Inverse problems in functional brain imaging

Joint detection-estimation in fMRI

Ph. Ciuciu[1,2]

philippe.ciuciu@cea.fr www.lnao.fr

1: CEA/NeuroSpin/LNAO

2: IFR49

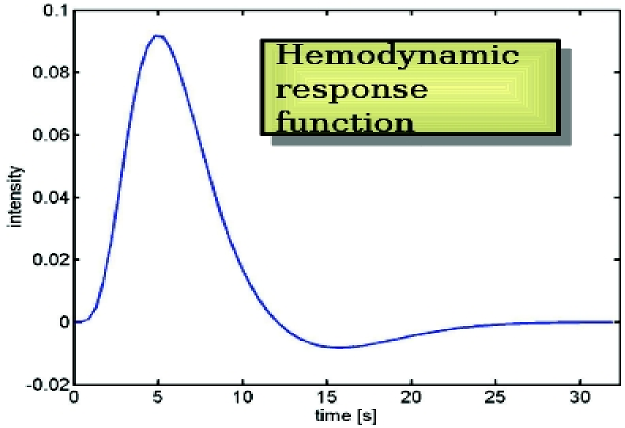
July 25, 2009

GDR -ISIS Summer school -Peyresq

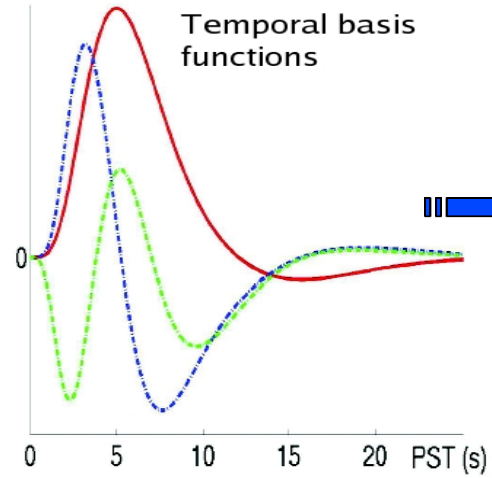
Classical fMRI analysis

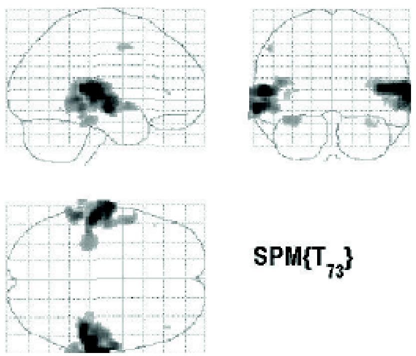
1. **Detect** and **localise** brain activations

Ex: In SPM **[Friston et al, 1994]**, the BOLD response is modelled with:



or

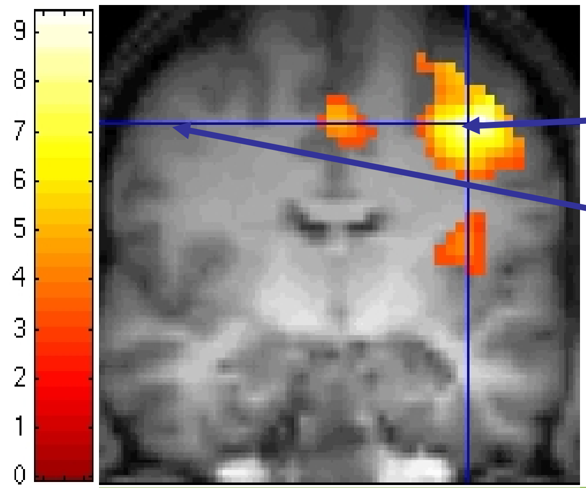




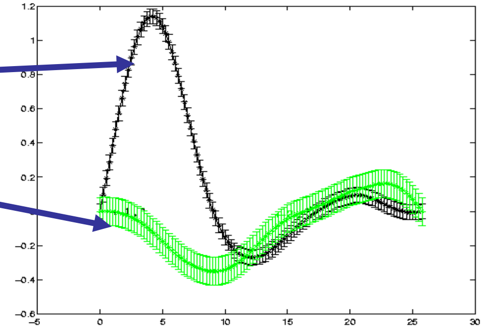
Compute statistical activation Maps

2. **Estimate** the dynamics of activation

[Goutte et al, IEEE TMI 2000; Marrelec et al, HBM 2003]

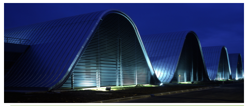


HRF estimates



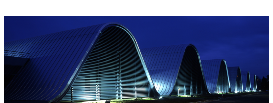
Time s

Probe brain dynamics



Classical fMRI analysis

- GLM limitations:
  - A single HRF shape is not physiologically appropriate
  - Studies report HRF variability:
    - within subject (between regions, sessions, conditions, trials)
      - **[Miezin et al., NIM 2000; Ciuciu et al., IEEE TMI 2003]**
      - **[Neumann et al, NIM 2003, 2006; Smith et al, NIM 2005]**
    - between subjects
      - **[Handwerker et al., NIM 2004; Aguirre et al., NIM 1998]**
    - between groups (infants, patients,...)
      - **[D'Esposito et al, NIM 1998, 2003; Richter and Richter, NIM 2003]**
    - ...



Motivations

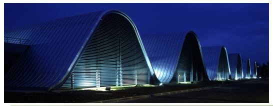
- Detection of brain activation and estimation of brain dynamics are addressed separately

- Any detection method supposes a given HRF shape

- Any estimation algorithm provides relevant results in activated voxels or regions only

→ Address these two problems simultaneously in a joint detection-estimation (JDE) framework

[Makni et al., IEEE SP 2005, ISBI'06, ICASSP'06, NIM, 2008; Vincent et al, ICASSP'07, EMBC'07, ISBI'08; Ciuciu et al, ISBI'08 ; Risser et al, sub. to NIPS'08; Vincent et al, sub. to IEEE TMI]



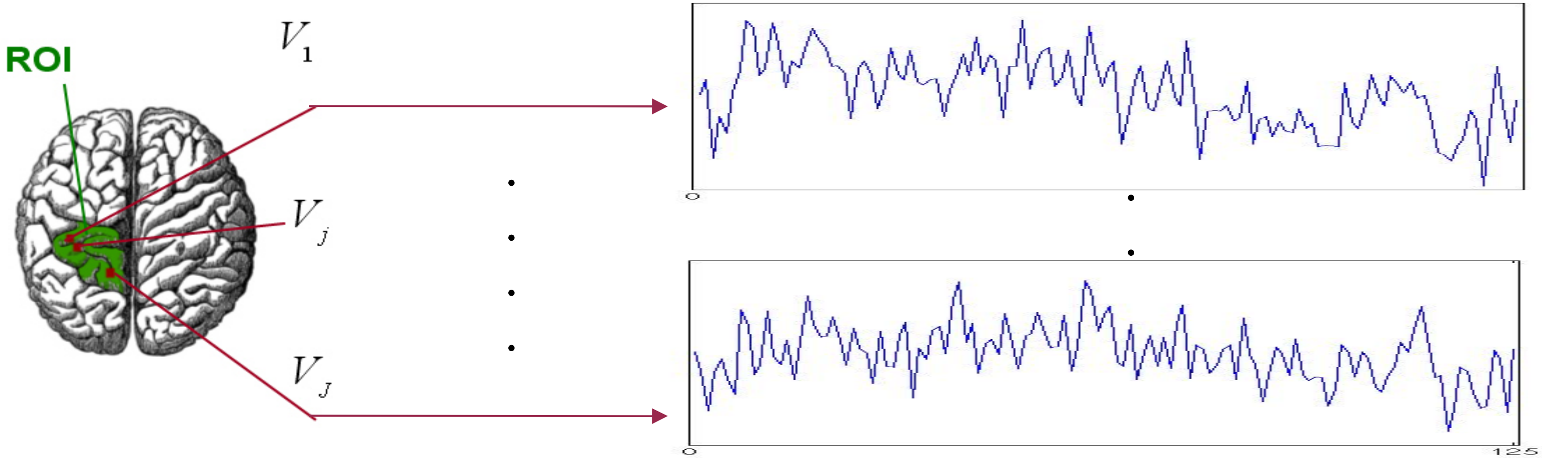
Forward BOLD signal model

[Makni, Ciuciu et al., IEEE SP 2005;
Makni et al, Ciuciu, NeuroImage, 2008]

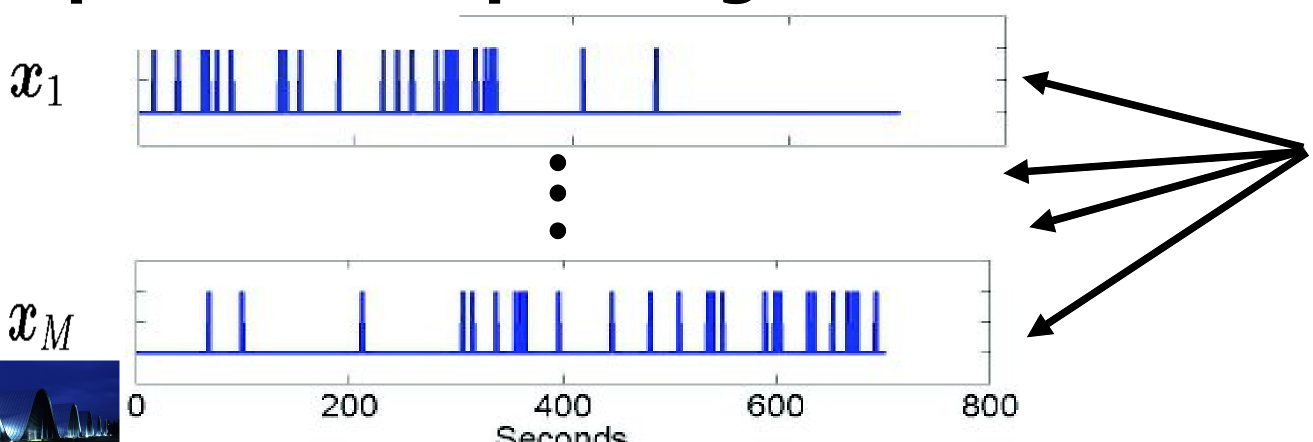
- **Region definition:** $\mathcal{R} = (V_j)_{j=1:J}$ ← voxels

Region of interest

fMRI time series in voxels of the ROI



- **Experimental paradigm:**

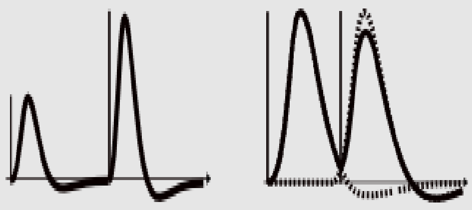


Arrival times of M stimuli (onsets)

Forward BOLD signal model

- **Temporal hypotheses:** for standard ISIs

Hyp. 1: **linearity**



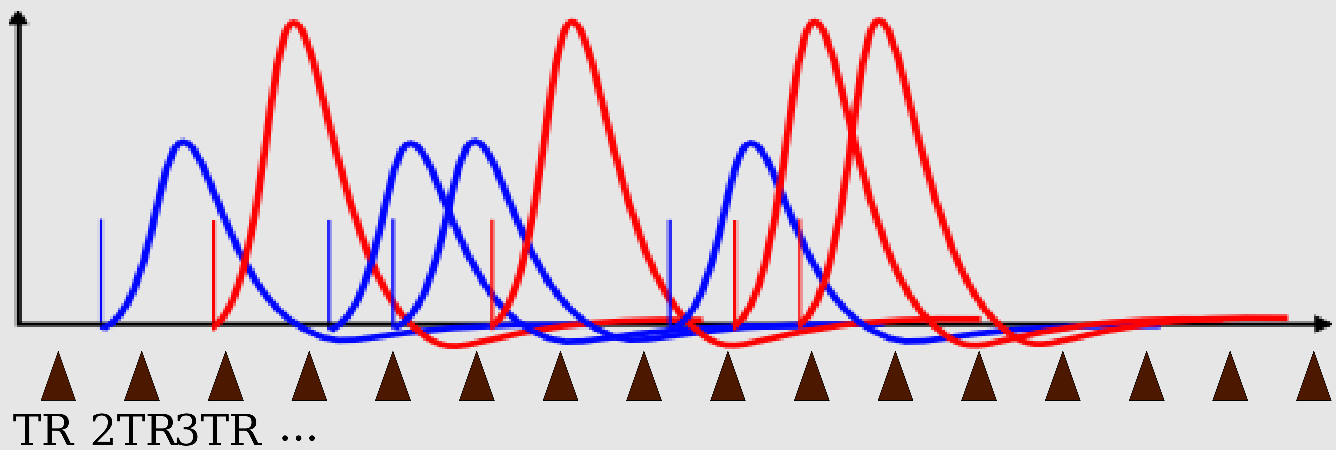
+

Hyp 2: **stationarity**

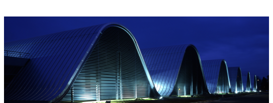
→ $\boldsymbol{h}$

Convolution kernel

Hyp 3: **additivity**



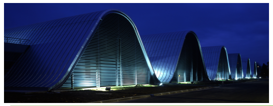
→ Stimulus-varying NRLs: $a^1 \neq a^2$



Forward BOLD signal model

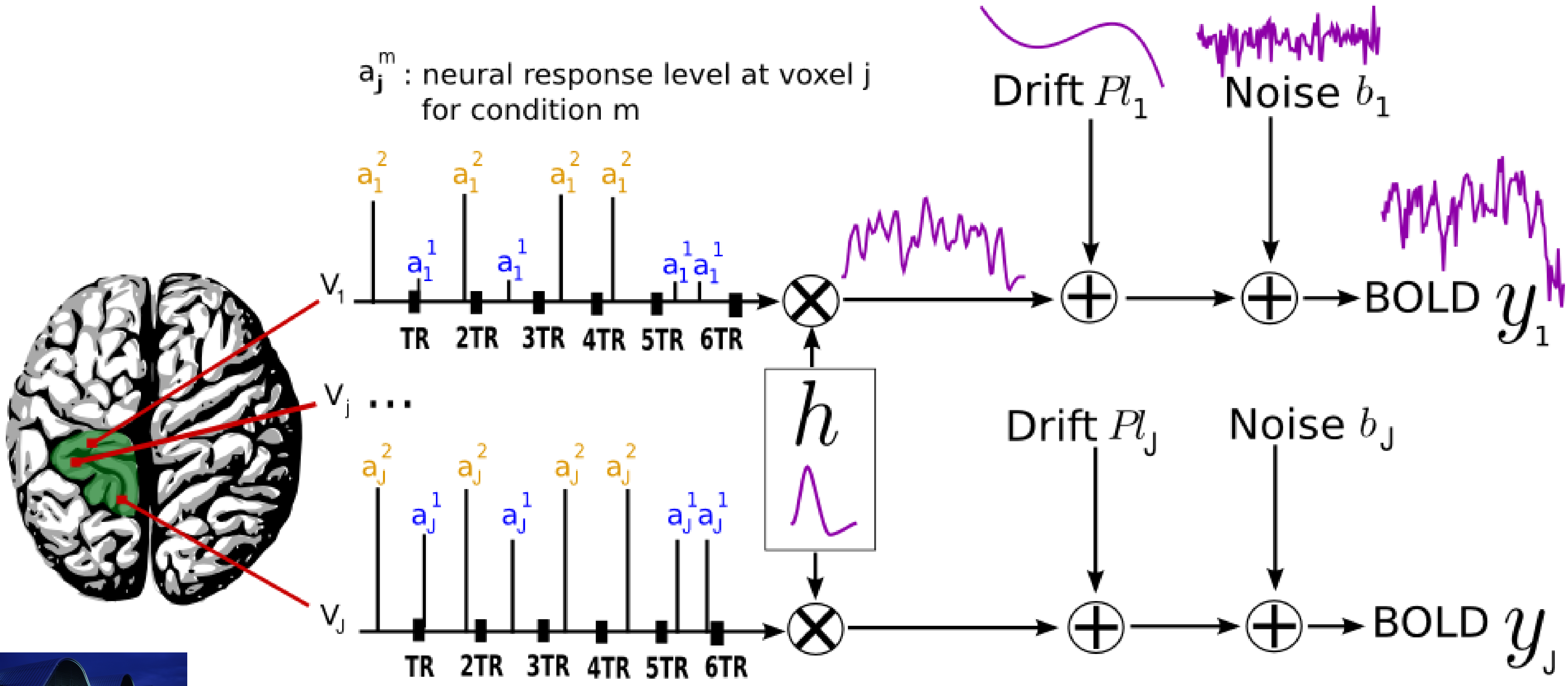
- **Spatial hypotheses**: functionally homogeneous ROI
  - Single HRF shape
  - Voxel-dependent magnitudes of the BOLD response

Neural Response Levels $\mathbb{a} = \{a_j^m\}_{j=1:J,m=1:M}$



Forward BOLD signal model

- **Spatial hypotheses**: functionally homogeneous ROI
  - Single HRF shape **[Makni, Ciuciu et al., IEEE SP 2005; Makni et al, Ciuciu, NeuroImage, 2008]**
  - Voxel-dependent magnitudes of the BOLD response



Forward BOLD signal model

Unknown parameters

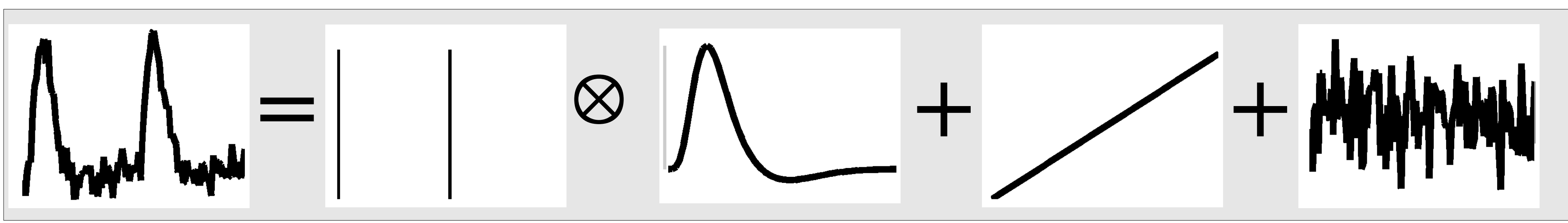
NRL in V_j and for condition m

HRF

Drift coefficients

Noise statistics in voxel V_j

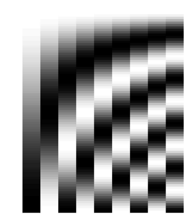
$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j$$



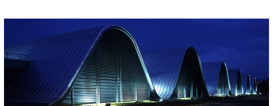
BOLD signal measured in voxel V_j

Arrival time of stimulus m

Orthonormal basis for low frequency drift modelling



Known parameters



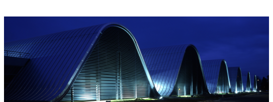
Bayes’ rule

likelihood

$$p(\boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta} \mid \mathbb{y}) = \frac{p(\mathbb{y} \mid \boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta})\, p(\boldsymbol{h} \mid \sigma_{\boldsymbol{h}}^2)\, p(\mathbb{l} \mid \sigma_l^2)\, p(\boldsymbol{a} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbb{y})}$$

How the data are generated from the parameters?

