VirginiaTech
Invent the Future

The Influence of EM Estimation of Missing Nodes in DCM on Model Ranking

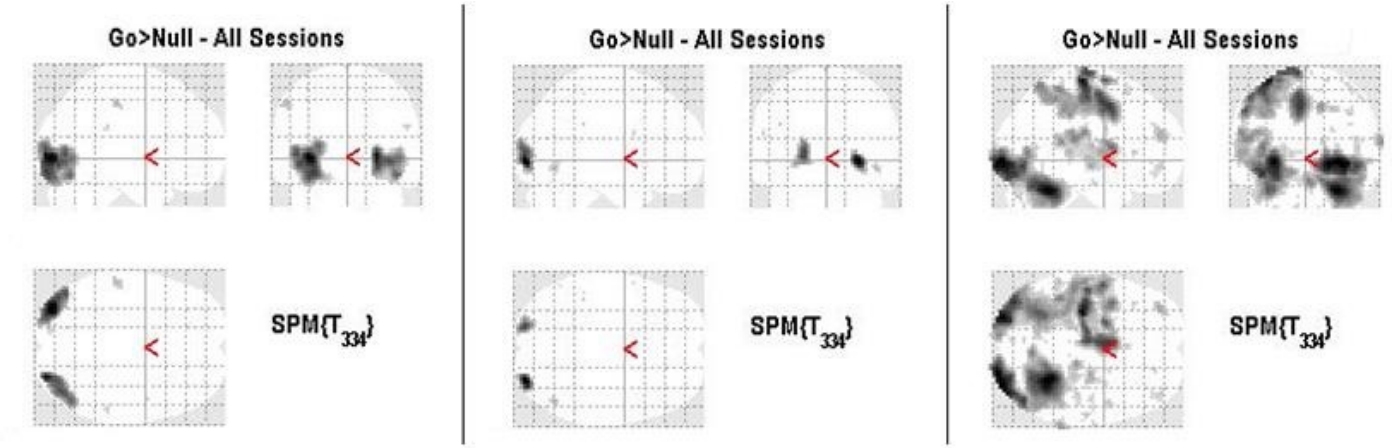
IEEE GlobalSIP – Dec 14, 2015

Shaza B. Zaghlool

Electrical and Computer Engineering Department – Biomedical Imaging Lab

Introduction

- Network Topology is used to model a cognitive process based on:
  - Prior information
  - Statistical analysis of measured signals
- Differences in network topology are due to:
  - Multiple Comparison Problem → Errors in 1st level analysis
  - Anatomical, functional, or measurement factors
  - Age, Gender, Disease (Brain damage, Stroke, Brain lesions)



DCM

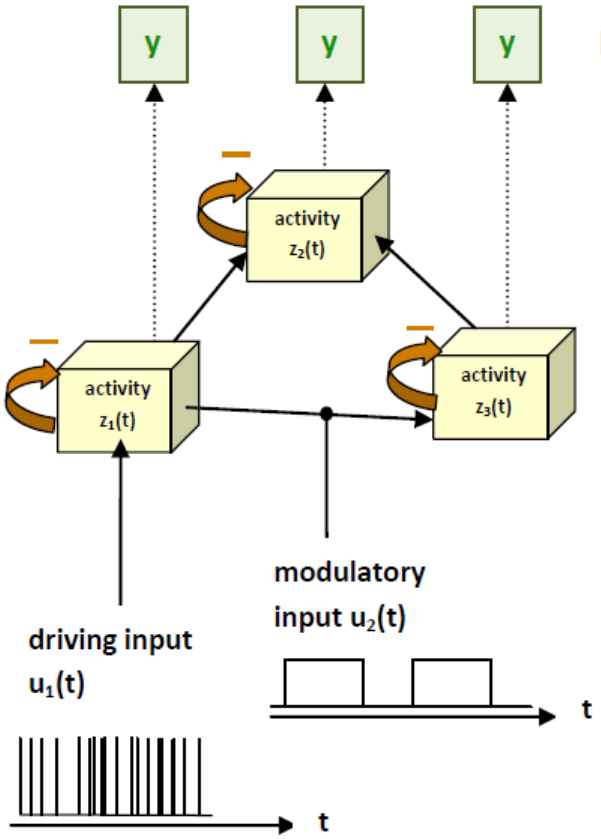
- Dynamic causal modeling (DCM)
  - Estimate the coupling among brain regions
  - Determine how subject responses to experimental changes affect that coupling

- Inferring model connectivity using Bayes Thereom:

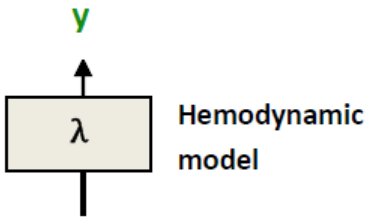
$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$Posterior = \frac{Likelihood * Prior}{Evidence}$$

DCM for fMRI



BOLD



Neuronal states

integration

Neural state equation $\dot{\mathbf{z}} = \left(\mathbf{A} + \sum \mathbf{u_j}\, \mathbf{B^j}\right)\mathbf{z} + \mathbf{Cu}$

