Resting-State Functional MRI Analysis by Graphical Model

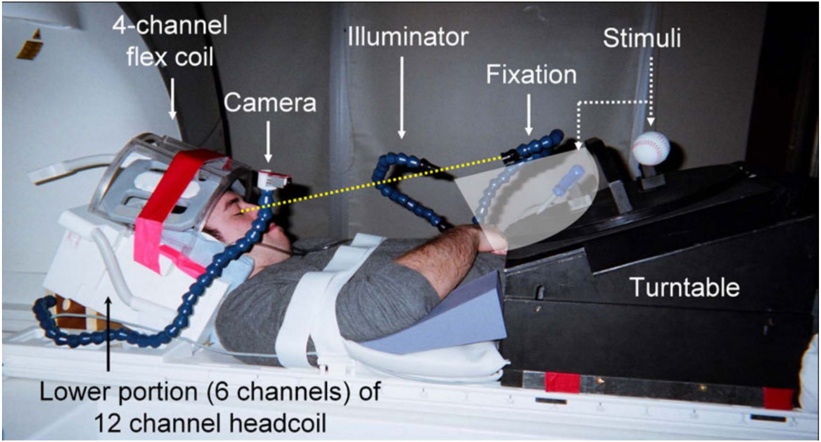
With Applications on Functional Network Estimation

Wei Liu

Scientific Computing and Imaging Institute
University of Utah
Advisor: Tom Fletcher

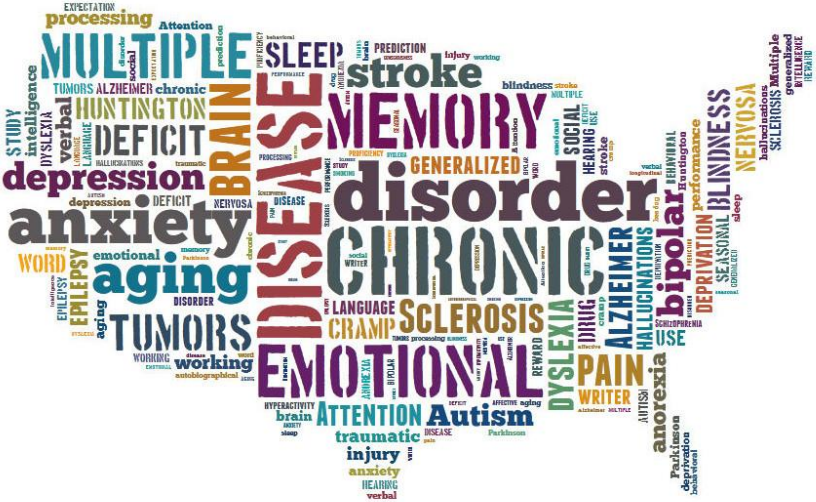
December 18, 2013

fMRI Experiments

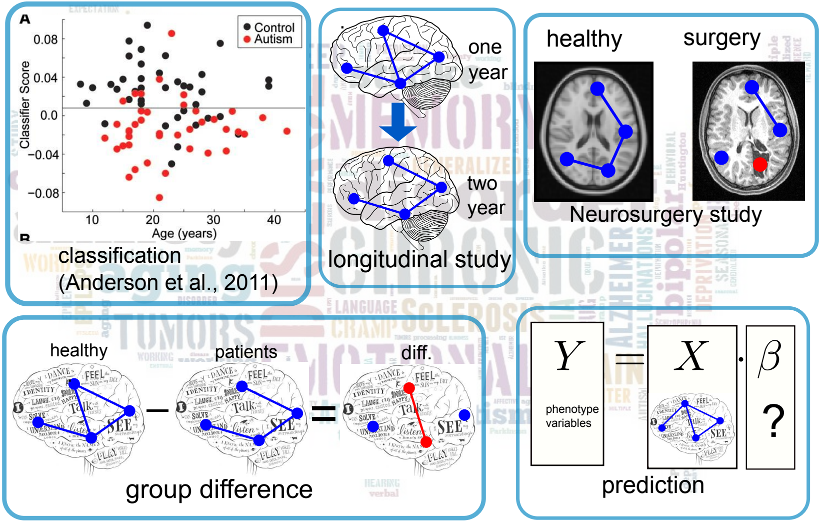


J C. Snow, Nature 2011

Clinical Applications of rs-fMRI



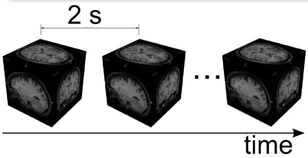
Clinical Applications of rs-fMRI

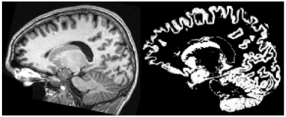


Outline

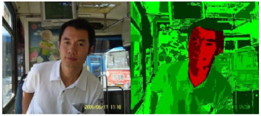
Functional MRI Data

- Blood oxygen level dependent (BOLD) indirectly measures neuronal activity.
- 3D volumes sampled at each time point.
- Fast scan, but noisy.
- Spatio-temporal dependency.





seg. of T1 MRI



seg. of natural images

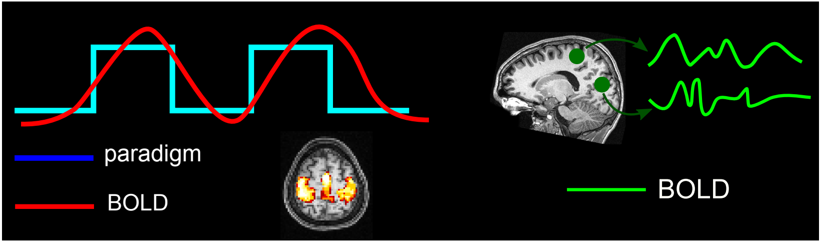
Task-based v.s. rs-fMRI

Task-based

- Experiment stimulus signal.
- Subjects undertake cognitive tasks.
- Linear regression between stimulus and BOLD.