Forward modeling



Likelihood definition

- **Main hypothesis: noise decorrelated in space**

→ fMRI time series are statistically independent in space:

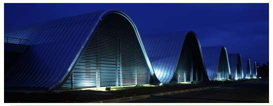
$$p(\mathbb{y} \,|\, \boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta}_0) = \prod_j p(\boldsymbol{y}_j \,|\, \boldsymbol{h}, \boldsymbol{a}_j, \boldsymbol{\ell}_j, \theta_{0,j})$$

$$\propto \prod_j f_{B_j}(\boldsymbol{y}_j - \Sigma_{m=1}^{M} a_j^m \boldsymbol{X}_m \boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j)$$

Temporal noise model: either white or serially correlated AR(1)

$\boldsymbol{b}_j \sim \mathcal{N}(\boldsymbol{0}, \epsilon_j^2 \boldsymbol{I})$ → $\theta_{0,j} = [\epsilon_j^2]$

$\boldsymbol{b}_j \sim \mathcal{N}(\boldsymbol{0}, \epsilon_j^2 \boldsymbol{\Lambda}_j^{-1})$ → $\theta_{0,j} = [\epsilon_j^2, \rho_j]$



[Makni et al, Ciuciu , NeuroImage 2008]

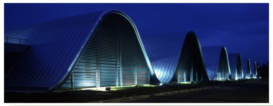
Bayes' rule

Prior

$$p(\boldsymbol{h}, \mathbb{a}, \mathbb{1}, \boldsymbol{\theta} \mid \mathbb{y}) = \frac{p(\mathbb{y} \mid \boldsymbol{h}, \mathbb{a}, \mathbb{1}, \boldsymbol{\theta})\, p(\boldsymbol{h} \mid \sigma_{\boldsymbol{h}}^2)\, p(\mathbb{1} \mid \sigma_l^2)\, p(\boldsymbol{a} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbb{y})}$$

What do we know about the parameters before the data are acquired?

Prior modeling

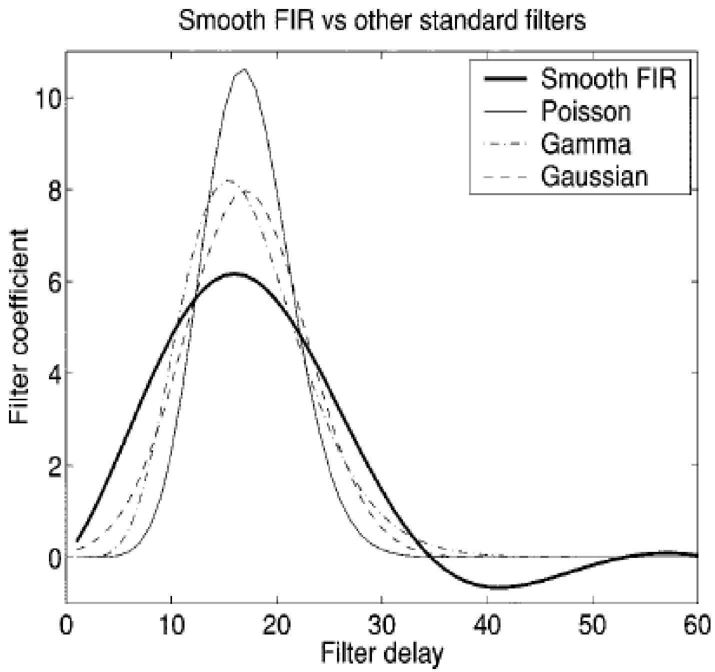


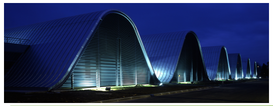
HRF prior modeling

- **Information about the HRF shape**
  - nonparametric approaches
  - Averaging: **[Buckner et al, 1996]**
  - Selective averaging: **[Dale et al, HBM 1997]**
  - FIR (Finite Impulse Response)
  - Regularized FIR (smoothing prior):

$$\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, \sigma_{\boldsymbol{h}}^2 \boldsymbol{\Sigma})$$

**[Marrelec, Ciuciu et al, IPMI'03
Ciuciu et al., IEEE TMI 2003]**





NRL prior modeling

- **Spatial information** [Vincent, Ciuciu et al, ICASSP'07]
  - Independence between conditions: $p(\mathrm{a}) = \prod_m p(\boldsymbol{a}^m \,|\, \boldsymbol{\theta}^m)$
  - **Spatial mixture model** (SMM) for each m

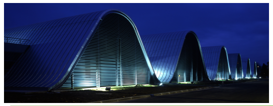
$$p(\boldsymbol{a}^m \,|\, \boldsymbol{\theta}^m) = \sum_{q^m = \boldsymbol{i}} \Big(\prod_{j=1}^{J} f_i(a_j^m \,|\, \boldsymbol{\theta}^m) \Big) \Pr(\boldsymbol{q}^m \,|\, \beta^m) \Big)$$

$$f_i(\cdot \,|\, \boldsymbol{\theta}^m) = f(\cdot \,|\, q_j^m = i, \boldsymbol{\theta}^m), \quad \forall j$$

  - Discrete Markov random Field (eg Potts/Ising):

$$Pr(\boldsymbol{q}^m \,|\, \beta^m) = \frac{1}{Z(\beta^m)} \exp\Big(\beta^m \sum_{j \sim k} I(q_j^m = q_k^m) \Big)$$

Partition function: $Z(\beta^m) = \sum_{\boldsymbol{q}^m} \exp\Big(\beta^m \sum_{j \sim k} I(q_j^m = q_k^m) \Big)$



Prior model

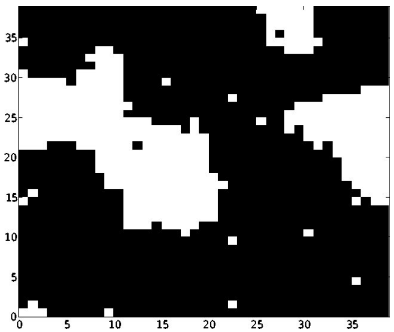
- Two-class Gaussian mixture **[Vincent et al., ICASSP' 07]**

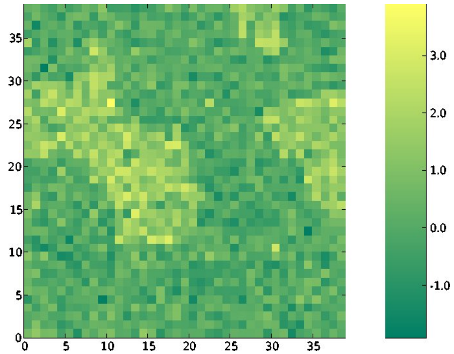
Non-activating voxels: $f_0(\cdot\,|\,\boldsymbol{\theta}^m) = \mathcal{N}(0, \sigma^2_{0,m})$

Activating voxels: $f_1(\cdot\,|\,\boldsymbol{\theta}^m) = \mathcal{N}(\mu_{1,m}, \sigma^2_{1,m})$

Unknown hyper-parameters: $\{\beta^m, \sigma^2_{0,m}, \mu_{1,m}, \sigma^2_{1,m}\}$

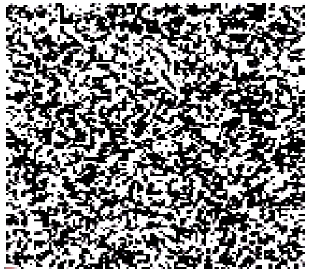
$\boldsymbol{q}^1$



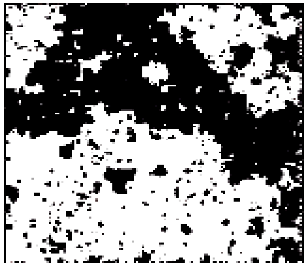


$\boldsymbol{a}^1$

- Influence of β

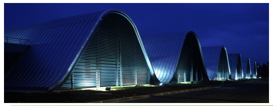


$\beta = 0.01$



$\beta = 0.45$

$\beta = 0.91$



Other statistical parameters

- **Noise parameters** [Makni et al, ISBI'06]

$$p(\boldsymbol{\theta}_0) = \prod_{j=1}^{J} p(\rho_j, \sigma^2_{\varepsilon_j}) = \prod_{j=1}^{J} \epsilon_j^{-1} \mathbb{1}_{(-1,1)}(\rho_j),$$

- **Mixture probabilities** [Makni et al, NIM 2008]
  - 2-class mixture(Jeffreys prior): $p(\boldsymbol{\lambda}) = \prod p(\lambda_m) \propto \prod_m \lambda_{1,m}^{1/2} \lambda_{0,m}^{1/2}$
  - 3-class mixture: $p(\boldsymbol{\lambda}) = \prod_m \mathcal{D}_3(\boldsymbol{\lambda}_m \,|\, \boldsymbol{\delta}), \quad \boldsymbol{\delta} = \delta \mathbf{1}_3$

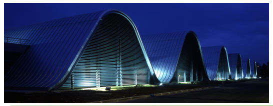
- **Mixture parameters** [Makni et al, NIM 2008]

Non-activating voxels

$$p(\boldsymbol{\sigma}_{0,m}) = \prod_m \sigma_{0,m}^{-1}$$

Activating voxels

$$p(\boldsymbol{\mu}_{1,m}, \boldsymbol{\sigma}_{1,m}) = \prod_m \mathcal{N}(\mu_{1,m}; 0, c)\, \mathcal{IG}(\sigma^2_{1,m}; a, b)$$

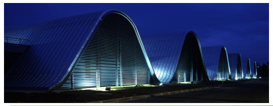


Bayes' rule

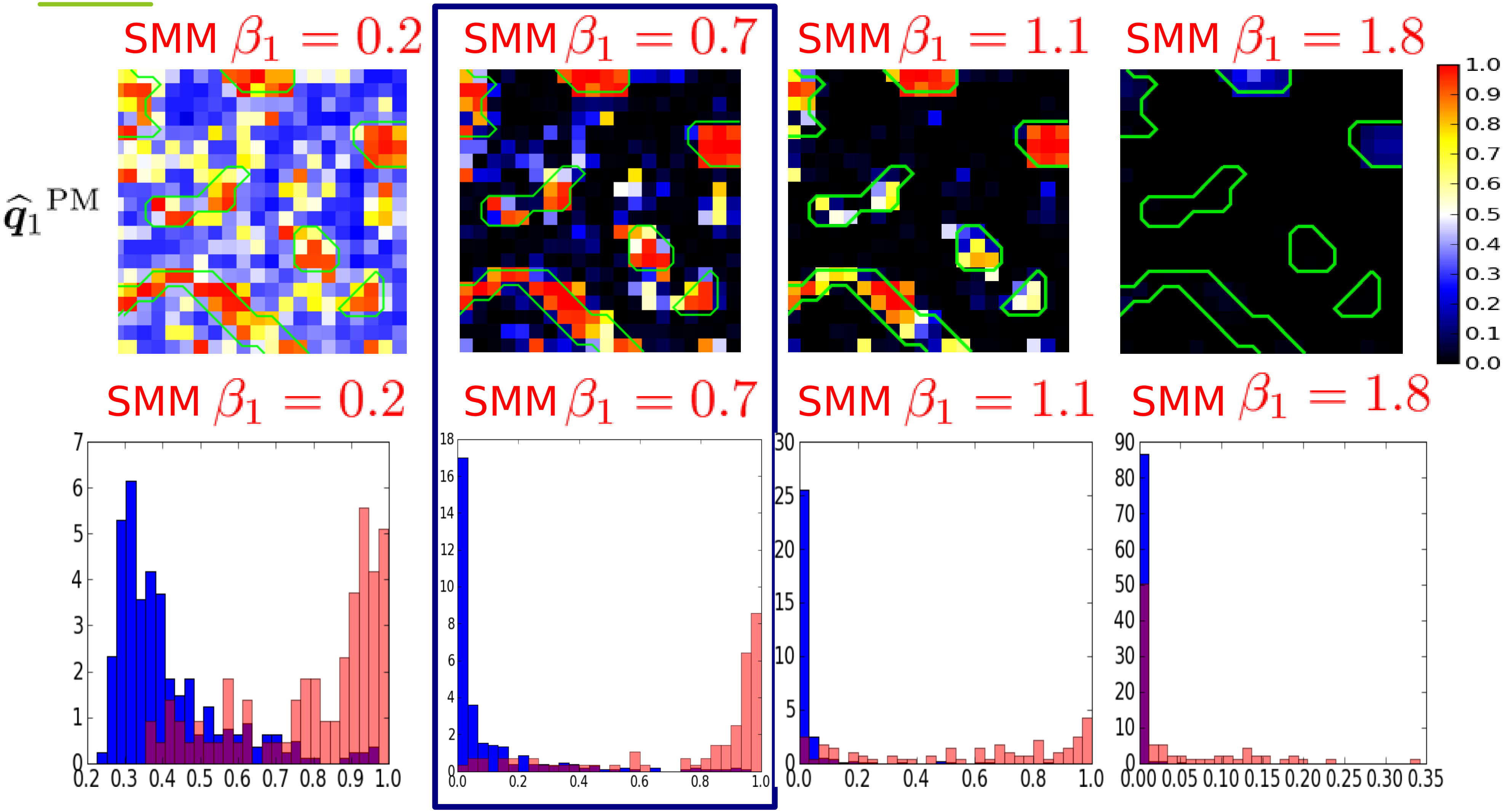
$$\boxed{p(\boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta} \mid \mathbb{y})} \propto p(\mathbb{y} \mid \boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta}) p(\boldsymbol{h} \mid \sigma_{\boldsymbol{h}}^2)\, p(\mathbb{l} \mid \sigma_l^2) p(\boldsymbol{a} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

What do we know about the HRF, the NRLs and the hyper-parameters given the data?

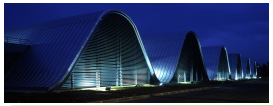
Keystone of learning scheme:
simulating realizations of $p(\boldsymbol{h}, \mathbb{a}, \boldsymbol{\theta} \mid \mathbb{y})$ **using Gibbs sampler**



