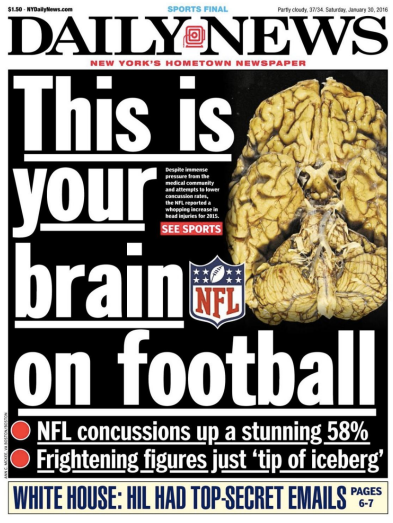
Application of Bayesian spatio-temporal models to the analysis of brain imaging data

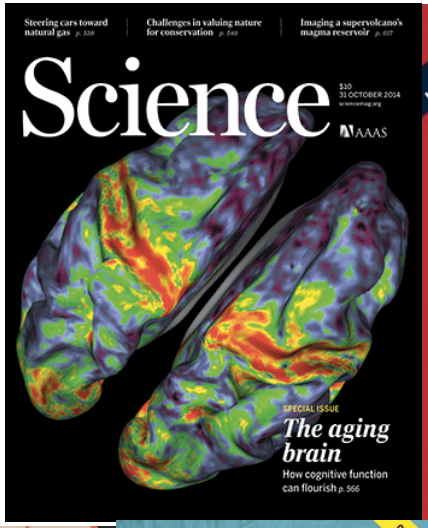
Michele Guindani

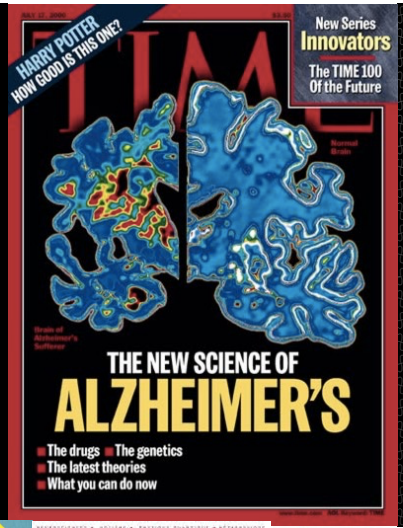
Department of Statistics
University of California, Irvine

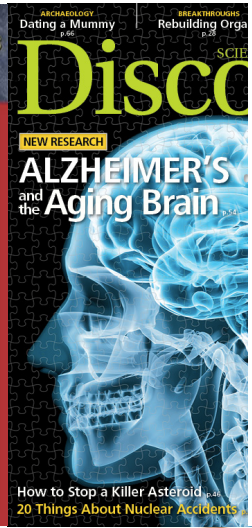
CBMS: Regional Conference On Spatial Statistics
University of California, Santa Cruz

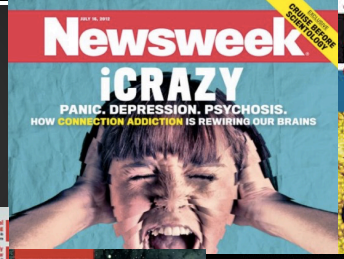
Brain Imaging in the News

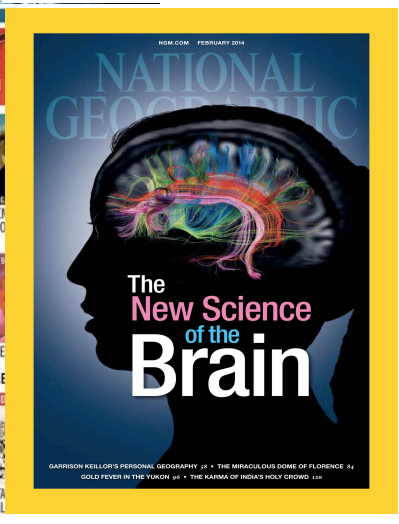












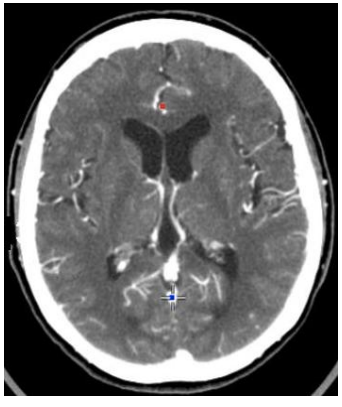
Brain Imaging Technologies

- Statistical methods play a crucial role in the analysis of brain imaging data
- Specific methods depend on the type of technology employed to obtain the data
- Brain imaging technologies can be separated into two major categories:
  - Structural Brain Imaging
  - Functional Brain Imaging
- Each category contains a number of specific technologies

Structural Brain Imaging

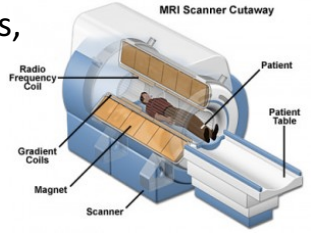
- Structural brain imaging studies the anatomical structure of the brain, e.g. in order to assist the diagnosis of brain injury, and the diagnosis of certain diseases.

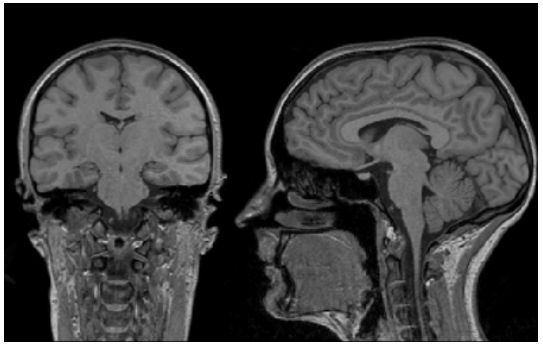
Computed Axial Tomography (CT)
provides images of bone, soft tissues and air. Quick/clinical evidence of trauma, e.g. stroke



http://www.cedars-sinai.edu

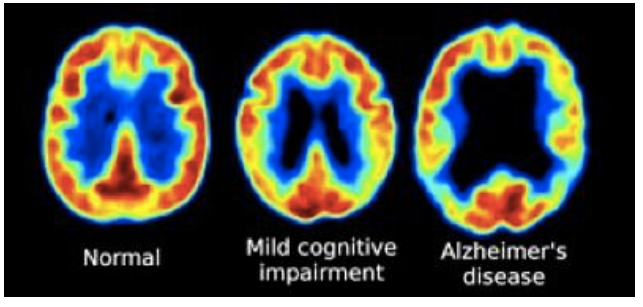
MRI SCAN
uses a powerful magnetic field to create detailed images of the organs (abnormalities, tumors, lesions, …)/no x-rays
most important diagnostic imaging modality





http://www.ucl.ac.uk/

Positron Emission Tomography (PET)
a radiotracer is injected, which release gamma waves detected by the scanner. Initially, for detecting brain activation; now for studying neurotransmitters, effect of pharmaceutical drugs, expression of genes in the brain

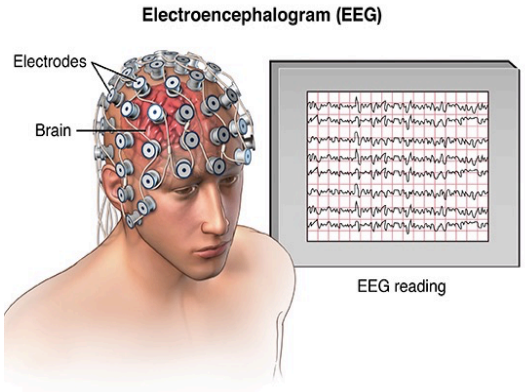


http://www.ed.ac.uk/

Functional Brain Imaging

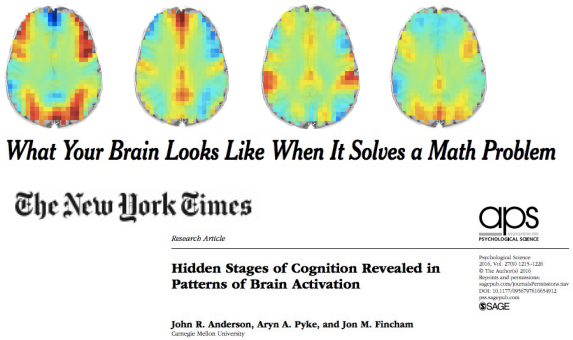
- Functional brain imaging is used to study brain functioning, in terms of its specialization (which parts of the brain respond to a given cognitive or sensory challenge) and integration (how different brain areas interact). Functional brain imaging is often employed in the study of cognitive and affective processes.

Electroencephaography (EEG)
records electrical activity of the brain. Typically noninvasive, with electrodes placed along the scalp. Used for diagnostic purposes (e.g. epilepsy, sleep disorders, stroke…)

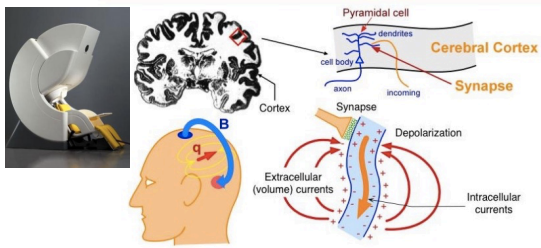


http://www.olavkrigolson.com/

Functional MRI (fMRI)
Noninvasive technique, used to map brain activity, showing which parts of the brain are involved in a particular mental process. Experiments conducted in MRI scanner



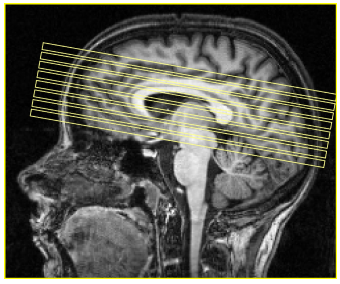
Magnetoencephalography (MEG)
Noninvasive technique, measures the magnetic field around the head to reconstruct the electrical neuronal activity in the brain. High temporal resolution, adaptable to many functions (sensory, motor, language, memory cortex)



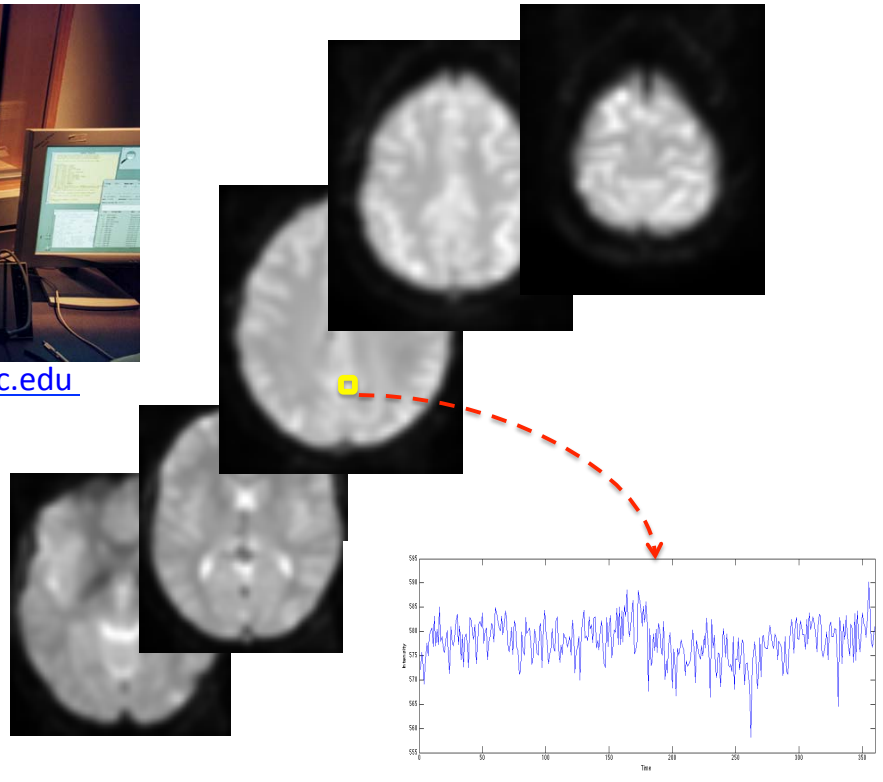
http://ilabs.washington.edu/

A primer in fMRI data...

Source: photos.uc.wisc.edu



Source: www.anc.ed.ac.uk



- ☞ Indirect measure of brain activity as changes in blood flow, typically collected during a sensorimotor task.
- ❑ Observed data ☞ time series of the blood oxygenation level dependent (BOLD) response, at each voxel in the brain.

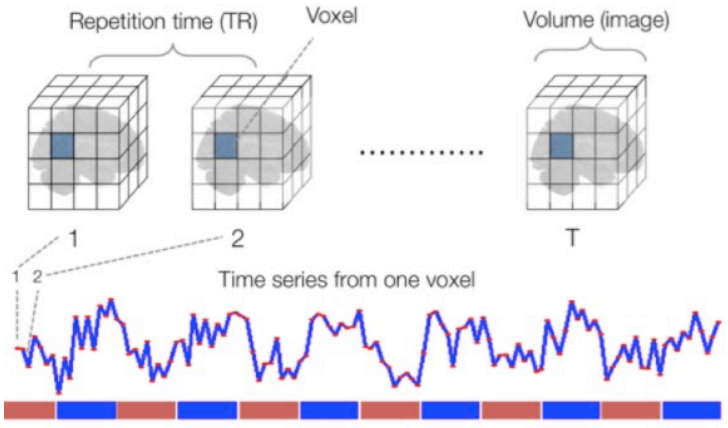
Characteristics of fMRI data

- Big data problem: a sequence of 3-D MR images ($\approx$ 100,000 voxel each) are typically collected in a single experiment (100-1,000 images over time) for a single subject
- A study may involve multiple subjects (typically, between 10-50)
- the signal is relatively weak
- the data exhibit a complicated spatio-temporal structure which can be predictive of psychological and clinical states.

Spatio-temporal resolution of fMRI data

The **temporal resolution** determines our ability to **separate brain events** in time

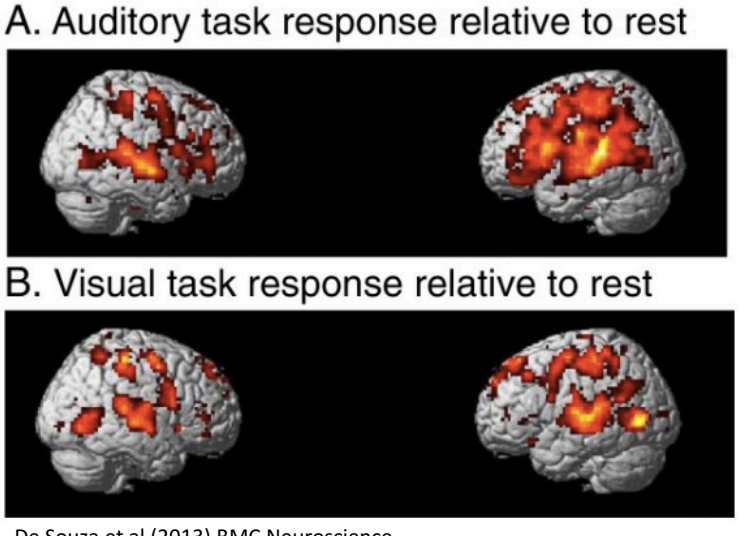
- It is determined by how quickly each image is acquired: the TR, repetition time between volumes, or TR, varies across studies but typical values have historically been about 2-3 seconds



Lindquist, Wager - Principles of fMRI

The **spatial resolution** allows to distinguish **changes in brain activity** across **different spatial locations**, e.g. in response to a task

- fMRI data have relatively high spatial resolution (3x3x5 mm corresponding to 64x64x30 voxels per image)
- Higher spatial resolution than PET, EEG and MEG but less than structural MRI
- Spatial resolution is impacted by preprocessing: slice-time correction, re-alignment, co-registration, normalization, smoothing



De Souza et al (2013) BMC Neuroscience

Major interests in fMRI studies

- There are several common objectives in the analysis of fMRI data:

1. Activation detection: identifying regions of the brain activated by a certain task

Major interests in fMRI studies

- There are several common objectives in the analysis of fMRI data:

1. Activation detection: identifying regions of the brain activated by a certain task

2. Connectivity: determining distributed networks that correspond to brain function:

Major interests in fMRI studies

- There are several common objectives in the analysis of fMRI data:

① Activation detection: identifying regions of the brain activated by a certain task

② Connectivity: determining distributed networks that correspond to brain function:

  - Functional Connectivity: undirected association between two or more fMRI time series.
  - Effective Connectivity: Directed influence of one brain region on the activity recorded in other brain regions.