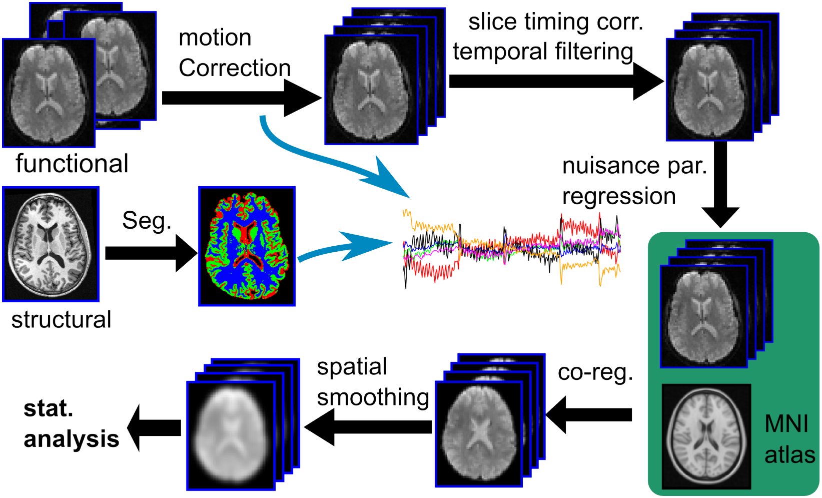
Resting-State

- No experiment paradigm signal.
- Subject stays in scanner. Eyes closed/open to a fixation cross.
- Correlation between two voxels.

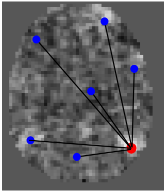


task resting-state

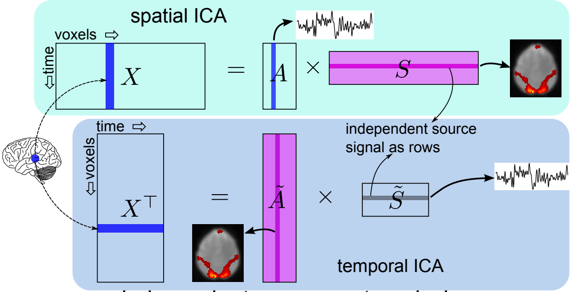
fMRI Processing Pipeline



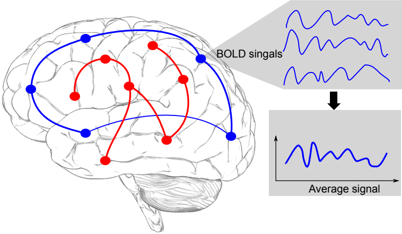
Network Analysis Methods



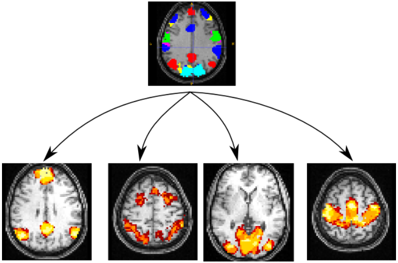
seed-based methods



independent component analysis

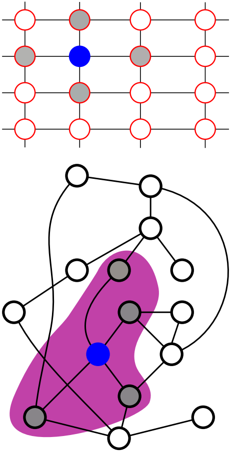


graph-based methods



segmentation/clustering

Markov Random Field



Definition

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$: undirected graph.
$s \in \mathcal{V}$: a node/site in $\mathcal{V}$.
$X = \{x_1, \ldots x_N\}$: Set of random variables defined on $\mathcal{G}$.
$\mathcal{N}_s$: Set of nodes neighboring s.
$(r, s) \in \mathcal{E} \Leftrightarrow r \in \mathcal{N}_s$.

Definition

A Markov Random Field (MRF) is a collection of variables X defined on graph $\mathcal{G}$ if for all $s \in \mathcal{V}$

$$P(s_s | X_{\mathcal{V}-s}) = P(s_s | x_{\mathcal{N}_s})$$

MRF and Gibbs Field

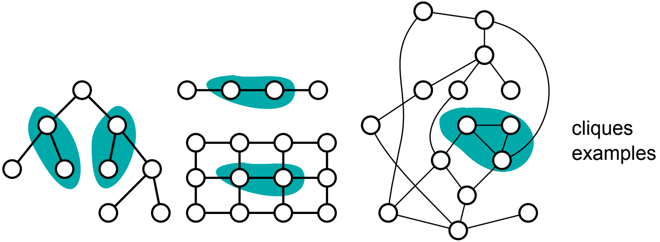
Theorem (Hammersley-Clifford, 1971)

X is an MRF on $\mathcal{G}$ if and only if X obeys Gibbs distribution in the following form

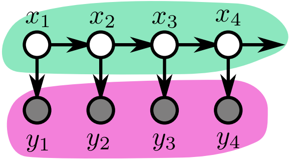
$$P(X) = \frac{1}{Z}\exp\left(-\frac{1}{T}U(X)\right),$$

$$U(X) = \sum_{c\in\mathcal{C}} V_c(X_c).$$

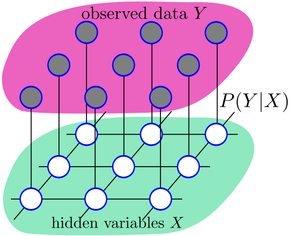
$Vc(X_c) = \beta\sum_{(r,s)\in\mathcal{V}}\psi(x_r, x_s)$: Ising, Potts model.

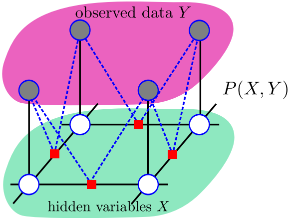


Hidden Markov Model: A Generative Model



- X is defined on MRF.
- Y is assumed to be generated from X.
- Inverse problem: Given Y, estimate X.





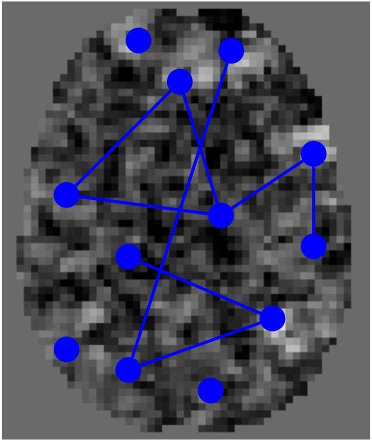
Other forms exist: conditional random field, but no Bayesian interpretation.

