A Bayesian Modeling Approach of Multiple-Subject fMRI Data

Marina Vannucci

Professor and Chair
Department of Statistics
Rice University
Houston, TX
USA

Joint work with Linlin Zhang (Rice U., USA) and Michele Guindani (MDACC, USA)

Banff - CA - February 4th, 2016

Outline of the Talk

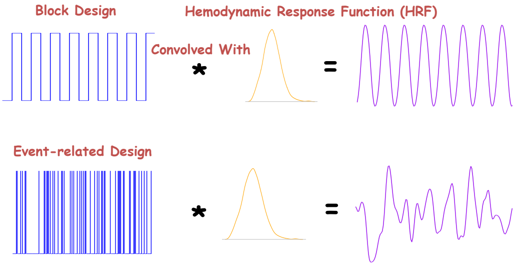
- Spatio-temporal Model for detection of task-related activation patterns (single subject):
  - Correlated Errors.
  - Network priors capturing structural dependencies.
- Extension to multiple subjects:
  - Nonparametric priors to capture the association among voxel time series within and across subjects
  - Variational Bayes algorithm for inference.

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

fMRI = BOLD Signal + Noise

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

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



Noise: caused by hardware and subjects themselves.

Model for Single Subject

$$Y_\nu = X_\nu \beta_\nu + \varepsilon_\nu, \quad \nu = 1, \ldots, V$$

- $Y_\nu = (Y_{\nu 1}, \ldots, Y_{\nu T})^T$ time-series data for voxel ν.
- X_ν the convolved design matrix $T \times p$ (set $p = 1$).
- Poisson HRF with voxel-dependent parameter λ_ν

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

- $\beta_\nu = (\beta_{\nu 1}, \ldots, \beta_{\nu p})^T$ vector of regression coefficients.
- $\varepsilon_\nu \sim N_T(0, \Sigma_\nu)$.

Modelling Challenges

Temporal Modelling:

- Choice and parameters of the HRF
  (Gossl *et. al* (2001), Woolrich *et al.* (2004), Quiros *et al.* (2010)).
- Noise structure - AR(q)
  (Woolrich *et al.* (2004), Penny *et al.* (2005), Lee *et al.* (2014)).

Spatial Modelling:

- Spatial priors on β_ν
  (GMRF in Gossl *et. al* (2001) & Quiros *et al.* (2010); Laplacian in Penny *et al.* (2005); SSBF in Flandrin and Penny (2007)).

  (*Spike-and-slab* in Smith and Fahrmeir (2007) & Kalus *et al.* (2013) & Lee *et al.* (2014)).

Temporal Modeling

Source of variability:

- Low-frequency noise, as head movement caused by slow rotation or translation during scanning.
- Physiological noise, as cardiac and respiratory cycle-related pulsations.

Introduce a shift component or model the noise as long-range

$$\gamma(\tau) \approx \psi \tau^{-\alpha} \text{ as } \tau \to \infty$$

(not negligible dependence between distant observations)

(Zarahn *et al.* (1997, NeuroIm) & Aguirre *et al.* (1997, NeuroIm); Fadili & Bullmore (2002, NeuroIm;2005, IEEETrSPr) and Jeong, Vannucci, and Ko (2013, Biometrics)).

$$\varepsilon_\nu \sim N_T(0, \Sigma_\nu), \quad \Sigma_\nu(i,j) = [\gamma(|i-j|)]$$

Use wavelet transforms to "whiten" the data.

- Wavelets decorrelate data from long-memory processes (Wornell & Oppenheim (1992, IEEETrSPr), Tewfik & Kim (1992, IEEETrITh), McCoy & Walden (1996, JCGS), Craigmile & Percival (2005, IEEETrITh), Ko and Vannucci (2006, JSPI)).

- Let W be the $T \times T$ matrix of the wavelet transform

$$Y^*_\nu = X^*_\nu \beta_\nu + \varepsilon^*_\nu, \quad \varepsilon^*_\nu \sim N_T(0, \Sigma^*_\nu),$$

with $Y^*_\nu = WY_\nu$, $X^*_\nu = WX_\nu$, $\varepsilon^*_\nu = W\varepsilon_\nu$.