Intrinsic connectivity → $A = \frac{\partial F}{\partial z} = \frac{\partial \dot{z}}{\partial z}$

Modulation of connectivity → $B^j = \frac{\partial^2 F}{\partial z \partial u_j} = \frac{\partial}{\partial u_j}\frac{\partial \dot{z}}{\partial z}$

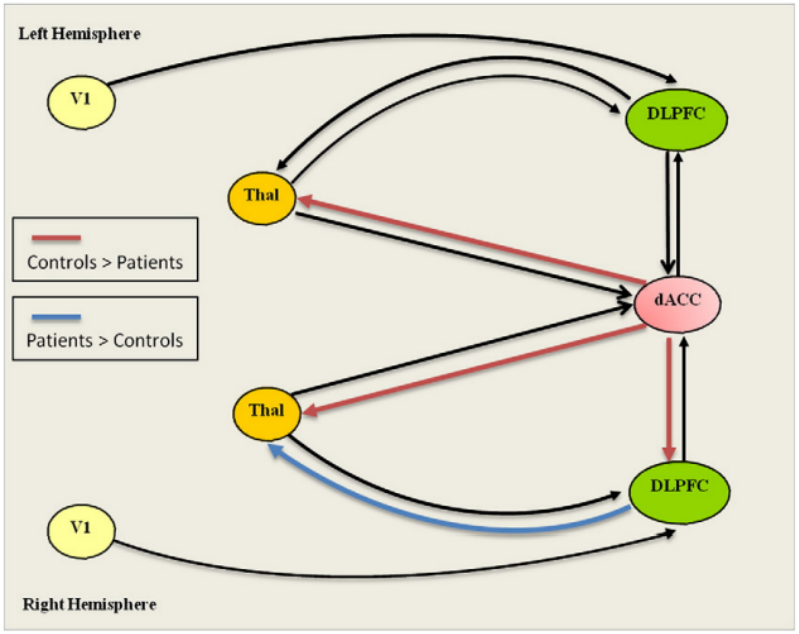
Direct Inputs → $C = \frac{\partial F}{\partial u} = \frac{\partial \dot{z}}{\partial u}$

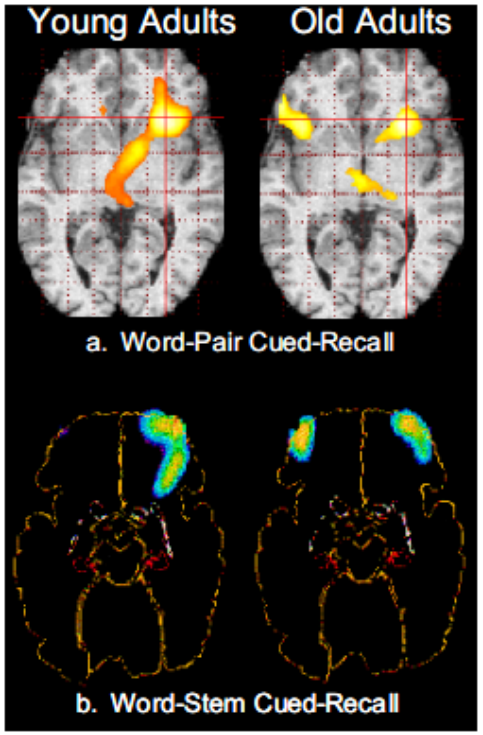
Neural state Vector

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ .. \\ \dot{z}_k \end{bmatrix} = \dot{z} = \frac{dz}{dt} = F(z, u, \theta^n)$$

DCM as a cognitive phenotyping tool

- The number of nodes from the first-level analysis can be informative of neuro-cognitive deficits (Cabeza et al. 2002)
- Other diseases (ie. schizophrenia) are thought to be related to connectivity or other parameters of the generative DCM model (Wagner et al. 2013)





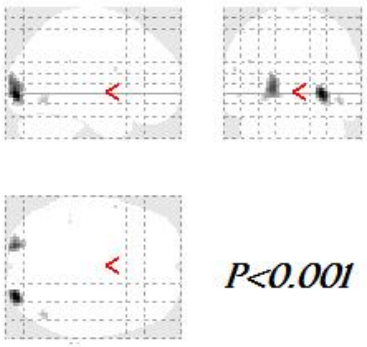
Research Problem

- Given a group of n subjects S_1, S_2, … S_n , the problem is to find the best DCM (M_1, M_2,M_m) that represents each subject
- We want to be able to compute an evidence matrix where every entry in the matrix represents the evidence that a certain model M_x fits a certain subject S_y

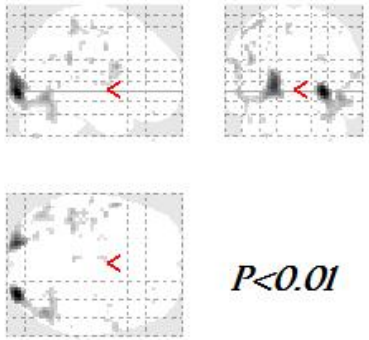
	S_1	S_2	…	S_n
M_1	0.85	0.77	…	0.66
M_2	0.72	0.91	…	0.93
…	…	…	…	…
M_m	0.54	0.63	…	0.89

- Bayes Factor: $$K = \frac{P(D|M_1)}{P(D|M_2)}$$
- Missing node = no activation detected from ROI → Evidence cannot be computed.
- Zero evidence assumption can be problematic.