Outline for section 2

Pairwise Connectivity With Spatial Coherence

The Goal

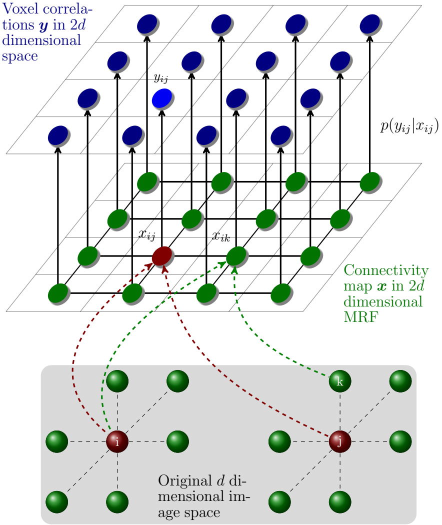
- The connectivity between each pair of voxels in single subject rs-fMRI.
- Spatial regularization by MRF, without changing signals.
- Learn the strength of the smoothness from the data.



Pairwise Connectivity With Spatial Coherence: the Model

Model Definition

- MRF defined on 6D graph.
- Pairwise connectivity variable $X \in \{0,1\}^N$, sample correlation Y.
- $(x_{ij}, x_{ik}) \in \mathcal{E} \Leftrightarrow j \in \mathcal{N}(k)$.
- $P(F(y_{ij})|x_{ij}) \sim \mathcal{N}(\mu, \sigma^2)$.



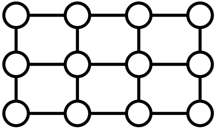
Statistical Inference

Questions

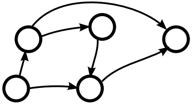
- Given x_{-s} and Y, what is $P(x_s)$.
- $x_s = \operatorname{argmax} P(X|Y)$
- $X^* = \operatorname{argmax} P(X|Y)$.

Algorithms

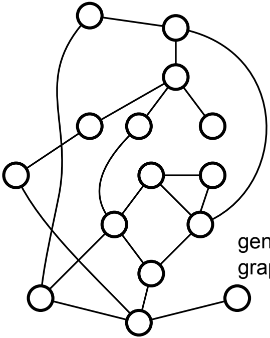
- Exact solutions for simple graphs (trees, chains): sum-product, max-sum, belief propagation.
- No exact solution for general graphs.



regular lattice

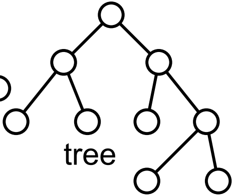


directed graph



general graph

chain



tree

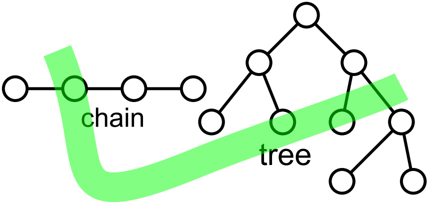
Statistical Inference

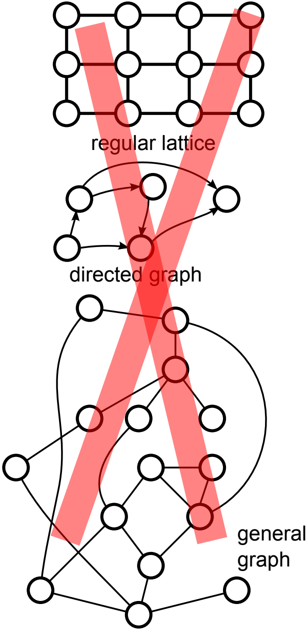
Questions

- Given x_{-s} and Y, what is $P(x_s)$.
- $x_s = \operatorname{argmax} P(X|Y)$
- $X^* = \operatorname{argmax} P(X|Y)$.

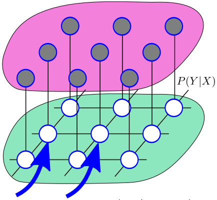
Algorithms

- Exact solutions for simple graphs (trees, chains): sum-product, max-sum, belief propagation.
- No exact solution for general graphs.



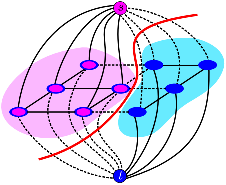


Approximate Inference

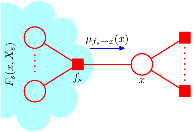


$x_s = \text{argmax} P(x_s|x_{\mathcal{N}(s)})$

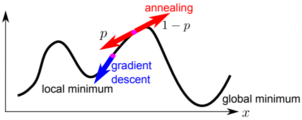
iterated conditional modes
= coordinate gradient descent



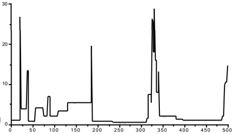
graph cuts: min (E) = min(Cuts)
= maxflow global optimum for
2 classes



And many other methods:
belief/expectation propagation,
Viterbi,...



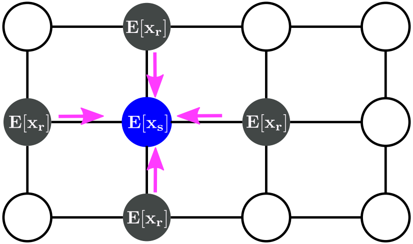
simulated annealing: non-zero
probability of going uphill



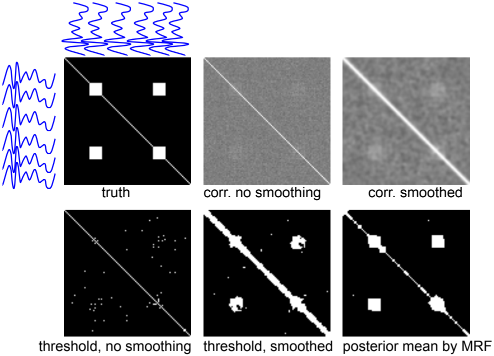
sampling from P(X|Y) to
approximate posterior

Variational Inference

- Assuming $Q(X) = \prod_s q_s(x_s)$.
- Search $Q(x)$ in a smaller space.
- $\log Q_s(x_s) = E_{r \neq s}[\log P(X, Y)] + \text{const}$



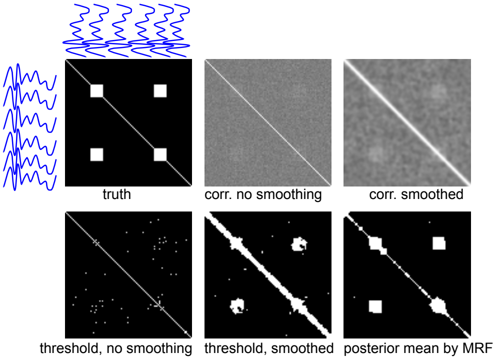
Experiments



Simulated data

- Construct 1D image. Connected voxels are added with sine wave signal.
- Spatial smoothing improves results, but increased false positive.