- $\Sigma^*_\nu = \mathcal{W}\Sigma_\nu\mathcal{W}' = \text{diag}[\psi_\nu, \alpha_\nu]$, with ψ_ν the innovation variance and $\alpha_\nu \in (0, 1)$ the long memory parameter (Vannucci and Corradi (1999, JRSSB)).

- Dirichlet Process (DP) prior on (ψ_ν, α_ν)

$$\begin{aligned}(\psi_\nu, \alpha_\nu)|G \sim & G;\\ G|\eta, G_0 \sim & DP(\eta, G_0);\\ G_0 = & \mathtt{IG}(a_0, b_0) \times \mathtt{Beta}(a_1, b_1)\end{aligned}$$

  This induces a "clustering" of spatially remote voxels that exhibit fMRI time series signals with similar characteristics, as an aspect of functional connectivity (Friston, K. (1994, HumBMap)).

- Complete temporal modelling with Uniform prior on voxel-dependent delay parameter λ_ν of the HRF,

$$\lambda_\nu \sim \mathtt{U}(u_1, u_2), \;\; \nu = 1, \ldots, V.$$

Spatial Modelling

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

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

  This prior selects activated voxels.
  If a voxel is not activated, then $\beta_\nu = 0$
  If a voxel is activated, then $\beta_\nu \sim N(0,\tau)$.

- Markov Random Field (MRF) prior on the voxel-dependent selection parameter γ_ν for activation

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

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

(George & McCulloch (1997, SINICA); Stingo and Vannucci (2011, Bioinfo))

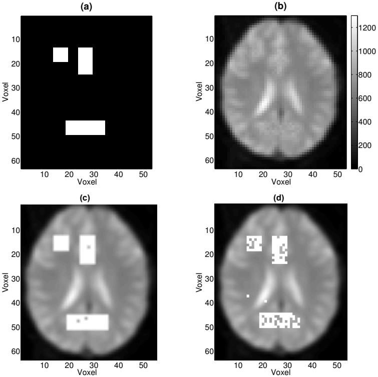
Posterior Inference

- MCMC:
  - MH schemes that use add-delete-swap moves on (γ_ν, β_ν) (George & McCulloch (1997, SINICA); Savitsky et al. (2011, STS))
  - Full conditionals on λ_ν
  - Sampling algorithms for nonparametric DP models on (ψ_ν, α_ν) (algorithm 8 of Neal (2000, JCGS))

- Hyperparameter settings:
  - DP prior with $\alpha = 1$
  - MRF prior $d = -2.5$ (10% base prob) and $e = 0.3$
  - Vague or non-informative priors otherwise
  - MCMC chains with 10,000 iterations, with 5,000 burn-in.

Synthetic Data

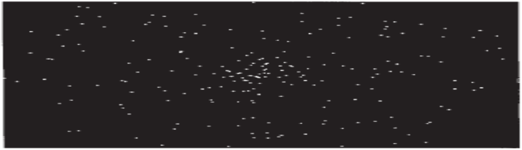
Simulate $Y_{syn} = Y + w$, where Y is simulated from our model and w is a selected slice (64×64) from the real fMRI study.



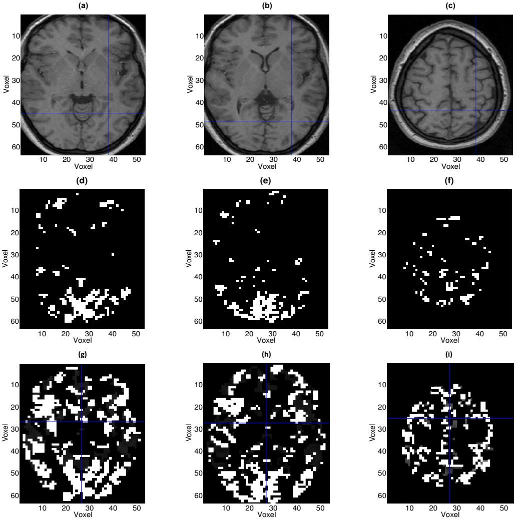
(a) True activation map; (b) First scan of synthetic data. (c) Our mehod. (d) SPM8 (Penny et al., 2005 - Bayes model, AR errors, spatial prior on β).

