_2 is $\Lambda_{v_1 v_2}$
- To estimate $P(\omega)$
  - via Shrinkage
  - Fiecas et al. (2010, NeuroImage) and Fiecas & Ombao (2011, AOAS)

Multi-scale Factor Analysis
Measures of Connectivity

Partial directed coherence between regions r_1 and r_2

$$C_{r_1 r_2}(\omega) = \frac{1}{|V_{r_1}|\,|V_{r_2}|} \sum_{i \in V_{r_1}, j \in V_{r_2}} \pi_{ij}^2(\omega)$$

where V_{r_1} and V_{r_2} are the voxel set within region r_1 and r_2 respectively.

Multi-scale Factor Analysis
Estimation Procedure

ALGORITHM.

STEP 1. Obtain $\mathbf{Z}(t)$ via source reconstruction methods or directly use the EEGs

STEP 2. Estimate factor loading $\mathbf{Q}$

STEP 3. Estimate the factor $\mathbf{f}_t$

STEP 4. Estimate the VAR order P and the coefficient matrices $\Phi^{\mathrm{f}}(\ell)$

STEP 5. Plug in estimates for the connectivity measures

Multi-scale Factor Analysis
Estimation

STEP 2. Estimating the factor loading $\mathbf{Q}_r$

- Compute the estimate of the covariance matrix of $\mathbf{Z}_r(t)$

$$\widehat{\Sigma}_{Z_r Z_r} = \mathbf{Z}_r' \, \mathbf{Z}_r / T$$

- Obtain the eigenvalue-eigenvector decomposition of $\widehat{\Sigma}_{Z_r Z_r}$.
  - Eigenvalues are: $\lambda_1 > \ldots > \lambda_{N_r}$
  - Corresponding eigenvectors (normalized) $\mathbf{V}_1, \ldots, \mathbf{V}_{N_r}$
- The estimator of $\mathbf{Q}_r$

$$\widehat{\mathbf{Q}}_r = [\mathbf{V}_1, \ldots, \mathbf{V}_{m_r}]$$

  where m_r is the dimension of the regional factor activity $\mathbf{f}_r(t)$.
- In practice, we would determine m_r based on the threshold of proportion of variation accounted by the factor.

Multi-scale Factor Analysis
Estimation

STEP 3. Estimating the factors

$$\widehat{\mathbf{f}}_r = \mathbf{Z}_r \widehat{\mathbf{Q}}_r, \; r = 1, \ldots, R$$

STEP 4. Estimating the VAR order and coefficient matrix $\Phi^{\mathrm{f}}(\ell)$

- Define $\widehat{\mathbf{f}}(t) = [\widehat{\mathbf{f}}_1(t)', \ldots, \widehat{\mathbf{f}}_R(t)']'$
- The optimal order minimizes the Akaike information criterion

$$\mathrm{AIC}(P) = \log|\widehat{\Sigma}_\eta| + \frac{2}{T} P m^2$$

where $\widehat{\Sigma}_\eta = T^{-1} \sum_{t=1}^{T} \widehat{\eta}(t)\widehat{\eta}(t)'$ is the residual covariance matrix.

Multi-scale Factor Analysis
Estimation

STEP 5. Estimating the connectivity measures

$$\widehat{\mathbf{Q}} = \text{Diag}\{\widehat{\mathbf{Q}}_1, \ldots, \widehat{\mathbf{Q}}_R\} \quad (1)$$

$$\widehat{\Phi}^Z(\ell) = \widehat{\mathbf{Q}}\widehat{\Phi}^f(\ell)\widehat{\mathbf{Q}}' \quad (2)$$

$$\widehat{\Phi}^Z(\omega) = \mathbf{I} - \sum_{\ell=1}^{P} \widehat{\Phi}^Z(\ell) \exp(-2\pi i \ell \omega / \Omega_s) \quad (3)$$

$$\widehat{\pi}^2_{ij}(\omega) = \frac{|\widehat{\Phi}^Z_{ij}(\omega)|^2}{(\widehat{\Phi}^Z_j(\omega))^H \widehat{\Phi}^Z_j(\omega)} \quad (4)$$

$$(5)$$

The estimate for directed connectivity from region from r_2 to r_1

$$\widehat{C}_{r_1 r_2}(\omega) = \frac{1}{|V_{r_1}||V_{r_2}|} \sum_{i \in V_{r_1}, j \in V_{r_2}} \widehat{\pi}^2_{ij}(\omega).$$