Experiments

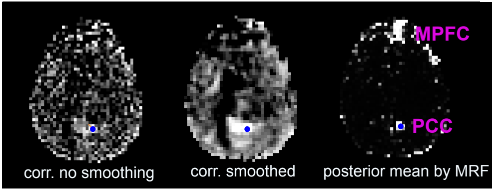


Simulated data

- Construct 1D image. Connected voxels are added with sine wave signal.
- Spatial smoothing improves results, but increased false positive.

Real data

- connectivity between a voxel in PCC and current slice.
- Detecting PCC-MPFC links in default mode network.



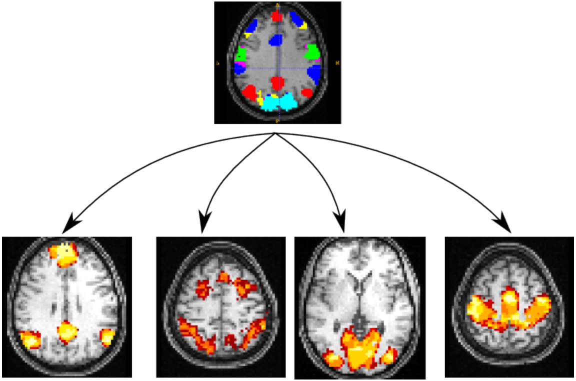
Outline for section 3

3. Consistent, Spatially Coherent Multiple Functional Networks
   - Problem Statement
   - A MAP framework with MRF prior
   - MCEM for Statistical Inference
   - Experiments on Synthetic Data and Real Data

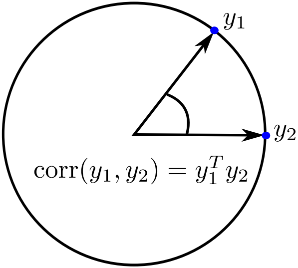
Consistent, Spatially Coherent Multiple Functional Networks

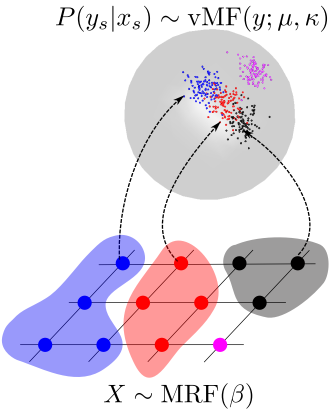
The Goal

- Partition the brain into multiple functional networks.
- Spatial coherence is respected.
- Parameter estimation.



MRF and vMF Models





Model

- $P(X) = (1/Z) \exp\left(-\beta \sum_{(r,s)\in\mathcal{E}} \psi(x_s, x_r)\right)$.
- $x_s \in \{1, \ldots, L\}$
- $P(y_s|x_s) = C_p(\kappa_l) \exp(\kappa_l \mu_l^\top y_s)$, $y_s \in S^{p-1}$ (von Mises-Fisher distribution).
- Solve $P(X|Y) \propto P(X) \cdot P(Y|X)$

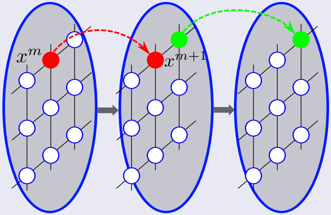
Statistical Inference by MCEM

Expectation Maximization

$$Q(\theta) = \mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$$

Monte Carlo Expectation Maximization (MCEM)

$$\mathbb{E}_{X|Y}[\log P(X, Y; \theta)] \approx \frac{1}{M} \sum_m \log P(X^m, Y; \theta)$$



Statistical Inference by MCEM

Expectation Maximization

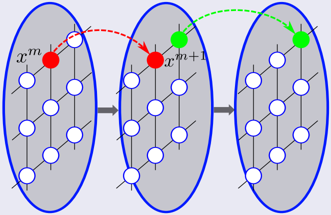
$Q(\theta) = \mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$

Pseudo Likelihood

$\log P(X^m; \theta) \approx \sum_{s \in \mathcal{V}} \log P(x_s | x_{\mathcal{N}_s}; \theta)$

Monte Carlo Expectation Maximization (MCEM)

$$\mathbb{E}_{X|Y}[\log P(X, Y; \theta)] \approx \frac{1}{M} \sum_m \log P(X^m, Y; \theta)$$



Statistical Inference by MCEM

Expectation Maximization

$Q(\theta) = \mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$

Pseudo Likelihood

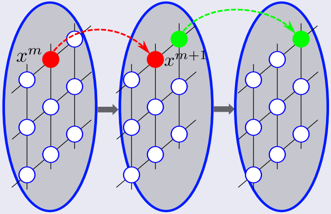
$\log P(X^m; \theta) \approx \sum_{s \in \mathcal{V}} \log P(x_s | x_{\mathcal{N}_s}; \theta)$

Parameter Estimation

$\hat{\mu}_l = \| R_l \|, R_l = \sum_{s \in \mathcal{V}_l} y_s$
$\hat{\kappa}_l \approx (pR_l - R^3)/(1 - R^2)$
$\hat{\beta} = \mathrm{argmax}_\beta \log P(X; \theta)$ by Newton's method.

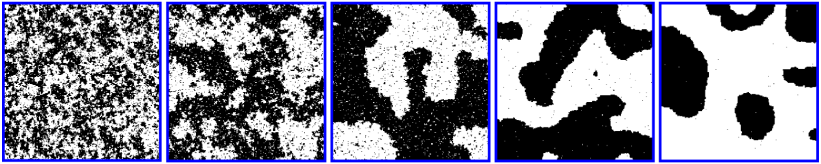
Monte Carlo Expectation Maximization (MCEM)

$$\mathbb{E}_{X|Y}[\log P(X, Y; \theta)] \approx \frac{1}{M} \sum_m \log P(X^m, Y; \theta)$$