Real Data

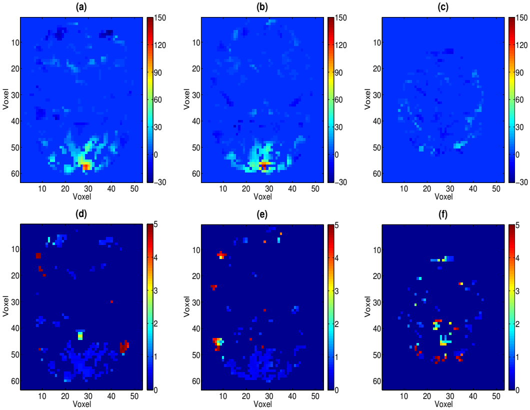
- Data available at `http://www.fil.ion.ucl.ac.uk/spm/data/attention/`
- A single subject, 4 different conditions
  “Fixation”, “Attention”, “No Attention”, and “Stationary”.
- Stimulus function: 1 for the images during the “Attention” and “No Attention” conditions and 0 during “Fixation”.
- Slices analysed: primary visual cortex (V1), motion-selective cortical area (V5), and posterior parietal cortex(PP).



Structural MRI images, posterior activation maps and SPM result.



Posterior mean maps of β and λ.



Extensions to Multiple Subjects

- Multiple subjects - computational challenges!
- Two stage modelling:

  Bowman *et al.* (2008, NeuroIm)
  (single-voxel estimates for each subject;
  ROI-based summaries to define “contrast” images)

  Sanyal and Ferreira (2012, NeuroIm)
  (GLM with indep errors to estimate the β’s)
  (2D wavelets at 2nd stage with spike-and-slab priors)

Regression Model for Multiple Subjects

Extend the single-subject regression model to multiple subjects

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

- $Y_{i\nu} = (Y_{i\nu 1}, \ldots, Y_{i\nu T})^T$, $T \times 1$ BOLD response data for the νth voxel in the ith subject
- $X_{i\nu}$, $T \times p$ design matrix, with Poisson HRF
- $\beta_{i\nu}$, $p \times 1$ vector of regression coefficients
- $\varepsilon_{i\nu}$, a long memory process
- We work in the wavelet domain

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

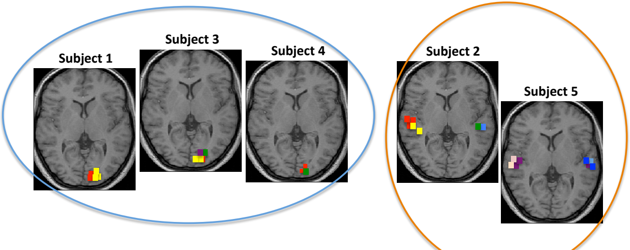
Spatial prior on β

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

Proposed prior: spike-and-slab nonparametric prior,

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

Clustering of subjects based on similar activation patterns; clustering of voxels within each subject according to similar activation strengths.



Hierarchical Dirichlet process (HDP) as slab distribution on $\beta_{i\nu}$

$$
\begin{aligned}
\beta_{i\nu}|\gamma_{i\nu}, G_i &\sim \gamma_{i\nu} G_i + (1-\gamma_{i\nu})\delta_0 \\
G_i|\eta_1, G_0 &\sim DP(\eta_1, G_0) \\
G_0|\eta_2, P_0 &\sim DP(\eta_2, P_0) \\
P_0 &= N(0, \tau)
\end{aligned}
$$