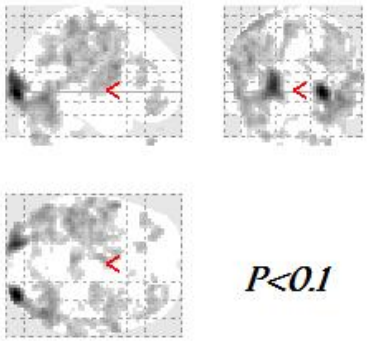
Dealing with Missing Nodes by tweaking the p-value



P<0.001



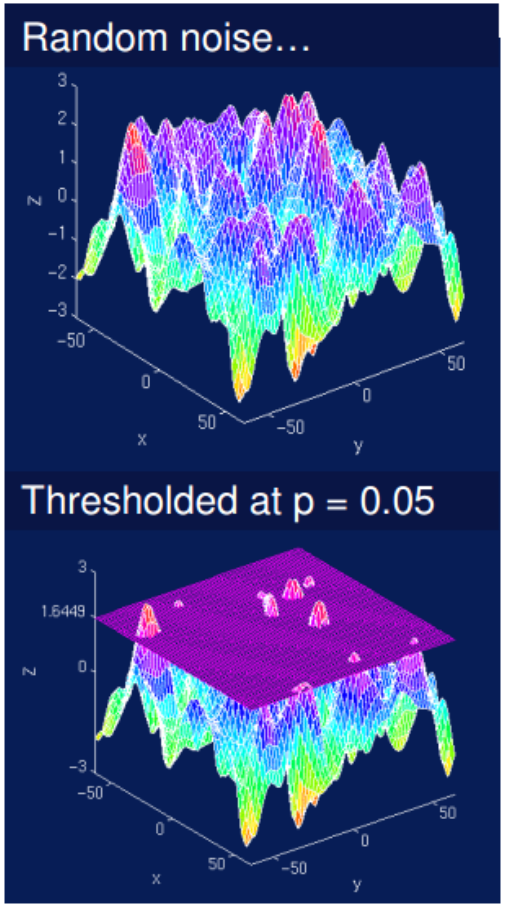
P<0.01



P<0.1

Good Specificity
(Risk of FN)

Poor Specificity
(Risk of FP)



Objectives

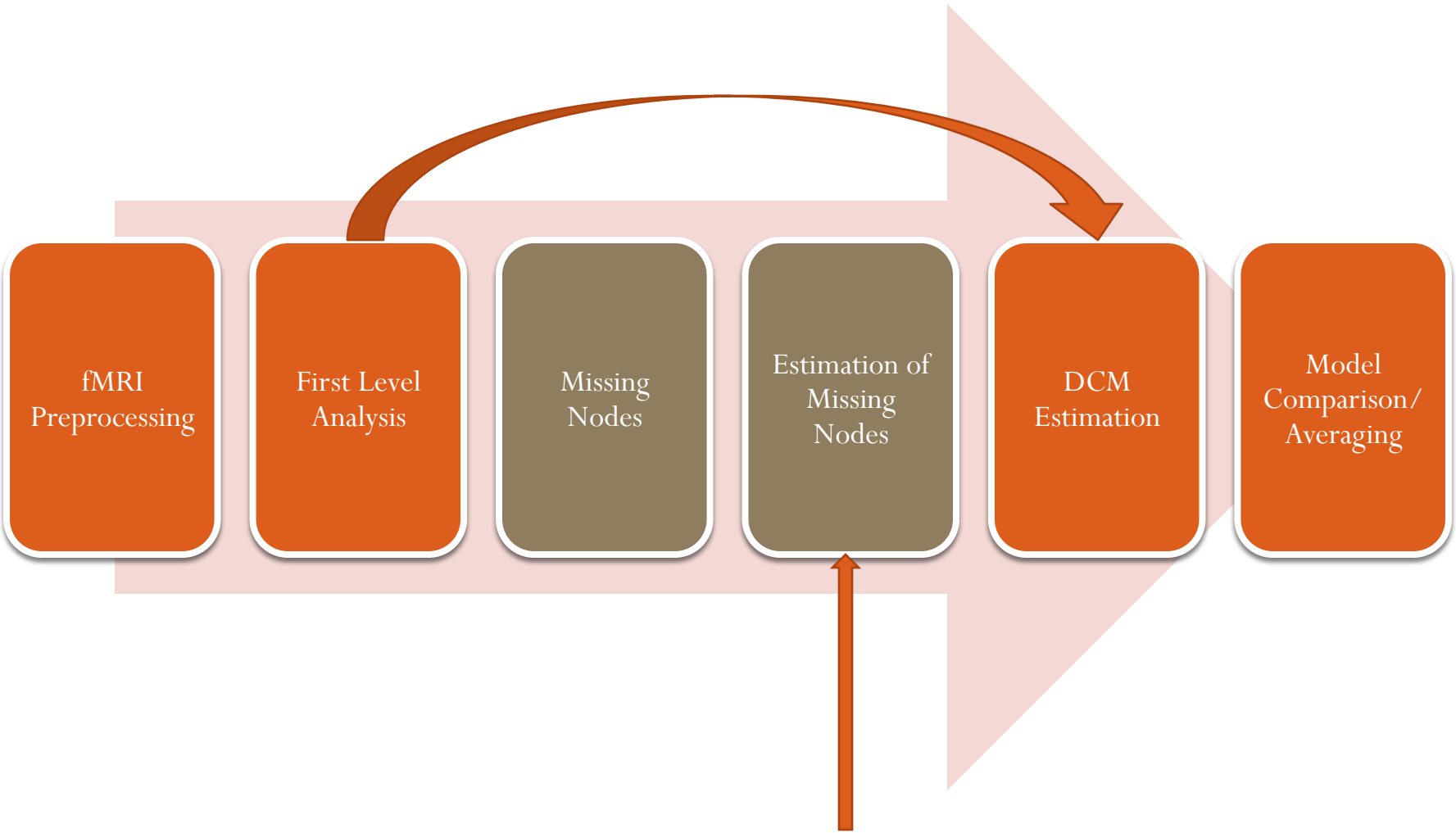
- Develop approach(es)to estimate the time courses associated with missing regions as a preprocessing step.