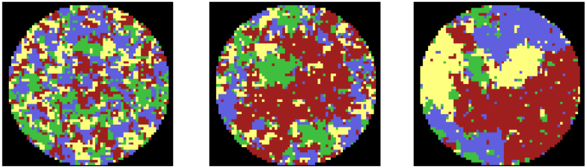
Simulation of Ising and Potts Models

Ising model



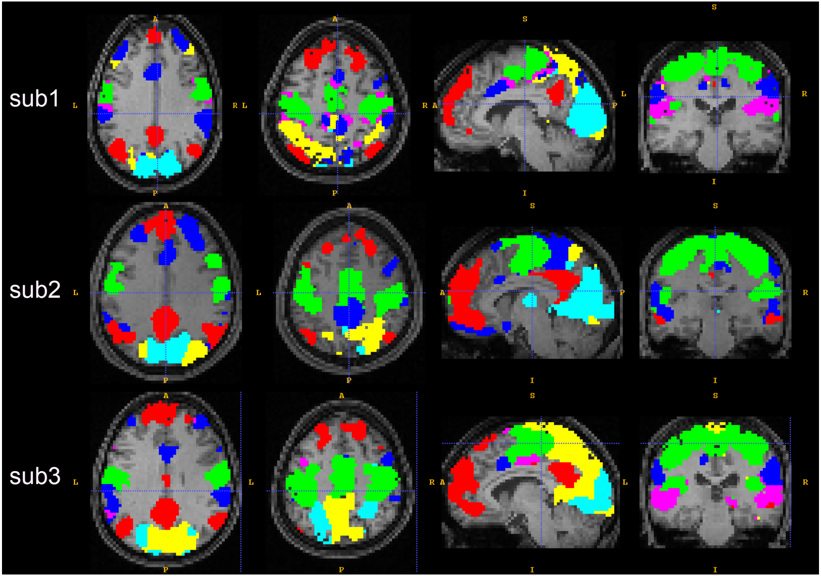
$\beta = 0.8$ $\beta = 0.88$ $\beta = 1.0$ $\beta = 1.5$ $\beta = 2.0$

Potts model



$\beta = 1.0$ $\beta = 1.1$ $\beta = 1.2$

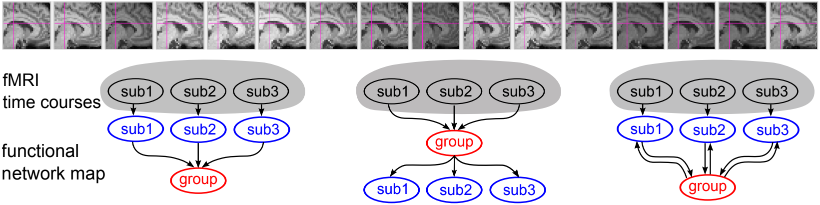
Experiments



visual (cyan), motor (green), executive control (blue), salience (magenta), dorsal attention (yellow), and default mode (red) networks

Outline for section 4

Hierarchical Model: a Better Approach



[Bellec, 2010]
[Van Den Heuvel, 2008]
[Esposito, 2005]

[Calhoun, 2001b]
[Beckmann, 2009]
[Filippini, 2009]

[Varoquaux, 2010, 2011]
Our method: HMRF

Existing Methods

- Bottom-up or top-down.
- Subject is estimated independently.
- Estimation is one way.

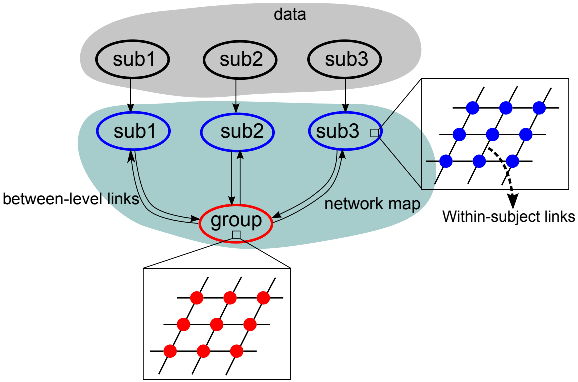
We Propose

- A hierarchical structure including group and subject.
- Jointly estimate both levels iteratively.
- Bayesian framework. Data driven. Parameter estimation.

Joint Estimation with MRF

Build a graph

- within-subject piecewise constant constraints.
- Between-subject (between-level) dependency.

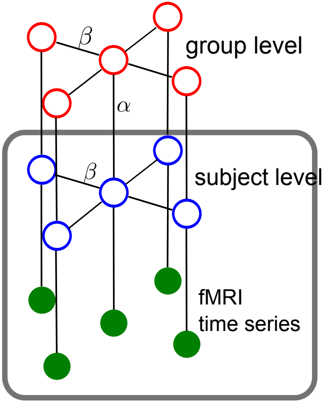


A Graphical View of the Hierarchical Model

$$\mathcal{G} = (\mathcal{V}_G, \mathcal{V}_1, \ldots, \mathcal{V}_J)$$

$$\mathcal{E} = \{(r,s) | (r,s) \in \mathcal{E}_G \text{ or } s \in \mathcal{V}_G, r \in \mathcal{V}_j \text{ or } (r,s) \in \mathcal{V}_j\}$$

$$U(X) = \sum_{(s,r)\in\mathcal{V}_G} \beta\psi(x_s, x_r) + \sum_{j=1}^{J} \left(\sum_{(s,\tilde{s})} \alpha\psi(x_s, x_{\tilde{s}}) + \sum_{(s,r)\in\mathcal{V}_j} \beta\psi(x_s, x_r) \right).$$



Likelihood

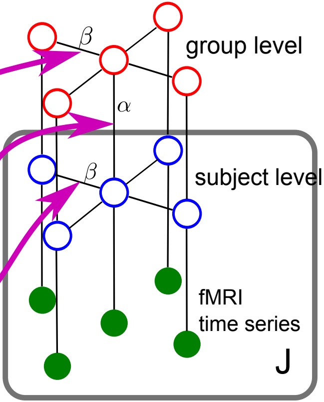
$P(y_s | x_s) \sim vMF(\mu, \kappa)$

A Graphical View of the Hierarchical Model

$$\mathcal{G} = (\mathcal{V}_G, \mathcal{V}_1, \ldots, \mathcal{V}_J)$$

$$\mathcal{E} = \{(r,s) | (r,s) \in \mathcal{E}_G \text{ or } s \in \mathcal{V}_G, r \in \mathcal{V}_j \text{ or } (r,s) \in \mathcal{V}_j\}$$

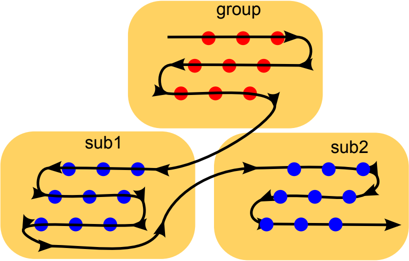
$$U(X) = \sum_{(s,r)\in\mathcal{V}_G} \beta\psi(x_s, x_r) + \sum_{j=1}^{J} \left(\sum_{(s,\tilde{s})} \alpha\psi(x_s, x_{\tilde{s}}) + \sum_{(s,r)\in\mathcal{V}_j} \beta\psi(x_s, x_r) \right).$$



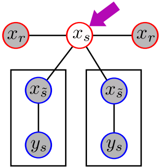
Likelihood

$P(y_s|x_s) \sim vMF(\mu, \kappa)$

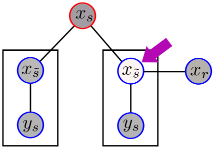
- Monte Carlo Sampling used to approximate $\mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$.
- Gibbs sampling also in a multi-level fashion.