Major interests in fMRI studies

- There are several common objectives in the analysis of fMRI data:

① Activation detection: identifying regions of the brain activated by a certain task

② Connectivity: determining distributed networks that correspond to brain function:

  - Functional Connectivity: undirected association between two or more fMRI time series.
  - Effective Connectivity: Directed influence of one brain region on the activity recorded in other brain regions.

③ Prediction about psychological or disease states, possibly by combining multiple types of data (integrative analyses/imaging genetics)

Models for Activation Detection

1. Single-subject modeling:

   - Detect activation of brain regions in response to a task
   - Take into account spatio-temporal dependences
   - Capture (functional/effective) connectivity, i.e. how different regions interact
   - High-dimensional volumes at each time point

Models for Activation Detection

1. Single-subject modeling:
   - Detect activation of brain regions in response to a task
   - Take into account spatio-temporal dependences
   - Capture (functional/effective) connectivity, i.e. how different regions interact
   - High-dimensional volumes at each time point
2. Multi-subject modeling:
   - Integrate information across subjects
   - Describe spatial correlation of voxel time series within and across subjects.

The General Linear Model of Friston et al. (1994)

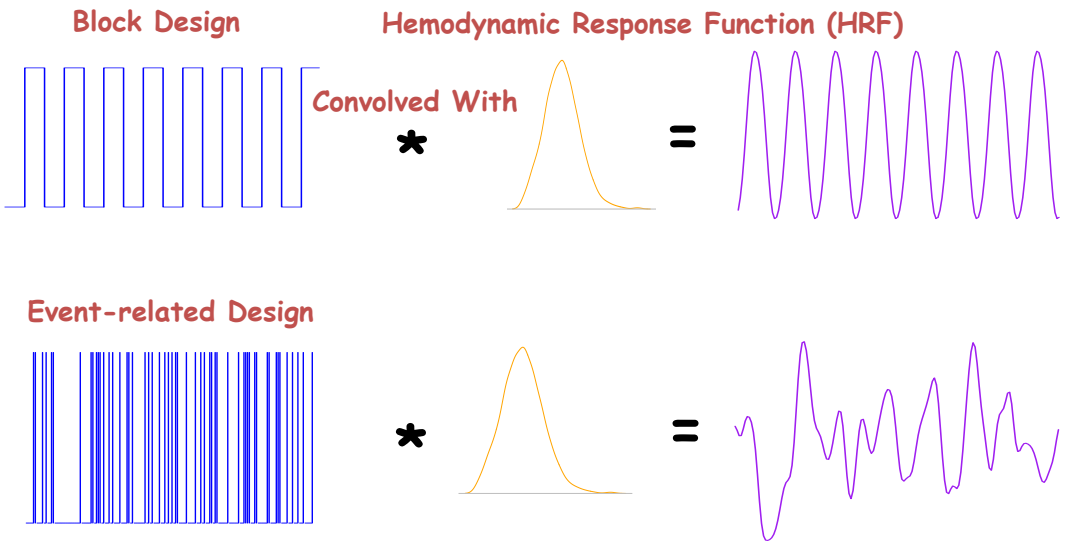
fMRI = BOLD Signal + Noise

The General Linear Model of Friston et al. (1994)

fMRI = BOLD Signal + Noise

BOLD Signal: convolution of stimulus function $v(t)$ and HRF $h(t)$ - time lapse between $v(t)$ and vascular response

$$x(t) = (v * h)(t) = \int_0^t v(s)h(t-s)ds$$



Noise: hardware and physiological

Model for a Single Subject

$$Y_v = X_v \beta_v + \varepsilon_v, \quad v = 1, \ldots, V$$

- $Y_v = (Y_{v1}, \ldots, Y_{vT})^T$ time-series data for voxel v.
- X_v the convolved design matrix $T \times p$ stimuli.
- $\beta_v = (\beta_{v1}, \ldots, \beta_{vp})^T$ vector of regression coefficients.
- The regression coefficients capture the dependence of the BOLD signal recorded at voxel v with the experimental stimulus (for a given HRF).
- $\varepsilon_v \sim N_T(0, \Sigma_v)$.

Taking into account spatial dependence

- Spatial correlation is expected in a voxel-level analysis of fMRI data because the response at a particular voxel is likely to be similar to the responses of neighboring voxels.
- In Bayesian modeling, spatial dependence between brain voxels (or regions) is captured by imposing spatial priors on the model parameters:
  - Gaussian Markov Random Field (GMRF)/Conditional Autoregressive (CAR) priors
  - Gaussian Process priors
  - Spatial Variable Selection priors

Gaussian Markov Random Field

- Many approaches use Gaussian Markov random field (GMRF) priors on the jth regression coefficient vector $\boldsymbol{\beta}_{(j)} = (\beta_{1,j}, \ldots, \beta_{V,j})^T$, $j = 1, \ldots, p$ (e.g. Gossl, Auer, Fahrmeir, 2001; Quirós, Diez, Gamerman, 2009),

$$p(\boldsymbol{\beta}_{(j)}|\lambda) \propto \exp(-\frac{1}{2}\lambda\boldsymbol{\beta}_{(j)}^T\mathbf{Q}\boldsymbol{\beta}_{(j)}),$$

with precision matrix $\mathbf{Q}$ having elements

$$\mathbf{Q}_{v,k} = \begin{cases} n_v, & v = k \\ -1, & v \sim k \\ 0, & \textit{otherwise}, \end{cases}$$

with n_v the number of neighbors of voxel v, and with $v \sim k$ denoting that voxels v and k are neighbors.

Gaussian Markov Random Field

- The prior with precision matrix **Q** is equivalent to

$$p(\boldsymbol{\beta}_{(j)}|\lambda) \propto \exp\{-\frac{1}{2}\lambda \sum_{v\sim k}(\beta_{v,j} - \beta_{k,j})^2\},$$

  from which we have

$$\beta_{v,j}|\beta_{-v,j}, \lambda \sim N(\frac{1}{n_v}\sum_{k\sim v}\beta_{k,j}, \frac{1}{n_v\lambda}),$$

  where $\beta_{-v,j} = \{\beta_{l,j}; l \neq v\}$.

- ⇨ Activity in a voxel is estimated by the response magnitude in the voxel as well as in neighboring voxel.
- ⇨ Expectation that activity takes form in regions (not isolated)

Gaussian Markov Random Field

- The matrix Q is of less than full rank $\Rightarrow$ Improper
- Advantages in this setting:

1. there is no need to specify a prior for the mean level of β, i.e. the proportion of activated voxels.
2. the intrinsic GMRF gives the maximum correlation possible, provided that the first-order neighborhood is used.

- For a review of GMRF priors, see “the bible” Banerjee, Carlin and Gelfand (2014)

"Laplacian prior" by Penny et al (2005)

- Penny et al (2005) considered a spatial prior on the regression coefficient vector $\beta_{(j)}$ of the type

$$\begin{aligned} \beta_{(j)} &\sim N(0, \alpha_j^{-1}(\mathbf{S}^T\mathbf{S})^{-1}) \\ \alpha_j &\sim Ga(a, b), \end{aligned}$$

  α_j is a spatial precision parameter.

  $\mathbf{S}$ a $V \times V$ is spatial kernel matrix equal to the Laplacian operator L ($= I - A$ where A is the adjacency matrix so that -1 if two voxels are adjacent).

- ⇨ If $\mathbf{w}_{(j)} = L\beta_{(j)}$ then $w_{v,j}$ is equal to the sum of the differences between $\beta_{v,j}$ and its neighbors, for $v = 1, \ldots, V$
- This is one of the spatial priors implemented in SPM.

“Laplacian prior” by Penny et al (2005)

- This prior enforces smoothness by penalizing differences between neighboring voxels and is a prior that is often used in the analysis of EEG.

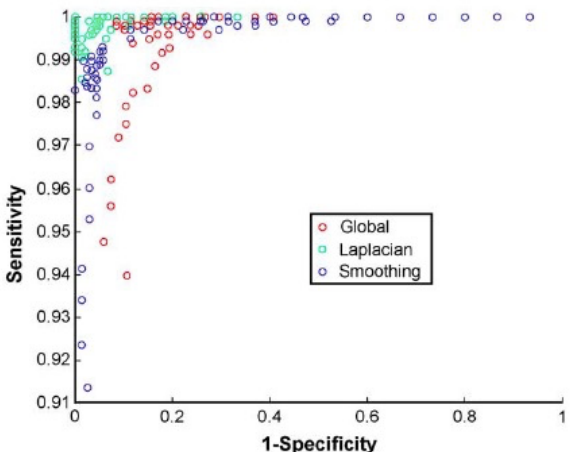
Simulation experiment

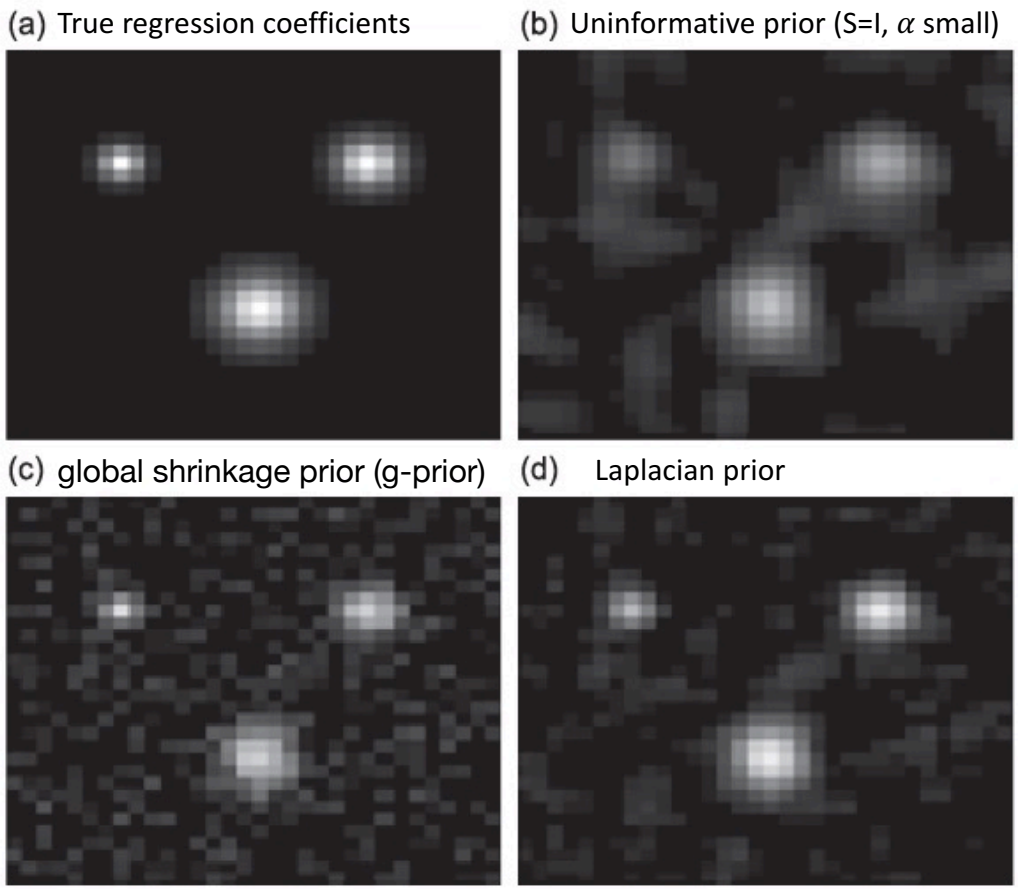
Block Design matrix **X** of dimension T x 2

T=40 time points 32 x 32 images

$\alpha_1 = \alpha_2 = 1$

The Laplacian prior is superior to the use of a global shrinkage prior in terms of Bayesian evidence, estimation accuracy and area under the ROC curve.





Penny et al. (2005) Bayesian fMRI time series analysis with spatial priors

“Laplacian prior” by Penny et al (2005)

Real Data

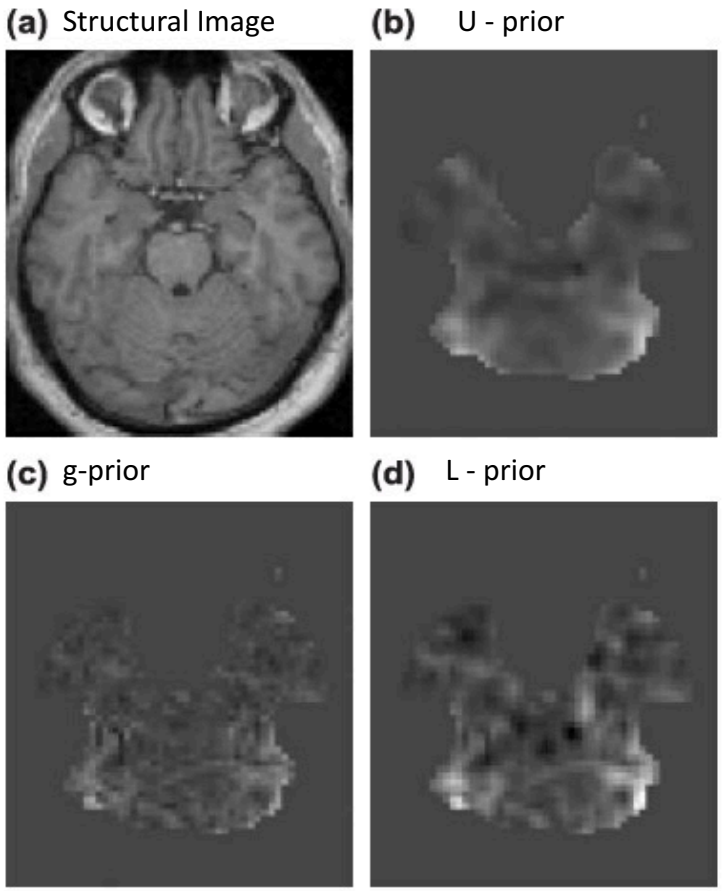
Event-related study where greyscale images of faces were presented for 500 ms, replacing a baseline of an oval chequerboard.

Whole brain images consisting of 24 transverse slices were acquired every 2 s resulting in a total T = 351 scans.

Analysis conducted on slices -24mm (MNI coordinates)

These maps show large responses in bilateral fusiform and occipital cortex as previously reported (Henson et al., 2002).

However, use of the Laplacian prior results in much higher estimates of effect size in these regions.



Penny et al. (2005) Bayesian fMRI time series analysis with spatial priors

- Graph-Laplacian priors have been considered by Harrison et al (2008); proper CAR by Derado et al. (2013).

Spatially adaptive CAR models

- Liu, Berrocal, Bartsch and Johnson (2016) have recently proposed a spatially adaptive CAR model to map out functionally relevant brain regions to aid pre-surgical evaluation of tumor resections

- Decision problem for pre-surgical data: False negatives (i.e., voxels falsely classified as null) are more dangerous to the patient than false positives (i.e., voxels falsely classified as activated or de-activated) since the former may result in the surgeon damaging healthy tissue vital to quality of life (Haller and Bartsch, 2009)

Spatially adaptive CAR models

They consider z-images (contrast summary images):

$$Y_v = \mu_v + \varepsilon_v \quad \varepsilon_v \sim N(0, \sigma_v)$$

with

$$\mu_v | \mu_{-v} \sim N\left(\sum_j \frac{w_{vj}}{w_{v+}} \mu_j, c_v \sigma_v^2 \right)$$

and

$$\log(\sigma_v^2) | \log(\sigma_{-v}^2) \sim N\left(\frac{w_{vj}}{w_{v+}} \log(\sigma_j^2), \frac{\lambda^2}{w_{v+}} \right),$$

with $w_{vj} = 1$ if $v \sim j$ and $w_{v+} = \sum_j w_{vj}$.