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

Other priors

- To capture the spatial correlation among voxels in each subject, use MRF prior on $\gamma_{i\nu}$

$$P(\gamma_{i\nu}|d, e, \gamma_{ik}, k \in N_{i\nu}) \propto \exp(\gamma_{i\nu}(d + e \sum_{k \in N_{i\nu}} \gamma_{ik})) \qquad (1)$$

  with $N_{i\nu}$ the set of neighboring voxels of voxel ν in subject i, $d \in (-\infty, \infty)$ the sparsity parameter, and $e > 0$ the smoothing parameter
- $\lambda_{i\nu} \sim \text{Uniform}(u_1, u_2)$
- $\psi_{i\nu} \sim \text{Inverse Gamma}(a_0, b_0)$
- $\alpha_{i\nu} \sim \text{Beta}(a_1, b_1)$

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

$$q(z|v) = \prod_{j=1}^{m} q(z_j|v_j) \tag{2}$$

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 $\lambda_{i\nu}$ and long memory parameter $\alpha_{i\nu}$?
- Solution: combine VB inference and importance sampling in the algorithm (Carbonetto and Stephens, 2012)

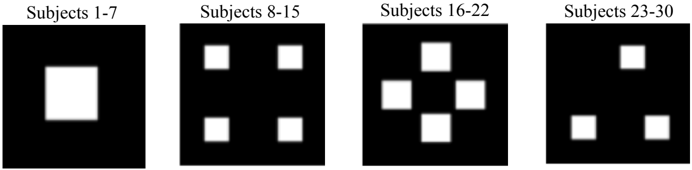
Simulation Study

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

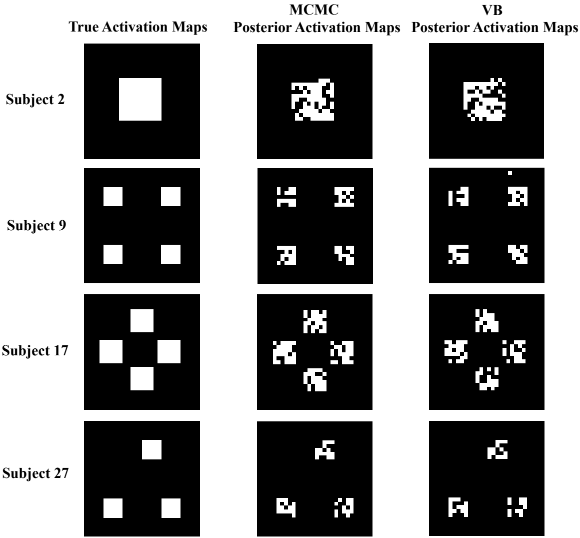
Parameter Setting:

- Four activation patterns:

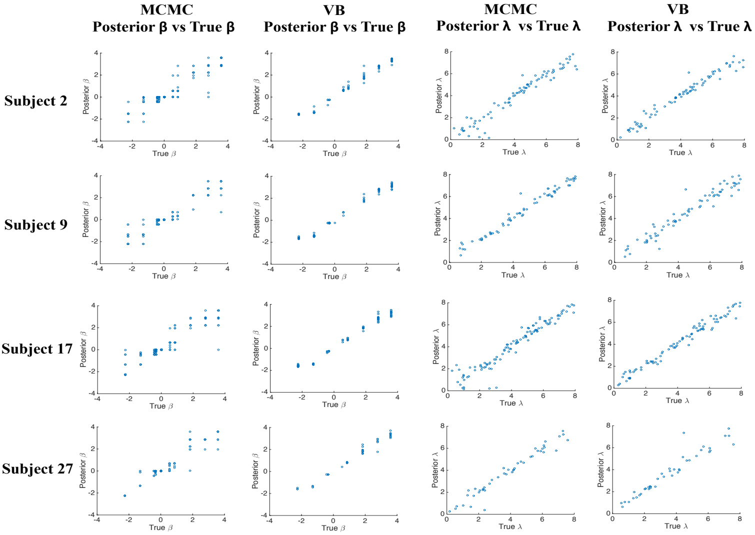


- $\beta_{i\nu}$' for active voxels: sampled from components $\phi_k \sim \text{Normal}(0, 1), k = 1, \ldots, 10$
- Delay parameters in HRF: $\lambda_{i\nu} \sim \text{Uniform}(0, 8)$
- Innovation variances: $\psi_{i\nu} \sim$ truncated Normal$(0, 1)$ on $(0, \infty)$
- Long memory parameters: $\alpha_{i\nu}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

