Hemodynamic-Informed Parcellation of fMRI Data in a Joint Detection Estimation Framework

Lotfi Chaari, Solveig Badillo, Thomas Vincent, Ghislaine Dehaene-Lambertz, Florence Forbes, Philippe Ciuciu

Hemodynamic-Informed Parcellation of fMRI Data in a Joint Detection Estimation Framework

Lotfi CHAARI, Solveig BADILLO, Thomas VINCENT, Ghislaine DEHAENE-LAMBERTZ, Florence FORBES and Philippe CIUCIU

Abstract

Identifying brain hemodynamics in event-related functional MRI (fMRI) data is a crucial issue to disentangle the vascular response from the neuronal activity in the BOLD signal. This question is usually addressed by estimating the so-called Hemodynamic Response Function (HRF). Voxelwise or region-/parcelwise inference schemes have been proposed to achieve this goal but so far all known contributions commit to pre-specified spatial supports for the hemodynamic territories by defining these supports either as individual voxels or a priori fixed brain parcels. In this paper, we introduce a Joint Parcellation-Detection-Estimation (JPDE) procedure that incorporates an adaptive parcel identification step based upon local hemodynamic properties. Efficient inference of both evoked activity, HRF shapes and *supports* is then achieved using variational approximations. Validation on synthetic and real fMRI data demonstrates the JPDE performance over standard detection estimation schemes and suggests it as a new brain exploration tool.

I. Introduction

Within-subject analysis in event-related BOLD fMRI mainly relies on (i) *the detection* of evoked activity to localize which parts of the brain are activated by a given stimulus type, and on (ii) *the estimation* of the dynamics of the brain response also known as the Hemodynamic Response Function (HRF). Most approaches to detect neural activity rely on a single *a priori* HRF model for the whole brain although there has been evidence that this response can vary between cortical regions and across subjects [1] and

Lotfi CHAARI is with the University of Toulouse, IRIT - INP-ENSEEIHT (UMR 5505), 2 rue Charles Camichel, BP 7122, Toulouse Cedex 7 France. E-mail: lotfi.chaari@enseeiht.fr. Thomas VINCENT and Florence FORBES are with the Mistis team at Inria Grenoble, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France. E-mail: {thomas.vincent,florence.forbes}@inria.fr. Philippe CIUCIU and Ghislaine DEHAENE-LAMBERTZ is with CEA/DSV/I^{2}BM/Neurospin and Parietal team, CEA Saclay, Bat. 145, Point Courrier 156, 91191 Gif-sur-Yvette cedex, France. E-mail: philippe.ciuciu@cea.fr.

that an accurate HRF model may significantly improve detection performance. To capture this variability, robust HRF estimation is necessary which can be achieved only in voxels or regions that elicit an evoked response to a given stimulus [2]. So far, many works have addressed this issue either by considering linear or nonlinear HRF models [3–5], parametric, semi-parametric or non-parametric (*i.e.* FIR models) descriptions [6–8], and by performing univariate (voxelwise) [4, 7], multivariate (regionwise) [9, 10] or even multiscale, *i.e.* spatially adaptive inference [11]. However, to the best of our knowledge, all these existing works assume the spatial support of the HRFs, either defined at the voxel or region-level, to be pre-specified. The proposed methodology takes place in the Joint Detection-Estimation (JDE) framework introduced in [9] and extended in [10, 12, 13] to account for spatial correlation between voxels. Standard JDE-based inference requires a pre-specified decomposition of the brain into functionally homogeneous parcels (groups of connected voxels) but with no guarantee of their optimality. These parcels should be small enough to guarantee the invariance of the HRF within each parcel, but large enough to contain reliable information for its inference [14]. Several attempts have been conducted to provide a robust parcellation such as in [15–19]. However, these approaches do not fully account for hemodynamics variability. Here, we introduce the concept of *hemodynamic territory* as a set of parcels which share a common HRF pattern. To determine such sets, we incorporate an additional layer in the JDE hierarchy, namely an adaptive parcel identification step based upon local hemodynamic properties. In this novel Joint Parcellation-Detection-Estimation (JPDE) model (Section II), for all the parcels of a given territory, HRFs are voxelwise but defined as local stochastic perturbations of the same HRF pattern. Then, hemodynamics estimation reduces to the identification of a limited number (say K) of such HRF patterns and parcel identification reformulates as a clustering problem where each voxel is assigned an HRF group among K. The HRF group assignment variables are governed by a hidden Markov model to enforce spatial correlation, *i.e.* favor group assignments that are spatially homogeneous. Finally, the overall scheme iteratively identifies hemodynamic territories as pairs of one HRF pattern and a set of parcels assigned to the corresponding HRF group.