Gibbs sampling
Schedule

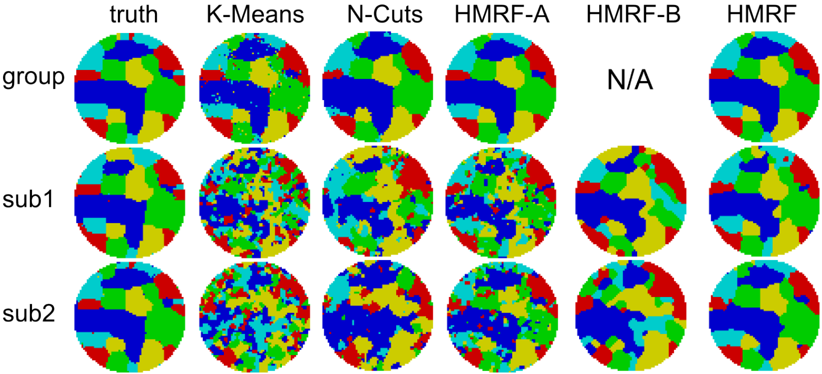


Sampling voxels
in group map

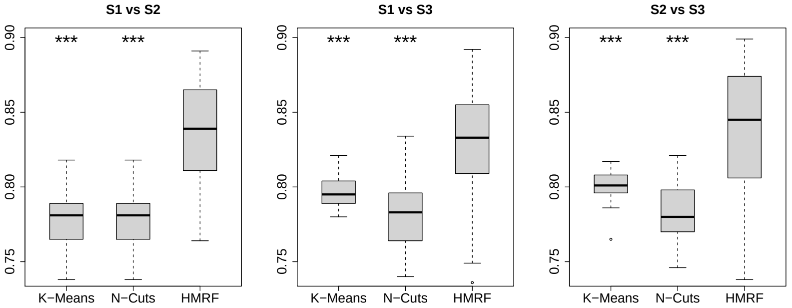


Sampling voxels
in subjects

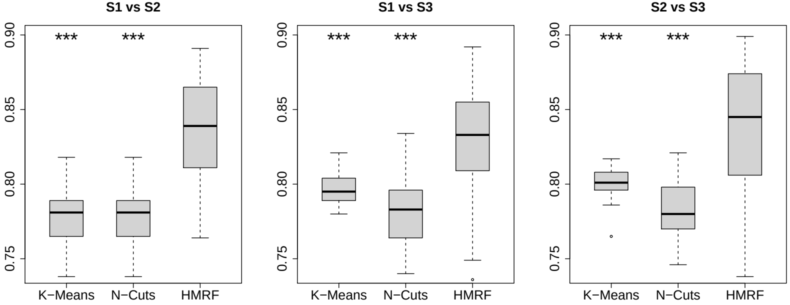
Synthetic Data: Estimation

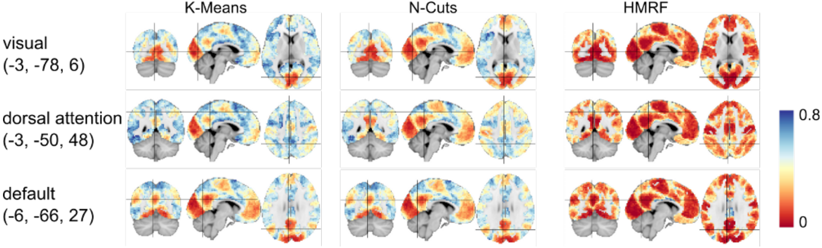


Cross-Session consistency

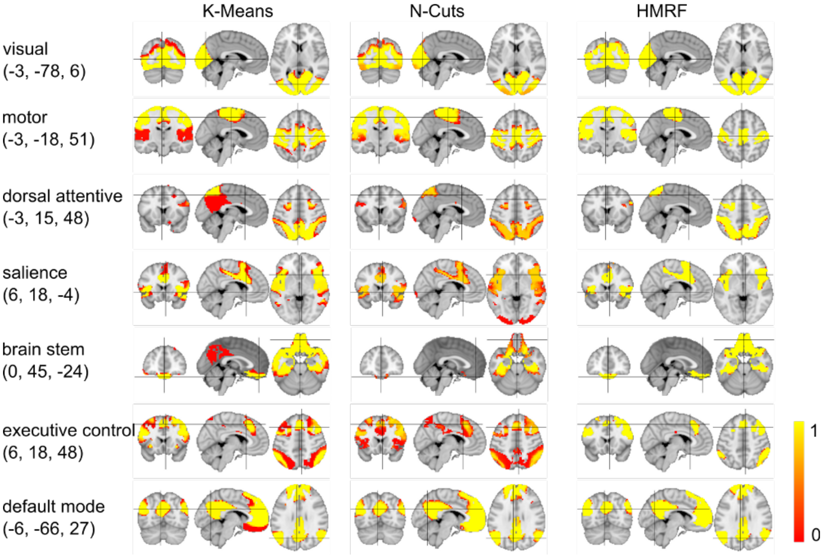


Cross-Session consistency

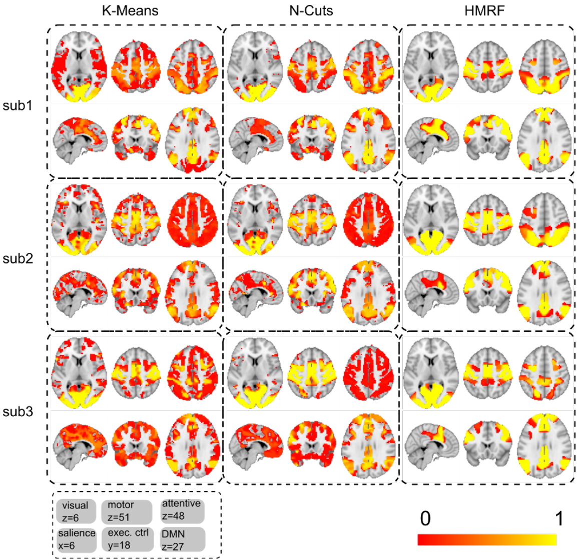




Bootstrapping: Group Mean Maps



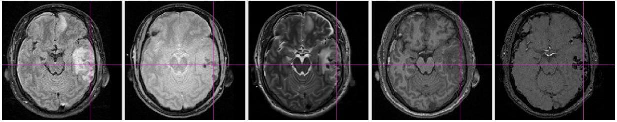
Bootstrapping: Subject Mean Maps

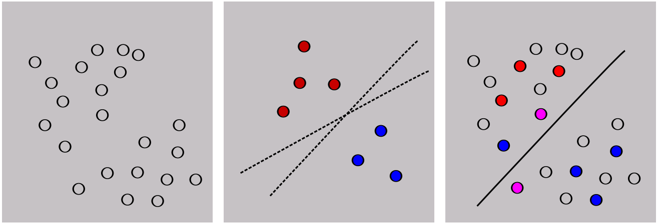


Outline for section 5

Traumatic Brain Injury Image Segmentation with Active Learning: Motivation

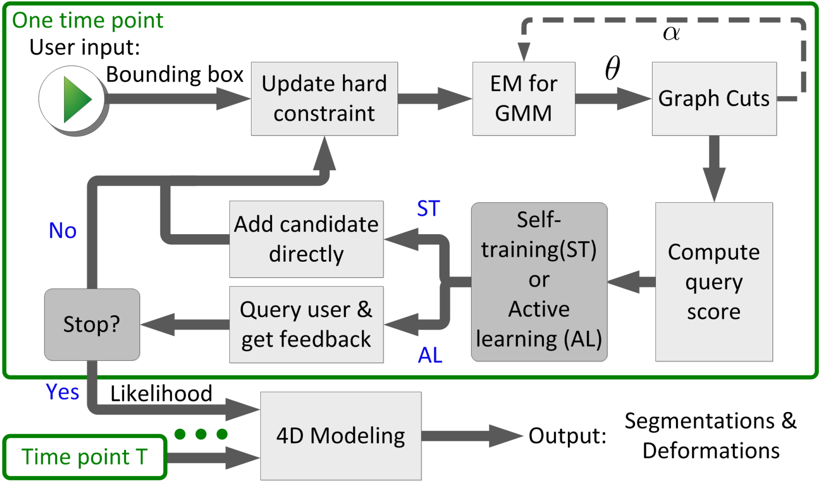
- Multi-modality, longitudinal, complex patterns.
- Existing methods: high false-positive/negative, 2D, single object.
- A slight user involvement significantly improves result.