Simulation results

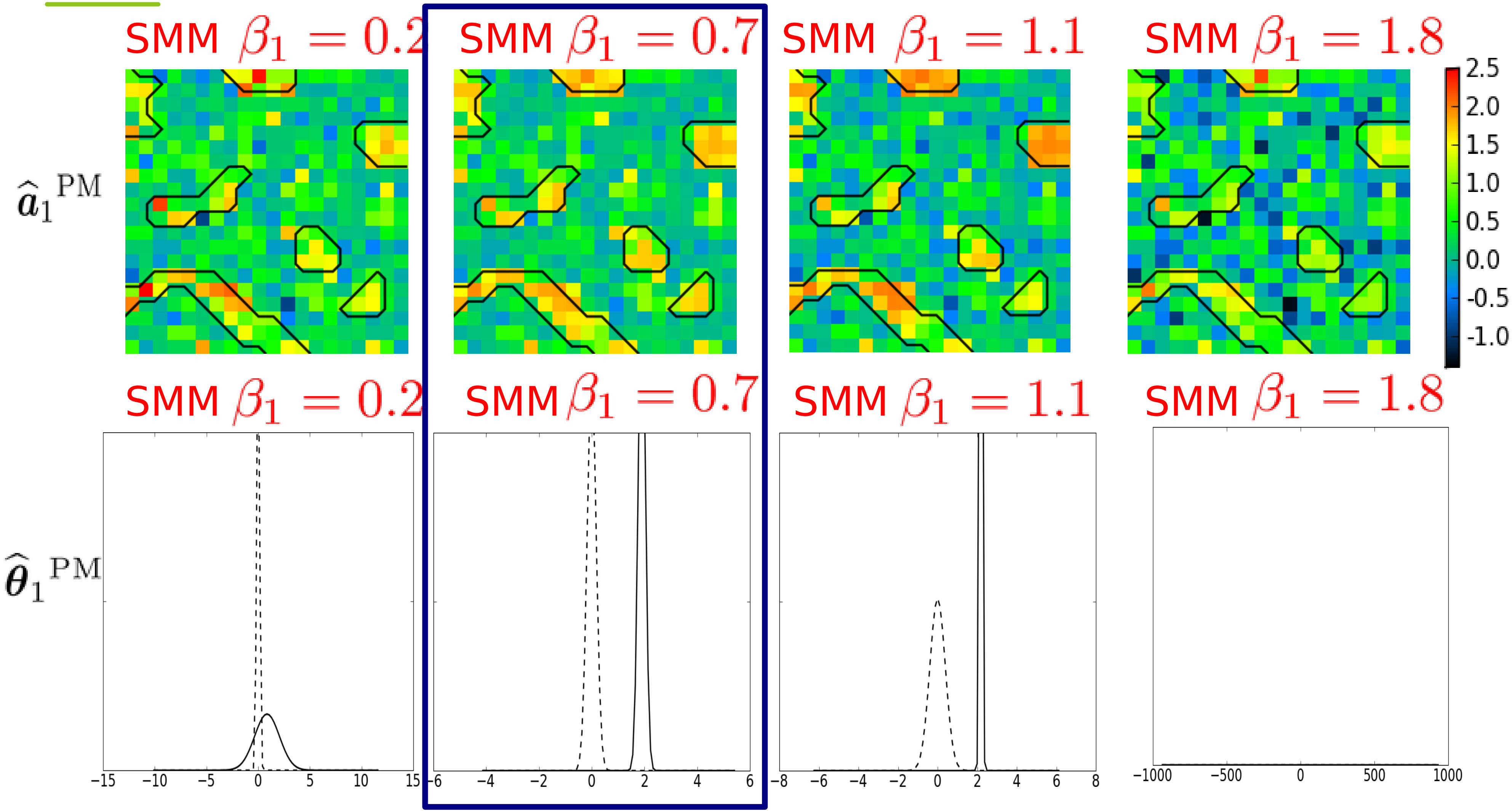


True activations   True inactivations

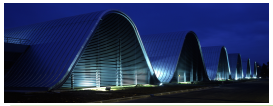


Strong influence of regularization parameter

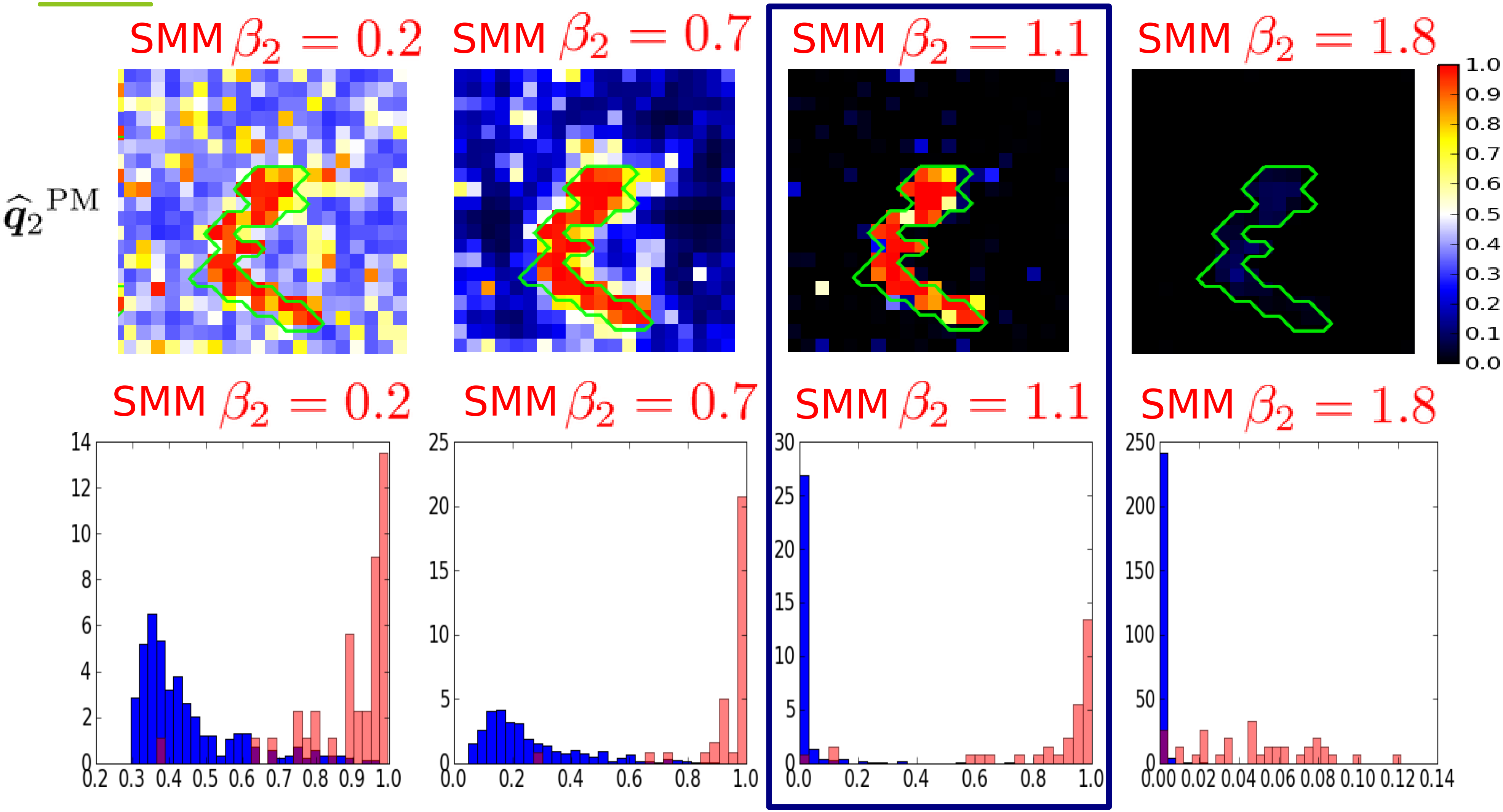
Simulation results



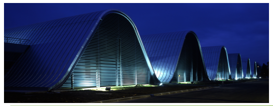
dash/continuous line: density of inactivating/activating voxels



Simulation results

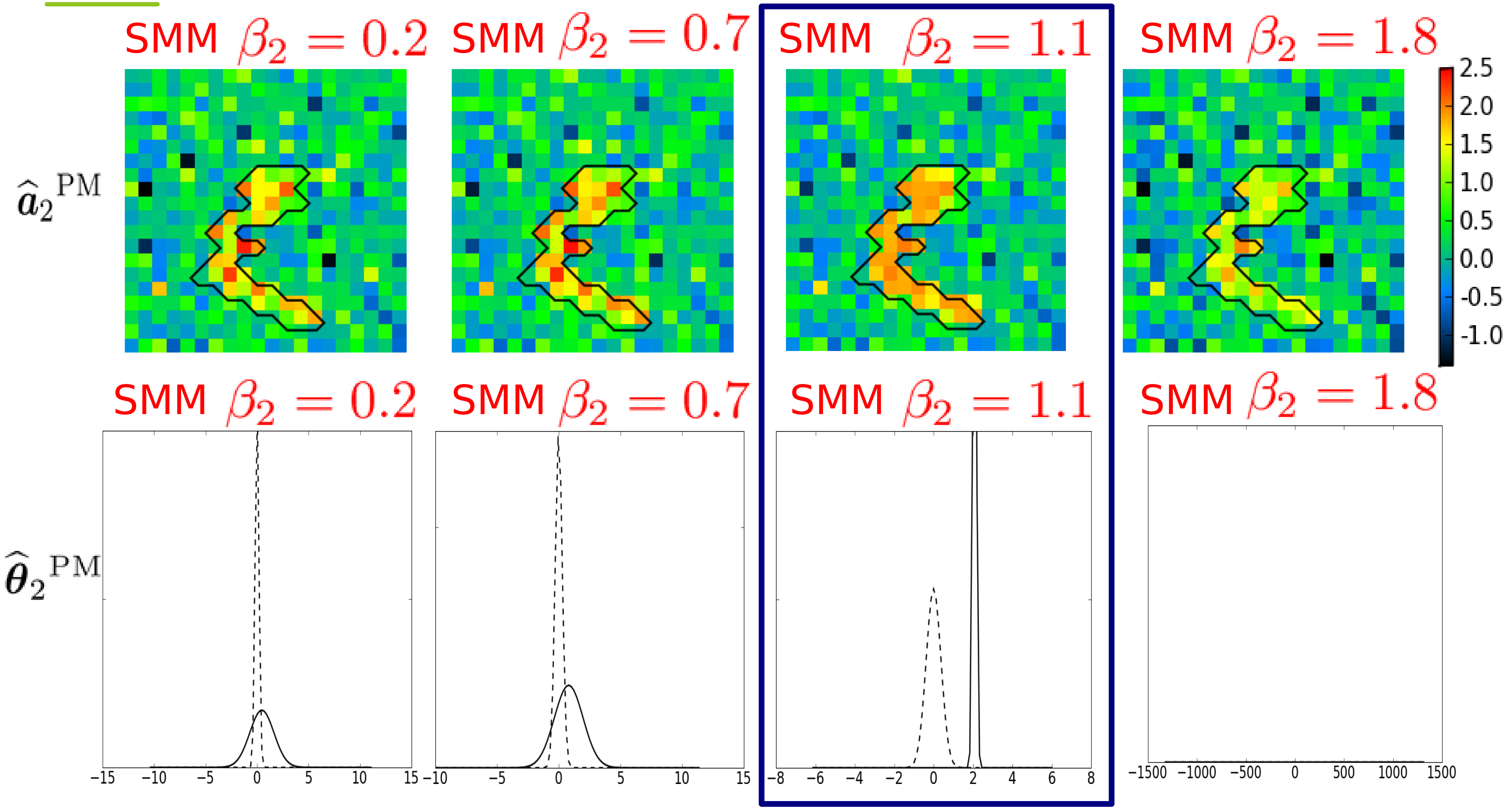


True activations True inactivations

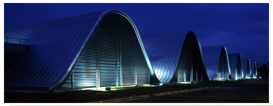


Best tuning: larger β_2 → lower SNR for $m = 2$

Simulation results

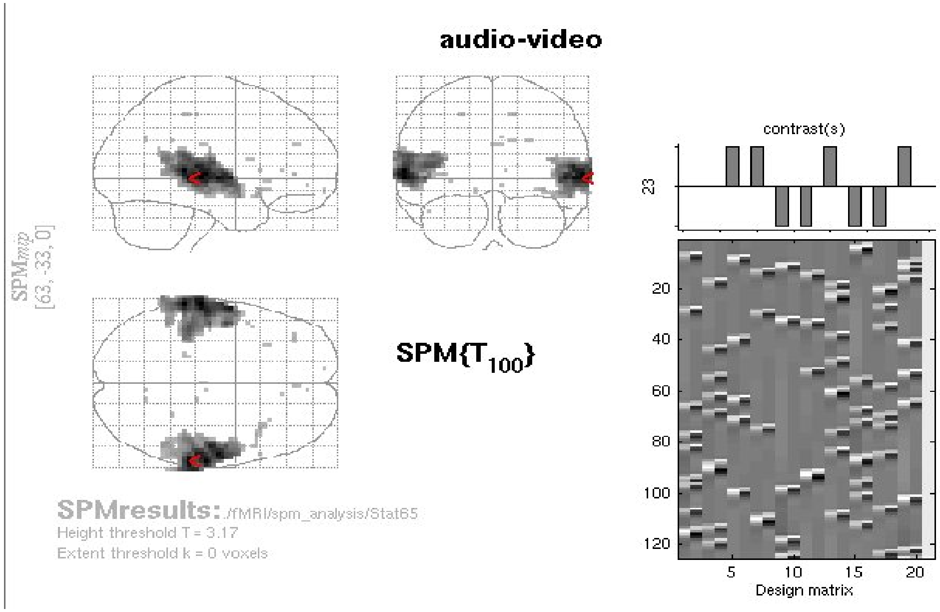


dash/continuous line: density of inactivating/activating voxels

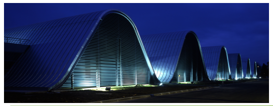


Real data sets

- Localizer fMRI experiment:
  - Experimental conditions under study: auditory and visual stimuli
  - Event-related paradigm :
    - Short stimuli duration
    - Inter-stimulus interval : ~3s to 10s
    - Randomised sequence
  - 125 scans with TR = 2.4s, scanning at 3T

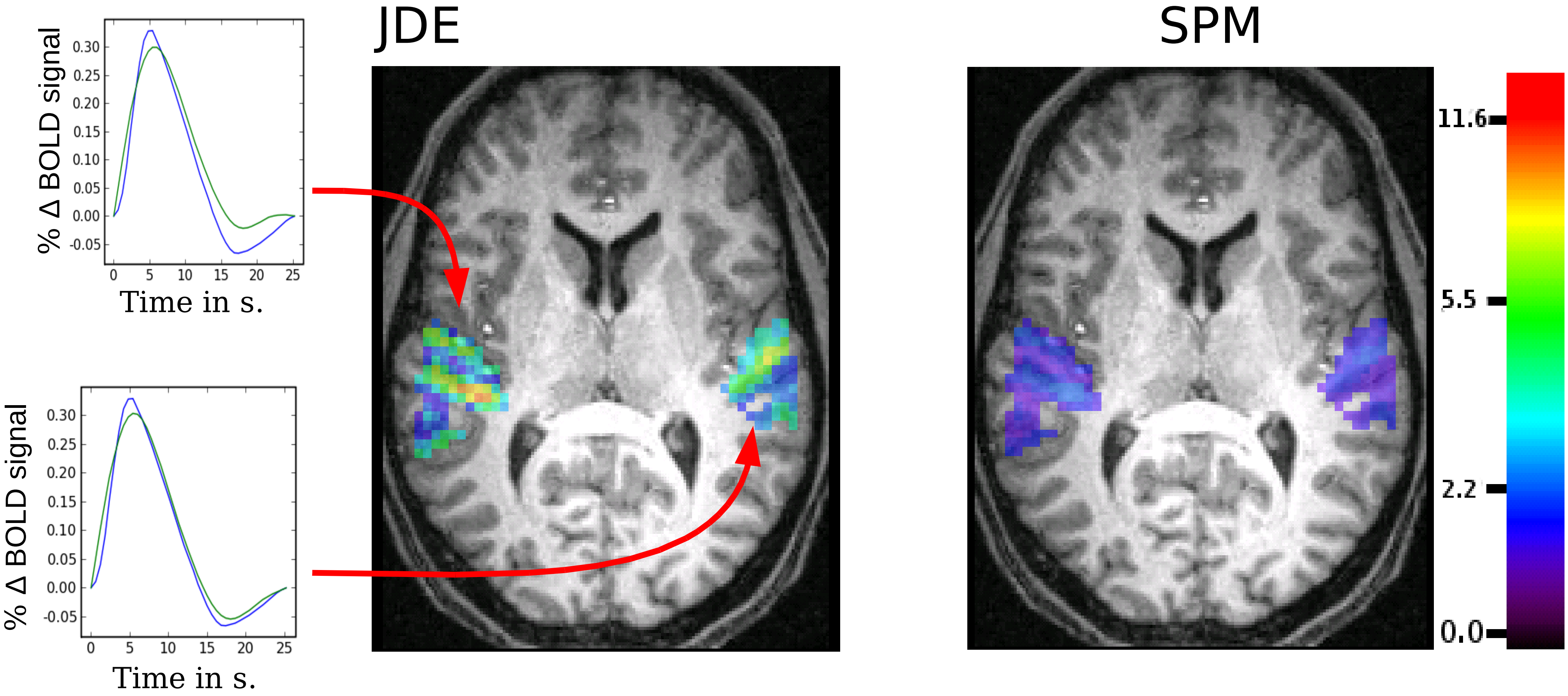


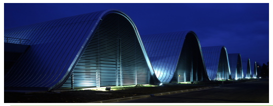
ROIs located in the primary auditory cortices



SPM vs. JDE

Auditory – Visual contrast: $\widehat{\boldsymbol{a}}^1 - \widehat{\boldsymbol{a}}^2$



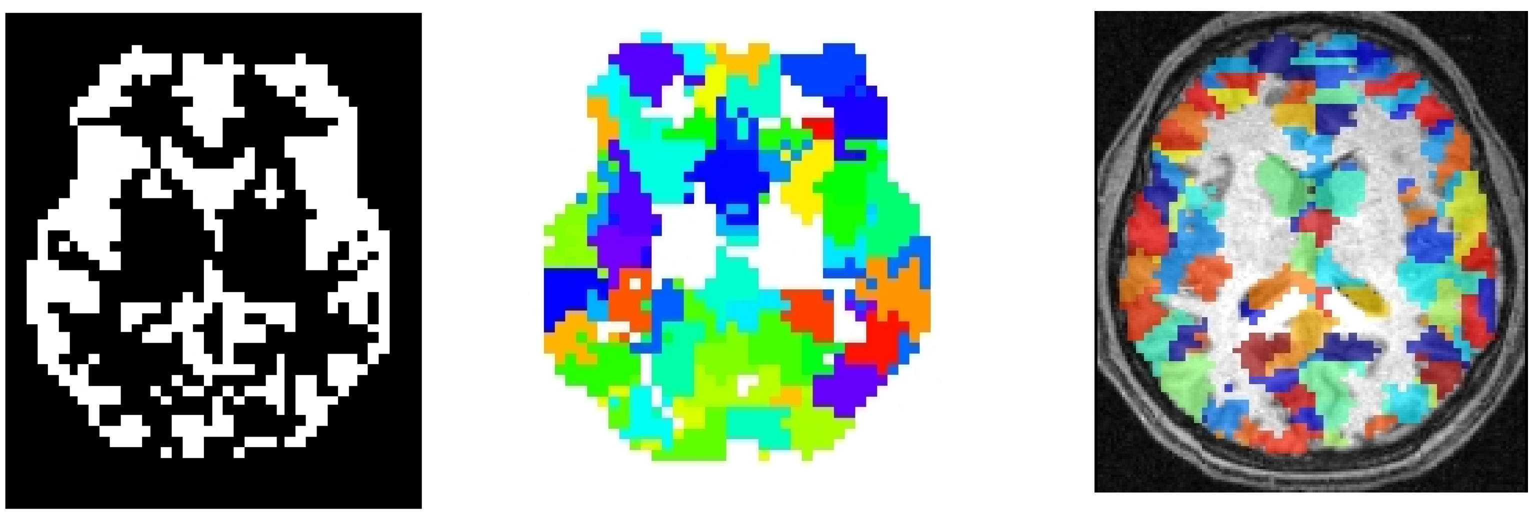


Bilateral activation detected along the gray matter from raw data sets (spatially unsmoothed)