The proposed approach thus makes the JDE framework fully adaptive and more flexible. It is based on a variational Expectation Maximization (EM) algorithm (Section III) to derive estimates of the HRF patterns, the response amplitude, the corresponding labels (activating/non-activating voxels) and the HRF group labels. Results on artificial and real fMRI data demonstrate that the JPDE approach outperforms the standard JDE (see Section IV).

II. A JOINT PARCELLATION-DETECTION-ESTIMATION MODEL

A. Observed and missing variables

We extend the parcel-based JDE model of [9, 10] to a whole-brain one, with a set of J voxels denoted by $\boldsymbol{\mathcal{V}}$, and recast it in a missing data framework. The resulting model is referred to as the Joint Parcellation-Detection-Estimation model (JPDE). At voxel j, the fMRI time series $\boldsymbol{y}_j$ is measured at times $\{t_n, n = 1{:}N\}$, where $t_n = nTR$, N being the number of scans and TR the time of repetition. The whole set of time series is denoted by $\boldsymbol{Y} = \{\boldsymbol{y}_j, j \in \boldsymbol{\mathcal{V}}\}$. The number of different stimulus types or experimental conditions is M. At each voxel j, we assume a voxel dependent HRF $\boldsymbol{h}_j \in \mathbb{R}^{D+1}$ with $\boldsymbol{H} = \{\boldsymbol{h}_j, j \in \boldsymbol{\mathcal{V}}\}$ the set of all HRFs. Each $\boldsymbol{h}_j$ is associated with a HRF group among K. These groups or HRF classes are specified by a set of hidden labels $\boldsymbol{Z} = \{z_j, j \in \boldsymbol{\mathcal{V}}\}$ where $z_j \in \{1 : K\}$ and $z_j = k$ means that voxel j belongs to the k-th group. An estimation of $\boldsymbol{Z}$ corresponds then to a partition of the brain into K hemodynamic territories whose connected components define a parcellation. The link to the observed BOLD data is specified via the following forward model:

$$\boldsymbol{y}_j = \boldsymbol{S}_j\boldsymbol{h}_j + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j, \quad \text{with} \quad \boldsymbol{S}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}_m \tag{1}$$

where $\boldsymbol{S}_j\boldsymbol{h}_j$ is the summation of the stimulus-induced components of the BOLD signal. The binary matrix $\boldsymbol{X}_m = \{x_m^{t_n - d\Delta t}, n = 1 \dots N, d = 0 \dots D\}$ is of size $N \times (D+1)$ and provides information on the stimulus occurrences for the m-th experimental condition, $\Delta t < TR$ being the sampling period of the unknown HRFs. The scalar a_j^m's are weights that model the transition between stimulations and the neuro-vascular response. They are generally referred to as Neural Response Levels (NRL). We denote by $\boldsymbol{A} = \{\boldsymbol{a}^m, m = 1 : M\}$ with $\boldsymbol{a}^m = \left\{a_j^m, j \in \boldsymbol{\mathcal{V}}\right\}$ the response amplitudes, a_j^m being the amplitude at voxel j for condition m. Similarly to the HRF's, each NRL is assumed to be in one of I groups specified by activation class assignment variables $\boldsymbol{Q} = \{\boldsymbol{q}^m, m = 1 : M\}$ where $\boldsymbol{q}^m = \left\{q_j^m, j \in \boldsymbol{\mathcal{V}}\right\}$ and q_j^m represents the *activation class* at voxel j for condition m. The number of classes considered here is $I = 2$ for non-activated ($i = 0$) and activated ($i = 1$) voxels. Finally, the rest of the signal is made of vector $\boldsymbol{P}\boldsymbol{\ell}_j$, which corresponds to low frequency drifts with $\boldsymbol{P}$ a $N \times O$ matrix, $\boldsymbol{\ell}_j \in \mathbb{R}^O$ a vector to be estimated and $\boldsymbol{L} = \{\boldsymbol{\ell}_j, j \in \boldsymbol{\mathcal{V}}\}$. Regarding the observation noise, the $\boldsymbol{b}_j$'s are assumed to be independent with $\boldsymbol{b}_j \sim \mathcal{N}(0, \boldsymbol{\Gamma}_j^{-1})$. The set of all unknown precision matrices is denoted by $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_j, j \in \boldsymbol{\mathcal{V}}\}$.