- Analyze the effect of the estimation schemes on:
  - Classification of subjects based on model evidence
  - Ranking of subjects

- Compare usage of estimation scheme with traditional methods
  - Using a more relaxed p-value
  - Excluding regions/subjects.

- Validation using real datasets.

Workflow

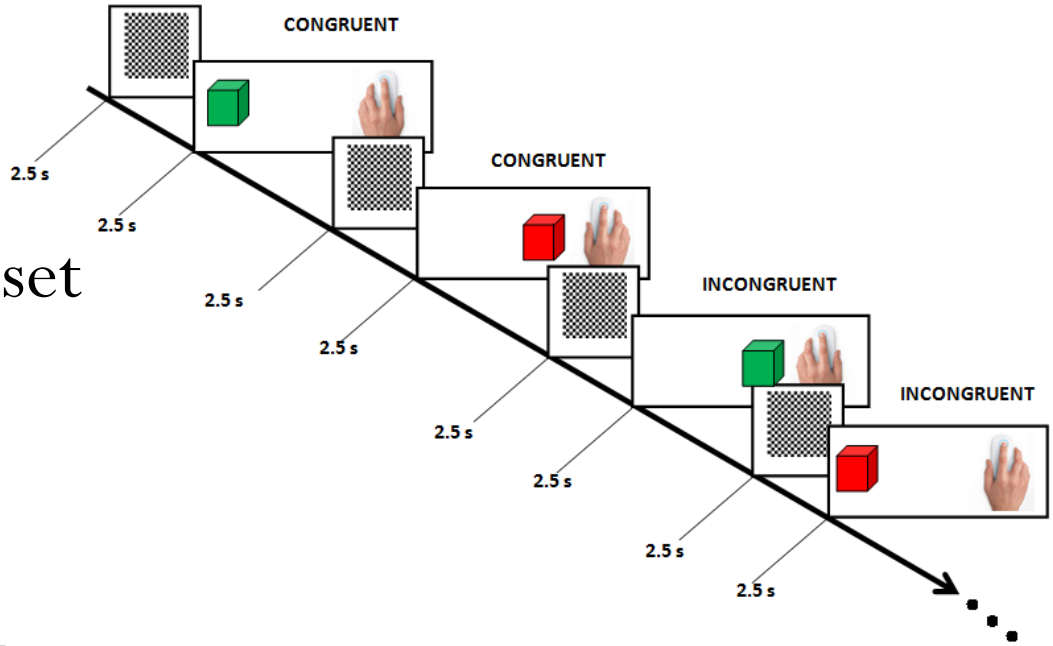

fMRI Preprocessing
First Level Analysis
Missing Nodes
Estimation of Missing Nodes
DCM Estimation
Model Comparison/ Averaging