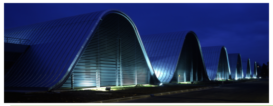
Whole brain analysis

- **First step:**
  Segmentation of the gray/white matter interface from the T1 MRI

- **Second step :**
  brain parcellation based upon functional similarities and spatial connectivty **[Flandin et al, ISBI'02; Thirion et al, HBM 2006]**



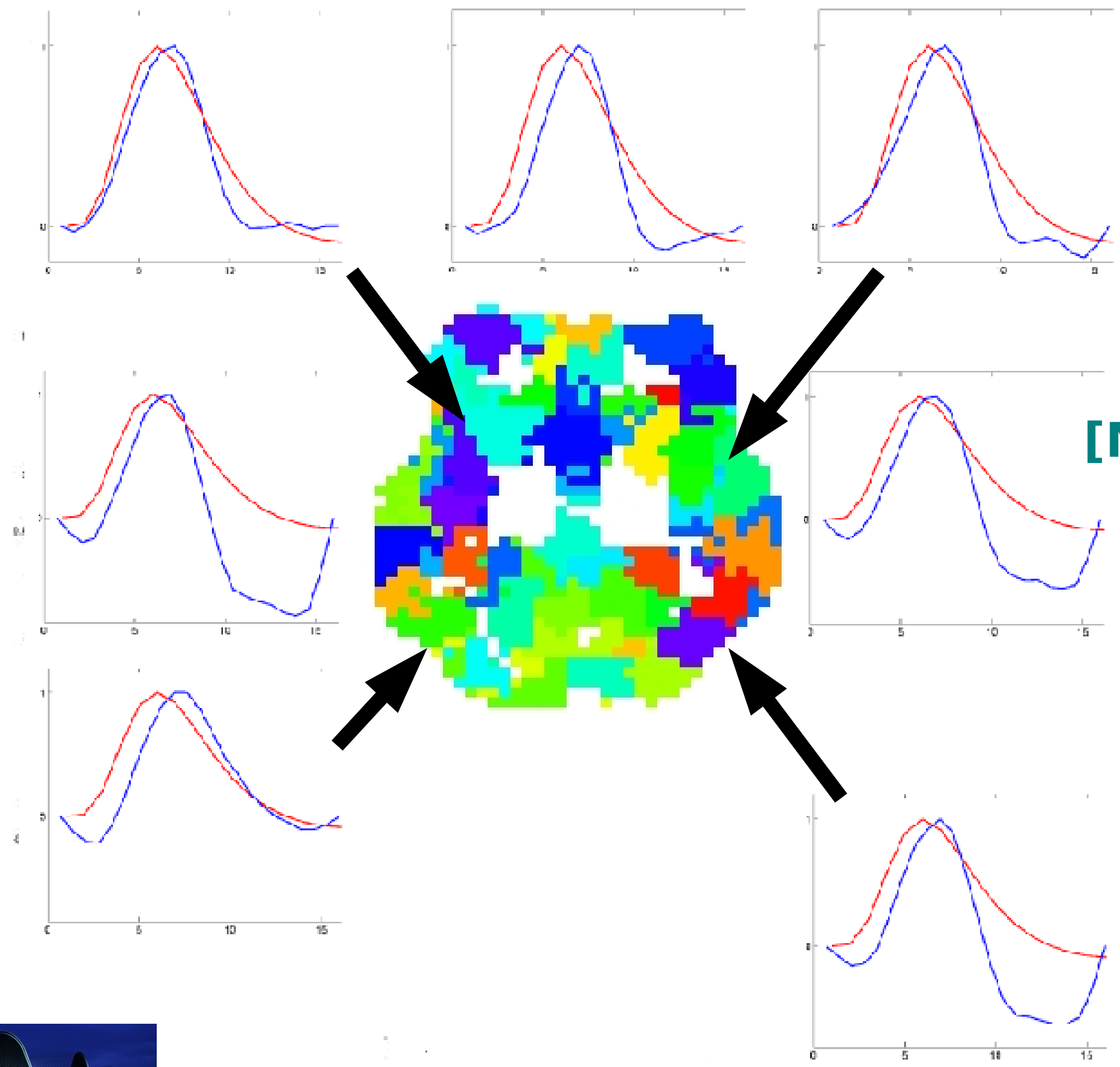
[Makni et al, Ciuciu, NIM 2008]



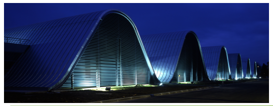
Sources of variability

Subject, region

[Birn et al, NIM 2001; Marrelec et al, HBM 2003; Menon et al NIM 2003; Neumann et al NIM 2003; Handwerker et al, NIM 2004]



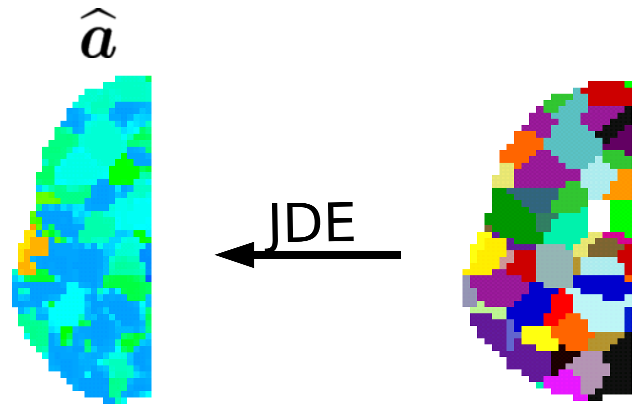
[Makni et al, Ciuciu, NIM 2008]



Sensitivity analysis

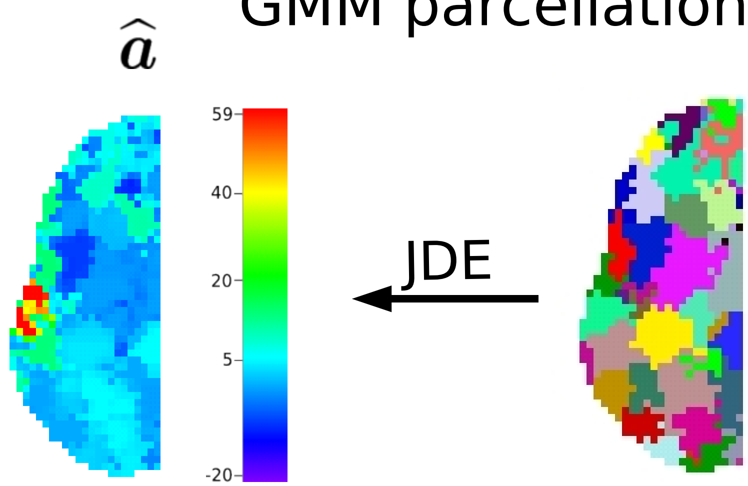
[Vincent, Ciuciu, ISBI 2008]

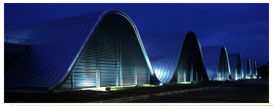
Random parcellation



Single JDE run

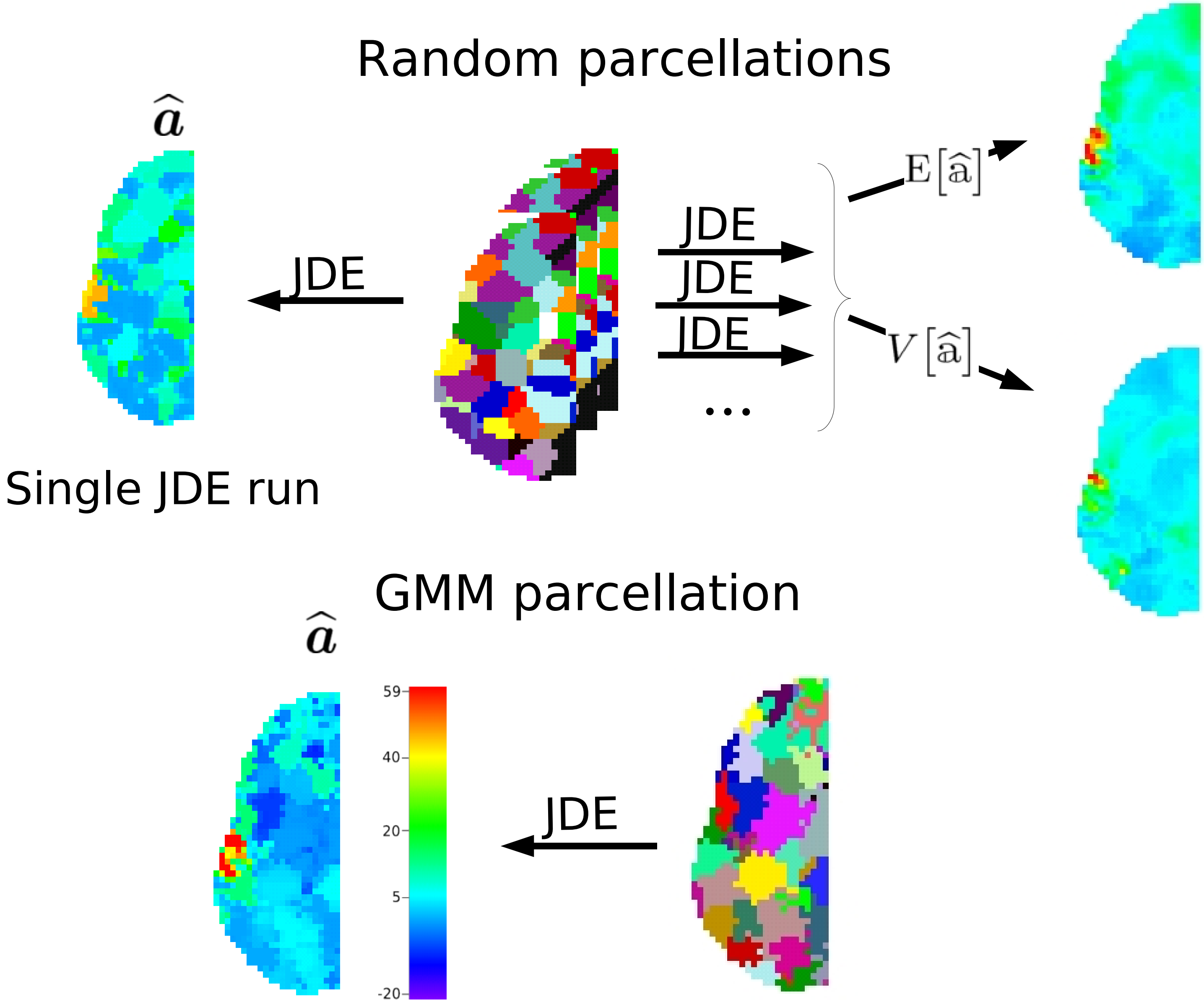
GMM parcellation

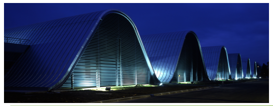




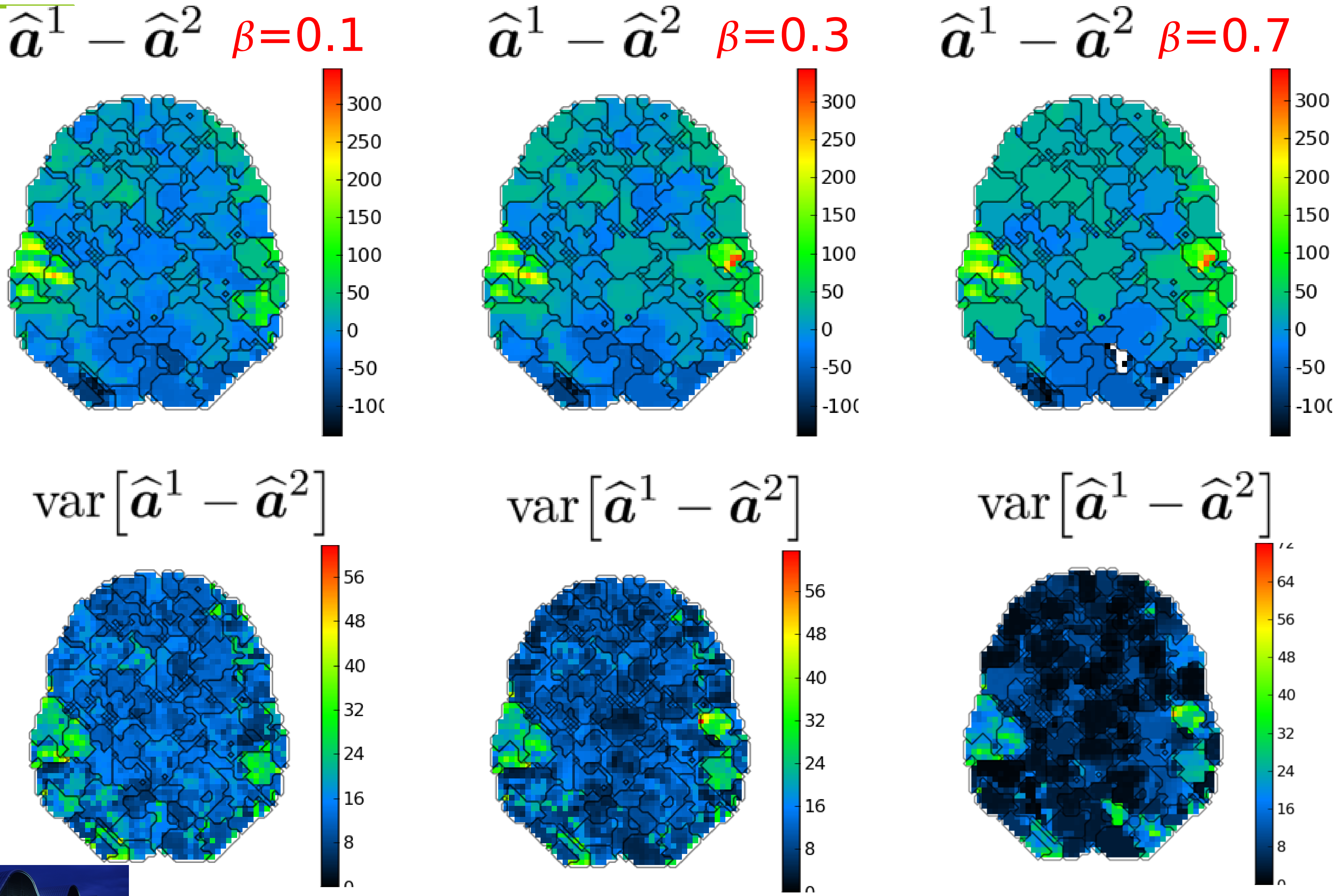
Sensitivity analysis

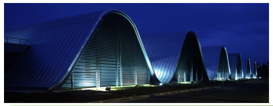
[Vincent, Ciuciu, ISBI 2008]





Whole brain: supervised SMM





Adaptive spatial regularization

- RW Metropolis-Hastings step for sampling β
  - Preliminary estimation of the partition function $Z(\beta)$

Importance sampling dentity:

$$\frac{Z(\beta)}{Z(\beta')} = \mathrm{E}_{\beta'}\left[\frac{\exp(\beta\,\mathcal{U}(\boldsymbol{q}))}{\exp(\beta'\,\mathcal{U}(\boldsymbol{q}))}\right] \text{ with } p_{\beta'}(\cdot) = Z(\beta')^{-1}\exp(\beta'\mathcal{U}(\cdot))$$

Practical implementation:

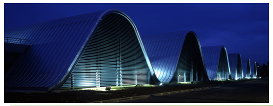
Tabulate $Z(\beta)$ over a fine grid $(\beta_0 = 0, \ldots, \beta_D)$

For $d = 1, \ldots, D$

- Generate $(\boldsymbol{q}_k)_{k=1:K}$ of $p_{\beta_{d-1}}$ (SW scheme)
- Compute $\log Z_{\mathrm{MCMC}}(\beta_d)$

$$\log Z_{\mathrm{MCMC}}(\beta_d) = \log Z_{\mathrm{MCMC}}(\beta_{d-1}) + \log\left(\frac{1}{K-I}\sum_{k=I+1}^{K}\exp\big((\beta_d - \beta_{d-1})\,\mathcal{U}(\boldsymbol{q}_k)\big)\right)$$

[Meng and Rubin, Biometrika 1998]



Hyper-parameter inference

- Random walk Metropolis-Hastings step for β

$$\alpha(\beta^{(t)} \to \beta^{(c)}) = \min(1, A_{t\to c})$$

$$A_{t\to c} = \frac{p(\beta^{(c)} \mid \boldsymbol{x}^{(t)})}{p(\beta^{(t)} \mid \boldsymbol{x}^{(t)})} \frac{g(\beta^{(t)}|\beta^{(c)})}{g(\beta^{(c)}|\beta^{(t)})}.$$