B. Hierarchical model of the complete data distribution

With standard additional assumptions [9, 10, 12, 13], the joint model distribution writes $p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z}) = p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H})\; p(\boldsymbol{A} \mid \boldsymbol{Q})\; p(\boldsymbol{Q})\; p(\boldsymbol{H} \mid \boldsymbol{Z})\; p(\boldsymbol{Z})$.

In the above expression, parameters have not been specified for simplicity. Each term in the product is then defined as follows.

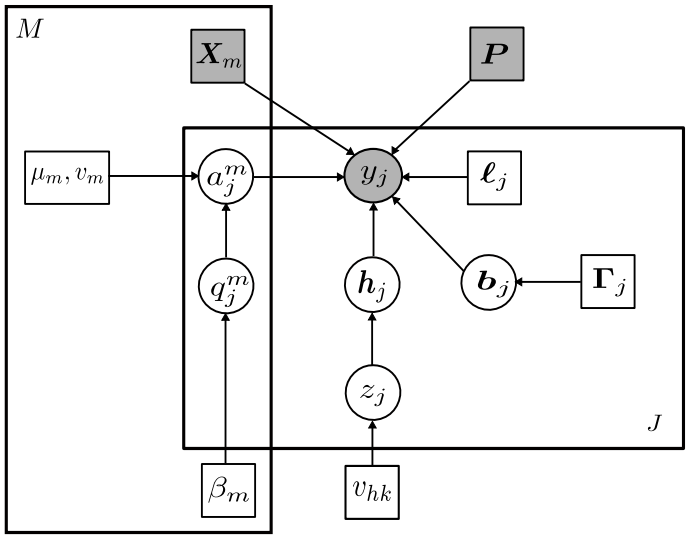

Fig. 1. Graphical model describing dependencies between *latent* and *observed* variables involved in the JPDE generative model with J voxels. Circles and squares indicate random variables and model parameters, respectively. Observed variables and fixed parameters are shaded. We used standard graphical notations where plates represent multiple similar nodes with their number given in the plate.

TABLE I

ACRONYMS USED IN THE JDE MODEL PRESENTATION AND INFERENCE.

Acronym	Definition
JDE	Joint Detection-Estimation
HRF	Hemodynamic Response Function
FIR	Finite Impulse Response
NRL	Neural Response Level
PPM	Posterior Probability Map
TTP	HRF Time-to-Peak: $\Delta t \times \arg\max_d \{h_{d\Delta t}\}_{d=0:D}$

1) Likelihood:

Akin to [9, 12, 13, 20], an autoregressive (AR) noise model has been adopted to account for serial correlations in fMRI time series. We will assume a noise model $\boldsymbol{b}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_j^{-1})$ with $\boldsymbol{\Gamma}_j = \sigma_j^{-2}\boldsymbol{\Lambda}_j$ where $\boldsymbol{\Lambda}_j$ is a tridiagonal symmetric matrix which depends on the AR(1) parameter ρ_j [9]: $(\boldsymbol{\Lambda}_j)_{1,1} = (\boldsymbol{\Lambda}_j)_{N,N} = 1$, $(\boldsymbol{\Lambda}_j)_{n,n} = 1 + \rho_j^2$ for $n \in \{2 : (N-1)\}$ and $(\boldsymbol{\Lambda}_j)_{n+1,n} = (\boldsymbol{\Lambda}_j)_{n,n+1} = -\rho_j$ for $n \in \{1 : (N-1)\}$. These parameters are assumed voxel-specific due to their tissue-dependence [7, 21].

TABLE II

NOTATIONS FOR VARIABLES AND PARAMETERS USED IN THE MODEL FOR A SET $\mathcal{V}$ OF J VOXELS.