Estimation Methods

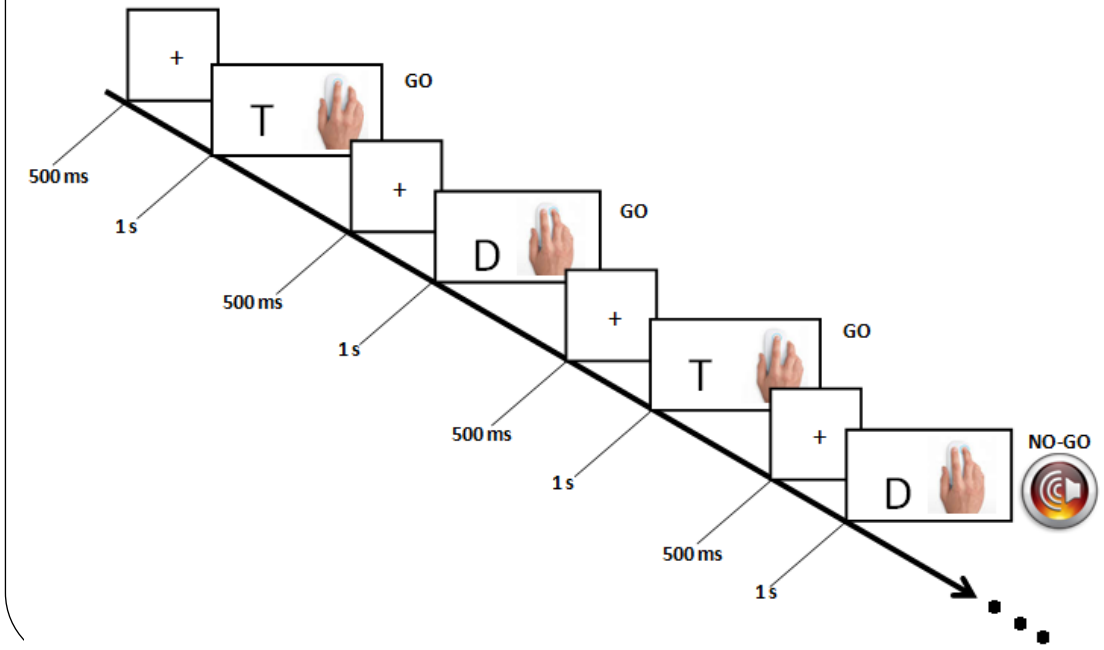
- Zero-Filling
- Noise-Filling (analogous to using high-p-value)
- Average-Filling
- Expectation Maximization

Real Datasets

- Simon Task Dataset



- Go-No/Go Task Dataset

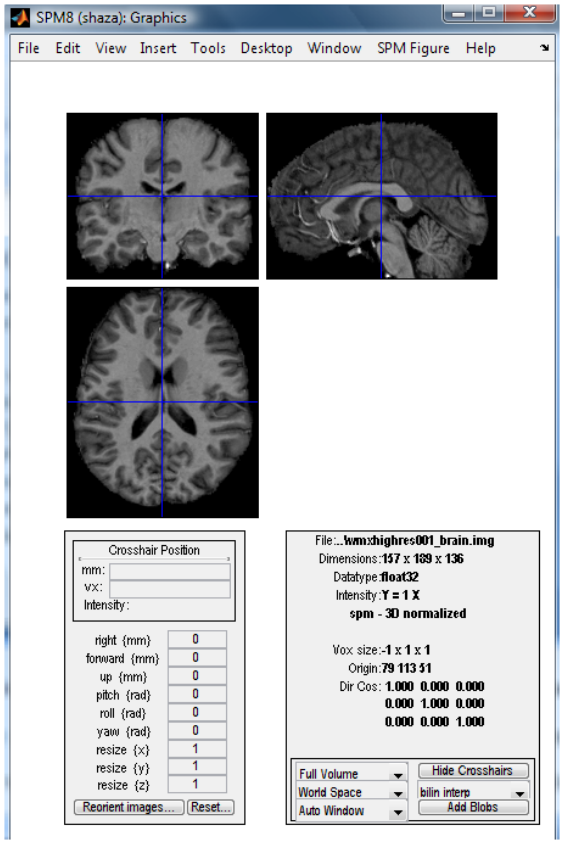


fMRI preprocessing using SPM

- Realignment of functional images to remove motion artifacts
- Slice-timing correction
- Co-registration
  - Between sessions/subjects
- Segmentation of structural image using default tissue probability maps as priors
- Registration with prior tissue probability maps from segmentation
- Spatial normalization
- Smoothing (8 mm kernel)

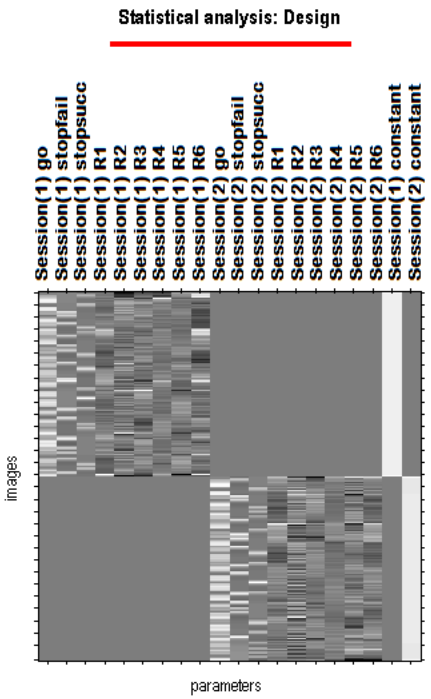
UCL SPM

By members & collaborators of the Wellcome Trust Centre for Neuroimaging



fMRI model specification and statistical analysis

- Categorical responses were modeled using the stimulus onset times and movement parameters from realignment
- Conditions were specified for each dataset
- Estimation of the GLM parameters was done using a Bayesian approach (using a VB algorithm)
- Contrast vectors were applied to the results to produce statistical parametric maps (SPMs)