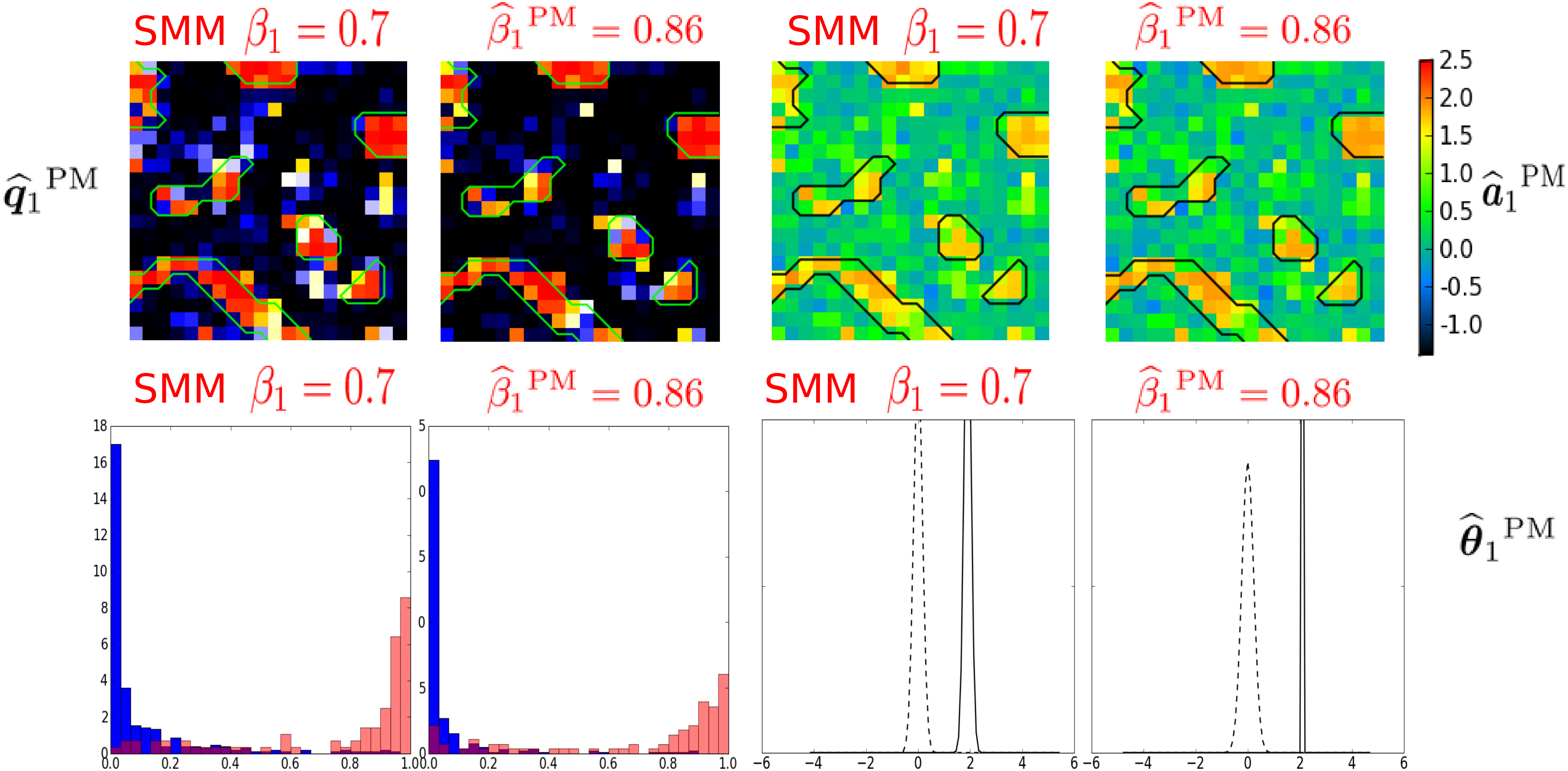
$$= \frac{Z(\beta^{(t)})}{Z(\beta^{(c)})} \exp\big(-(\beta^{(c)} - \beta^{(t)})U(\boldsymbol{x}^{(t)})\big) B_{t\to c}$$

- Special case:

$$g(\cdot|x) \sim \mathcal{N}_{[0,\beta_{\max}]}(x, \xi^2) \Longrightarrow B_{t\to c} = \frac{\mathtt{erf}(-\xi^{-2}\beta^{(c)})}{\mathtt{erf}(-\xi^{-2}\beta^{(t)})} \frac{\mathtt{erf}(\xi^{-2}(\beta_{\max} - \beta^{(t)}))}{\mathtt{erf}(\xi^{-2}(\beta_{\max} - \beta^{(c)}))}.$$

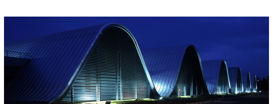
- Alternative: Gibbs sampling on the discrete grid

Supervised vs. unsupervised

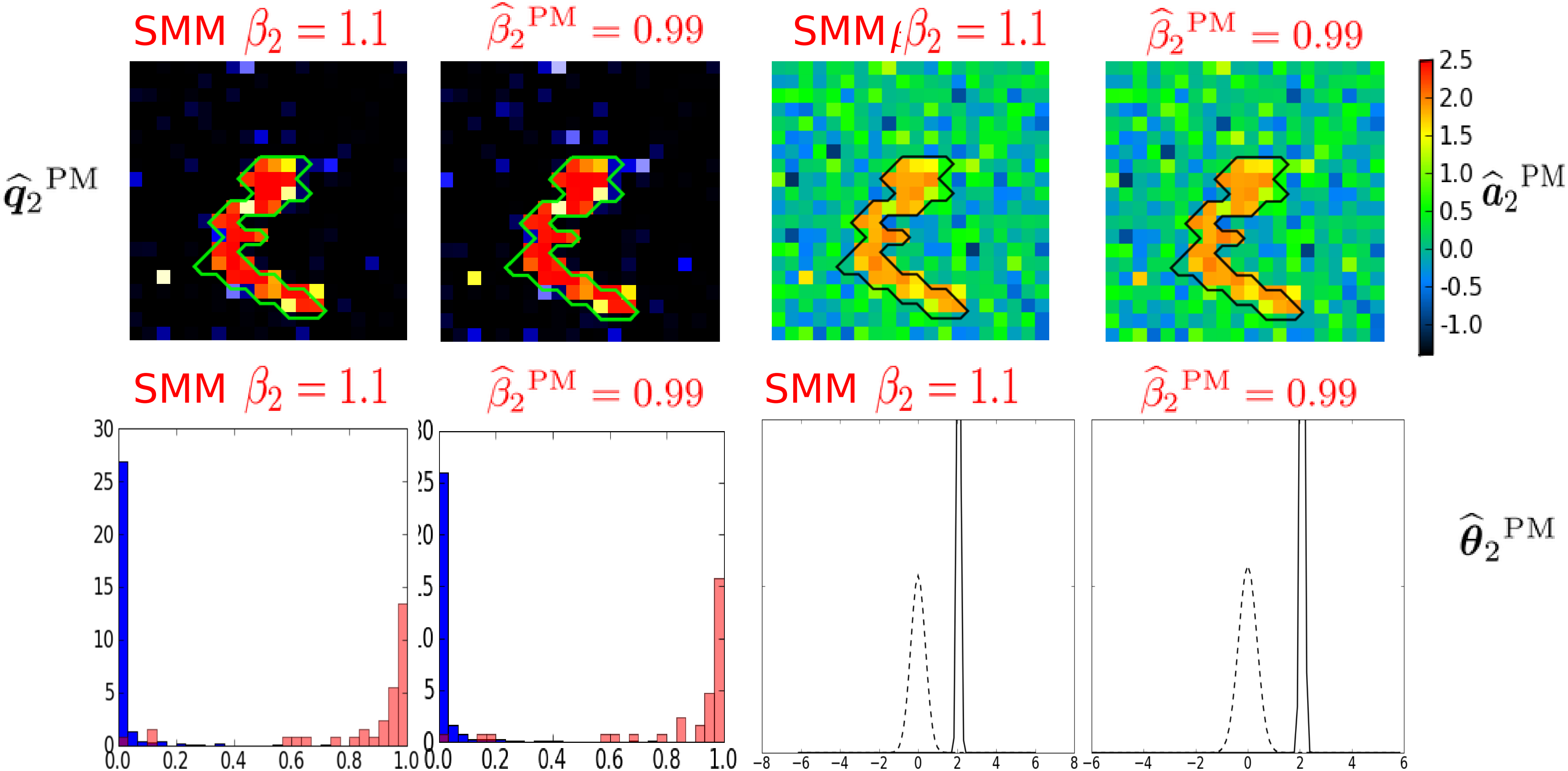


USMM

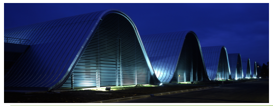
Interest of unsupervised spatial mixture models



Supervised vs. unsupervised

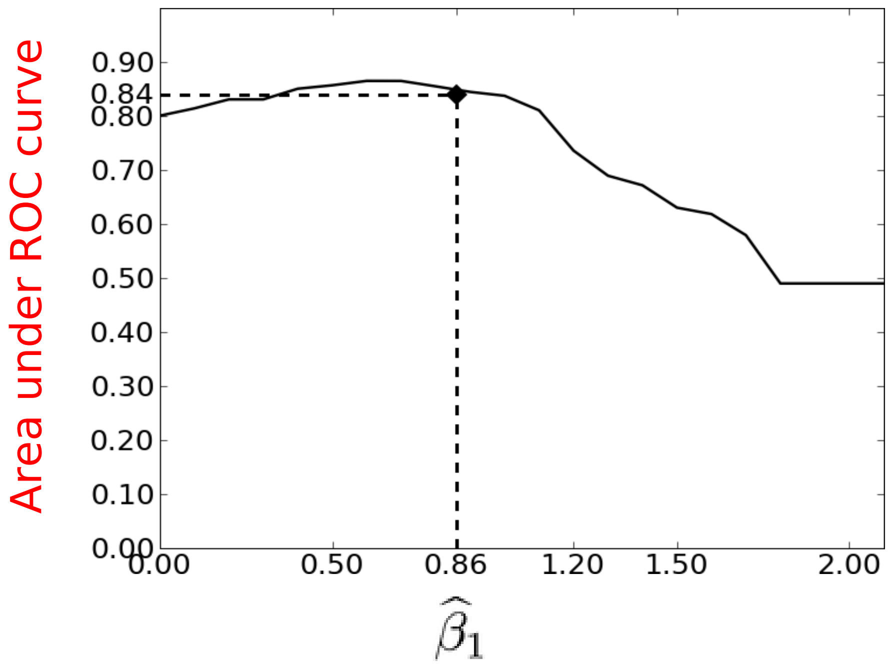


Best tuning: larger β_2 → lower SNR for $m = 2$

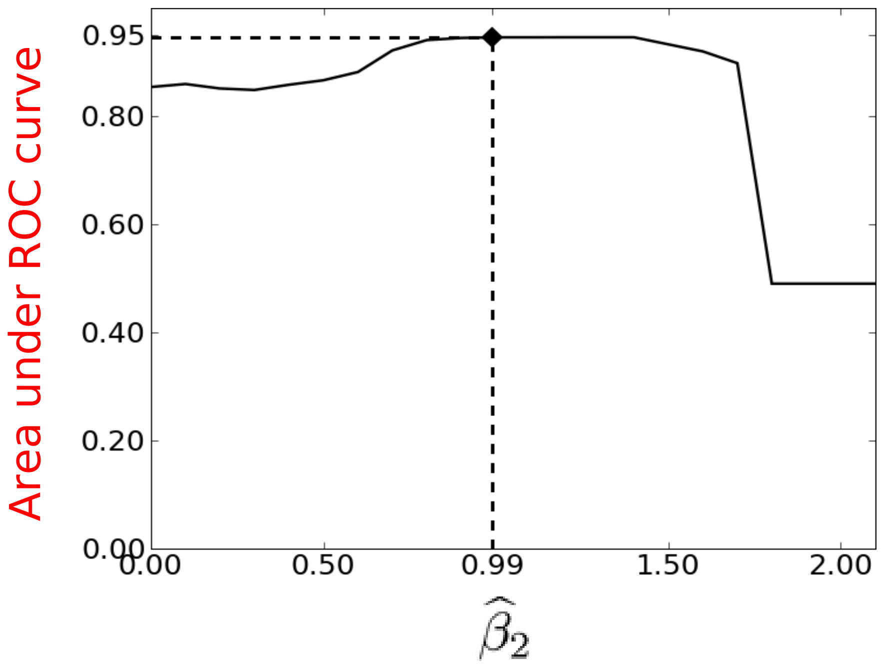


Area under R0C curves

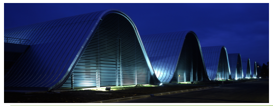
$m = 1$



$m = 2$



nearly optimal unsupervised settings



Remarks

- Starting point for path sampling:

$$Z(0) = 2^J \text{ for Ising fields}$$

$$Z(0) = C^J \text{ for Potts fields with } C \text{ states}$$

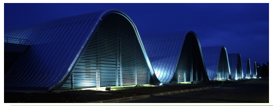
- PF dependence on topological configuration

$$Z(\beta) = f(J, C) \text{ where } C = |\mathcal{C}| = \text{number of cliques}$$

  - Path sampling for multiple data defined on the same grid
  - Which solution to adopt in other situations?

- Numerical cost of path sampling:
  - fluctuates with D (number of grid points)
  - fluctuates with MRFs $(\boldsymbol{x}^k)_{k=1:K}$ and β value
  - depends on MRF sampling algorithm (Gibbs, SW, PD)

[Higdon, JASA 1998]



Adaptive spatial regularization

- Remarks:

$$Z(0) = 2^J \text{ for Ising fields}$$

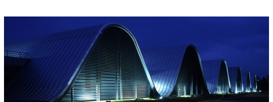
$$Z(0) = C^J \text{ for Potts fields with } C \text{ states}$$

$$Z(\beta^m) = Z(\beta^n) \quad \forall n \neq m$$

$$Z(\beta) = f(J, C) \text{ where } C = |\mathcal{C}| = \text{number of cliques}$$

⇒ Extrapolation methods required for ROI of variable size and shape

Parcel-dependent regularization factor β



Linear extrapolation of $Z(\beta)$

- Linear interpolation technique: **[Trillon and Idier, Eusipco 2008]**
  - Reference grids:

$$(\mathcal{G}_p)_{p=1:P} \Longrightarrow (\log \widehat{Z}_{\mathcal{G}_p}(\beta_k))_{p=1:P}, \forall \beta_k = k\Delta\beta$$

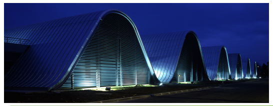
  - Linear regression:

$$\forall \beta_k, (A_{\beta_k}, B_{\beta_k}) = \underset{(A,B)\in\mathbb{R}^2}{\arg\min} \sum_{p=1}^{P} \| \log \widehat{Z}_{\mathcal{G}_p}(\beta_k) - Ac_p - B\|^2$$

  - Application of linear interpolation to test grid:

$$\forall \beta_k, \quad \log \widetilde{Z}_{\mathcal{T}}(\beta_k) = A_{\beta_k} c_{\mathcal{T}} + B_{\beta_k}$$

→ Require homogeneity of the reference grids and regular grids !!



Bilinear extrapolation of $Z(\beta)$

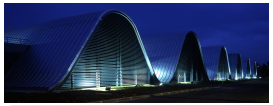
- Bilinear extension: **[Risser, Idier, Ciuciu, ICIP 2009]**
  - Bilinear regression:

$$\forall \beta_k, (A_{\beta_k}, B_{\beta_k}, D_{\beta_k}) = \underset{(A,B,D)\in\mathbb{R}^3}{\arg\min} \sum_{p=1}^{P} \| \log \widehat{Z}_{\mathcal{G}_p}(\beta_k) - A c_p - B s_p - D \|^2$$

  - Application of bilinear interpolation to test grid:

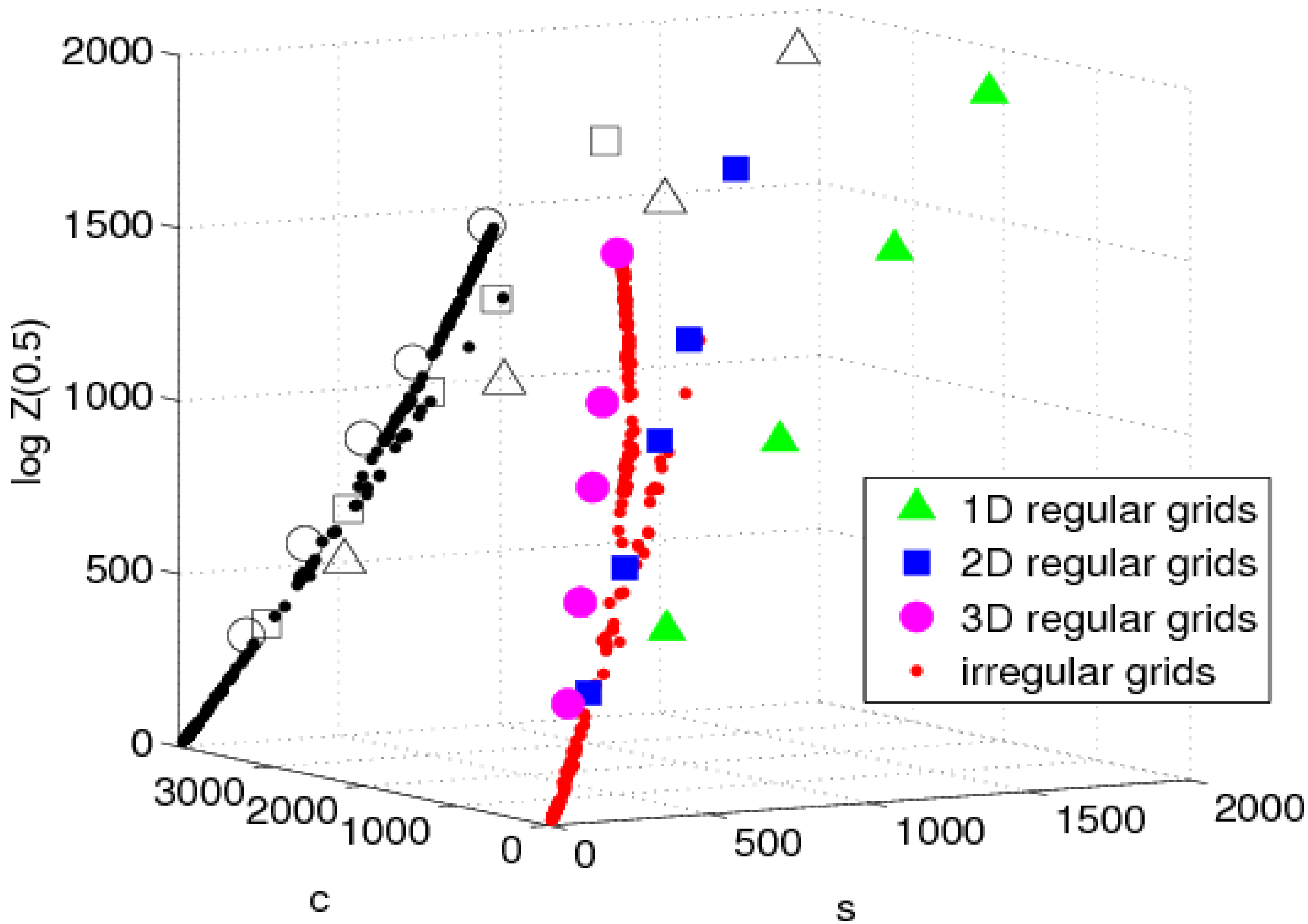
$$\forall \beta_k, \quad \log \widetilde{Z}_{\mathcal{T}}(\beta_k) = A_{\beta_k} c_{\mathcal{T}} + B_{\beta_k} s_{\mathcal{T}} + D_{\beta_k}$$

→ Still homogeneous reference set but applicable to
non regular grids
Multiple PFs involved in the extrapolation process

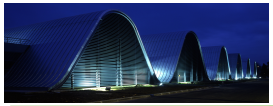


Bilinear dependence of $Z(\beta)$

- Illustration



At a fixed number of C, the larger S the larger $\log Z(\beta)$

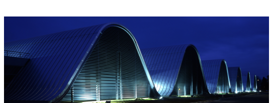


Comparison of linear/bilinear

Mean approximation error over Regular & Irregular test fields.
Errors given in percentage.