	Notation	Definition	Dimension
Variables	$\boldsymbol{y}_j \in \mathbb{R}^N$	Observed BOLD signal at voxel j	N
	$\boldsymbol{Y} = \{\boldsymbol{y}_j, j \in \mathcal{V}\} \in \mathbb{R}^{J \times N}$	Observed BOLD signals	$J \times N$
	$\boldsymbol{b}_j \in \mathbb{R}^N$	Acquisition noise vector at voxel j	N
	$\boldsymbol{h}_j = \{h_{d\Delta t}, d = 0 : D\} \in \mathbb{R}^{D+1}$	HRF sampled at Δ_t	$D+1$
	$\boldsymbol{H} = \{\boldsymbol{h}_j, j \in \mathcal{V}\} \in \mathbb{R}^{J \times (D+1)}$	HRFs	$J \times (D+1)$
	$a_j^m \in \mathbb{R}$	NRL at voxel j for condition m	1
	$\boldsymbol{a}^m = \{a_j^m, j \in \mathcal{V}\} \in \mathbb{R}^J$	NRLs for condition m	J
	$\boldsymbol{A} = \{\boldsymbol{a}^m, m = 1 : M\}$	NRLs	$M \times J$
	$q_j^m \in \{0, 1\}$	Activation class for voxel j and condition m	1
	$\boldsymbol{q}^m = \{q_j^m, j \in \mathcal{V}\} \in \{0, 1\}^J$	Activation classes for condition m	J
	$\boldsymbol{Q} = \{\boldsymbol{q}^m, m = 1 : M\}$	Activation classes	$M \times J$
	$z_j \in \{1 : K\}$	HRF group for voxel j	1
	$\boldsymbol{Z} = \{z_j, j \in \mathcal{V}\}$	HRF groups	J
Unknown Parameters	$\boldsymbol{\ell}_j \in \mathbb{R}^O$	Low frequency drifts for voxel j	O
	$\boldsymbol{L} = \{\boldsymbol{\ell}_j, j \in \mathcal{V}\}$	Low frequency drifts	$J \times O$
	$\boldsymbol{\Gamma}_j \in \mathbb{R}^{N \times N}$	Noise precision matrix for voxel j	$N \times N$
	$\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_j, j \in \mathcal{V}\}$	Noise precisions matrices	$J \times N \times N$
	$\boldsymbol{\mu}_m = \{\mu_{m0}, \mu_{m1}\} \in \mathbb{R}^2$	Mixture model means for NRLs in condition m	2
	$\boldsymbol{\mu} = \{\boldsymbol{\mu}_m, m = 1 : M\}$	NRLs means	$M \times 2$
	$\boldsymbol{v}_m = \{v_{m0}, v_{m1}\} \in \mathbb{R}_+^2$	Mixture model variances for NRLs in condition m	2
	$\boldsymbol{v} = \{\boldsymbol{v}_m, m = 1 : M\}$	NRLs variances	$M \times 2$
	$\beta_m \in \mathbb{R}_+$	Potts regularization parameter for condition m	1
	$\boldsymbol{\beta} = \{\beta_m, m = 1 : M\}$	regularization parameters	M
	$\beta_z \in \mathbb{R}_+$	Potts regularization parameter for HRF groups	1
	$\boldsymbol{\nu} = \{\nu_k, k = 1 : K\} \in \mathbb{R}_+^K$	HRF prior parameters (isotropic covariances)	K
	$\bar{\boldsymbol{h}} = \{\bar{\boldsymbol{h}}_k, k = 1 : K\} \in \mathbb{R}^{(D+1) \times K}$	HRF patterns	$(D+1) \times K$
	$\sigma_h^2 \in \mathbb{R}_+$	HRF smoothing parameter	1
Fixed	$\boldsymbol{X}_m \in \{0, 1\}^{N \times (D+1)}$	Binary stimulus occurrence matrix for condition m	$N \times (D+1)$
	$\boldsymbol{P} \in \mathbb{R}^{N \times O}$	Low frequency orthonormal function basis	$N \times O$