Spatially adaptive CAR models

The model is completed by a prior on c_v:

$$c_v = \frac{p_v}{1 - p_v} \quad v = 1, \ldots, V$$

with $p_v \sim Beta(\alpha, \beta)$.

⇨ The posterior full conditional distribution of μ_v is given by:

$$\mu_v | y_v, \mu_{-v}, p_v, \sigma_v \sim N(p_v y_v + (1 - p_v)\bar{\mu}_v, p_v \sigma_\nu^2),$$

with $\bar{\mu}_v = \sum_j \frac{w_{vj}}{w_{v+}} \mu_j$.

⇨ Smoothing more in some regions ($p_v < 0.5$), and less in others ($p_v > 0.5$), e.g. at the interface between activated and null regions.

Gaussian Process Priors

- Harrison and Green (2010) propose the use of a GPP for the regression coefficients in the GLM model:

$$\beta_{(j)} \sim N(\mathbf{0}, K_j(\cdot)) \quad j = 1, \ldots, p$$

  where $K_j(\cdot)$ is an appropriately defined covariance function.

- Continuous random field *vs* areal grid
- Greater flexibility, possibility to make predictions at locations where measurements have not been taken.

Gaussian Process Priors

- Assuming stationarity, Harrison and Green (2010) consider

$$K(v, v') = \sigma^2 \exp\left(-\frac{d_{ij}^2}{2\phi}\right),$$

  with $d_{ij} = |v - v'|$ is the modulus of the distance between two voxels, and σ^2 and ϕ are the variance and range parameters.

- In order to be able to fit the model they consider a mean-field approximation of the posterior distribution.

Spatial Variable Selection priors

- The task of selecting activated voxels is a variable selection problem ⇨ identification of the nonzero $\beta_{v,j}$

Spatial Variable Selection priors

- The task of selecting activated voxels is a variable selection problem ⇨ identification of the nonzero $\beta_{v,j}$
- Mixture prior (Spike-and-slab prior) on β_v

$$\beta_v \sim \gamma_v \, N(0, \tau) + (1 - \gamma_v) \, \delta_0, \qquad v = 1, \ldots, V,$$

This prior selects activated voxels.

Spatial Variable Selection priors

- The task of selecting activated voxels is a variable selection problem ⇨ identification of the nonzero $\beta_{v,j}$

- Mixture prior (Spike-and-slab prior) on β_v

$$\beta_v \sim \gamma_v \, N(0, \tau) + (1 - \gamma_v) \, \delta_0, \qquad v = 1, \ldots, V,$$

This prior selects activated voxels.

If a voxel is not activated, then $\gamma_v = 0$ & $\beta_v = 0$

Spatial Variable Selection priors

- The task of selecting activated voxels is a variable selection problem ⇨ identification of the nonzero $\beta_{v,j}$
- Mixture prior (Spike-and-slab prior) on β_v

$$\beta_v \sim \gamma_v \, N(0, \tau) + (1 - \gamma_v) \, \delta_0, \qquad v = 1, \ldots, V,$$

This prior selects activated voxels.

If a voxel is not activated, then $\gamma_v = 0$ & $\beta_v = 0$

If a voxel is activated, then $\gamma_v = 1$ & $\beta_v \sim N(0, \tau)$.

Modeling the prior probabilities of activation γ_v

Modeling the prior probabilities of activation γ_v

Kalus, Sämann, Fahrmeir (2014) specified a spatial probit model for the prior probabilities of activation:

$$P(\gamma_v = 1) = \Phi(\alpha_v)$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_V)^T$ following a first order intrinsic Gaussian Markov random field (IGMRF) prior:

$$p(\boldsymbol{\alpha}|\xi^2) \propto (\xi^2)^{-(V-1)/2} \exp\left(-\frac{1}{2\xi^2}\boldsymbol{\alpha}^T \mathbf{Q}\boldsymbol{\alpha}\right),$$

with Q as before and ξ^2 a variance parameter determining the degree of smoothness.

Modeling the prior probabilities of activation γ_v

- Zhang et al. (2014) have proposed a Markov Random Field (MRF) prior on the voxel-dependent γ_v for activation:

$$P(\gamma_v | d, e, \gamma_k, k \in N_v) \propto \exp(\gamma_v (d + e \sum_{k \in N_v} \gamma_v))$$

  with N_v the set of neighbors of voxel v, $d \in R$, $e > 0$.

- Here, d is a sparsity parameter controlling the expected prior number of activated voxels and e is a smoothing parameter affecting the probability of identifying a voxel as active according to the activation of its neighbors.
- Smith and Fahrmeir (2007), Lee et al (2014) considered similar priors to incorporate also anatomical prior information.

Temporal modeling

- Let's reconsider the General Linear Model of Friston (1994):

$$Y_v = X_v \beta_v + \varepsilon_v, \quad v = 1, \ldots, V$$

- Some of the temporal correlation of fMRI data is captured via the modeling of the HRF, e.g. Poisson HRF:

$$h_{\lambda_\nu}(t) = \exp(-\lambda_\nu)\lambda_\nu^t / t!$$

with voxel-dependent hemodynamic delay parameter λ_ν.

- In addition, the temporal correlation is captured through the modeling of the noise component ε_v.

Temporal modeling

- The most common choice is to impose an autoregressive structure of order q (AR(q)) on ϵ_v,

$$\epsilon_{v,t} = \sum_{j=1}^{q} w_{v,j}\epsilon_{v,t-j} + z_{v,t},$$

with $\mathbf{w}_v = (w_{v,1}, \ldots, w_{v,q})^T$ a $q \times 1$ vector of AR coefficients and $z_{v,t}$ a white noise, assuming prior distributions on the AR coefficients (Penny et al 2003, Woolrich et al, 2004, Penny et al 2005, Lee et al 2014)

Temporal Modeling: Long-range dependences

- Alternatively, as in Zhang et al. (2014), one can model the noise as a long-memory process:

$$\varepsilon_v \sim N_T(0, \Sigma_v), \;\; \Sigma_v(i,j) = [\gamma(|i-j|)]$$
$$\gamma(\tau) \approx \psi \tau^{-\alpha} \text{ as } \tau \to \infty$$

  (not negligible dependence between distant observations)

- fMRI time series behave like fractal time series and exhibit long-range autocorrelations (long memory) in time (Bullmore et al, 2004)

- Use wavelet transforms to “whiten” the data.

- Wavelets decorrelate data from long-memory processes
- Let W be the $T \times T$ matrix of the wavelet transform

$$Y_v^* = X_v^* \beta_v + \varepsilon_v^*, \quad \varepsilon_v^* \sim N_T(0, \Sigma_v^*),$$

with $Y_v^* = WY_v$, $X_v^* = WX_v$, $\varepsilon_v^* = W\varepsilon_v$.

- $\Sigma_v^* = \mathcal{W}\Sigma_v\mathcal{W}' = \text{diag}[\psi_v, \alpha_v]$, with ψ_v the innovation variance and $\alpha_v \in (0, 1)$ the long memory parameter.

❑ The variance covariance is greatly simplified by the wavelet transform.

A Spatio-Temporal Nonparametric Bayesian Variable Selection Model of fMRI Data for Clustering Correlated Time Courses

Basic Modeling Ingredients

- Zhang et al (2014) $Y_\nu = (Y_{\nu 1}, \ldots, Y_{\nu T})^T$, $T \times 1$ consider

$$Y_v = X_v \beta_v + \varepsilon_v, \quad v = 1, \ldots, V$$

with X_ν, $T \times p$ design matrix, p is the number of experimental conditions, Poisson HRF

- A spike-and-slab prior on β_ν

$$\beta_\nu \sim \gamma_\nu N(0, \tau) + (1 - \gamma_\nu)\, \delta_0, \quad \nu = 1, \ldots, V,$$

with Markov Random Field (MRF) prior on the voxel-dependent γ_v for activation

$$P(\gamma_v | d, e, \gamma_k, k \in N_v) \propto \exp(\gamma_v (d + e \sum_{k \in N_v} \gamma_v))$$

Temporal priors

- Uniform prior on voxel-dependent delay parameter λ_v of the HRF,

$$\lambda_v \sim \mathrm{U}(u_1, u_2), \quad v = 1, \ldots, V.$$

Temporal priors

- Uniform prior on voxel-dependent delay parameter λ_v of the HRF,

$$\lambda_v \sim \mathtt{U}(u_1, u_2), \quad v = 1, \ldots, V.$$

- A Dirichlet Process (DP) prior can be used on the parameters of the long-memory process, (ψ_v, α_v)

$$(\psi_v, \alpha_v)|G \sim G; G|\eta, G_0 \sim DP(\eta, G_0);$$
$$G_0 = \mathtt{IG}(a_0, b_0) \times \mathtt{Beta}(a_1, b_1)$$

We induce clustering of the fMRI time series, which can be seen as a way to detect functional connectivity

Posterior Inference

MCMC approach

- Jointly update (β, γ) with *add-delete-swap* algorithm

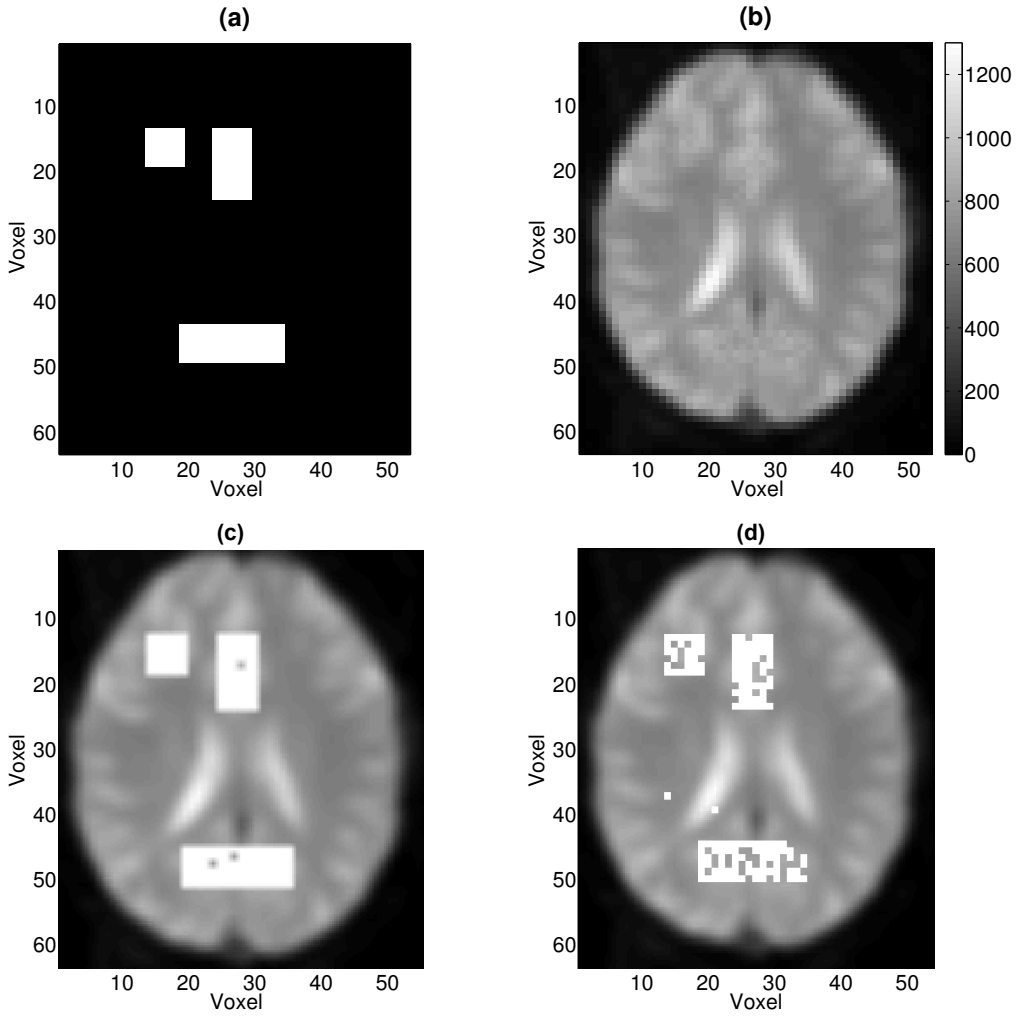
  In each iteration, randomly choose among the three moves

  - *Add*: choose a voxel which is currently not activated, and propose to activate it
  - *Delete*: choose a voxel which is currently activated, and propose it as inactive
  - *Swap*: propose to swap the activation status of two voxels

  (Savitsky et al., 2011)
- Update λ with Metropolis-Hastings (MH) schemes
- Jointly update (ψ, α) by using Gibbs sampling with auxiliary parameters, proposed by Neal 2000 for Dirichlet process (DP) mixture models

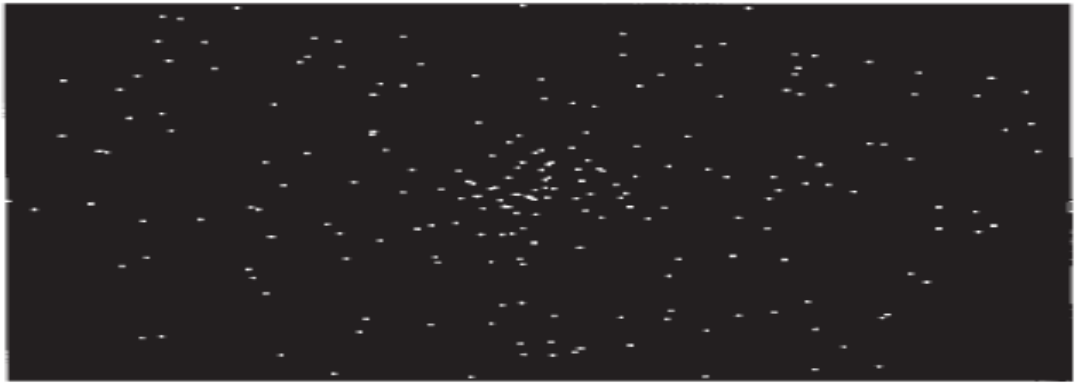
Synthetic Data

Simulate $Y_{syn} = Y + w$, where Y is simulated from our model and w is the selected slice from the real fMRI study



Real Data

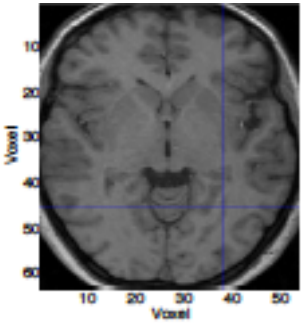
- The real fMRI data is available at `http://www.fil.ion.ucl.ac.uk/spm/data/attention/`
- A single subject, 4 different conditions "Fixation", "Attention", "No Attention", and "Stationary"
- Stimulus function: 1 for condition "Attention" and "No Attention", and 0 for condition "Fixation"
- We analyze primary visual cortex (V1), motion-selective cortical area (V5), and the posterior parietal cortex(PP)