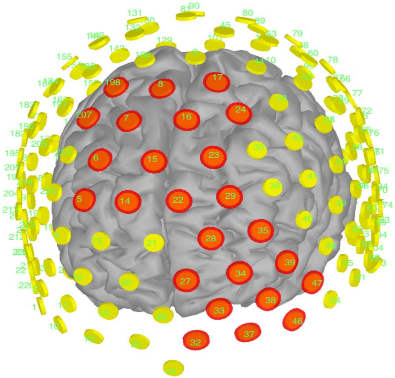
Multi-scale Factor Analysis
EEG Connectivity Analysis

Data description

- Single-subject, male college student
- Number of channels 194 channels (out of 256)
- Sampling rate 1000 Hz
- $T = 1000$ (1 second)
- Regions ($R = 14$): left/right prefrontal (LPF, RPF), frontal (LF, RF), central (LC, RC), parietal (LP, RP), temporal (LT, RT), occipital (LO, RO) and limbic (LL, RL).

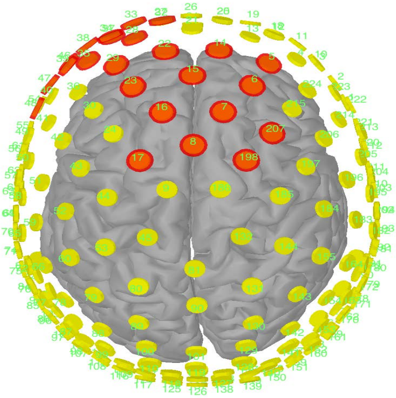
Multi-scale Factor Analysis

Exploratory - EEG and factor plots



SMA and Left Pre-frontal channels

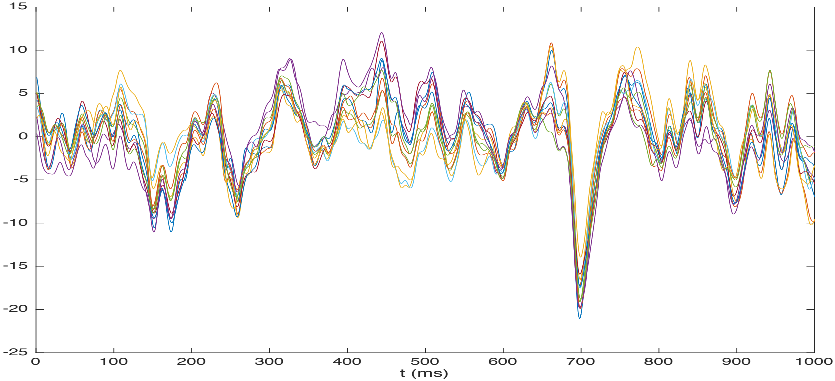
Multi-scale Factor Analysis
Exploratory - EEG and factor plots