Denoting $\overline{\boldsymbol{y}}_j = \boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j - \boldsymbol{S}_j\boldsymbol{h}_j$, the likelihood can be factorized over voxels as follows:

$$p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}, \boldsymbol{\Gamma}) \propto \prod_{j \in \mathcal{V}} \det\boldsymbol{\Lambda}_{\mathrm{j}}^{1/2} \sigma_{\mathrm{j}}^{-\mathrm{N}} \exp\big(-\frac{\overline{\boldsymbol{y}}_{\mathrm{j}}^{\mathrm{t}} \boldsymbol{\Lambda}_{\mathrm{j}} \overline{\boldsymbol{y}}_{\mathrm{j}}}{2\sigma_{\mathrm{j}}^2}\big). \tag{2}$$

2) Neuronal response levels:

The NRLs are assumed to be statistically independent across conditions: $p(\boldsymbol{A}; \boldsymbol{\theta}_{\boldsymbol{a}}) = \prod_{m=1}^{M} p(\boldsymbol{a}^m; \boldsymbol{\theta}_m)$

where $\boldsymbol{\theta}_{\boldsymbol{a}} = \{\boldsymbol{\theta}_m, m = 1 : M\}$ and $\boldsymbol{\theta}_m$ gathers the parameters for the m-th condition. A mixture model is then adopted by using the allocation variables q_j^m to segregate non-activated voxels ($q_j^m = 0$) from activated ones ($q_j^m = 1$). For the m-th condition, and conditionally to the assignment variables $\boldsymbol{q}^m$, the NRLs are assumed to be independent: $p(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m; \boldsymbol{\theta}_m) = \prod_{j \in \boldsymbol{\mathcal{V}}} p(a_j^m \,|\, q_j^m; \boldsymbol{\theta}_m)$ with $p(a_j^m \,|\, q_j^m = i; \boldsymbol{\theta}_m) \sim \mathcal{N}(\mu_{mi}, v_{mi})$ and $\boldsymbol{\theta}_m = \{\mu_{mi}, v_{mi}, i = 0, 1\}$. We also denote $\boldsymbol{\mu} = \{\boldsymbol{\mu}_{mi}, m = 1 : M, i = 0, 1\}$ and $\boldsymbol{v} = \{\boldsymbol{v}_{mi}, m = 1 : M, i = 0, 1\}$. For non-activating voxels ($i = 0$) we set $\mu_{m0} = 0$, for all $m = 1 : M$. The other parameters are unknown and have to be estimated.

3) Activation classes:

As in [10], we assume prior independence between the M experimental conditions regarding the activation class assignments. It follows that $p(\boldsymbol{Q}) = \prod_{m=1}^{M} p(\boldsymbol{q}^m; \beta_m)$ where we assume in addition that $p(\boldsymbol{q}^m; \beta_m)$ is a Markov random field, namely a Potts model. Such prior modeling assumption is consistent with the physiological properties of the fMRI signal where the activity is known to be correlated in space [7, 22]. Here, the prior Potts model with interaction parameter β_m [10] is expressed, omitting the intractable normalizing constant, as:

$$p(\boldsymbol{q}^m; \beta_m) \propto \exp\big(\beta_m U(\boldsymbol{q}^m)\big) \quad \text{with} \quad U(\boldsymbol{q}^m) = \sum_{(j,j') \in \boldsymbol{\mathcal{V}}^2, j \sim j'} I(q_j^m = q_{j'}^m) \tag{3}$$

and where for all $(a, b) \in \mathbb{R}^2$, $I(a = b) = 1$ if $a = b$ and 0 otherwise. The notation $j \sim j'$ means that the summation is over all neighboring voxels. The neighboring system may cover a 3D scheme through the brain volume. The unknown parameters are denoted by $\boldsymbol{\beta} = \{\beta_m, m = 1 : M\}$. In what follows, we will consider a 6-connexity 3D neighboring system.

4) HRF groups: In order to promote parcellation connexity, we also introduce here a spatial Markov prior, namely a K-class Potts model with interaction parameter β_z:

$$p(\boldsymbol{Z}; \beta_z) \propto \exp\big(\beta_z U(\boldsymbol{Z})\big) \quad \text{with} \quad U(\boldsymbol{Z}) = \sum_{(j,j') \in \boldsymbol{\mathcal{V}}^2, j \sim j'} I(z_j = z_{j'}). \tag{4}$$

It results from the above model that neighboring voxels tend to belong to the same HRF group, sharing thus the same HRF pattern.