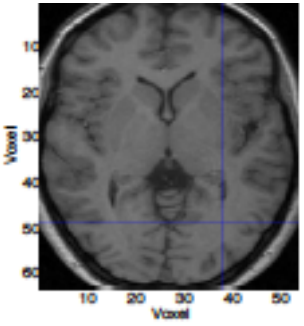
Posterior Estimation

Structural MRI images

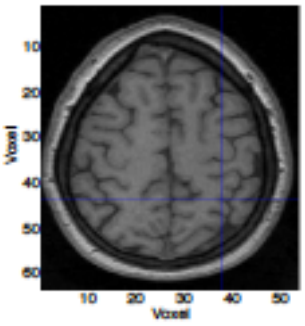
V1



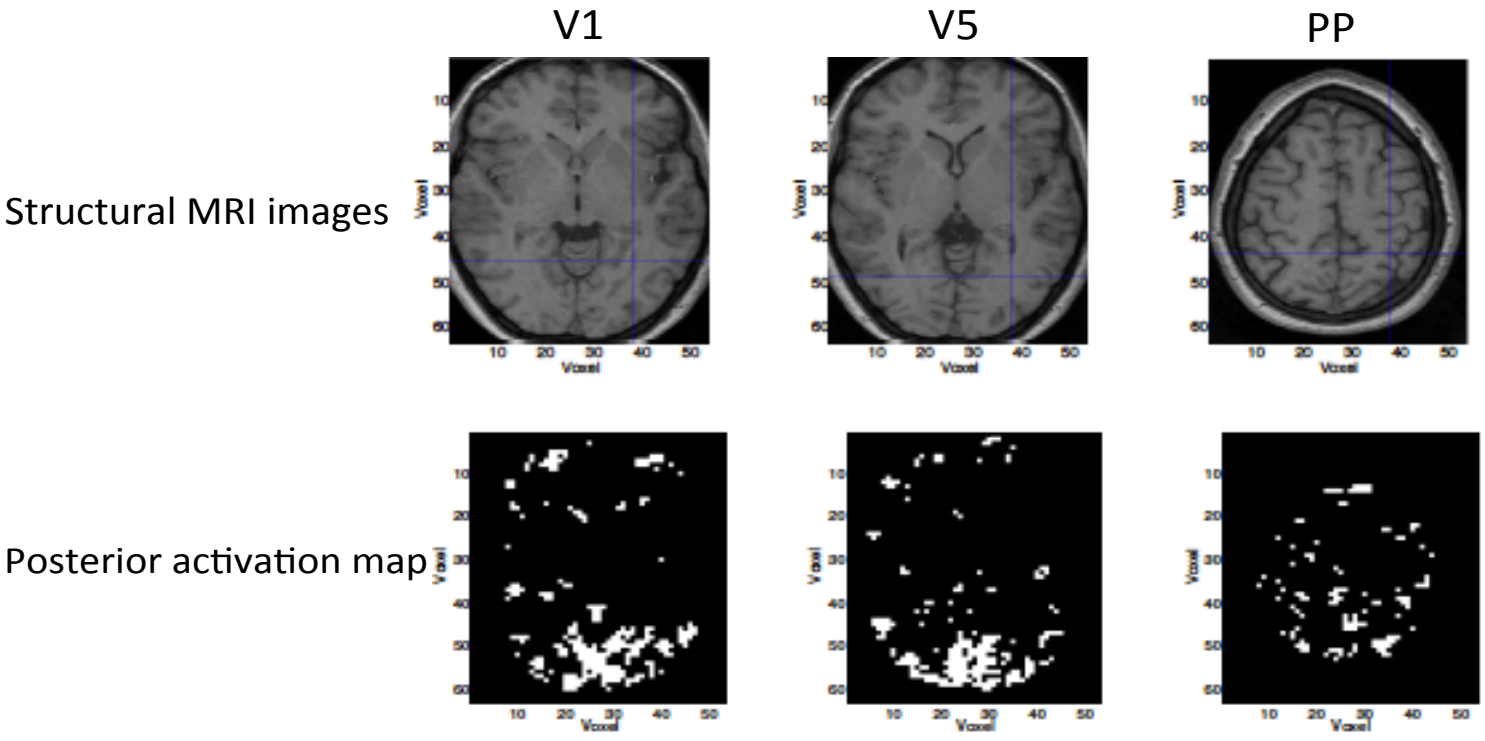
V5



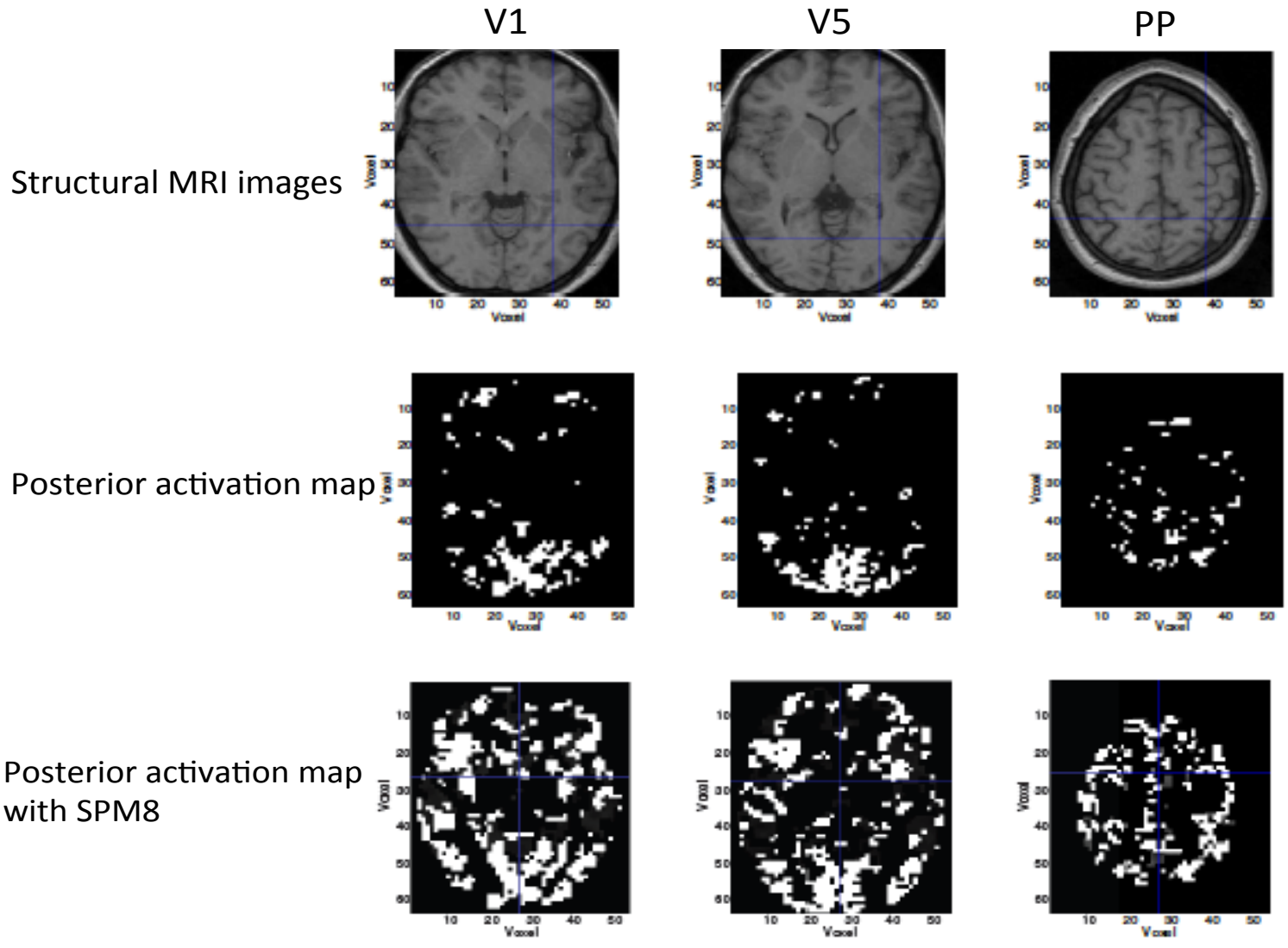
PP



Posterior Estimation

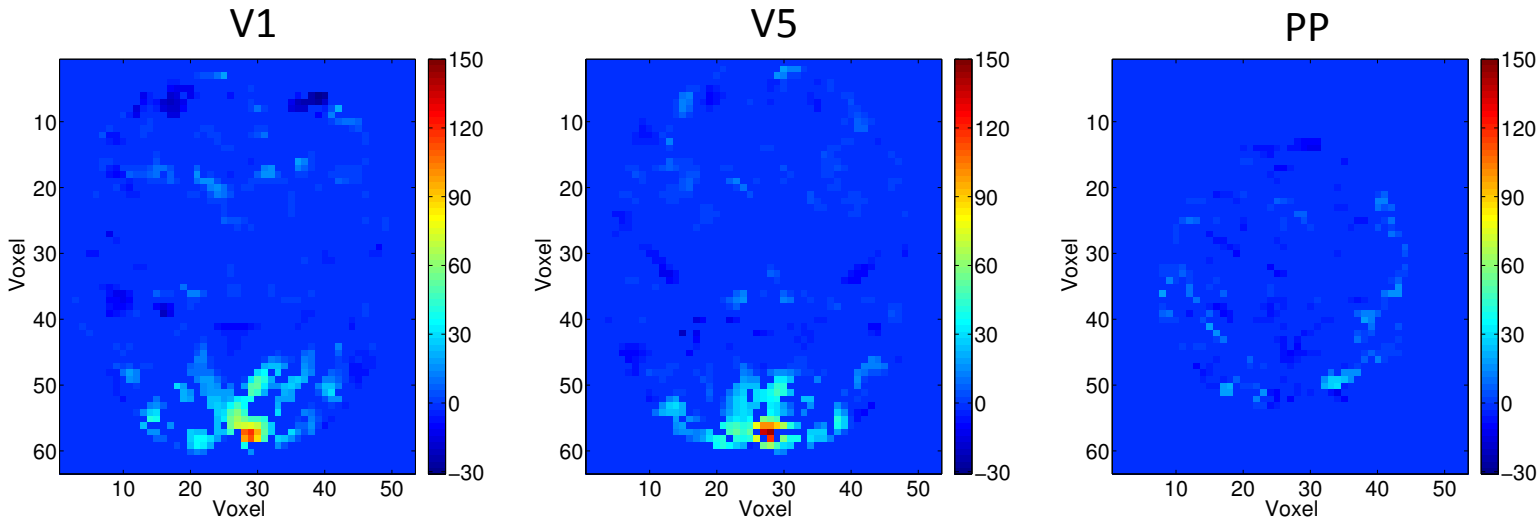


Posterior Estimation

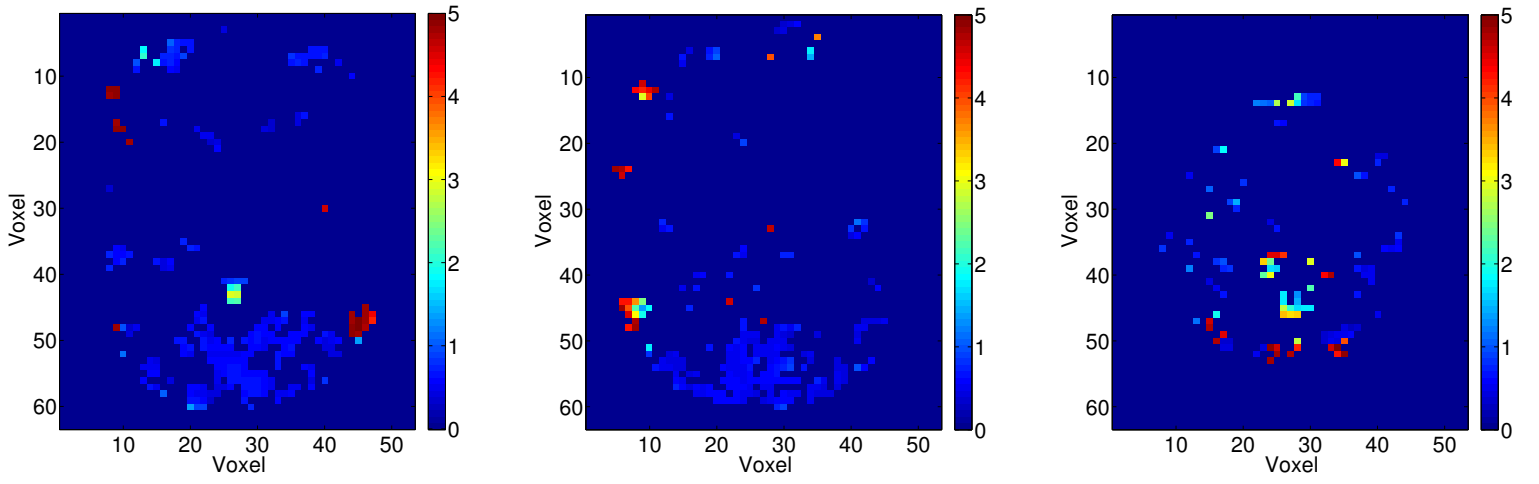


Posterior Estimation

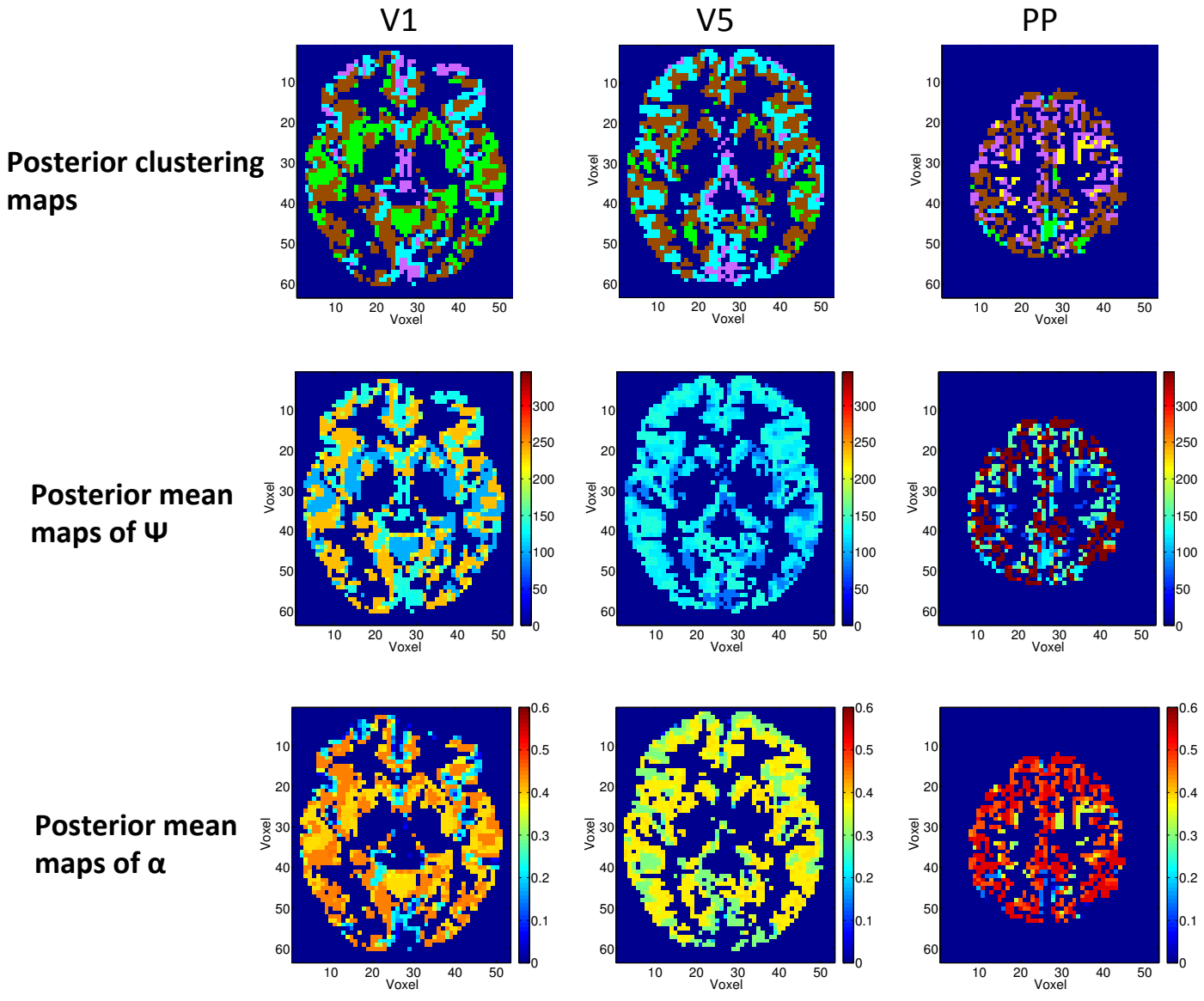
Posterior mean maps of β



Posterior mean maps of λ



Posterior Estimation



Multi-Subject Modeling

Multi-Subject Modeling: Challenges

- It is computationally challenging to fit complex spatio-temporal models to voxel-wise data on multiple subjects
- Strategies: group voxels into regions of interests (ROIs) and "summary statistics" are calculated for each ROI.