SMA and Left Pre-frontal channels

Multi-scale Factor Analysis

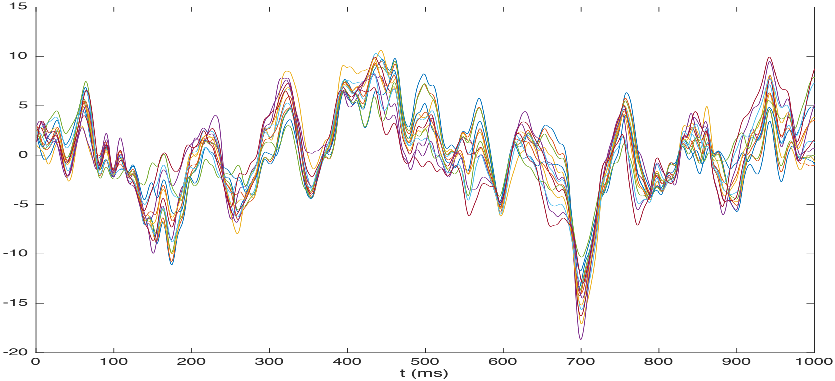
Exploratory - EEG and factor plots



EEG from the Left Pre-frontal channels

Multi-scale Factor Analysis

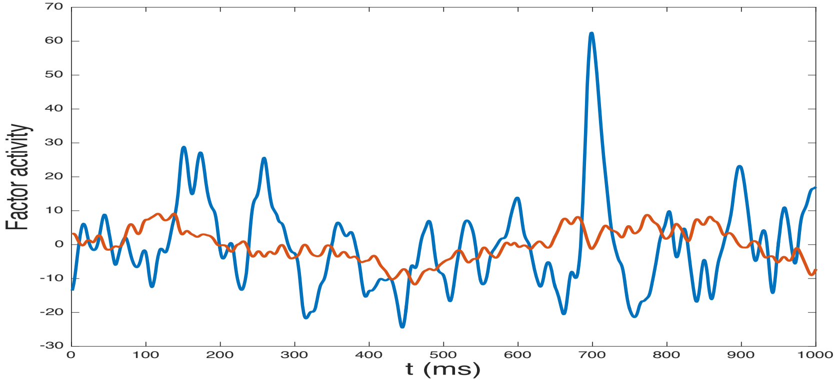
Exploratory - EEG and factor plots



EEG from the SMA channels

Multi-scale Factor Analysis

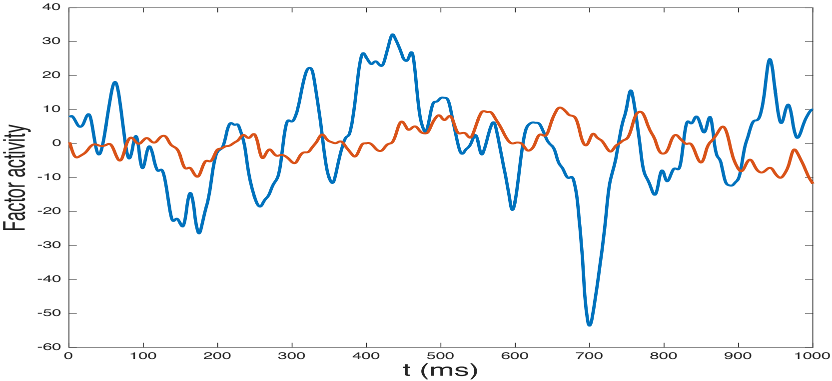
Exploratory - EEG and factor plots



Factors from the Left Pre-frontal channels

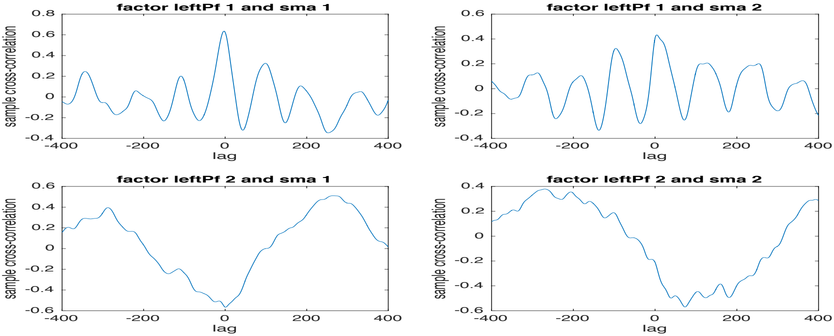
Multi-scale Factor Analysis

Exploratory - EEG and factor plots



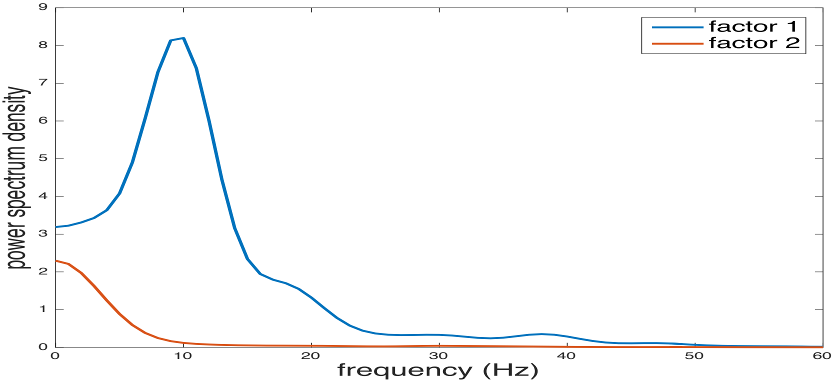
Factors from the SMA region

Multi-scale Factor Analysis
Exploratory - EEG and factor plots



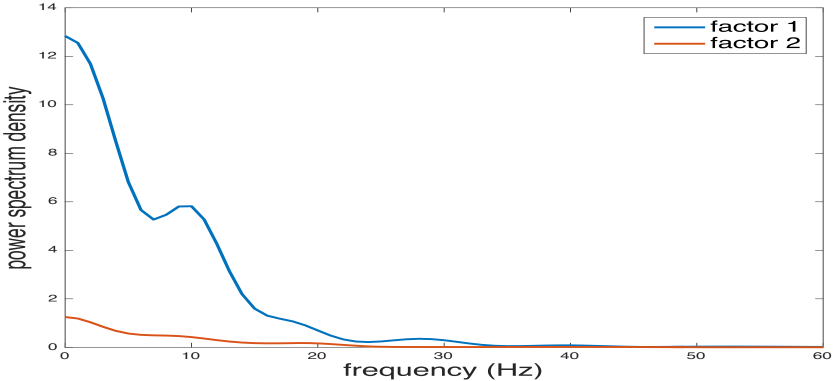
Cross-correlation between SMA and Left-PF channels

Multi-scale Factor Analysis
Exploratory - EEG and factor plots



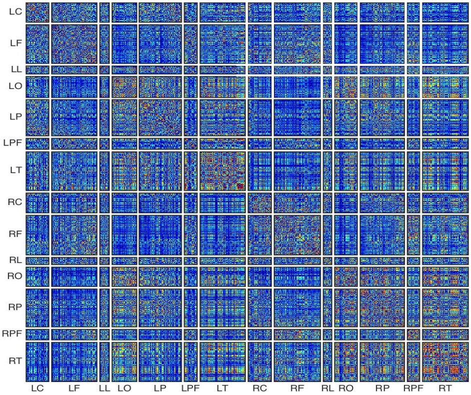
Estimated power spectrum: factors for left Pf region

Multi-scale Factor Analysis
Exploratory - EEG and factor plots



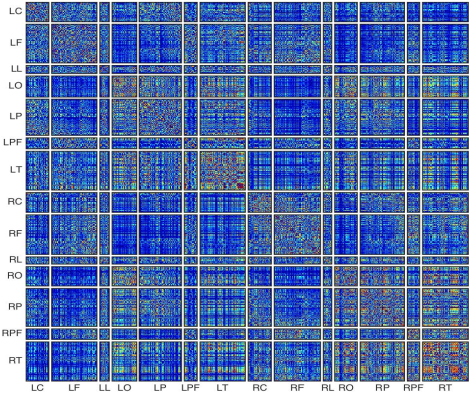
Estimated ower spectrum: factors for SMA region

MULTI-SCALE FACTOR ANALYSIS
PDC AT THE ALPHA BAND



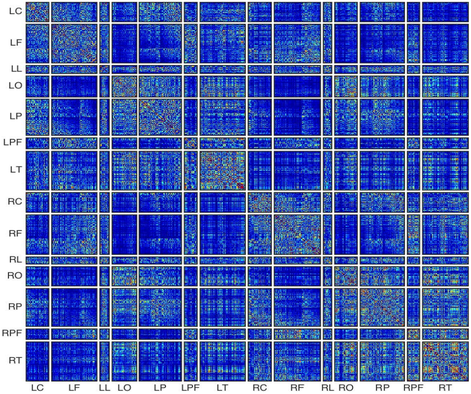
PDC at the alpha band (8-12 Hz)

Multi-scale Factor Analysis
PDC at the beta band



PDC at the beta band (12-30 Hz)

Multi-scale Factor Analysis
PDC at the gamma band



PDC at the gamma band (30-50 Hz)

Multi-scale Factor Analysis
EEG Analysis: Results

- Even with a small number of factors ($m_r = 3$) was able to capture most of the variation within each region (over 85%)
- Modular organization - connectivity between channels/voxels in the same region is generally higher than between channels/voxels from different regions
- Connectivity at the alpha (8-12 Hz) and beta (12-30 Hz) stronger than at the gamma band (30-50 Hz) – resting state!
- Pronounced directed outflows: LC $\rightarrow$ LP; LP $\rightarrow$ RL.

Multi-scale Factor Analysis
Future Work

- Combining information across trials
- Combining information across subjects
- Testing for differences in PDC across patient groups; experimental conditions
- Associations between physiology and behavior
- Non-stationarity (factor analysis)
  - Dynamic factor models (Forni, Hallin, Lippi & Reichlin)
  - Evolutionary factor analysis (Motta & Ombao (2012))
- Visualization with M. Genton and Y. Sun (KAUST)

ACKNOWLEDGEMENT

Thanks to the organizers – especially to Linglong Kong – for this exciting workshop!