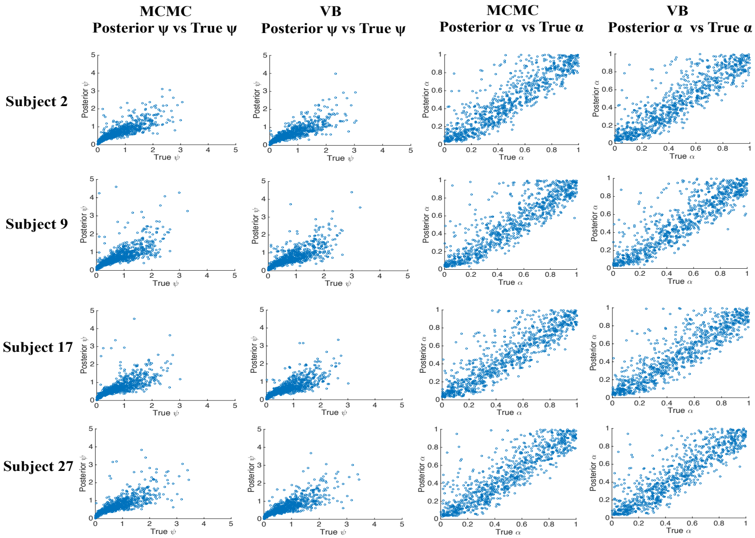
Activation Detection



Estimation of β and λ



Estimation of ψ and α



Comparison of MCMC and VB

MCMC

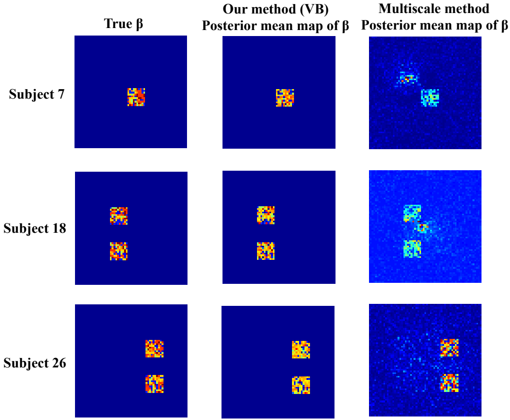
- Good performance on activation detection
- Good estimation results for model parameters
- Very expensive computation

VB

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

Synthetic Data

$Y_{syn} = Y + w$, with Y simulated from our model and w selected slice from real fMRI data. 27 subjects, 3 activation patterns.

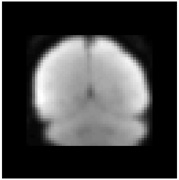


Comparison with two-stage method of Sanyal & Ferreira (2012).

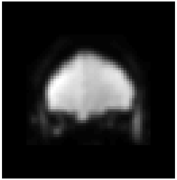
Case Study

- Real fMRI data collected by Dr. Versace's lab at MD Anderson cancer center
  - Data Dimension: 27 subjects, 286 time points, 2 slices of interest, 64 $\times$ 64 voxels per slice

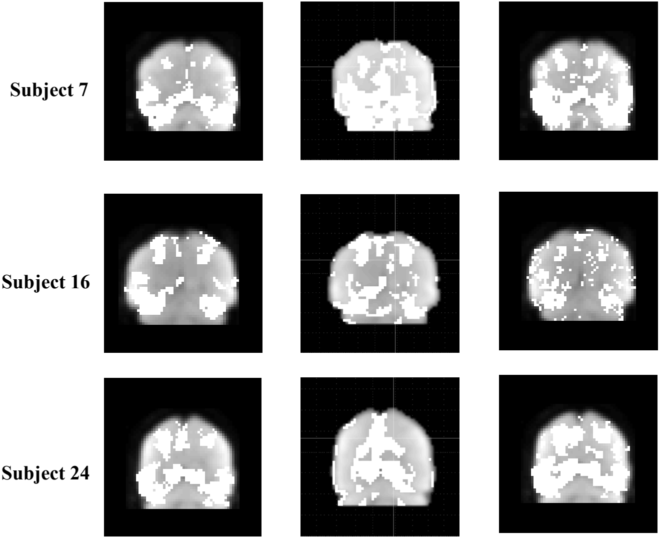
Occipital Slice
(y = -60 mm)