⇨ Frequentist Group Analysis: GLM-based estimates of the regression parameters obtained at the voxel level are treated as summary statistics at the group level.

Bayesian two-stage multi-subject analysis

- Bowman (2008) propose a two-stage model:

① First stage: obtain estimates of the voxel-wise regression coefficients in a GLM

② Second stage: combine all localized estimates from all voxels into regions of interests (ROI), and build a hierarchical model for the contrast fMRI BOLD response vector associated with stimulus j in region $g = 1, \ldots, G$, e.g.

$$\boldsymbol{\beta}_{igj} = (\beta_{ig(1)j}, \ldots, \beta_{ig(v_g)j})^T$$

⇨ Simplification of the spatio-temporal structure to conduct multi-subject inference.

Clustering Activation Profiles in multi-subject fMRI studies (Zhang et al, 2016)

Regression Model for Multiple Subjects

Extend the single-subject regression model to multiple subjects

$$Y_{iv} = X_{iv}\beta_{iv} + \varepsilon_{iv},\ \varepsilon_{iv} \sim N_T(0, \Sigma_{iv})$$

- $Y_{iv} = (Y_{iv1}, \ldots, Y_{ivT})^T$, $T \times 1$ BOLD response data for the vth voxel in the ith subject
- X_{iv}, $T \times p$ design matrix, with Poisson HRF
- β_{iv}, $p \times 1$ vector of regression coefficients
- ε_{iv}, a long memory process
- We work in the wavelet domain

$$Y^*_{iv} = X^*_{iv}\beta_{iv} + \varepsilon^*_{iv},\ \varepsilon^*_{iv} \sim N_T(0, \Sigma^*_{iv}), \quad \Sigma^*_{iv} \approx \text{diag}(\psi_{iv}, \alpha_{iv})$$

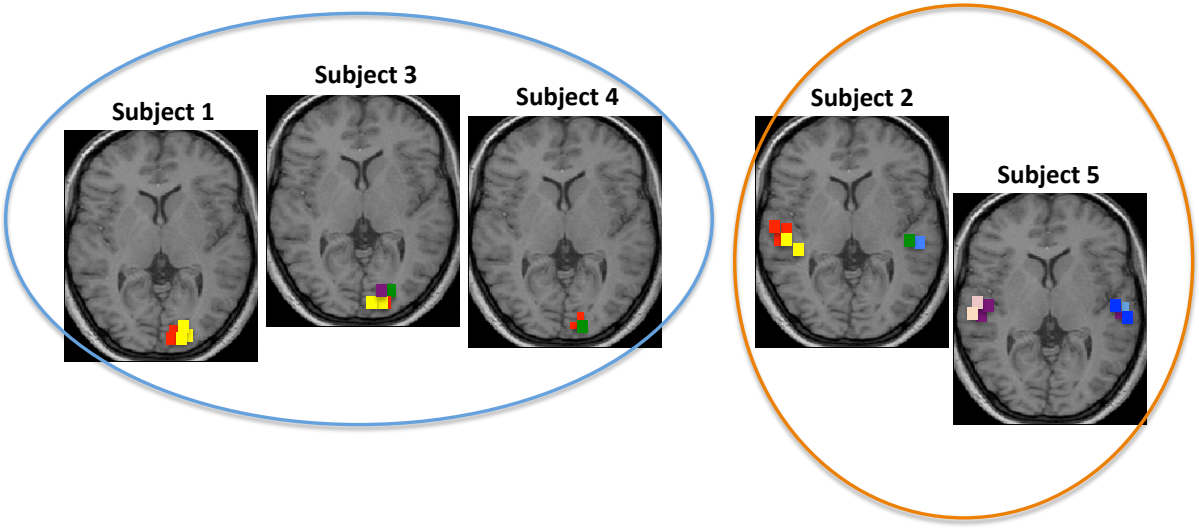
Spatial prior on β

- Objective: capture correlation of voxel time series within and across subjects.

Spatial prior on β

- Objective: capture correlation of voxel time series within and across subjects.

⇨ Clustering of subjects showing similar activation patterns, at the same type capturing spatial correlation among potential activations of distant voxels within each subject



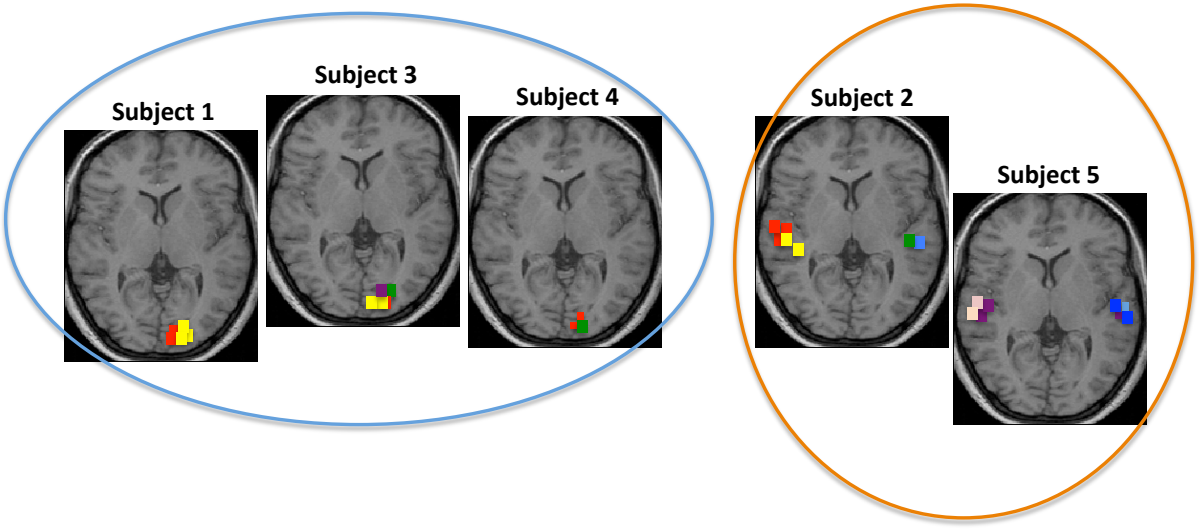
Spatial prior on β

- Objective: capture correlation of voxel time series within and across subjects.

⇨ Clustering of subjects showing similar activation patterns, at the same type capturing spatial correlation among potential activations of distant voxels within each subject

Proposal: multi-subject spike-and-slab nonparametric prior,

$$\beta_{iv}|\gamma_{iv}, G_i \sim \gamma_{iv} G_i + (1 - \gamma_{iv})\delta_0$$



Hierarchical Dirichlet process (HDP) as slab distribution on β_{iv}

$$
\begin{aligned}
\beta_{iv}|\gamma_{iv}, G_i &\sim \gamma_{iv} G_i + (1-\gamma_{iv})\delta_0 \qquad \textit{within-subject} \\
G_i|\eta_1, G_0 &\sim DP(\eta_1, G_0) \qquad \textit{between-subject}
\end{aligned}
$$

$$
\begin{aligned}
G_0|\eta_2, P_0 &\sim DP(\eta_2, P_0) \\
P_0 &= N(0, \tau)
\end{aligned}
$$

- η_1, η_2: concentration parameters, controlling the variability
- P_0: base measure, generating the global components which are shared within and across subjects

Spatially dependent selection & other priors

- To capture the spatial correlation among voxels in each subject, use MRF prior on γ_{iv}

$$P(\gamma_{iv}|d, e, \gamma_{ik}, k \in N_{iv}) \propto \exp(\gamma_{iv}(d + e \sum_{k \in N_{iv}} \gamma_{ik}))$$

  with N_{iv} the set of neighboring voxels of voxel v in subject i, $d \in R$ the sparsity parameter, and $e > 0$ the smoothing parameter

- $\lambda_{iv} \sim$ Uniform(u_1, u_2)
- $\psi_{iv} \sim$ Inverse Gamma(a_0, b_0)
- $\alpha_{iv} \sim$ Beta(a_1, b_1)

Posterior Inference

Strategy I: MCMC approach

- Jointly update (β, γ) with a combination of *add-delete-swap* steps and a Gibbs sampler proposed by Teh et al, 2006 (based on the Chinese restaurant franchise)
- Update λ with Metropolis-Hastings (MH) schemes
- Update ψ with Gibbs sampling or MH algorithm
- Update α with MH algorithm

Strategy II: Variational Bayes (VB) approach

- Basic idea: approximate the true posterior distribution p with a variational distribution q

$$q(z|v) = \prod_{j=1}^{m} q(z_j|v_j)$$

  so that the Kullback-Leibler (KL) divergence between q and p is minimized (Jordan et al. 1999)

- Challenge: What are the optimal variational parameters for the delay parameter λ_{iv} and long memory parameter α_{iv}?
- Solution: combine VB inference and importance sampling in the algorithm (Carbonetto and Stephens, 2012)

MCMC vs VB: Simulation Study

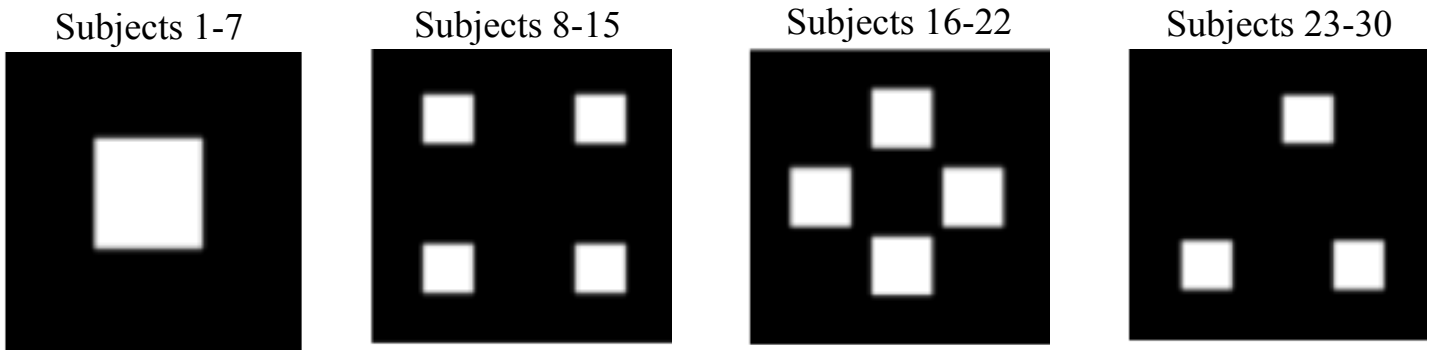
MCMC vs VB: Simulation Study

We simulate

- 30 subjects, event-related design
- one slice with 30×30 voxels for each subject
- 4 different activation patterns
- 10 different components from which nonzero $\beta_{i\nu}$'s are generated

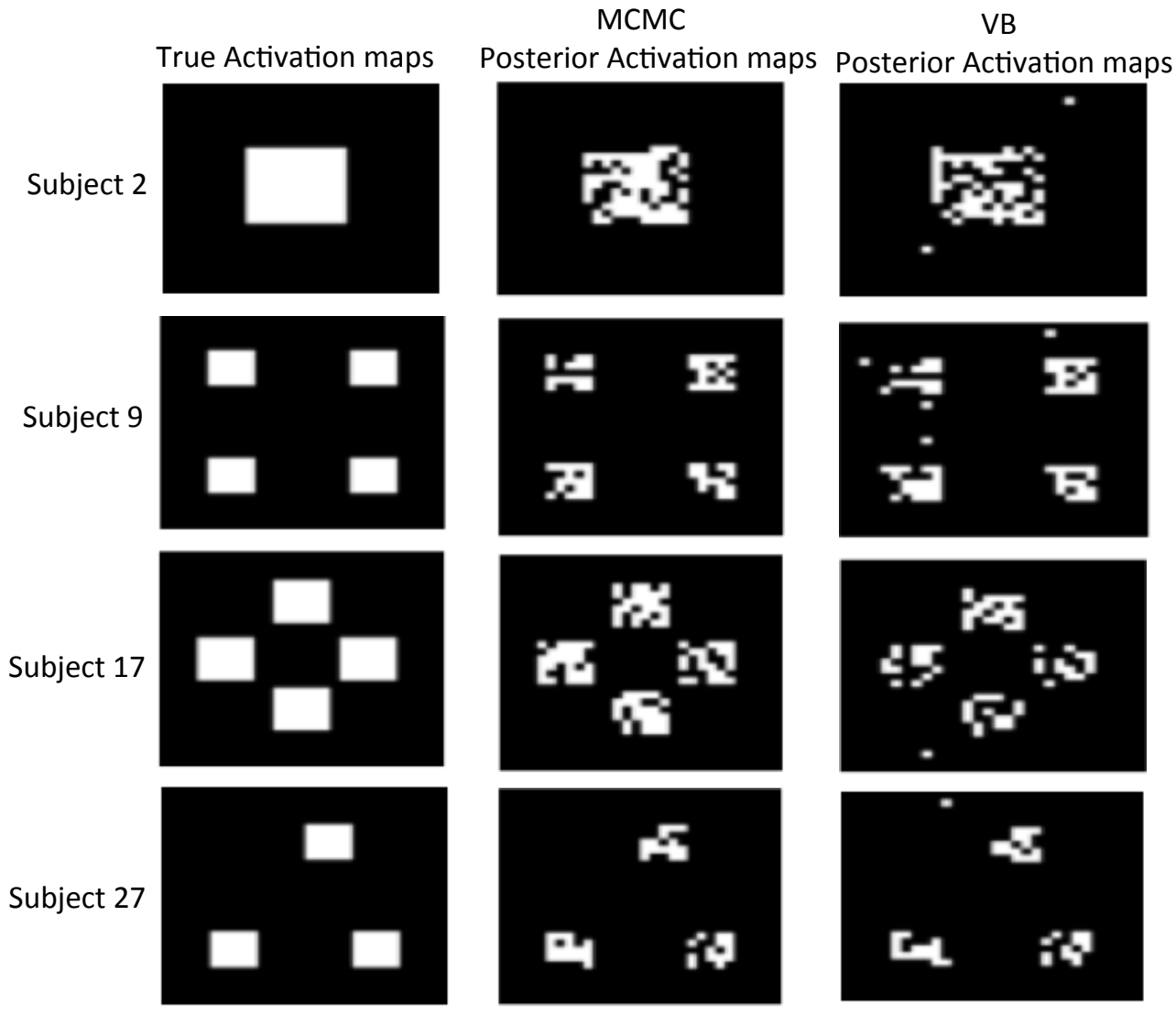
Parameter Setting:

- Four activation patterns:

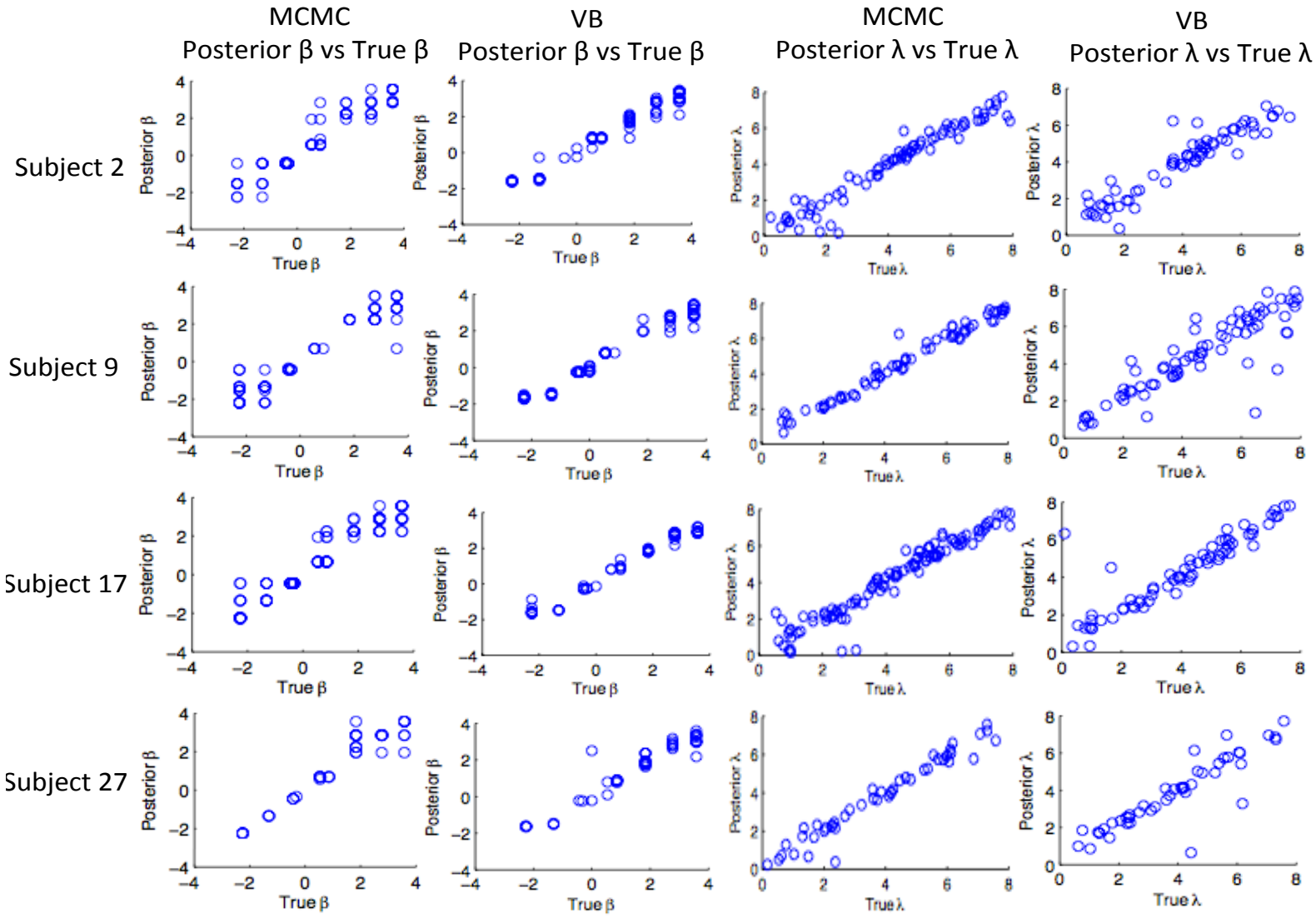


- $\beta_{i\nu}$' for active voxels: sampled from components $\phi_k \sim \text{Normal}(0,1), k = 1,\ldots,10$
- Delay parameters in HRF: $\lambda_{i\nu} \sim \text{Uniform}(0,8)$
- Innovation variances: $\psi_{i\nu} \sim$ truncated Normal$(0,1)$ on $(0,\infty)$
- Long memory parameters: $\alpha_{i\nu} \sim \text{Uniform}(0,1)$

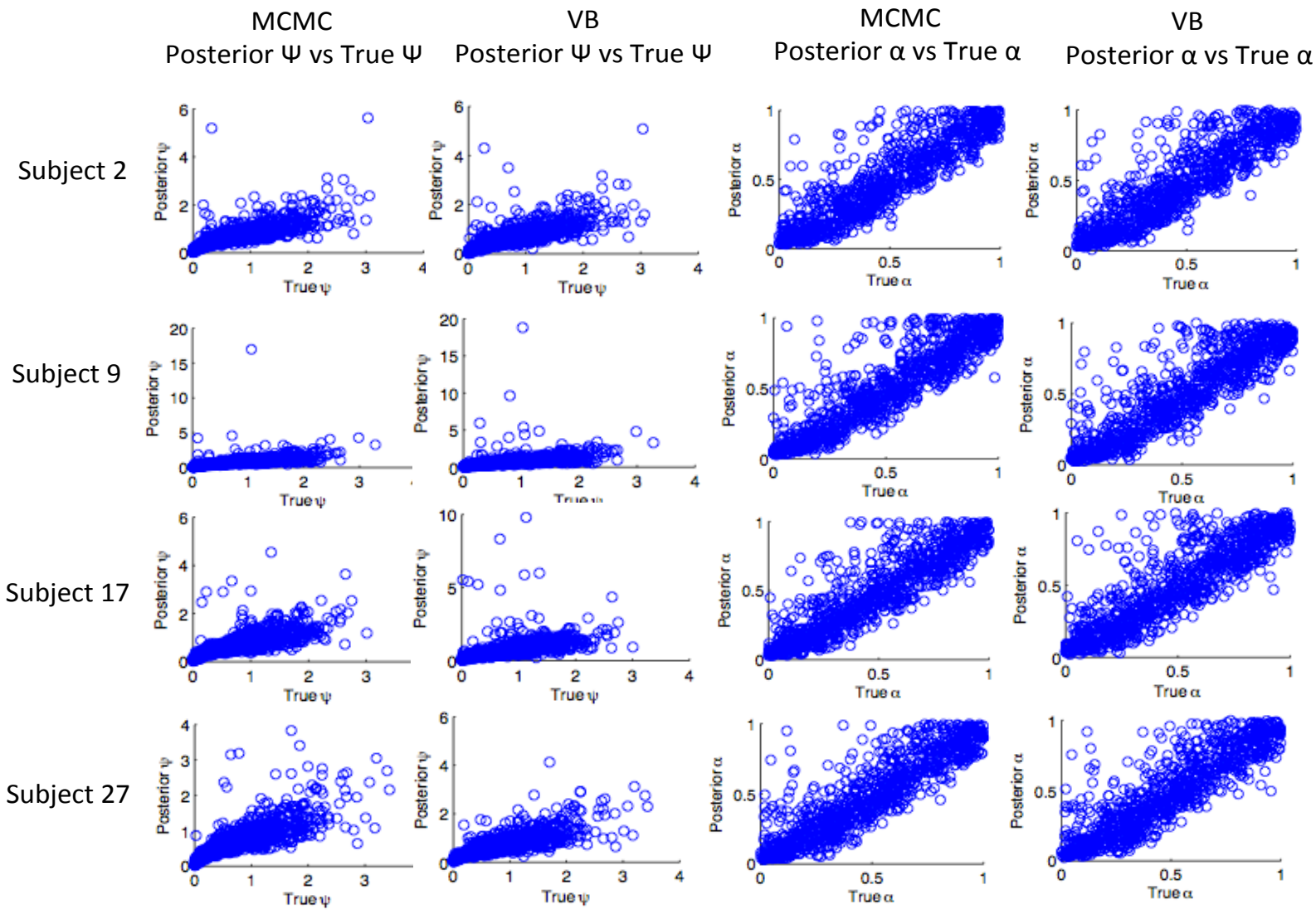
Activation Detection



Estimation of β and λ



Estimation of ψ and α



MCMC vs VB: Simulation Study

MCMC

- Good performance on activation detection
- Good estimation results for model parameters
- Very expensive computation

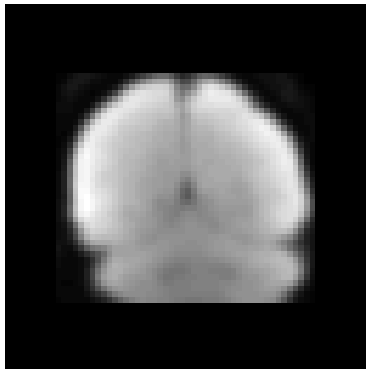
VB

- Good performance on activation detection, with a slightly higher FPR
- Good estimation results for model parameters
- Much more computationally efficient (32 fold faster)

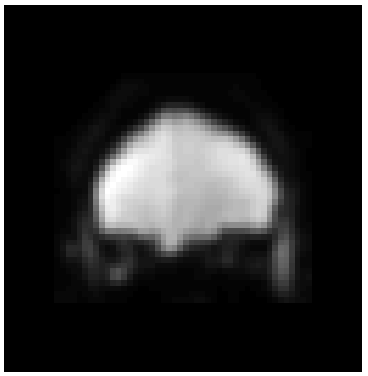
Case Study

- Real fMRI data collected by Versace's lab (MDACC):
  - Data Dimension: 27 subjects, 286 time points, 2 slices of interest, 64 $\times$ 64 voxels per slice

Occipital Slice
(y = -60 mm)

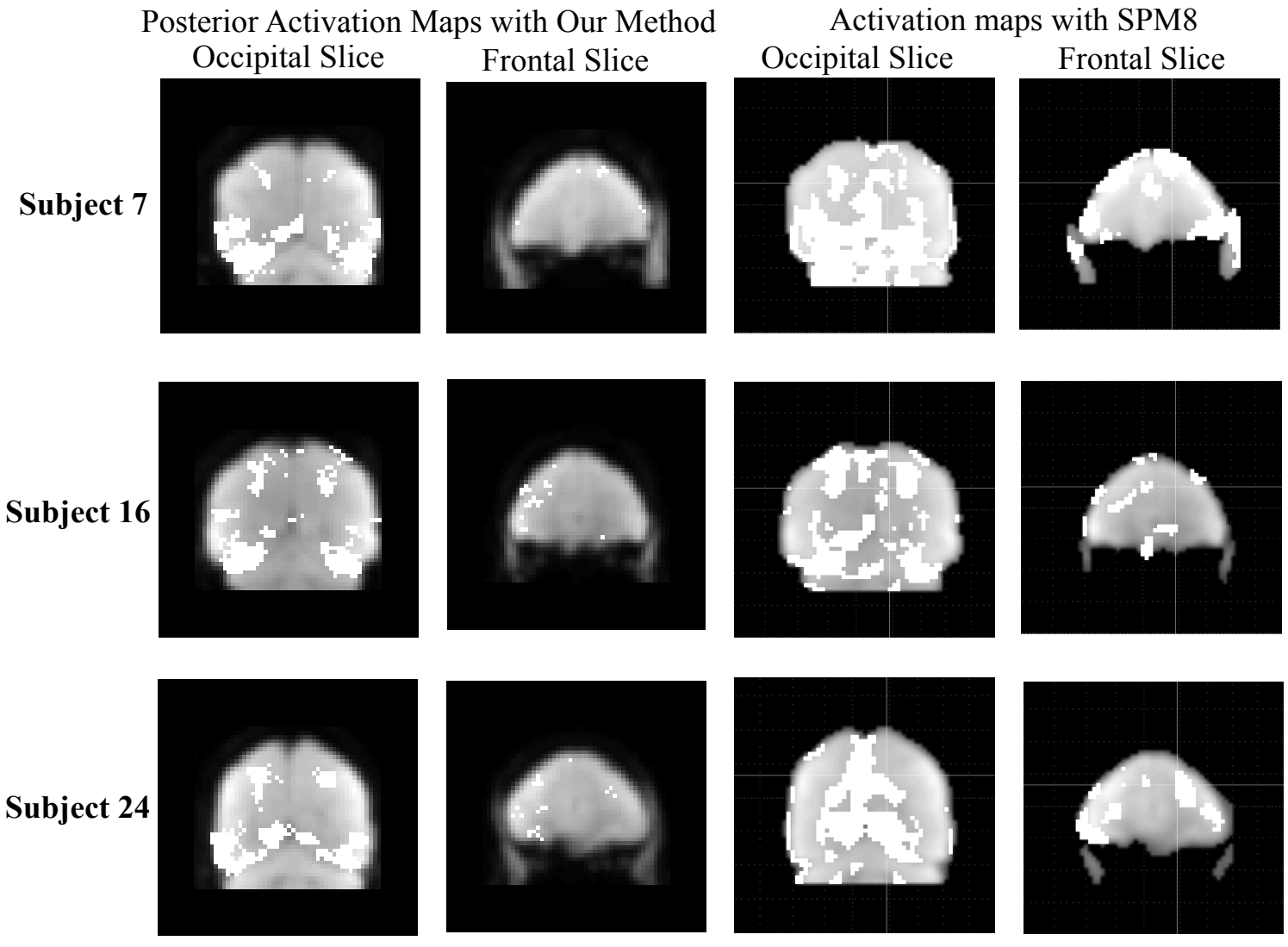


Frontal Slice
(y = +38 mm)

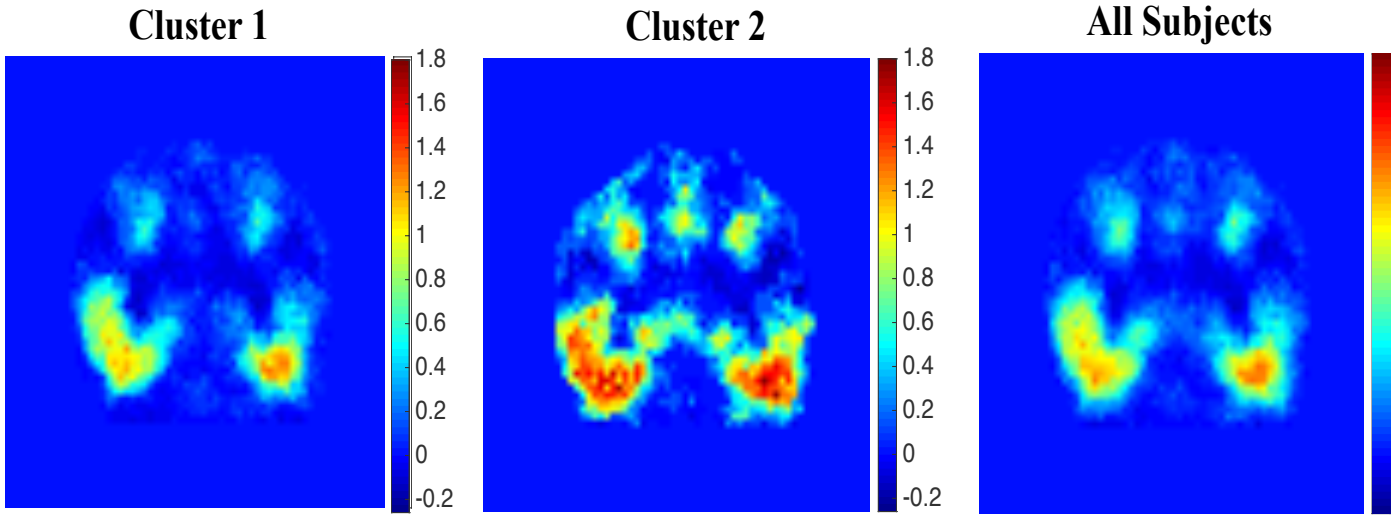


  - Event-related design
  - Goal: detecting brain activity in response to visual scenes

Results of activation detection



Results of subject-level clustering



Two groups of subjects characterized by different levels of activations (reward sensitivity)

Summary

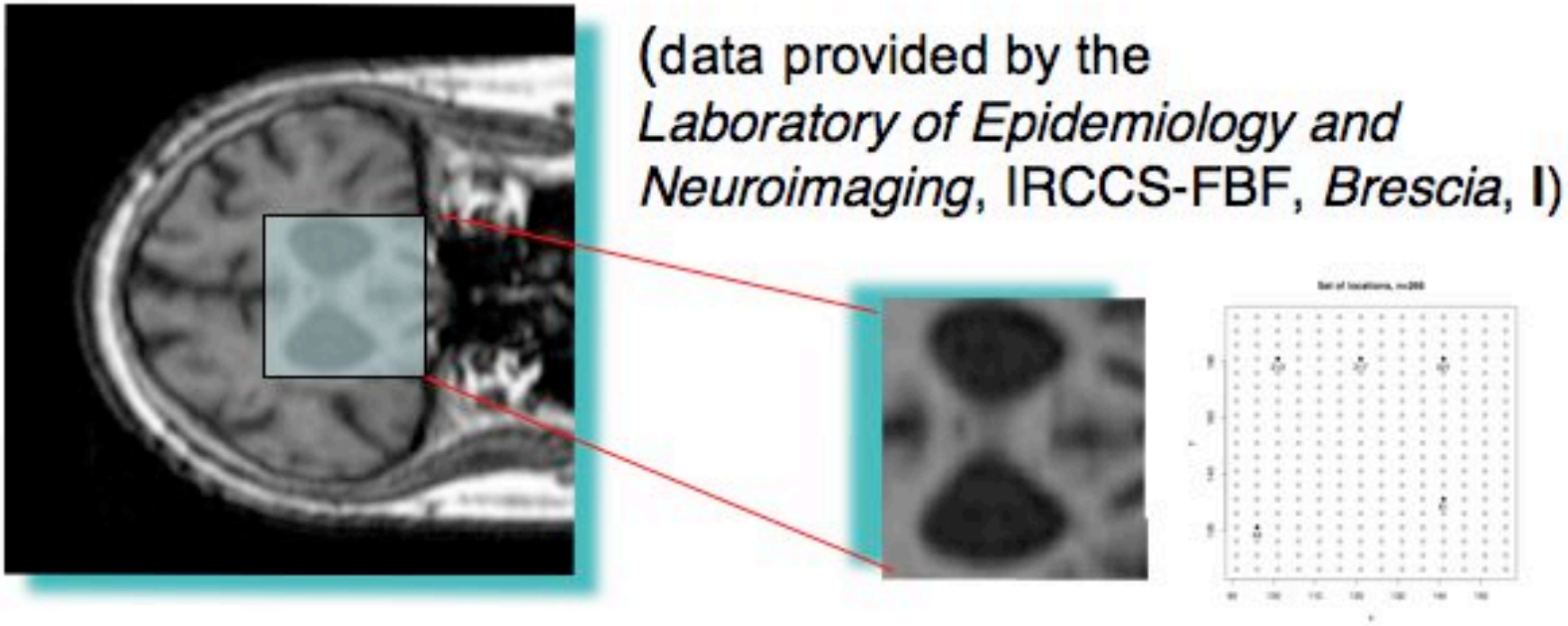
- Priors to take into account structural feature in the data
- Improve performance of activation detection
- Capture the association among voxel time series within and across subjects via Bayesian nonparametric models
- Bayesian approximation methods alternative to full MCMC

Zhang, L., Guindani, M., Versace, F., Engelmann, J.M. and Vannucci, M. (2016). A Spatio-Temporal Nonparametric Bayesian Model of Multi-Subject fMRI Data. *Annals of Applied Statistics*, 10(2), 638-666.