- Computer active, user passive (less burden).





unlabeled | supervised learning | semi-supervised active learning

An Active Learning Framework

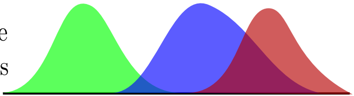


Prior:

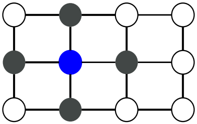
$$p(\mathbf{z}) = \frac{1}{C}\exp\left(\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\log\pi_{nk} + \beta\sum_{(n,m)}\langle z_n, z_m\rangle\right)$$

K = 2 (FG)
K = 3 (BG)

z_{nk}: indicator variable
(n,m): neighbor voxels



mixture model

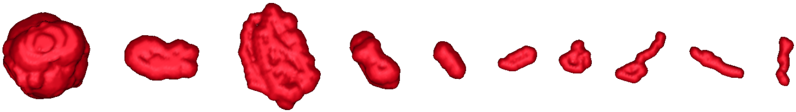


MRF

Likelihood: $p(\mathbf{x}|\mathbf{z}) = \prod_n p(x_n|z_n) = N(x_n;\mu,\sigma^2)$

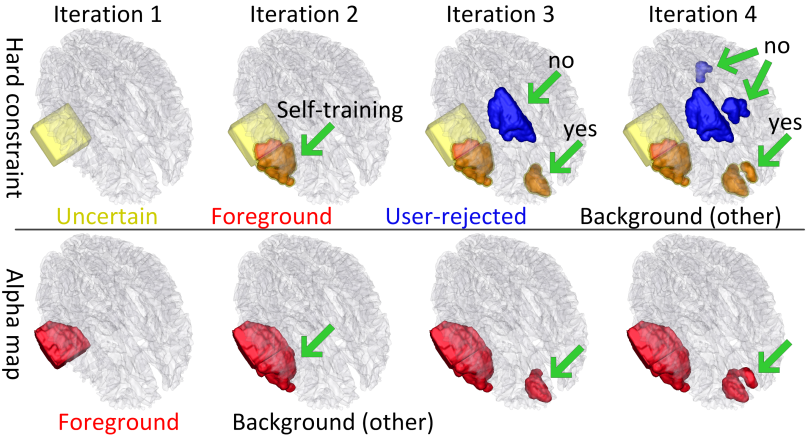
Query Score

- Voxel posterior $\rightarrow$ object query score.
- Prefer large objects, blob-like objects.



$$q(R_i) = \left(\sum_{n \in R_i} p(\alpha = 1 | \boldsymbol{x}_n) \right) \Big/ |\{n : n \in \mathcal{B}(R_i)\}|.$$