5) HRF patterns: In contrast to [9, 10, 13] where a unique HRF shape is considered for a whole parcel, the distribution of $\boldsymbol{h}_j$ is expressed, for each voxel j, conditionally to the HRF group variable z_j: $p(\boldsymbol{H}|\boldsymbol{Z}) = \prod_{j \in \boldsymbol{\mathcal{V}}} p(\boldsymbol{h}_j \,|\, z_j)$ with $p(\boldsymbol{h}_j \,|\, z_j = k) \sim \mathcal{N}(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)$. Here, the mean vector $\bar{\boldsymbol{h}}_k$ can be seen as the HRF pattern for group k and $\bar{\boldsymbol{\Sigma}}_k$ regulates the stochastic perturbations around $\bar{\boldsymbol{h}}_k$. In practice, we will consider $\bar{\boldsymbol{\Sigma}}_k = \nu_k I_{(D+1)}$. In addition, smooth $\bar{\boldsymbol{h}}_k$'s are favored by controlling their second order

derivatives with the following prior: $\bar{\boldsymbol{h}}_k \sim \mathcal{N}(\boldsymbol{0}, \sigma_h^2 \boldsymbol{R})$ with $\boldsymbol{R} = (\Delta t)^4 \ (\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2)^{-1}$ where $\boldsymbol{D}_2$ is the second-order finite difference matrix and σ_h^2 is a hyperparameter to be estimated or fixed. Moreover, $\bar{\boldsymbol{h}}_{k0} = \bar{\boldsymbol{h}}_{kD\Delta t} = 0$ as in [9, 10, 13].

Using notation $\bar{\boldsymbol{h}} = \left\{\bar{\boldsymbol{h}}_k, k = 1 : K\right\}$ and $\boldsymbol{\nu} = \{\nu_k, k = 1 : K\}$, the parameters are then denoted by $\boldsymbol{\Theta} = \left\{\boldsymbol{\Gamma}, \boldsymbol{L}, \boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\beta}, \beta_z, \sigma_h^2, \boldsymbol{\nu}, \bar{\boldsymbol{h}}\right\}$ and belong to a set $\underline{\boldsymbol{\Theta}}$. In the following developments, the $\bar{\boldsymbol{h}}_k$'s will be considered as random parameters following the prior specified above with hyperparameter σ_h^2.

III. Variational EM estimation

We propose to use an EM framework to deal with the missing data $\boldsymbol{A} \in \mathcal{A}$, $\boldsymbol{H} \in \mathcal{H}$, $\boldsymbol{Q} \in \mathcal{Q}$, $\boldsymbol{Z} \in \mathcal{Z}$. From a methodological viewpoint, the adopted approach will be similar to the one in [12, 13]. For the sake of self contentedness, we recall below the main principle (see also [23]).

A. Variational Expectation-Maximization

Let $\mathcal{D}$ be the set of all probability distributions on $\mathcal{A} \times \mathcal{H} \times \mathcal{Q} \times \mathcal{Z}$. EM can be viewed [24] as an alternating maximization procedure of a function $\mathcal{F}$ on $\mathcal{D}$, for all $\tilde{p} \in \mathcal{D}$,

$$\mathcal{F}(\tilde{p}, \boldsymbol{\Theta}) \quad = \quad \mathrm{E}_{\tilde{p}}\big[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{\Theta})\big] + \mathcal{G}(\tilde{p}) \tag{5}$$

where $\mathrm{E}_{\tilde{p}}\big[.\big]$ denotes the expectation with respect to $\tilde{p}$ and $\mathcal{G}(\tilde{p}) = -\mathrm{E}_{\tilde{p}}\big[\log \tilde{p}(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z})\big]$ is the entropy of $\tilde{p}$. This function is called the free energy. It can be equivalently expressed in terms of the log-likelihood as $\mathcal{F}(\tilde{p}, \boldsymbol{\Theta}) = \log p(\boldsymbol{Y} \,|\, \boldsymbol{\Theta}) - KL(\tilde{p} \,||\, p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta}))$ where $KL(\tilde{p} \,||\, p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta}))$ is the Kullback-Leibler (KL) divergence between $\tilde{p}$ and $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta})$ with

$$KL(\tilde{p} \,||\, p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta})) = \int \tilde{p}(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z}) \log \frac{\tilde{p}(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z})}{p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta})} \, d\boldsymbol{A} \, d\boldsymbol{H} \, d\boldsymbol{Q} \, d\boldsymbol{Z}. \tag{6}$$