- *Forward look:* Neuroimaging/Clinical Imaging/Data Integration are promising field for application and testing of Bayesian techniques

Complex Models for Structural Brain Imaging

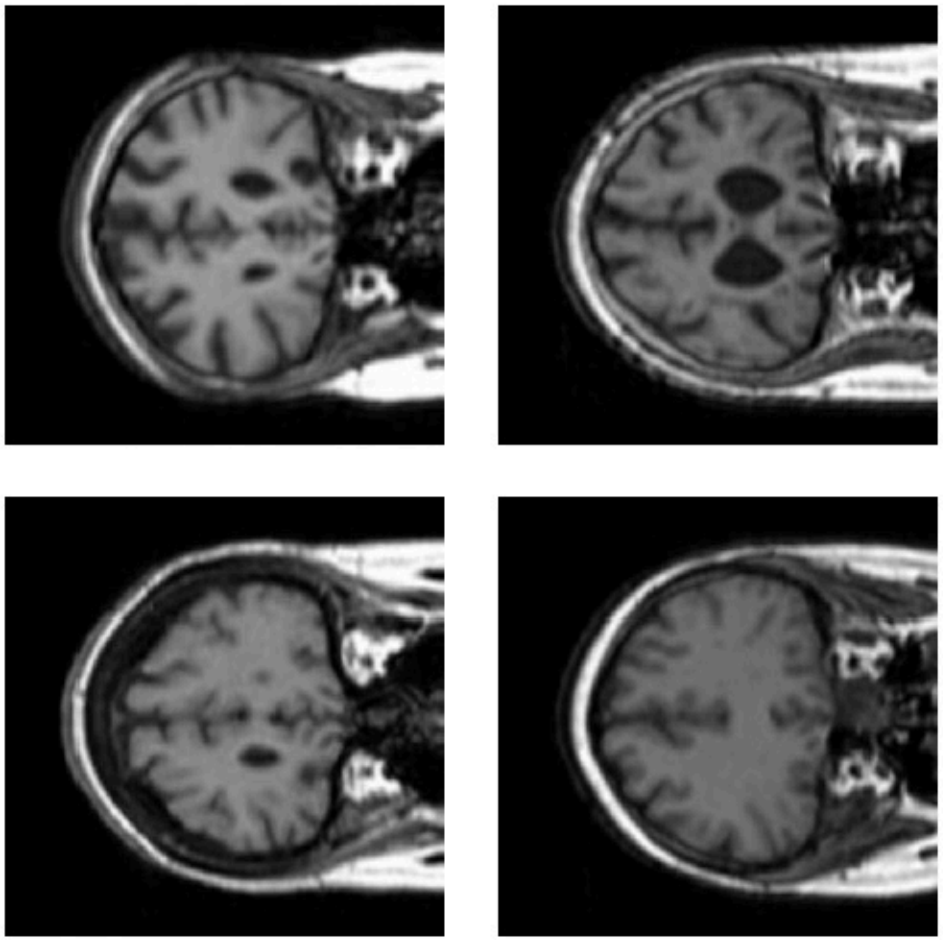
Amydgalar-hyppocampal section



(data provided by the *Laboratory of Epidemiology and Neuroimaging*, IRCCS-FBF, *Brescia*, I)

T=18 patients, 8 sick, 10 healthy

The effect of Alzheimer's disease on the amygdalar-hippocampal complex



- Envisage spatial processes typical of "healthy" and "impaired" regions/brains.
- Describe the data by means of a collection of *base* processes.

A collection of base processes...

- One way of describing the heterogeneity of images by enforcing selection of *base processes* is by using a Spatial Dirichlet Process approach (Gelfand and Kottas, 2005):

- The SDP can be described by a mixture model for the observables,

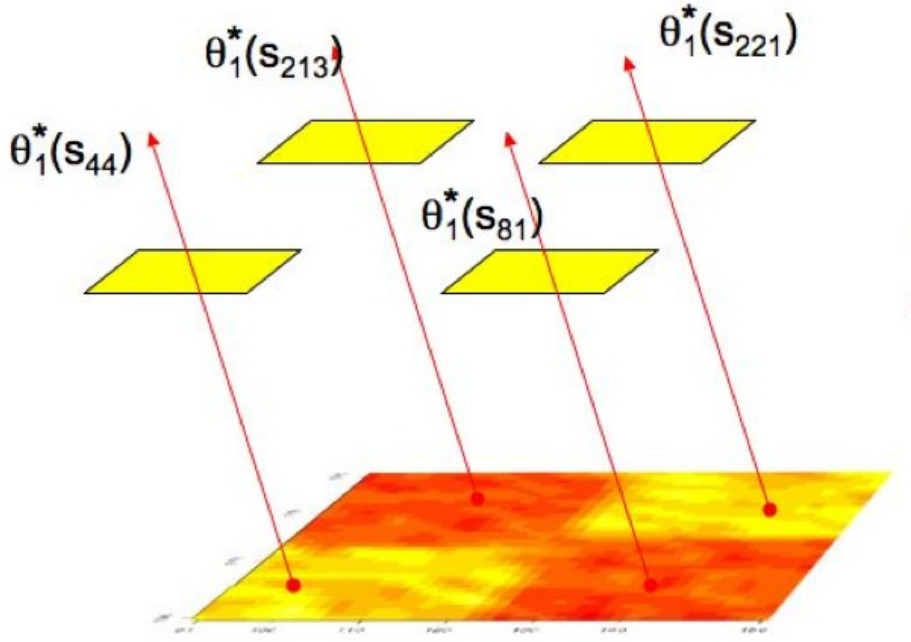
$$Y_i|G \overset{i.i.d.}{\sim} \sum_{j=1}^{k} p_j N(y|\theta_j^*, \sigma^2 I),$$

for some base processes $\theta_j^*(s), s \in D$, when assuming $k = \infty$.

The mixture model enforces a clustering of the observations into a number of components.

Global allocation rules

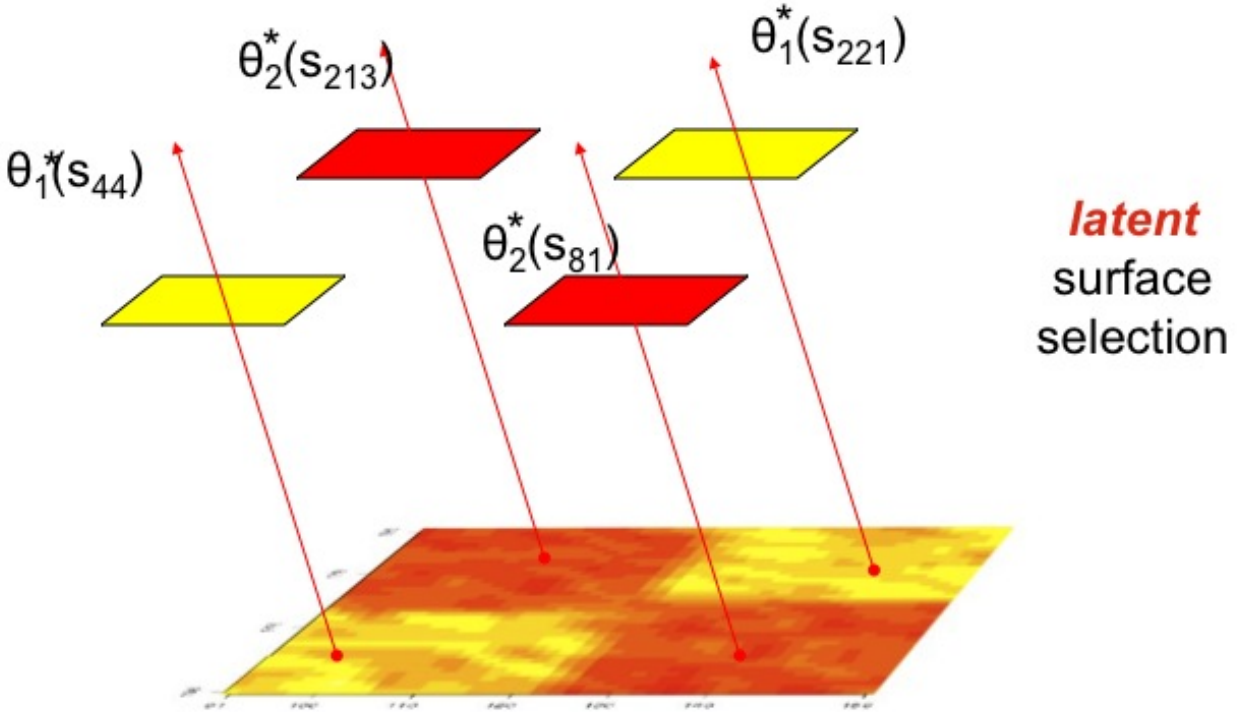
The Spatial DP dictates a global sampling: each observation is matched to an entire “base” curve



⇨ The induced global smoothing either increases the number of mixture components (species, kernel...) or tends to fit an “average” curve.

Local allocation rules

Instead we would like to allow different observations to be matched to different “base curves”, so that each observation is the result of the composition of several acting processes



Spatially dependent mixtures.

- We can obtain a local surface selection (and fewer clusters/base surfaces) by enabling the weights p_j to vary spatially:

$$G(s) = \sum_{j=1}^{k} p_j(s)\, \delta_{\theta_j^*(s)}$$

Spatially dependent mixtures.

- We can obtain a local surface selection (and fewer clusters/base surfaces) by enabling the weights p_j to vary spatially:

$$G(s) = \sum_{j=1}^{k} p_j(s)\, \delta_{\theta_j^*(s)}$$

- We model a hidden factor $\gamma(s)$ that at each location selects one of the base clusters available:

$$\gamma_i = (\gamma_i(s_1), \ldots, \gamma_i(s_n))$$

where $\gamma_i(s) \in \{1, \ldots, k\}$.

Hybrid Dirichlet priors

- Petrone, Guindani and Gelfand (JRSSB, 2009) define a the Hybrid Dirichlet Process as follows:

$$G_n(\cdot) = \sum_{j_1=1}^{k} \cdots \sum_{j_n=1}^{k} p(j_1, \ldots, j_n)\, \delta_{\theta^*_{j_1}(s_1),\ldots,\theta^*_{j_n}(s_n)},$$

where the weights $p(\cdot)$ models the distribution of the latent allocation factor

$$p(j_1, \ldots, j_n) = P(\gamma_i(s_1) = j_1, \ldots, \gamma_i(s_n) = j_n | p),$$

and are *consistent* for any choice of $(s_1, \ldots s_n)$.

Hybrid Dirichlet priors

- Let $k < \infty$. We assume that the weights are Dirichlet distributed with parameters centered around a prior distribution q

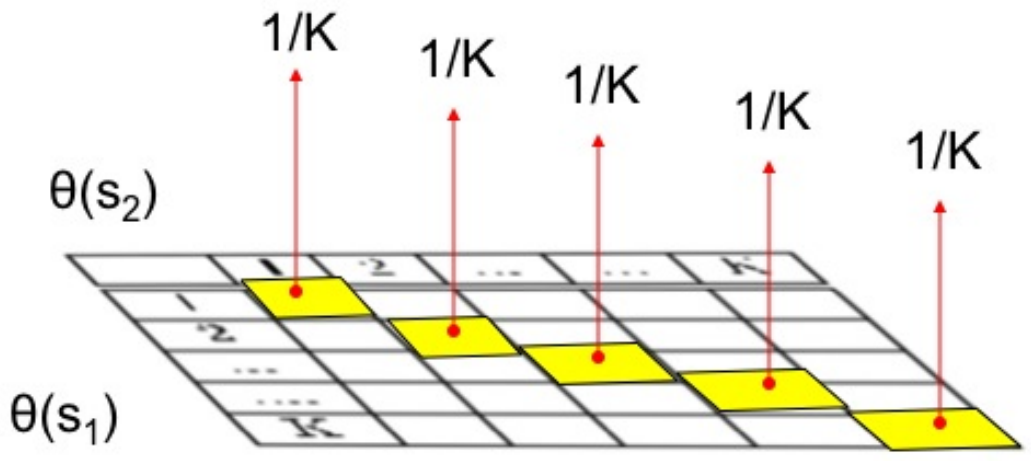
$$P(\gamma_i(s_1) = j_1, \ldots, \gamma_i(s_n) = j_n) = E(p(j_1, \ldots, j_n)) = q(j_1, \ldots j_m)$$

- Special cases are obtained by choosing q appropriately.

Global allocation factor

- If *q* doesn't allow different choices of components at 2 locations, then the hybrid reduces to a finite spatial DP (DP_k).