Test grid		Scheme / Reference grid B=bilinear, L=linear / R=regular, I=irregular			
		B / R	B / I	L / R	L / I
regular	small	**0.747**	3.84	5.55	93.0
regular	medium	1.30	**0.991**	7.27	6.37
regular	large	1.59	**1.31**	9.18	7.18
irregular	$\beta = 0.2$	6.85	**1.29**	23.6	83.9
irregular	$\beta = 0.4$	0.984	**0.264**	7.71	8.28
irregular	$\beta = 0.5$	1.73	**1.27**	1.64	1.52

Improved performance with the bilinear approach for small and irregular fields
Approximation accuracy depends on β



Min/max extrapolation of $Z(\beta)$

- Fast extrapolation technique: **[Risser, Ciuciu et al, MLSP 2009]** **[Risser, Ciuciu et al, MICCAI 2009]**
  - Reference grids:

$$(\mathcal{G}_p)_{p=1:P} \Longrightarrow (\log \widehat{Z}_{\mathcal{G}_p}(\beta_k))_{p=1:P}, \forall \beta_k = k\Delta\beta$$

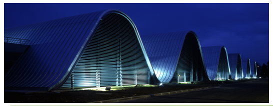
  - Grid selection: Min/max criterion

$$\mathcal{G}_{\text{ref}} = \underset{(\mathcal{G}_p)_{p=1:P}}{\arg\min}\, \mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) \text{ with } \mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p) = \| \log \widehat{Z}_{\mathcal{T}}(\beta) - \log \widetilde{Z}_{\mathcal{T}}(\beta) \|^2$$

  - Extrapolation: $\log \widetilde{Z}_{\mathcal{T}}(\beta) = \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_{\text{ref}}}} \Big(\log \widehat{Z}_{\mathcal{G}_{\text{ref}}}(\beta) - \log L \Big) + \log L$
  - Maximal error: $\mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) = \| \log L \big[(s_{\mathcal{T}} - 1) - \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}} (s_{\mathcal{G}_p} - 1) \big] \|^2$

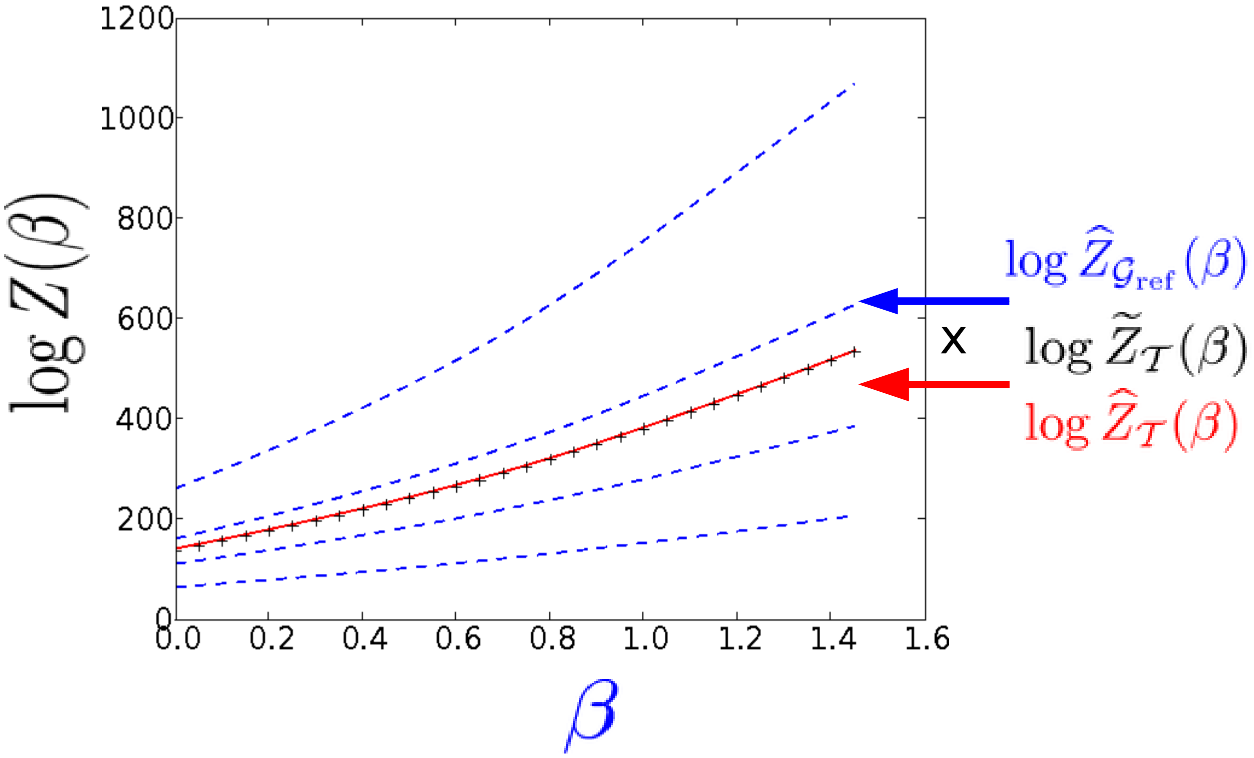
Single PF estimate involved in the extrapolation

The more different the reference grids the smaller the approximation error



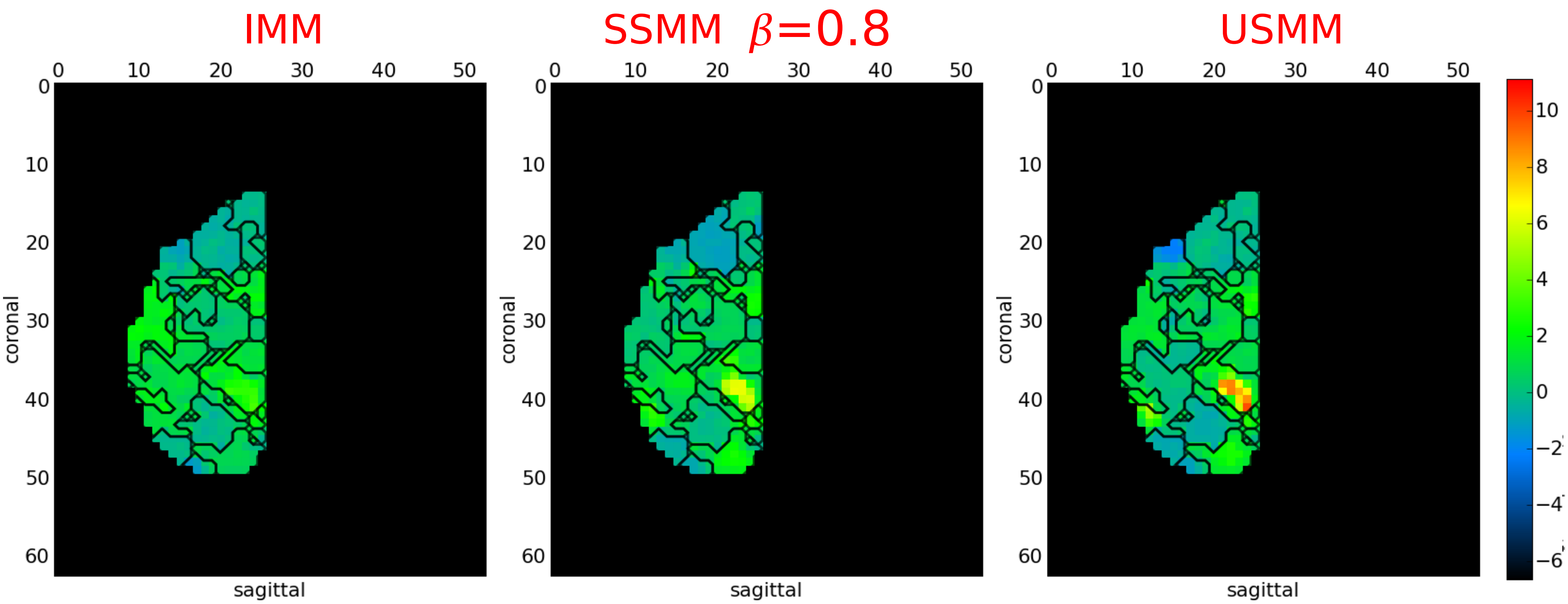
Min/max extrapolation of $Z(\beta)$

- Fast extrapolation technique:
  - Grid selection:
    $$\mathcal{G}_{\text{ref}} = \underset{(\mathcal{G}_p)_{p=1:P}}{\arg\min}\, \mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) \text{with } \mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p) = \| \log Z_{\mathcal{T}}(\beta) - \log \widetilde{Z}_{\mathcal{T}}(\beta) \|^2$$
  - Maximum error: $\mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) = \| \log L\big[(s_{\mathcal{T}} - 1) - \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}(s_{\mathcal{G}_p} - 1)\big] \|^2$



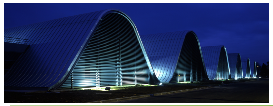
Half whole brain analysis

Normalized contrast: Audit (Calculation – Sentence)

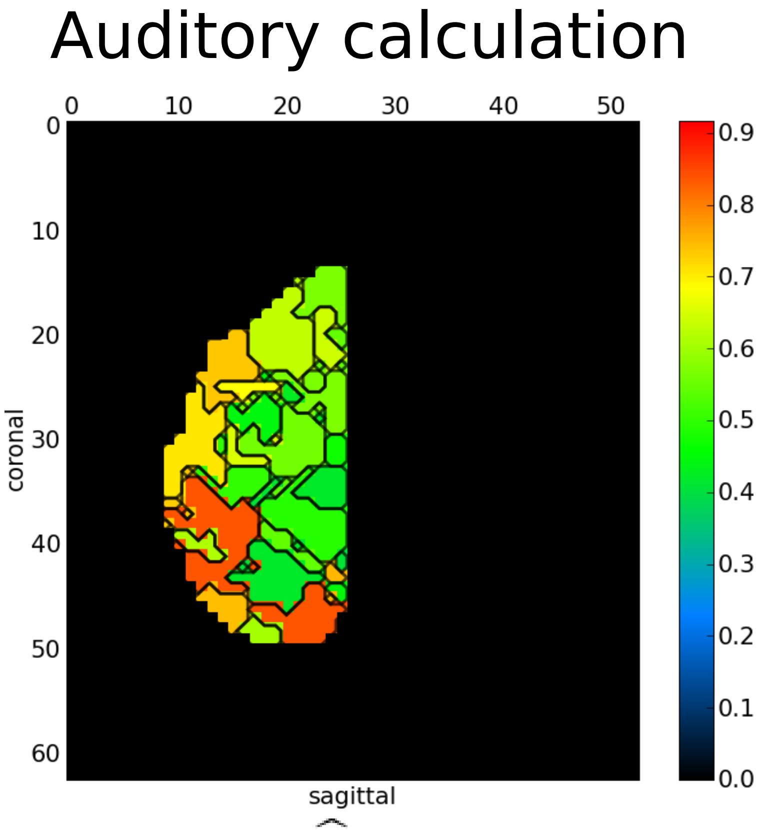


Activations enhanced in the parietal cortex using U/SSMM
Coherence with sulcal anatomy & literature

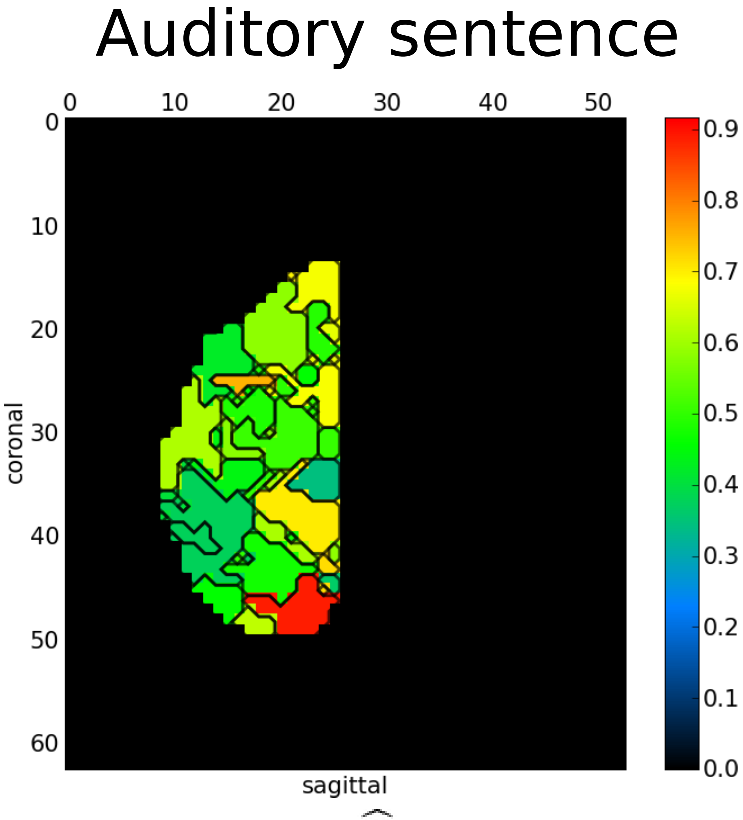
[Vincent, Ciuciu et al, in rev. IEEE TMI 2009]