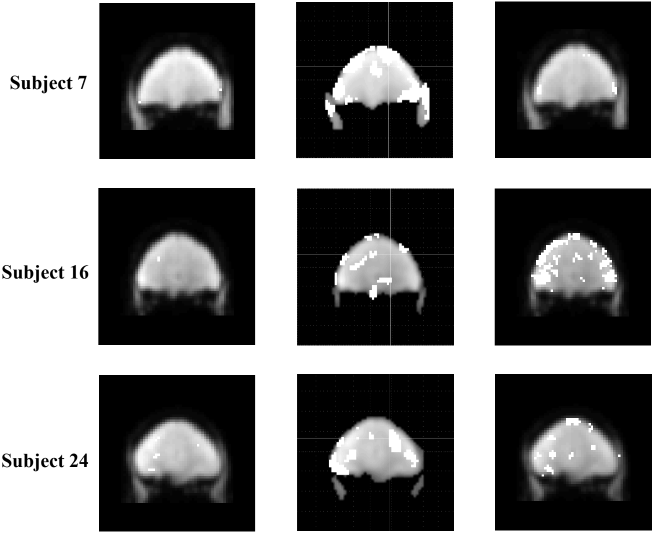
Frontal Slice
(y = +38 mm)



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



Our method; SPM8 (Penny et al., 2005); single-sub (Zhang et al., 2014).



Our method; SPM8 (Penny et al., 2005); single-sub (Zhang et al., 2014).

Results of subject-level clustering

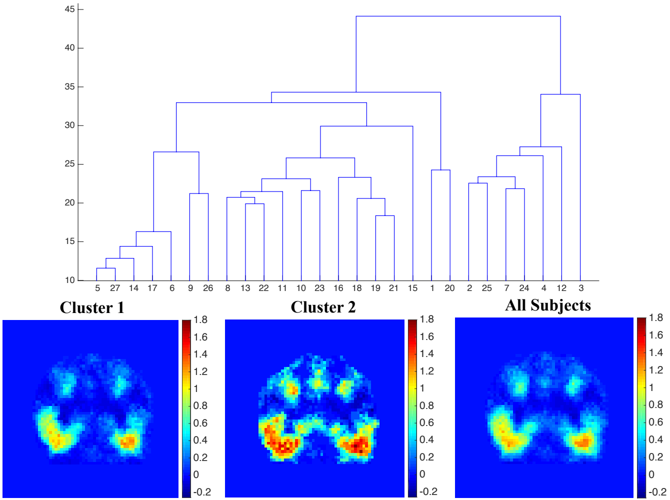


Figure : Occipital slice. Group-level β maps.

Conclusions

- Flexible class of spatio-temporal models for fMRI data.
- Incorporate several features into one modeling framework
- Use 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
- Compare VB and MCMC

Related Papers

Zhang, L., Guindani, M., Versace, F. and Vannucci, M. (2014) A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage* 95, 162–175.

Zhang, L., Guindani, M., Engelmann, J.M.,Versace, F., Vannucci, M. (2015). A Bayesian spatio-temporal model for multi-subject fMRI data. *Annals of Applied Statistics*, revised.

Zhang, L., Guindani, M., Vannucci, M. (2014) Bayesian Models for fMRI Data Analysis. *WIREs Computational Statistics*, 7, 21-41.

THANKS!

- Linlin Zhang, Ph.D. 2015, at Schlumberger, Houston, TX
- Michele Guindani, Assistant Prof, MD Anderson Cancer Center, Houston, TX
- Jeffrey M. Engelmann, Department of Behavioral Science, MD Anderson Cancer Center, Houston, TX
- Francesco Versace, Department of Behavioral Science, MD Anderson Cancer Center, Houston, TX