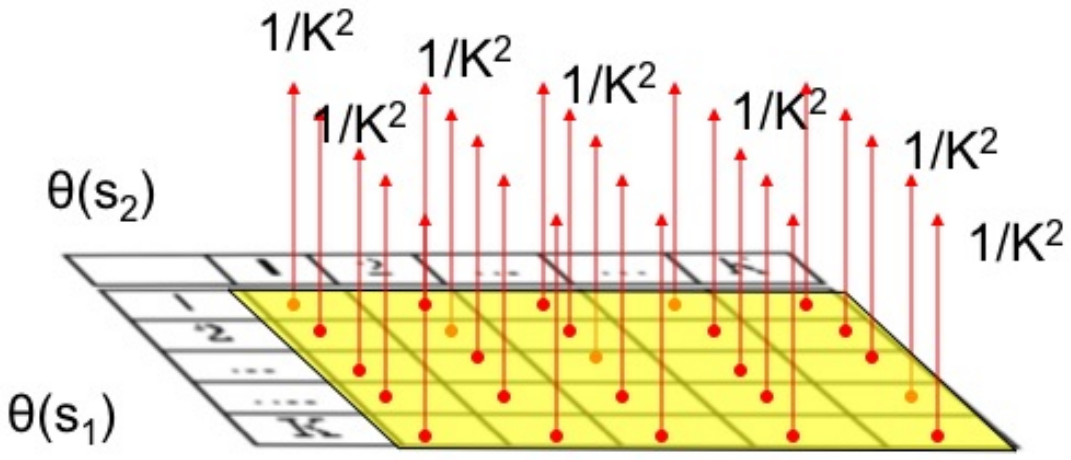
$$q_k(j_{1,}, \ldots, j_2) = 1/k \quad j_1 = \ldots = j_n = j,\ j = 1, \ldots, K$$
$$0 \quad \text{otherwise}$$



Independent local effects

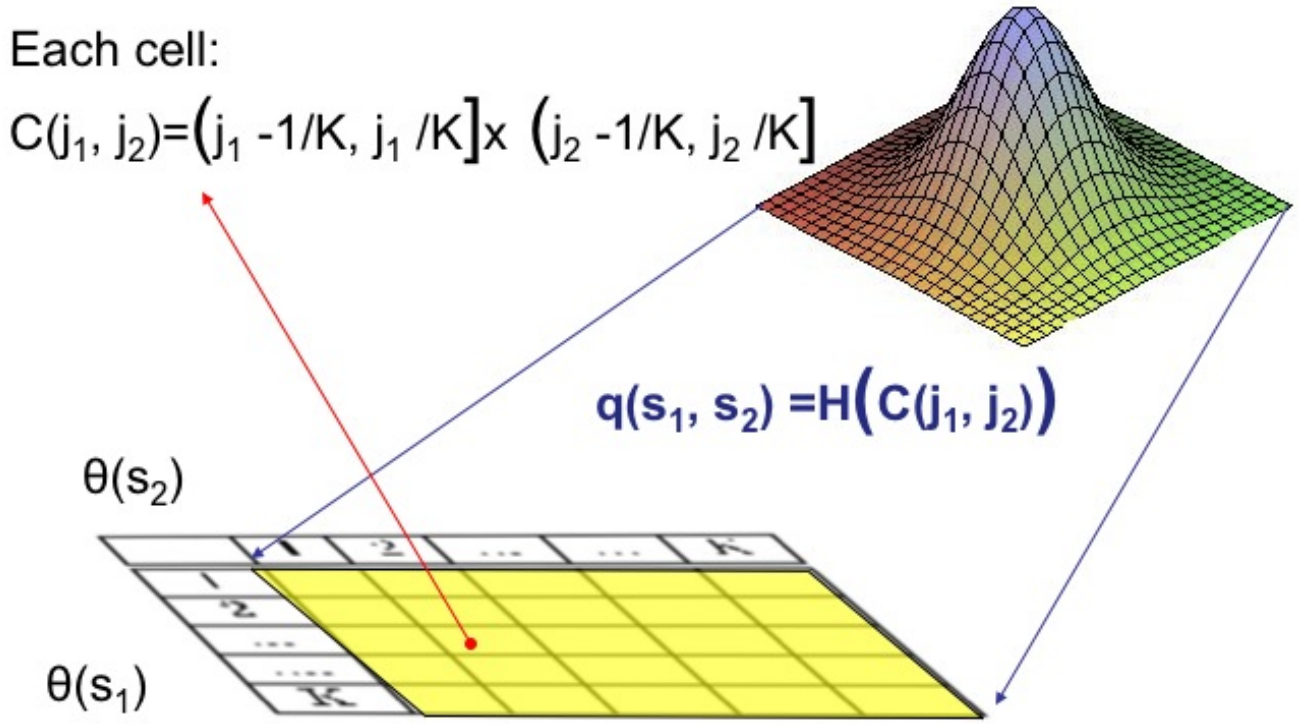
- The factor γ acts independently at each location. There's local allocation but it is not driven by spatial information.

$$q_k(j_1, \ldots, j_2)=1/K^n \qquad j_i=1,\ldots,K,\ i=1,\ \ldots,\ n$$

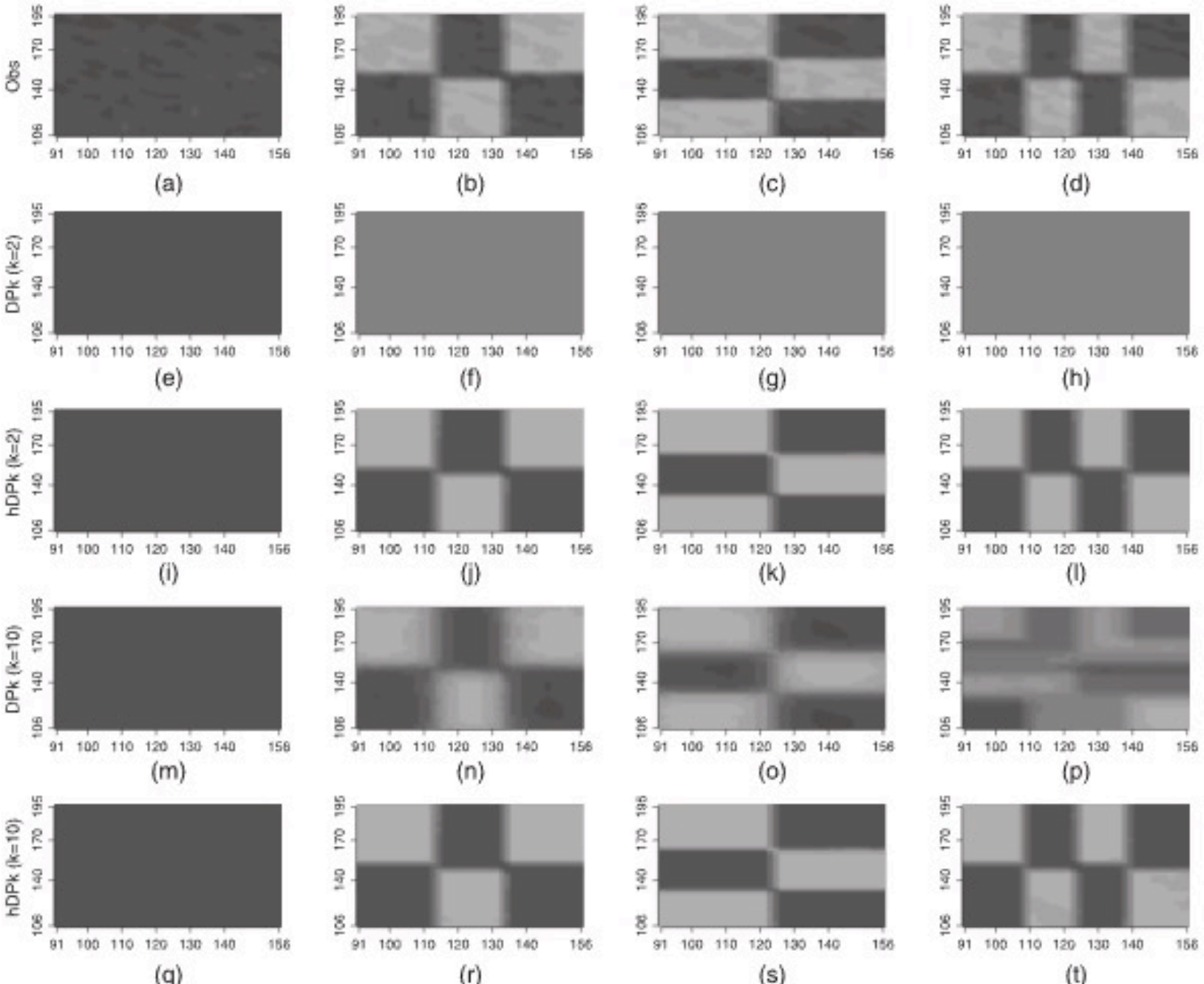


Spatially dependent local effects

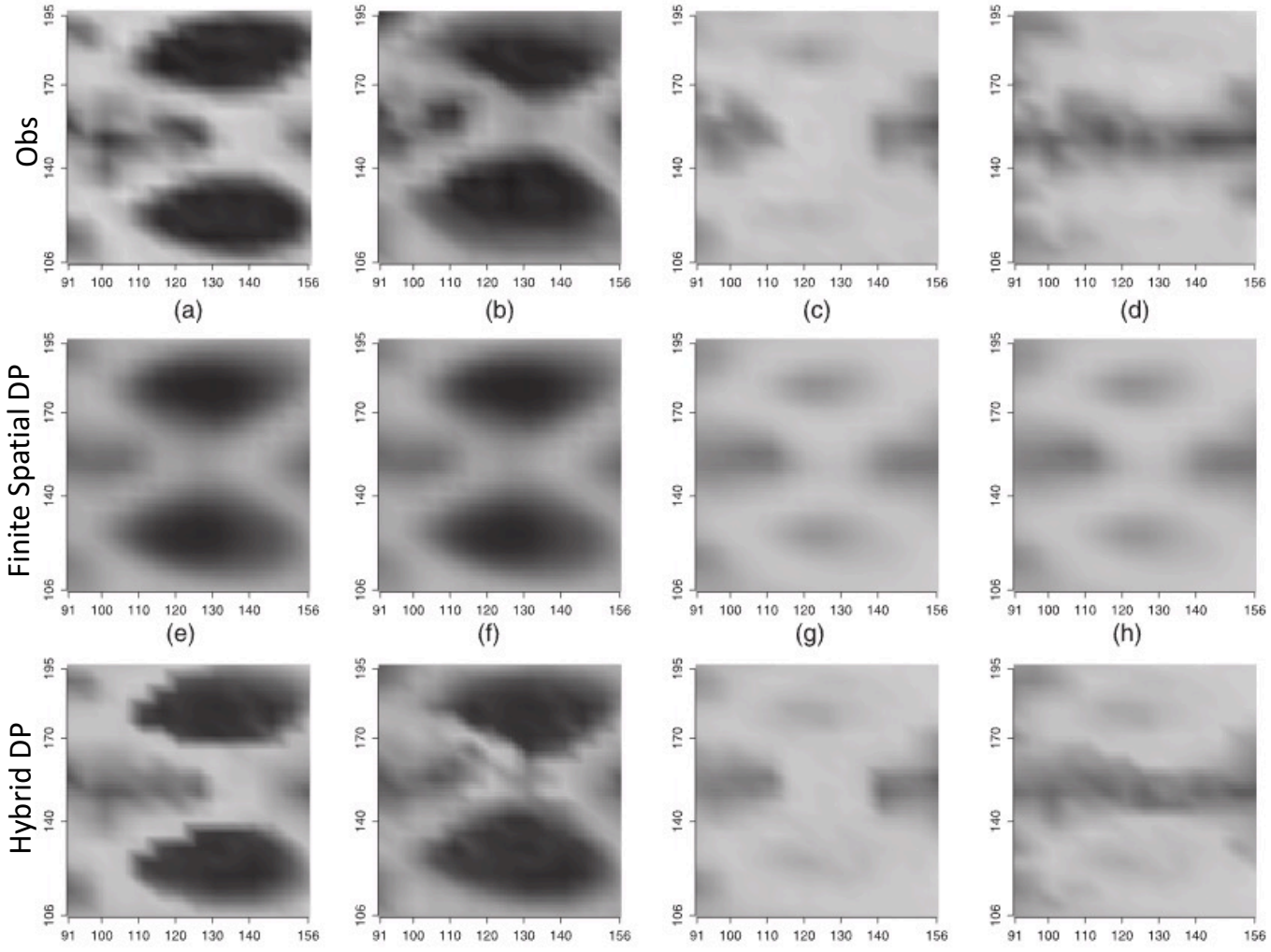
- As an intermediate case, we can consider *q* as the copula *H* of a latent Gaussian process. The choice of the base surfaces at each location is governed by the spatial properties of the latent process.

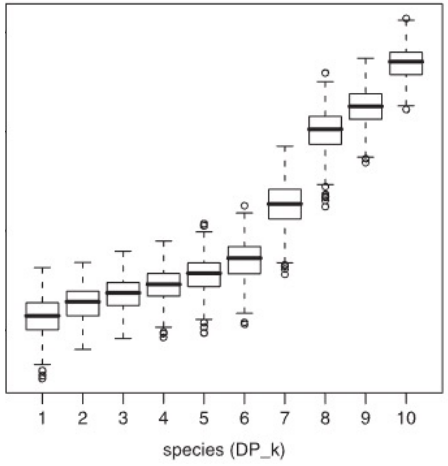


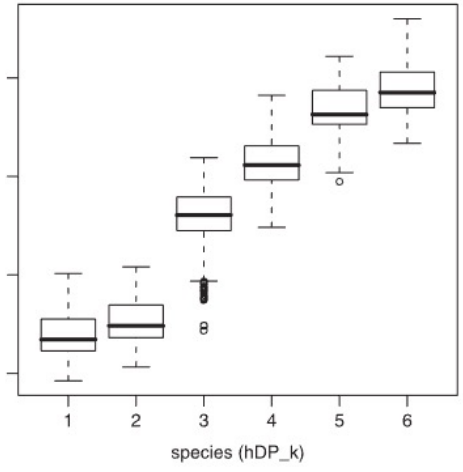
Application to Simulated data



Application to real data







Summary

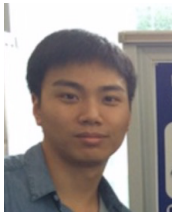
- The spatial resolution of structural/functional brain images and the heterogeneity of individual configurations require the consideration of complex spatial models
- The models can be extended to investigate the relationship between imaging features and health outcomes (imaging biomarkers)
- Computational challenges as the dimension increases and the number of subjects increases (e.g. ADNI repositories)
- For a review: Zhang L, Guindani M, Vannucci M (2015). *Bayesian models for functional magnetic resonance imaging data analysis.* WIREs Computational Statistics, 7, 21-41.

Acknowledgements

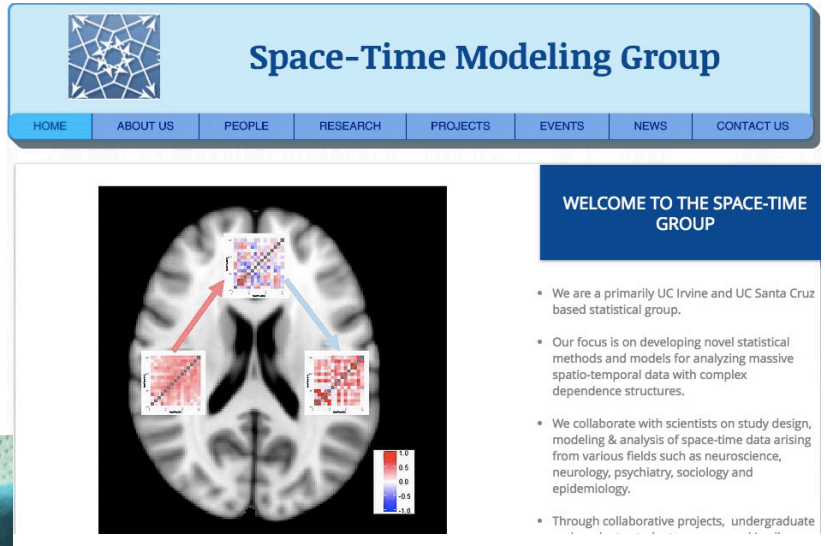
Marina Vannucci, Rice

Sharon Chiang,
PhD from Rice
MD Baylor

Ryan Warnick, Rice
NSF fellowship



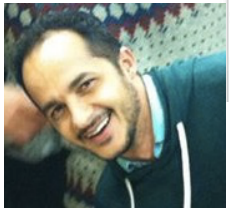
Eric Kook,
Rice



Linlin Zhang,
Phd from Rice
Schlumberger,
Savage Prize Runner Up

Duncan Wadsworth,
PhD from Rice
Microsoft

Qiwei Li,
PhD from Rice
UT Southwestern



Ronaldo Guedes
University of Verona

Francesco Versace,
MD Anderson Cancer Center

Alberto Cassese
Maastricht University

Space Time modeling group (UCI/UCSC)
http://www.spacetimemodeling.org

For more information:
http://www.micheleguindani.info

NSF SES-1659921

Postdoctoral Opportunity at Rice

- Contact Marina Vannucci (marina@rice.edu)
- The successful candidates will work with faculty members and graduate students in both Statistics and Computer Science to develop innovative methodologies for data science. Active research areas at Rice include Bayesian statistics and Machine learning, among others. Funding are limited to citizen and permanent residents. Positions can start any time.
- More information here: `https://www.cs.rice.edu/~nakhleh/RTG`
- and `https://datascience.rice.edu/postdoctoral-research-associate-positions`