Adaptive spatial regularization



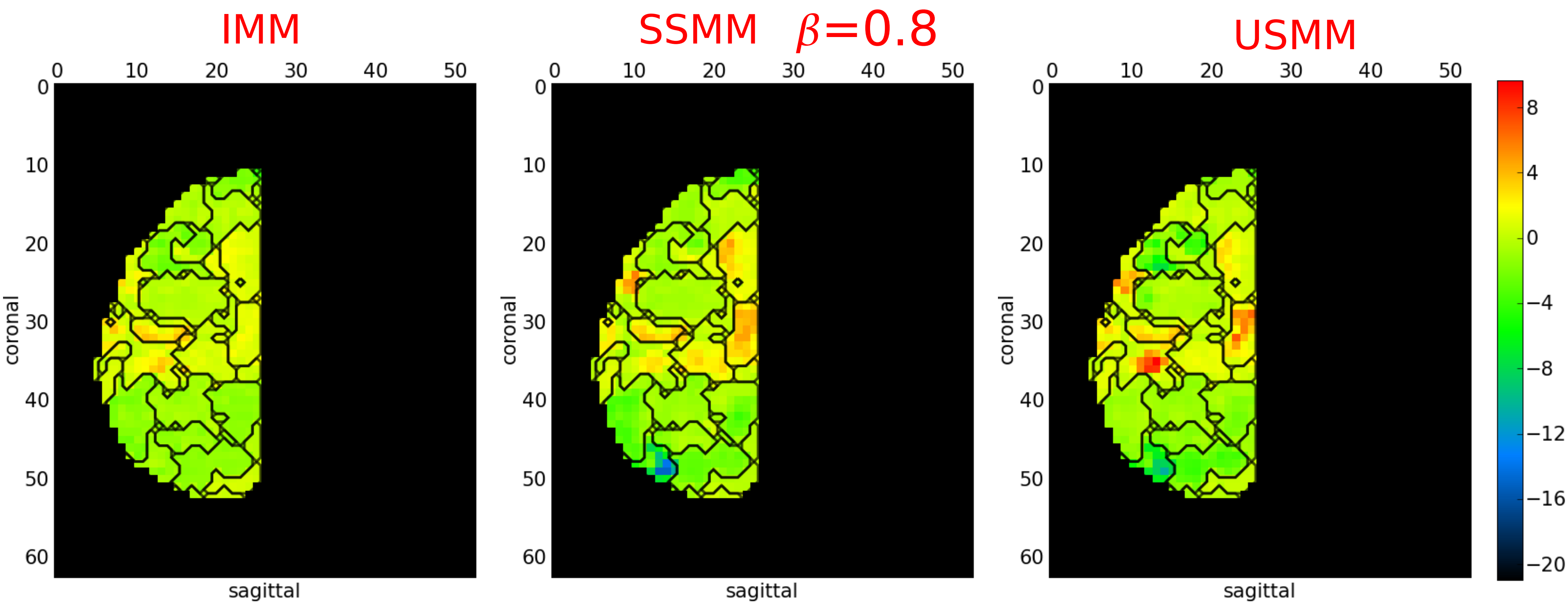
$\widehat{\beta_1}$



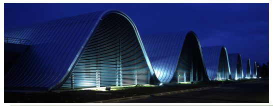
$\widehat{\beta_2}$

Half whole brain analysis

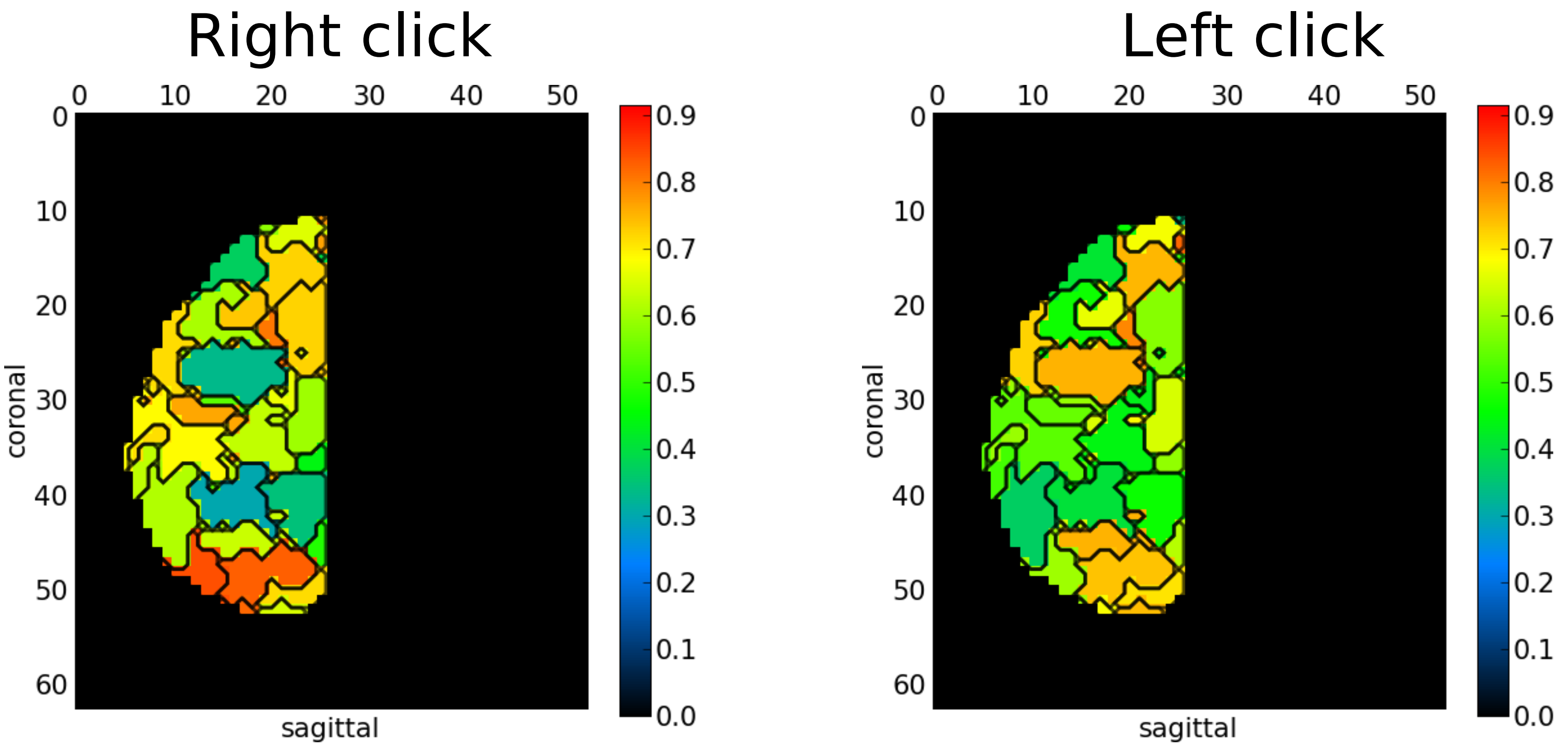
Normalized contrast: Right – left “auditory clicks”



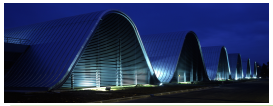
Only USMM provides more sensitive activation in the left motor cortex



Adaptive spatial regularization



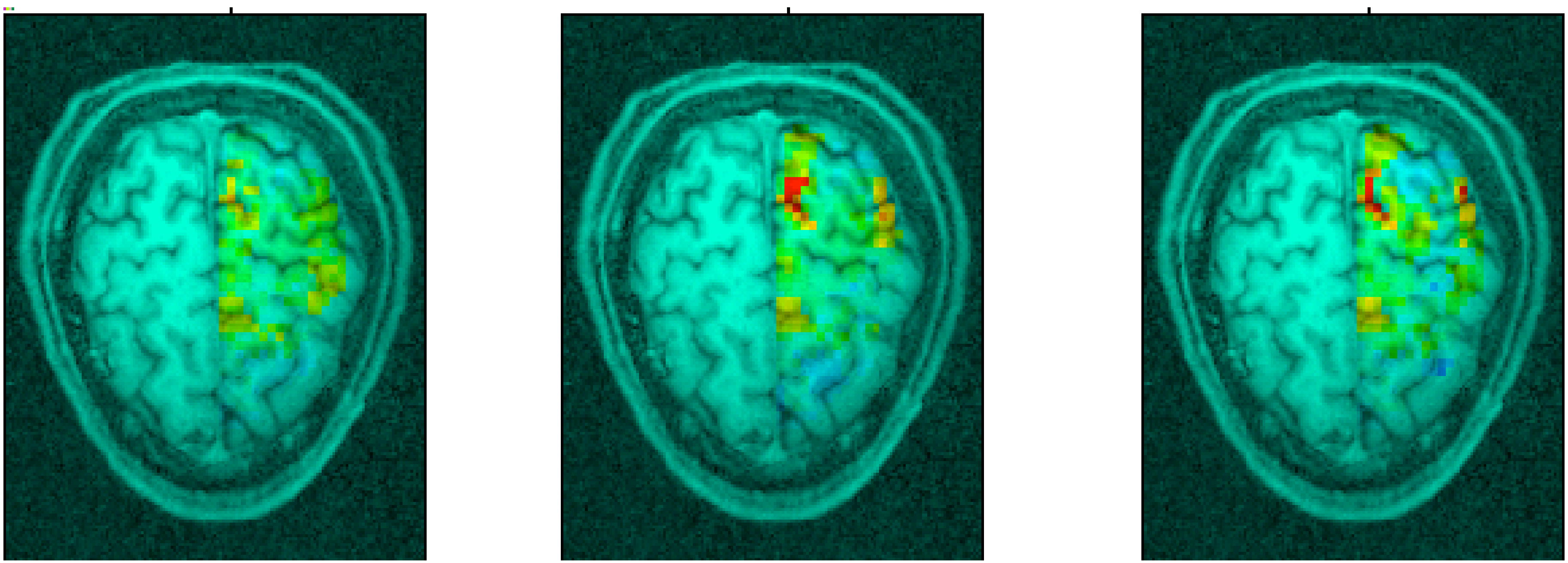
$\widehat{\beta}_1$ $\widehat{\beta}_2$



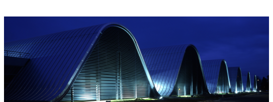
Whole brain analysis

Normalized contrast: Audit. (Calculation – Sentence)

IMM SSMM β=0.8 USMM



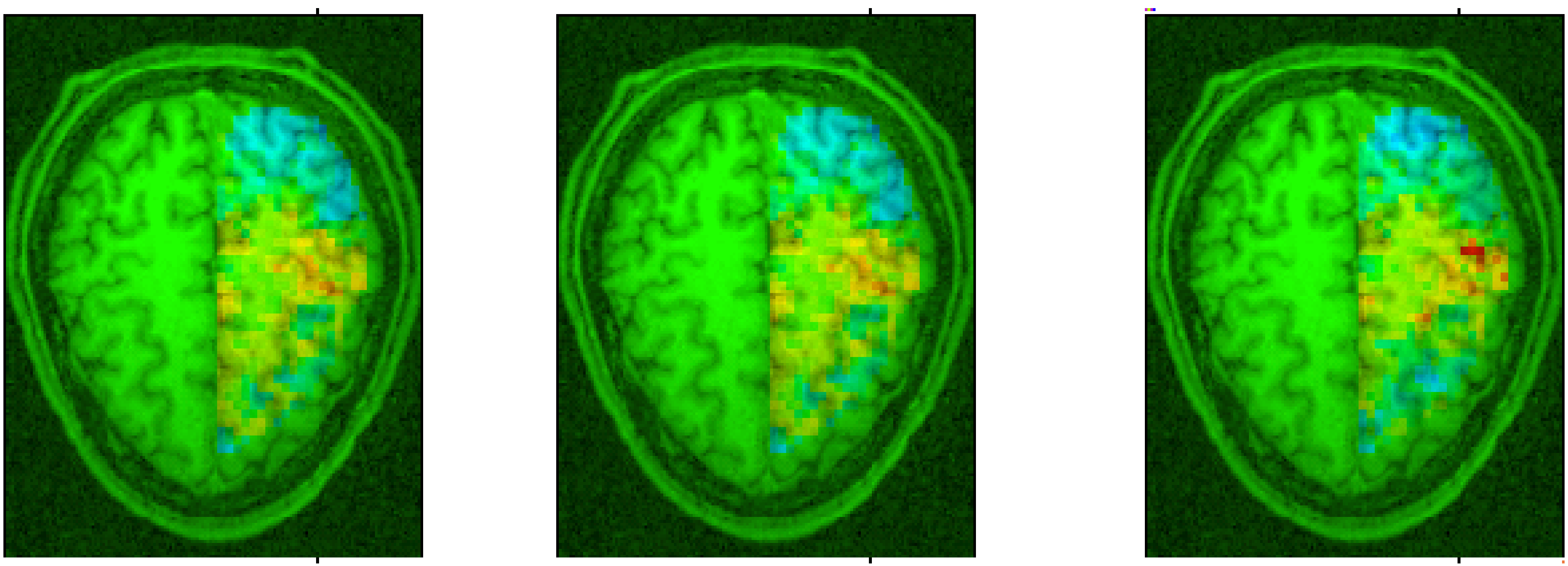
Activations enhanced in the parietal cortex using U/SSMM
Coherent with sulcal anatomy



Half whole brain analysis

Normalized contrast: Left – right "auditory clicks"

IMM SSMM β=0.8 USMM



Only USMM provides more sensitive activation in the motor cortex

Scale ambiguity

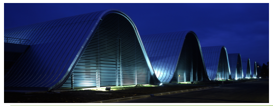
- Identifiability problem

[Veit and Idier, TS 2009]
[Ciuciu et al, GRETSI 2007]

$$p(\mathrm{y} \,|\, \boldsymbol{h}, \mathrm{a}, \mathbb{1}, \boldsymbol{\theta}_0) = p(\mathrm{y} \,|\, \boldsymbol{h}/s, s\mathrm{a}, \mathbb{1}, \boldsymbol{\theta}_0), \forall s \neq 0$$

- A common issue to all bilinear inverse problems:
  - Blind source separation
  - Blind deconvolution
  - Joint detection-estimation

- Bayesian inference: proper priors help in solving this ambiguity



Gibbs sampling

$$1.\ \mathbb{a}^{(k+1)} \leftarrow \mathbb{A} \sim f_{\mathbb{A}|\boldsymbol{H},\mathbb{Y},\boldsymbol{\Theta}}(\mathbb{a}\,|\,\boldsymbol{h}^{(k)},\mathbb{y},\boldsymbol{\theta}^{(k)})$$

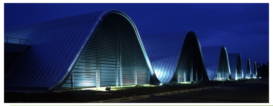
$$2.\ \boldsymbol{h}^{(k+1)} \leftarrow \boldsymbol{H} \sim f_{\boldsymbol{H}|\mathbb{A},\mathbb{Y},\boldsymbol{\Theta}}(\boldsymbol{h}\,|\,\mathbb{a}^{(k+1)},\mathbb{y},\boldsymbol{\theta}^{(k)})$$

$$3.\ \boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\Theta} \sim f_{\boldsymbol{\Theta}|\mathbb{A},\boldsymbol{H},\mathbb{Y}}(\boldsymbol{\theta}\,|\,\mathbb{a}^{(k+1)},\boldsymbol{h}^{(k+1)},\mathbb{y})$$

- Slow evolution of the scale

  → Slow convergence of Gibbs sampling
  Wrong mixing properties of the MC

- Solutions available in the literature:
  - Do nothing: assume an implicit shape/scale decorrelation
  - Normalize at a fixed arbitrary scale at each iteration
    (deterministic transformation incompatible with target density)



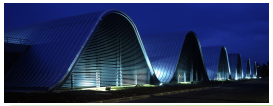
Scale sampling

- Introduce an additional step in Gibbs sampling

$$\mathbb{A} = \mathbb{A}_{\mathrm{old}} \times S, \quad \boldsymbol{H} = \boldsymbol{H}_{\mathrm{old}} / S$$

  - Up to now: $S = \|\boldsymbol{H}\|$
  - Alternative: S is a random variable to be sampled

according to which pdf?

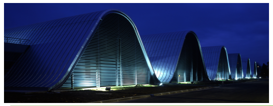


Scale sampling (cont'd)

- General principle: make a change of variable

$$
\begin{aligned}
(s, \boldsymbol{v}) &= \phi(\boldsymbol{a}, \boldsymbol{h}) \\
&= \big(\underbrace{a_1/a_1^{\text{old}}}_{s}, \underbrace{a_2/a_1, \ldots, a_J/a_1}_{v_1, \ldots, v_{J-1}}, \underbrace{a_1 h_1, \ldots, a_1 h_P}_{v_J, \ldots, v_{J+P-1}}\big).
\end{aligned}
$$

$$
f_{S,\boldsymbol{V} \mid \Theta}(s) \propto |s|^{J-P-1} \, f_{\mathbb{A},\boldsymbol{H} \mid \Theta}(\phi^{-1}(s, \boldsymbol{v}))
$$

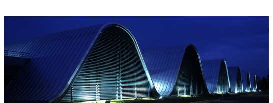


Scale sampling (cont'd)

- General principle: make a change of variable

$$
\begin{aligned}
(s, \boldsymbol{v}) &= \phi(\boldsymbol{a}, \boldsymbol{h}) \\
&= \big(\underbrace{a_1/a_1^{\text{old}}}_{s}, \underbrace{a_2/a_1, \ldots, a_J/a_1}_{v_1, \ldots, v_{J-1}}, \underbrace{a_1 h_1, \ldots, a_1 h_P}_{v_J, \ldots, v_{J+P-1}}\big).
\end{aligned}
$$

$$
f_{S \mid \boldsymbol{V}, \Theta}(s) \propto f_{S, \boldsymbol{V} \mid \Theta}(s) \propto |s|^{J-P-1} \, f_{\mathbb{A}, \boldsymbol{H} \mid \Theta}(\phi^{-1}(s, \boldsymbol{v}))
$$



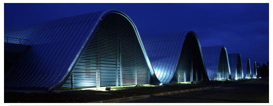
Scale sampling (cont'd)

- General principle: make a change of variable

$$
\begin{aligned}
(s, \boldsymbol{v}) &= \phi(\boldsymbol{a}, \boldsymbol{h}) \\
&= \big(\underbrace{a_1/a_1^{\text{old}}}_{s}, \underbrace{a_2/a_1, \ldots, a_J/a_1}_{v_1,\ldots,v_{J-1}}, \underbrace{a_1 h_1, \ldots, a_1 h_P}_{v_J,\ldots,v_{J+P-1}}\big).
\end{aligned}
$$

- S is independent of the data $\mathbb{Y}$

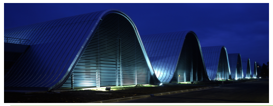
$$
f_{S \mid \boldsymbol{V}, \mathbb{Y}, \Theta}(s) = f_{S \mid \boldsymbol{V}, \Theta}(s) \propto f_{S, \boldsymbol{V} \mid \Theta}(s) \propto |s|^{J-P-1} f_{\mathbb{A}, \boldsymbol{H} \mid \Theta}(\phi^{-1}(s, \boldsymbol{v}))
$$



Modified Gibbs sampling

Given $\mathbb{A}^{(k)}, \boldsymbol{h}^{(k)}, \boldsymbol{\theta}^{(k)}$

1. $\mathbb{a}_{\text{old}}^{(k+1)} \leftarrow \mathbb{A} \sim f_{\mathbb{A}|\boldsymbol{H},\mathbb{Y},\boldsymbol{\Theta}}(\mathbb{a} \,|\, \boldsymbol{h}^{(k)}, \mathbb{y}, \boldsymbol{\theta}^{(k)})$
2. $\boldsymbol{h}_{\text{old}}^{(k+1)} \leftarrow \boldsymbol{H} \sim f_{\boldsymbol{H}|\mathbb{A},\mathbb{Y},\boldsymbol{\Theta}}(\boldsymbol{h} \,|\, \mathbb{a}_{\text{old}}^{(k+1)}, \mathbb{y}, \boldsymbol{\theta}^{(k)})$
3. $s \leftarrow S \sim f_{S\,|\,rest}(s) \propto |s|^{MJ-P-1} f_{\mathbb{A},\boldsymbol{H}|\Theta}(s\mathbb{a}_{\text{old}}^{(k+1)}, \boldsymbol{h}_{\text{old}}^{(k+1)}/s \,|\, \boldsymbol{\theta})$
   $\mathbb{a}^{(k+1)} \leftarrow s\, \mathbb{a}_{\text{old}}^{(k+1)}$ and $\boldsymbol{h}^{(k+1)} \leftarrow \boldsymbol{h}_{\text{old}}^{(k+1)}/s$
4. $\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\Theta} \sim f_{\boldsymbol{\Theta}|\mathbb{A},\boldsymbol{H},\mathbb{Y}}(\boldsymbol{\theta} \,|\, \mathbb{a}^{(k+1)}, \boldsymbol{h}^{(k+1)}, \mathbb{y})$