Methods

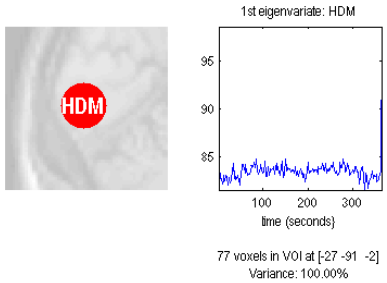
- VB were used for the estimation of the parameters for each DCM.
- RMSE measured the difference between the computed parameters before and after estimation of missing data. (between the SPM computed parameters with the full data, and the computed SPM parameters vector after estimating the missing data)

$$\theta = \{A, B, C, h\} \qquad RMSE = \sqrt{\frac{1}{N}\sum(\hat{\theta} - \theta)^2}$$

- Mutual Information was computed between the initial BOLD signal and the estimated BOLD signal after missing data estimation.

$$MI(x,y) = \sum_{y}\sum_{x} p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

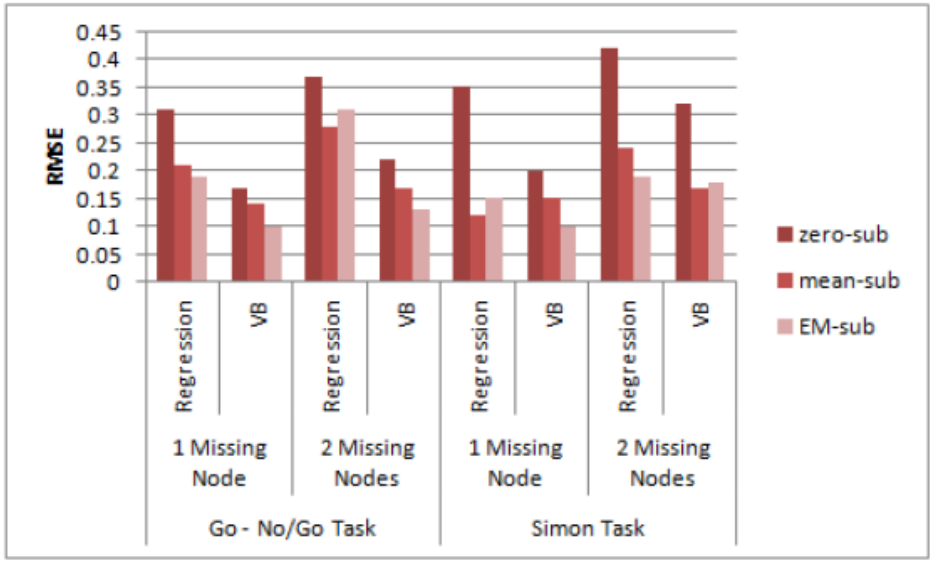
Methods



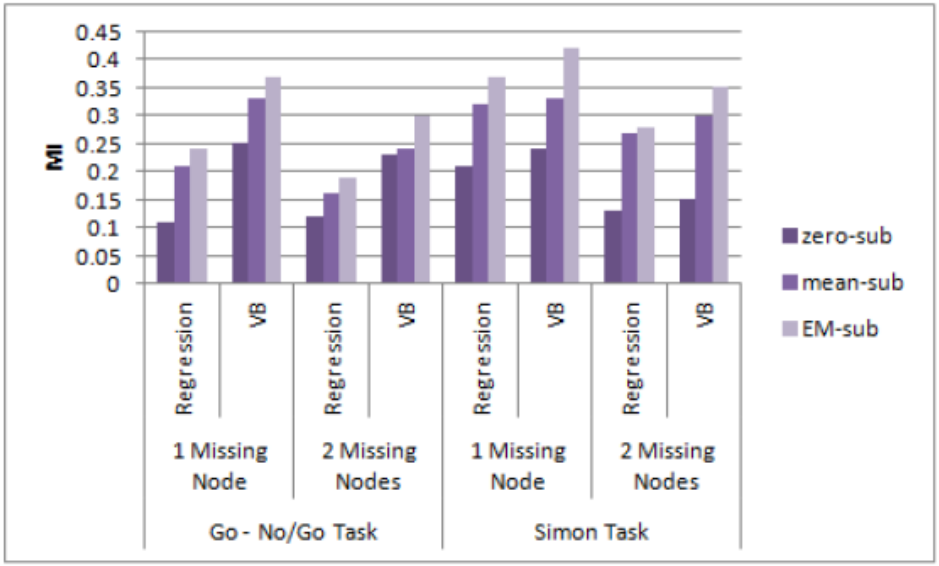
- VOI time series from the datasets were based on centers of peak activation (8mm sphere). Node signal=1st eigenvariate of VOI.
- 4 nodes were considered for each Go/No-Go subject and 3 nodes for each Simon subject.
  - P-value was tweaked to force nodes to drop
  - Any extra nodes were ignored for all subjects
  - Missing nodes were estimated
- Estimation of parameters for all subjects using VB
- EM-substitution vs. mean-substitution vs. zero-substitution