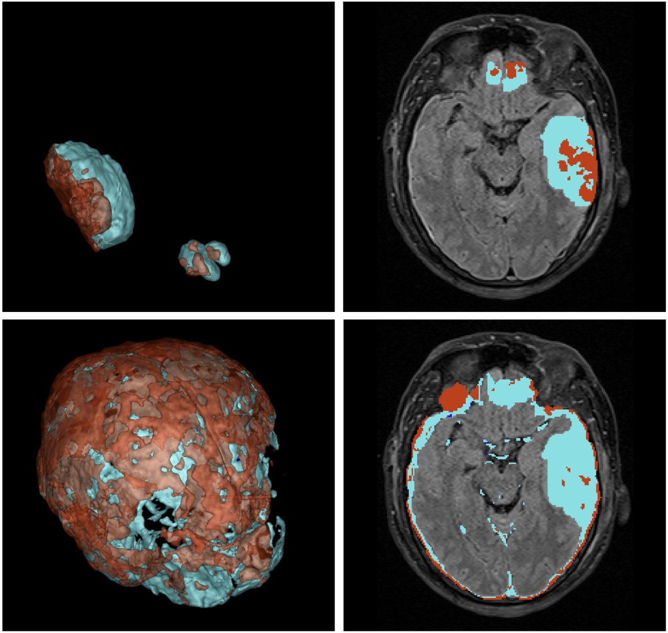
Active Learning Iteration



Qualitative comparison

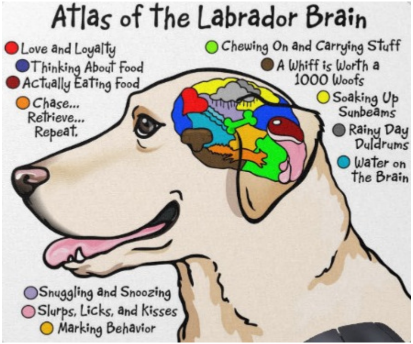
Our method

GrabCut



Conclusions

- Graphical models (directed, undirected) are good for learning and inference of multivariate distribution.
- Model the within- and between-subject soft constraints for rs-fMRI.
- within- and between-class regularization for TBI segmentation.
- Other vision problem: Stereo matching, etc.
- Exact inference difficult. Approximate solution sufficiently good.



Thank you.
This is the end of the talk.

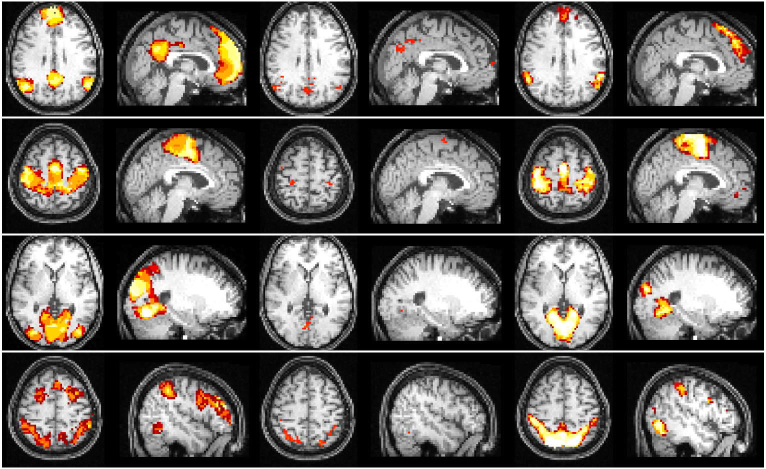


Figure : Comparison of the overlap of the label maps estimated by our MCEM approach, group ICA and single subject ICA on 16 subjects. Left: MCEM methods. Middle: single subject ICA. Right: group ICA. Color map ranges from 8 (red) 16 (yellow).

Data: Normalized rs-fMRI, initial group label map
Result: MC samples of label maps $\{X^m, m = 1, \dots, M\}$, parameters $\{\beta, \mu, \sigma\}$

while $\mathbb{E}_{P(Y|X)}[\log P(Y, X; \theta)]$ *not converged* **do**
  repeat
    foreach $s \in \mathcal{V}_G$ **do** Draw sample of x_s from $P(x_s | x_{-s}, y_s; \theta)$;
    foreach $j = 1 \dots J$ **do**
      foreach $s \in \mathcal{V}_j$ **do** Draw sample of x_s from $P(x_s | x_{-s}, y_s; \theta)$;
    end
    Save sample X^m after B burn-ins;
  until $B + M$ *times*;
  foreach $l = 1 \cdots L$ **do**
    Estimate $\{\mu_l, \kappa_l\}$ by maximizing $(1/M) \sum_{m=1}^{M} \log P(Y | X^m; \theta)$;
  end
  Estimate β by maximizing $(1/M) \sum_{m=1}^{M} \log P(X^m; \theta)$;
end

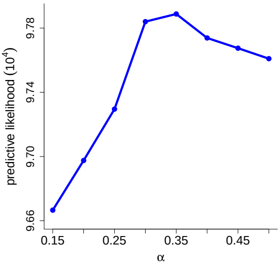
Inter-level Links Estimation

Bayesian Cross-Validation

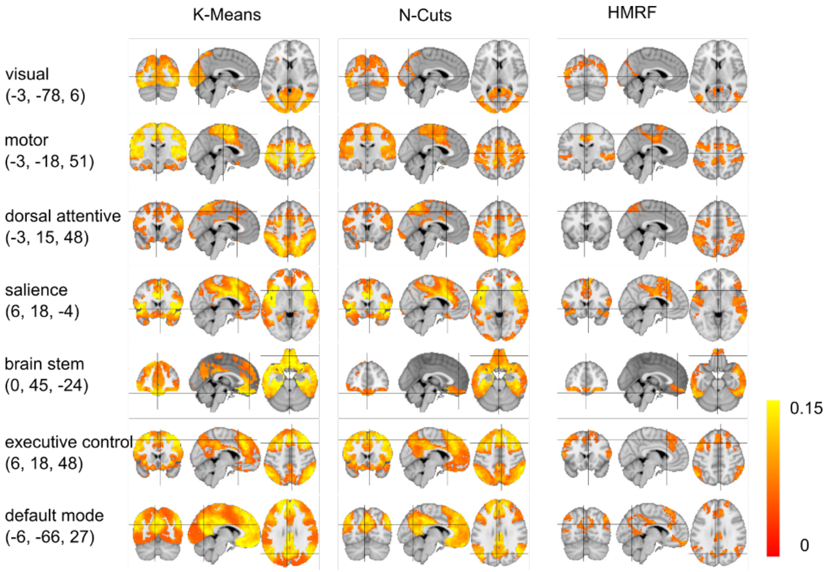
$\alpha = \operatorname{argmax} P(Y_t|Y;\alpha,\theta_t)$

$$\begin{aligned} &P(Y_t|Y;\alpha,\theta_t) \\ &= \int P(Y_t|X_t;\theta_t)P(X_t|Y;\alpha)\,\mathrm{d}X_t \\ &\approx (1/M)\sum_m P(Y_t|X_t^m;\alpha,\theta_t), \\ &\quad X_t^m \sim P(X_t|Y;\alpha). \end{aligned}$$

X_t's are generated within MCEM.



Bootstrapping: Subject Variance Maps



Synthetic Data: Monte Carlo Test