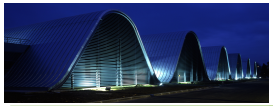


Examples

- Gaussian priors on $(\boldsymbol{H}, \mathbb{A})$ → $f_{S^2 \mid rest} \sim GIG(\lambda, \alpha, \beta)$
- Gamma priors on $(\boldsymbol{H}, \mathbb{A})$ → $f_{S \mid rest} \sim GIG(\lambda, \alpha, \beta)$
- Still valid for Gaussian mixtures on $\mathbb{A}$
- JDE framework:

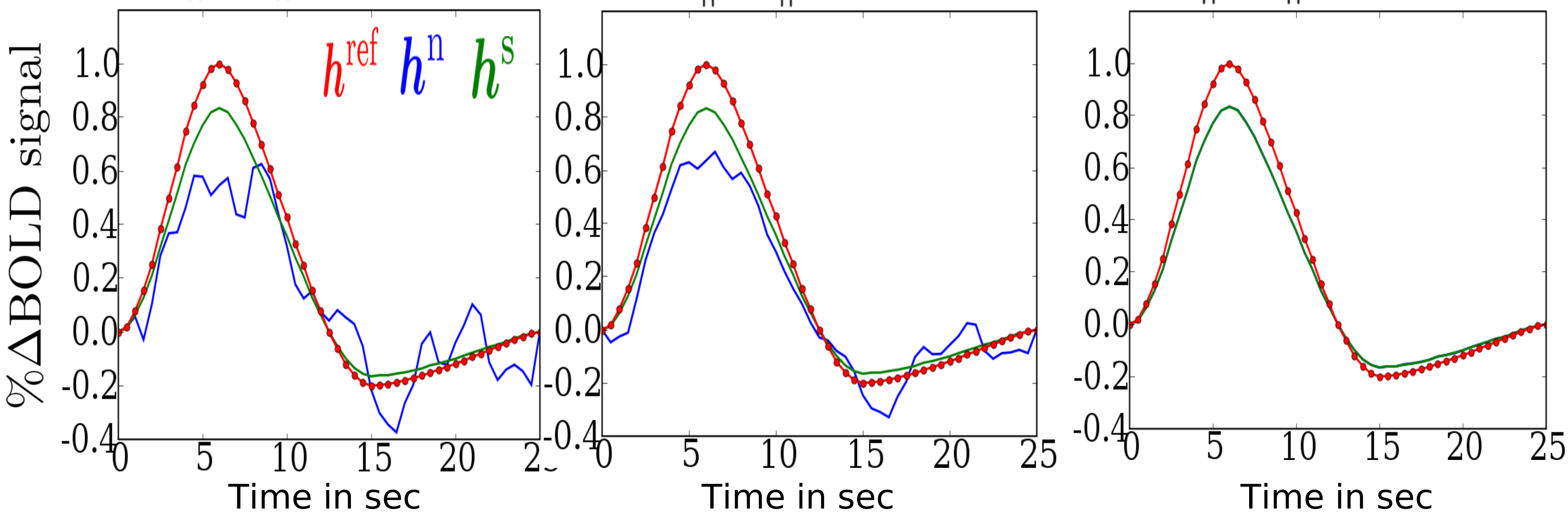
$$\lambda = (P - M(J+1))/2, \quad \alpha = (\boldsymbol{h}_{\text{old}})^{\text{t}} \boldsymbol{R}^{-1} \boldsymbol{h}_{\text{old}} / \sigma_{\boldsymbol{h}}^2,$$

$$\beta = \frac{\|\boldsymbol{\mu}_{\text{old}}\|^2}{2\sigma_\mu^2} + \sum_{j=1}^{J} ((\boldsymbol{a}_j)_{\text{old}})^{\text{t}} \Sigma_j^{-1} (\boldsymbol{a}_j)_{\text{old}} \text{ with } \Sigma_j = \text{diag}_M [v_{q_j^m}^m]$$

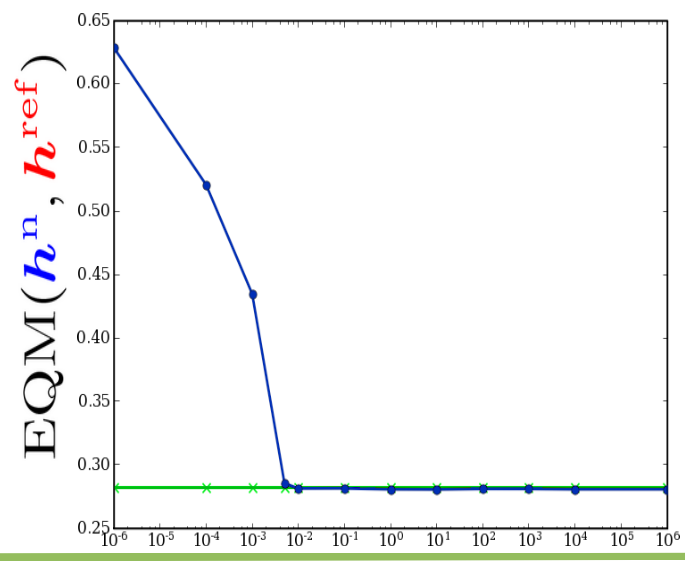


Illustrations

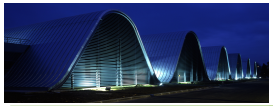
$\|\boldsymbol{h}^{\mathrm{n}}\| = 10^{-4}$ $\|\boldsymbol{h}^{\mathrm{n}}\| = 10^{-3}$ $\|\boldsymbol{h}^{\mathrm{n}}\| = 10^{-2}$



$$\mathrm{EQM}(\boldsymbol{h}^{\mathrm{n}}, \boldsymbol{h}^{\mathrm{ref}}) = \|\boldsymbol{h}^{\mathrm{n}} - \boldsymbol{h}^{\mathrm{ref}}\|^2$$

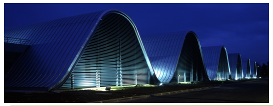


$\|\boldsymbol{h}^{\mathrm{n}}\|$



Conclusions

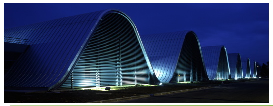
- The joint detection-estimation framework:
  - directly accounts for different sources of variability
  - provides both region-based HRF time courses and contrast maps
  - embeds unsupervised spatial regularization
  - avoids using spatial filtering of fMRI datasets
  - depends on an input parcellation: **[Vincent et al, ISBI'08]**
  - gives improved RFX maps in comparison with SPM
- First release of Pyhrf package (v 1.0)
  - downloadable at http://launchpad.net (nipy project)



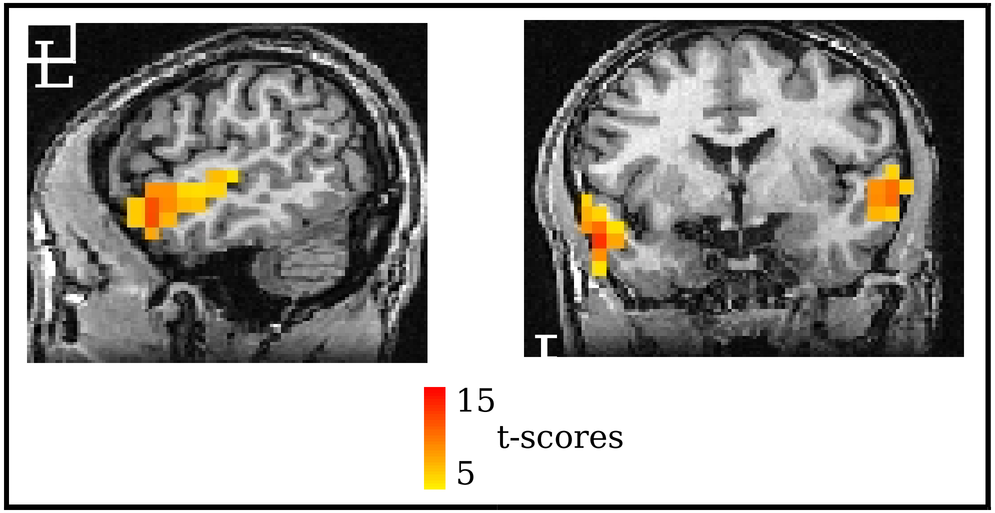
Ongoing works

- Neuro-dynamics models (habituation effect)
- Validation at the group level
- Model comparison and selection
- Extension to the cortical surface for EEG/fMRI fusion
- Application to neonate fMRI datasets

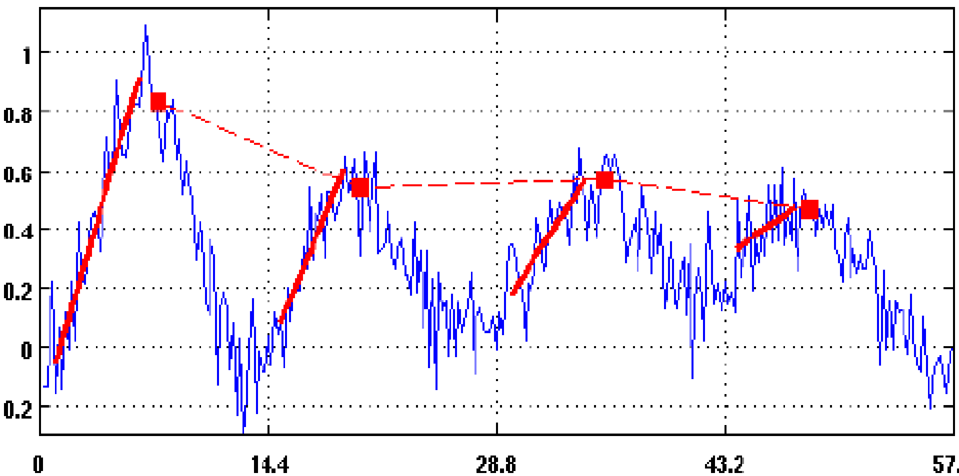
→ collab: G. Dehaene (INSERM U562)



Forward model with habituation

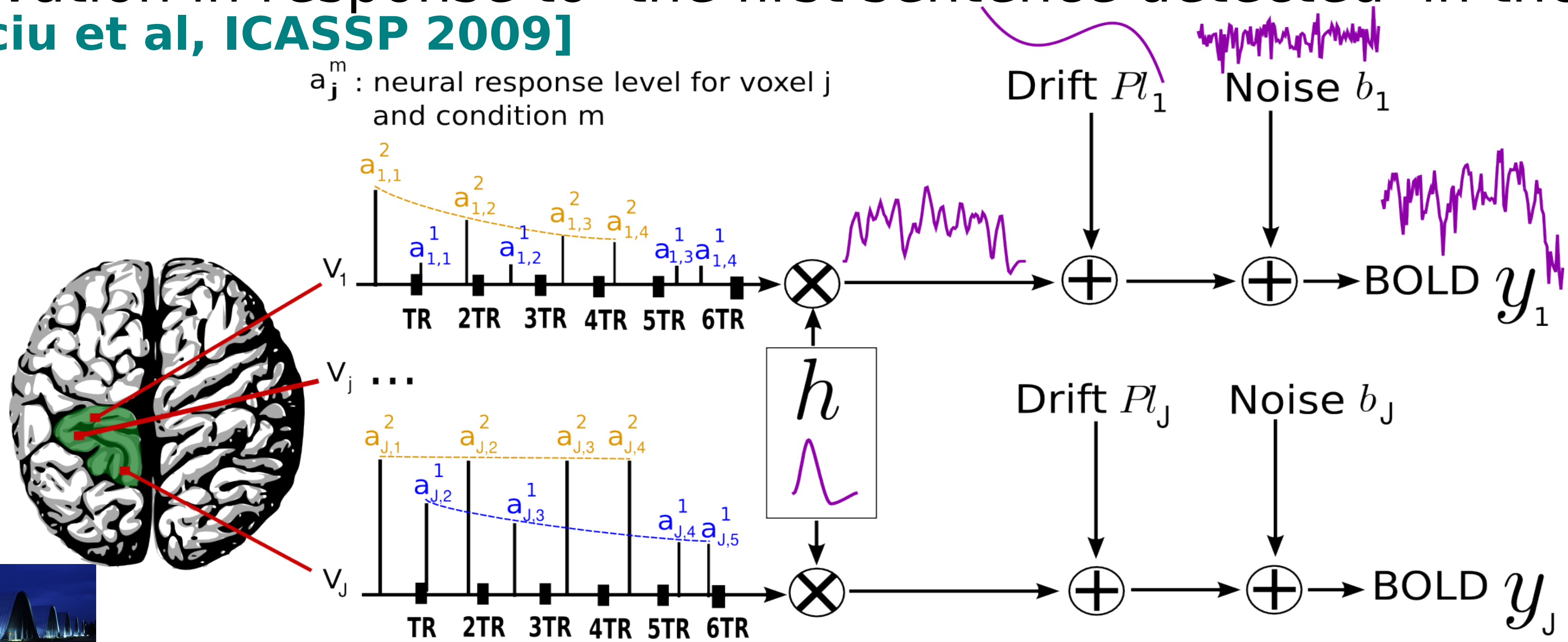


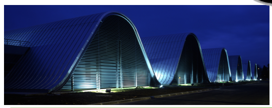
[Rabrait, Ciuciu et al, JMRI 2008]



Activation in response to the first sentence detected in the STS/G

[Ciuciu et al, ICASSP 2009]





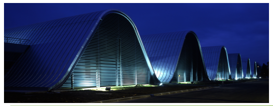
Acknowlegdments

- This work was partly supported by grants from *Région Ile-de-France*

- Special thanks to:

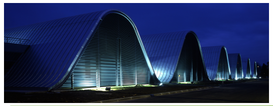
Thomas Vincent
Salima Makni
Anne-Laure Fouque
Laurent Risser
Marion Soumoy
Stéphane Sockeel

Jérôme Idier
Sophie Donnet
Thomas Veit



References

Aguirre et al (1998). Neuroimage 8:360–369.
Birn et al (2000). Neuroimage 14:817–826.
Birn et al (2005). Neuroimage 27:70–82
Boynton et al (1996). J Neurosci 16(13):4207–4221.
Buckner et al (1996). Proc Natl Acad Sci U.S.A. 93:14878–14883.
Buckner (1998). Hum Brain Mapp 6:373–377.
Burock & Dale (2000). Hum Brain Mapp 11:249–260.
Buxton et al (2004). Neuroimage 23:S220–S233.
Chen et al (2004). Neurcomputing 61:395–400.
Ciuciu et al (2003). IEEE Trans Medical Imag 22:1235–1251.
Donnet et al (2006). NeuroImage. 31:1169-1176.
Deneux & Faugeras (2006). NeuroImage. 32(4): 1669-1689.
D'Esposito et al (1999). Neuroimage 10:6–14.
D'Esposito et al (2003). Nat Rev Neurosci 4:863–872.
Ford et al (2005). Neuroimage 26:922–931
Friston et al (1994). Hum Brain Mapp 1:153–171.
Friston et al (1998). Neuroimage 7:30–40
Friston et al (2000). Neuroimage 12:466–477.
Genovese (2000). J Amer Statist Assoc 1995:691–719.
Gibbons et al (2004). Neuroimage 22:804–814.
Glover et al (1999). Neuroimage 9:416-429.
Gössl et al (2001a). Biometrics 57:554–562.
Gössl et al (2001b). Neuroimage 14:140–148.
Goutte et al (2000). IEEE Trans Medical Imag 19:1188–1201.
Handwerker et al (2004). Neuroimage 21:1639–1651.
Hansen et al (2004). Neuroimage 23:233–241.
Henson et al (2002). Neuroimage 15:83–97.



References

Kershaw et al (1999). IEEE Trans Medical Imag 18:1138–1153.
Kruggel & Von Cramon (1999a). Magn Reson Med 42:787–797.
Kruggel & Von Cramon (1999b). Hum Brain Mapp 8:259–271.
Kruggel et al (2000). Neuroimage 12:173–183.
Liao et al (2002). Neuroimage 16:593–606.
Logothetis (2003). J Neurosci 23:3963–3971.
Logothetis et al (2001). Nature 412:150–157.
Makni et al (2005). IEEE Trans Signal Process 53(9):3488–3502.
Marrelec et al (2003) Hum Brain Mapp 19:1–17.
Marrelec et al (2004). IEEE Trans Medical Imag 23:959–967.
McGonigle et al (2000). Neuroimage 11:708–734.
Miezin et al (2000). Neuroimage 11:735–759.
Neumann et al (2003). Neuroimage 19:784–796.
Neumann et al (2006). Neuroimage 32(3): 1185-1194.
Penny et al (2003). NeuroImage. 19(3): 727-744.
Penny et al (2005). NeuroImage. 24(2): 350-362.
Rajapakse et al (1998). Hum Brain Mapp 6:283–300.
Richter & Richter (2003). Neuroimage 20:1122–1131.
Riera et al (2004). NeuroImage 21(2): 547--567.
Saad et al (2001). Hum Brain Mapp 13:74–93.
Smith et al (2005). Hum Brain Mapp 24:248–257.
Vazquez et al (1998). Neuroimage 7:108–118.
Woolrich et al (2004a). Neuroimage 21:1732–1747.
Woolrich et al (2004b). Neuroimage 21:1748–1761.
Woolrich et al (2004c). IEEE Trans Medical Imag 23:213–231.
Woolrich et al (2005). IEEE Trans Medical Imag 24(1): 1-11
Worsley & Friston (1995). Neuroimage 2:173–181.
Worsley et al (2002). NeuroImage. :15(1): 1-15.