MI increases, RMSE decreases

Comparison of RMSE and MI



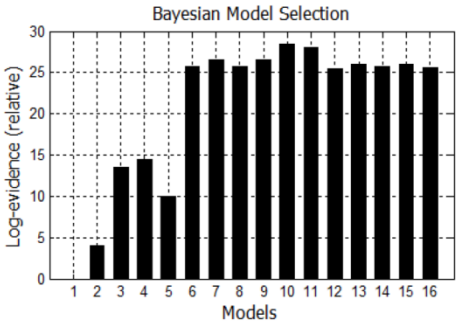
Average RMSE between initial and final parameters

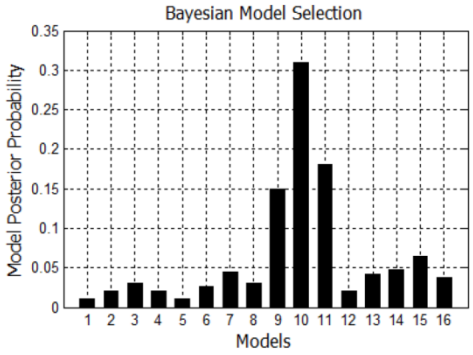


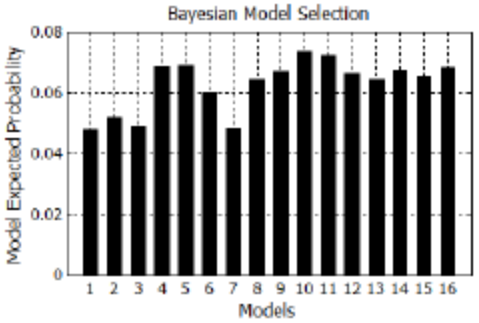
Average MI between predicted and measured response

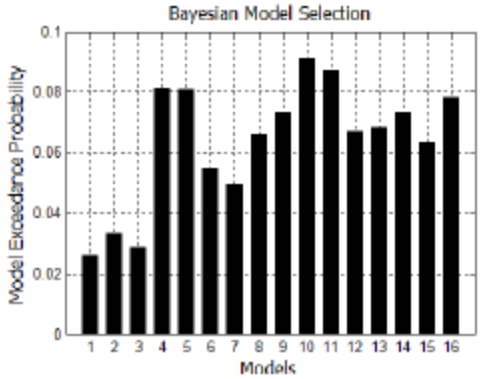
Comparing the different models by specifying alternative models for Go-No/Go Dataset