Hence maximizing the free energy with respect to $\tilde{p}$ amounts to minimizing the Kullback-Leibler divergence between $\tilde{p}$ and the posterior distribution of interest $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta})$. Since the KL divergence is always non-negative, and because the KL divergence of the posterior distribution to itself is zero, it follows easily that the maximum free energy over all $\tilde{p} \in \mathcal{D}$ is the log-likelihood. The link to the EM algorithm follows straightforwardly. At iteration (r), denoting the current parameter values by $\boldsymbol{\Theta}^{(r-1)}$, the alternating procedure proceeds as follows:

$$\textbf{E-step: } \tilde{p}_{A,H,Q,Z}^{(r)} = \underset{\tilde{p} \in \mathcal{D}}{\arg\max} \ \mathcal{F}(\tilde{p}, \boldsymbol{\Theta}^{(r-1)}) = p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta}^{(r-1)})) \tag{7}$$

$$\textbf{M-step: } \boldsymbol{\Theta}^{(r)} = \underset{\Theta \in \underline{\boldsymbol{\Theta}}}{\arg\max} \ \mathcal{F}(\tilde{p}_{A,H,Q,Z}^{(r)}, \boldsymbol{\Theta}) = \mathrm{E}_{\tilde{p}^{(r)}}\left[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{\Theta})\right] . \tag{8}$$

However, the optimization step in Eq. (7) leads to $\tilde{p}^{(r)}_{A,H,Q,Z} = p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta}^{(r-1)})$, which is intractable for our model. Hence, we resort to a variational EM (VEM) variant in which the intractable posterior is approximated by constraining the space of possible $\tilde{p}$ distributions in order to make the maximization procedure tractable. In that case, the free energy optimal value reached is only a lower bound on the log-likelihood. The most common variational approximation consists of optimizing over the distributions in $\mathcal{D}$ that factorize as a product of four pdfs on $\mathcal{A}$, $\mathcal{H}$, $\mathcal{Q}$ and $\mathcal{Z}$, respectively.

B. Variational Joint Parcellation-Detection-Estimation

We resort to an iterative variational EM procedure as in [13]. At each iteration (r), with $\boldsymbol{\Theta}^{(r-1)}$ denoting the current parameter values, the intractable posterior $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{\Theta}^{(r-1)})$ is approximated as a product of four pdfs, $\widetilde{p}^{(r)}_H$, $\widetilde{p}^{(r)}_A$, $\widetilde{p}^{(r)}_Q$ and $\widetilde{p}^{(r)}_Z$ respectively on $\mathcal{A}$, $\mathcal{H}$, $\mathcal{Q}$ and $\mathcal{Z}$. Our E-step becomes then an approximate E-step, which is decomposed into four sub-steps that consist of updating the four pdfs above in turn using four equivalent expressions of $\mathcal{F}$ when $\tilde{p}$ factorizes. At iteration (r) with current estimates denoted by $\widetilde{p}^{(r-1)}_H, \widetilde{p}^{(r-1)}_A, \widetilde{p}^{(r-1)}_Q$, $\widetilde{p}^{(r-1)}_Z$ and $\boldsymbol{\Theta}^{(r-1)}$, the updating rules become:

$$\textbf{VE-H:}\;\; \widetilde{p}^{(r)}_H = \arg\max_{p_H} \mathcal{F}(\widetilde{p}^{(r-1)}_A \; p_H \; \widetilde{p}^{(r-1)}_Q, \widetilde{p}^{(r-1)}_Z, \boldsymbol{\Theta}^{(r-1)})$$

$$\textbf{VE-A:}\;\; \widetilde{p}^{(r)}_A = \arg\max_{p_A} \mathcal{F}(p_A \; \widetilde{p}^{(r)}_H \; \widetilde{p}^{(r-1)}_Q \; \widetilde{p}^{(r-1)}_Z, \boldsymbol{\Theta}^{(r-1)})$$