- Fixed Effects Analysis
- Random Effects Analysis





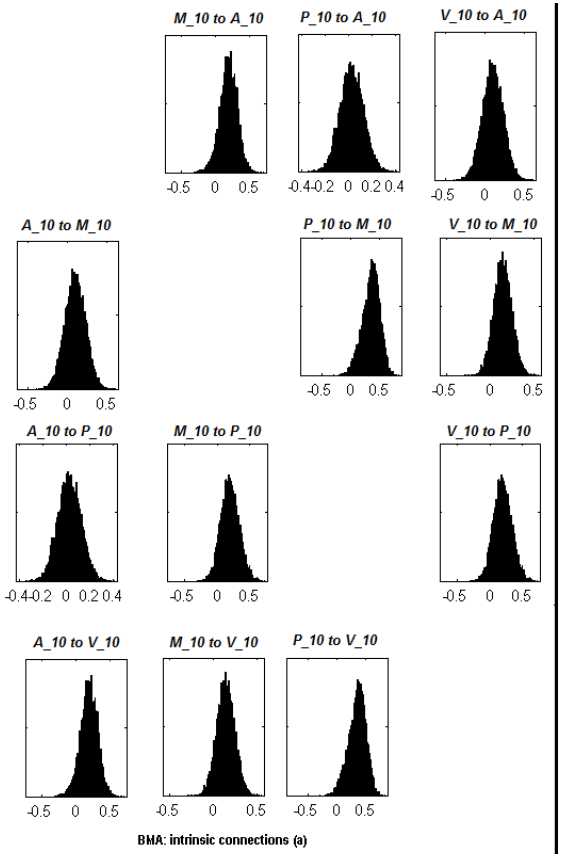


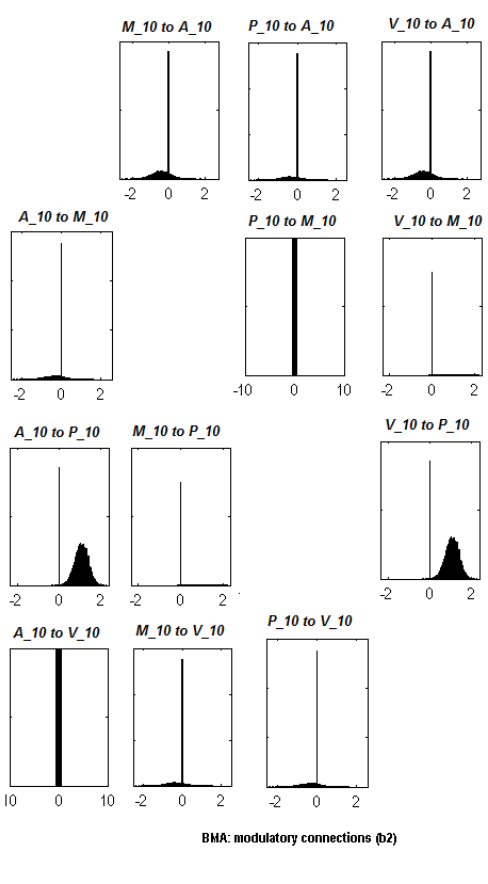


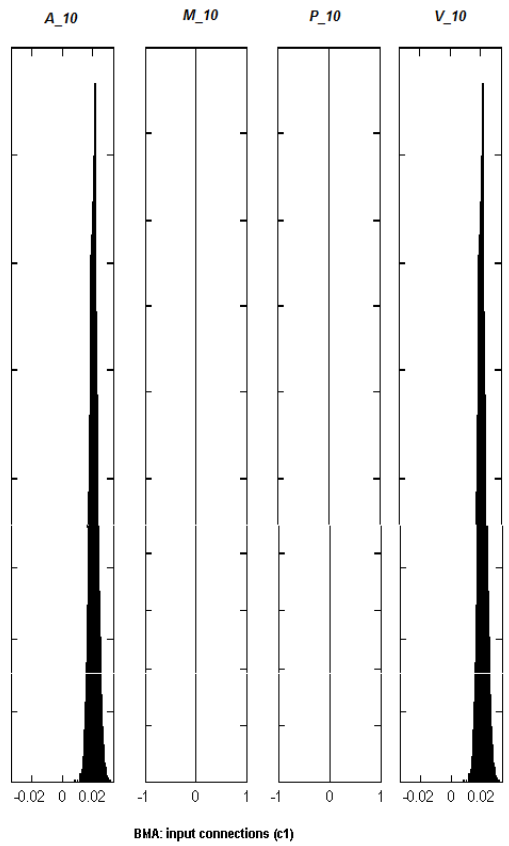
BEST MODEL FOR EACH SUBJECT (GO/NO GO TASK) USING FFX AND RFX

Subject	Missing Node(s)	Best Model FFX	Best Model RFX
1	yes	10	10
2	yes	10	10
3	no	6	6
4	no	8	9
5	yes	11	10
6	no	9	9
7	no	7	7
8	no	10	8
9	yes	12	10
10	yes	10	10
11	no	13	13
12	no	6	6
13	no	11	11
14	no	8	8
15	yes	10	10
16	no	7	7
17	no	6	6
18	yes	10	10
19	yes	9	9
20	no	7	7
21	no	8	9

Bayesian model averaging over all 16 models for Go-No/Go dataset







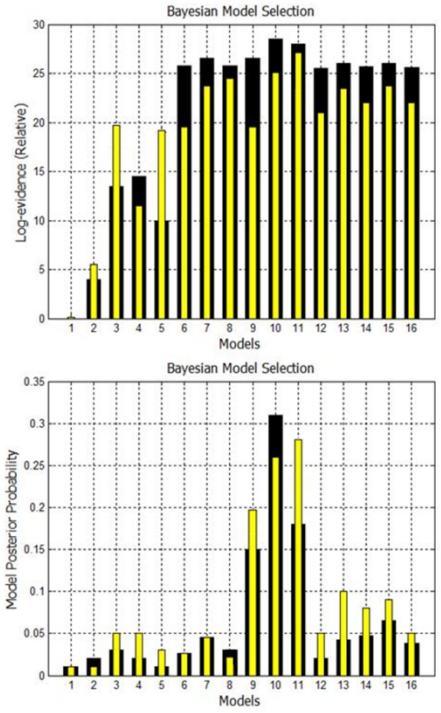
Finding nodes using a less conservative p-value

- Explore whether the evidence is higher for an EM estimated missing node or increasing the p-value until a noise point eventually emerges nearby

- For the Go/No-Go Task Dataset a p-value of 0.001 was set
  - 8/21 total subjects had 1 or more missing nodes.
  - The p-value was increased in increments of 0.005 up to 0.1 Family-wise corrected.

Finding nodes using a less conservative p-value

- Black bars are for estimation of missing nodes with EM.
- Yellow bars are for using the higher p-values to get the missing nodes.



INCREASED P-VALUES TO GET MISSING NODES IN GO-NO/GO TASK DATASET

Subject	Missing nodes at p-value 0.001	Highest p-value to force node to emerge
1	Yes	0.025
2	Yes	0.045
3	No	
4	No	
5	Yes	0.05
6	No	
7	No	
8	No	
9	Yes	0.035
10	Yes	0.1
11	No	
12	No	
13	No	
14	No	
15	Yes	0.01
16	No	
17	No	
18	Yes	0.1
19	Yes	0.05
20	No	
21	No	

Conclusions

- Missing data approaches can be used as a prior step in DCM to compute missing nodes.
- EM yields the highest classification accuracy using a simple loss function and highest model evidence for various dataset sizes and varying numbers of model choice.
- In real data, computation of missing nodes and model evidence was possible in 100% of subjects compared to 62% and 48% if no preprocessing was performed.
- The ability to compute the model evidence for all cases improves the ranking of subjects and Bayesian model averaging.

Thank you

Questions???