$$\textbf{VE-Q:}\;\; \widetilde{p}^{(r)}_Q = \arg\max_{p_Q} \mathcal{F}(\widetilde{p}^{(r)}_A \; \widetilde{p}^{(r)}_H \; p_Q \; \widetilde{p}^{(r-1)}_Z, \boldsymbol{\Theta}^{(r-1)})$$

$$\textbf{VE-Z:}\;\; \widetilde{p}^{(r)}_Z = \arg\max_{p_Z} \mathcal{F}(\widetilde{p}^{(r)}_A \; \widetilde{p}^{(r)}_H \; \widetilde{p}^{(r)}_Q \; p_Z, \boldsymbol{\Theta}^{(r-1)}).$$

In other words, the factorization is used to maximize the free energy by alternately maximizing it with respect to $\tilde{p}_H$, $\tilde{p}_A$, $\tilde{p}_Q$ and $\tilde{p}_Z$ while keeping the other distributions fixed. The steps above can then be equivalently written in terms of minimizations of some Kullback-Leibler divergences. The properties of the latter lead to the following solutions:

$$\textbf{VE-H:}\;\; \widetilde{p}^{(r)}_H(\boldsymbol{H}) \propto \exp\left(\mathrm{E}_{\widetilde{p}^{(r-1)}_A \widetilde{p}^{(r-1)}_Z}\left[\log p(\boldsymbol{H} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}; \boldsymbol{\Theta}^{(r-1)}\right]\right) \tag{9}$$

$$\textbf{VE-A:}\;\; \widetilde{p}^{(r)}_A(\boldsymbol{A}) \propto \exp\left(\mathrm{E}_{\widetilde{p}^{(r)}_H \widetilde{p}^{(r-1)}_Q}\left[\log p(\boldsymbol{A} \,|\, \boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{Q}; \boldsymbol{\Theta}^{(r-1)})\right]\right) \tag{10}$$

$$\textbf{VE-Q:}\;\; \widetilde{p}^{(r)}_Q(\boldsymbol{Q}) \propto \exp\left(\mathrm{E}_{\widetilde{p}^{(r)}_A}\left[\log p(\boldsymbol{Q} \,|\, \boldsymbol{Y}, \boldsymbol{A}; \boldsymbol{\Theta}^{(r-1)})\right]\right) \tag{11}$$

$$\textbf{VE-Z:}\;\; \widetilde{p}^{(r)}_Z(\boldsymbol{Z}) \propto \exp\left(\mathrm{E}_{\widetilde{p}^{(r)}_H}\left[\log p(\boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{H}; \boldsymbol{\Theta}^{(r-1)})\right]\right) \;. \tag{12}$$

The corresponding **M-step** is (since $\boldsymbol{\Theta}$ and $\mathcal{G}(\widetilde{p}^{(r)}_{A,H,Q,Z})$ are independent, see Eq. (5)):

$$\textbf{VM:}\;\; \boldsymbol{\Theta}^{(r)} = \arg\max_{\boldsymbol{\Theta}} \; \mathrm{E}_{\widetilde{p}^{(r)}_A \widetilde{p}^{(r)}_H \widetilde{p}^{(r)}_Q \widetilde{p}^{(r)}_Z}\left[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z}; \boldsymbol{\Theta})\right] \;. \tag{13}$$

Compared to [13], this implies adding an E-sub-step for the HRF group assignments ($\widetilde{p}_Z^{(r)}$ updating) and specifying its impact on the other E-sub-steps. The **E-Q** sub-step ($\widetilde{p}_Q^{(r)}$ updating) is not actually impacted by the HRF groups addition and can be found in [13]. The **E-A** sub-step ($\widetilde{p}_A^{(r)}$ updating) is also very close to the one involved in [13]: similar updating formulas are obtained by replacing the HRF of [13] by voxel dependent HRFs.

It follows from standard algebra that at all iterations (r), $\tilde{p}_H^{(r)}$ and $\widetilde{p}_A^{(r)}$ are both Gaussian distributions: $\widetilde{p}_H^{(r)} = \prod_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{h_j}^{(r)}$ and $\widetilde{p}_A^{(r)} = \prod_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{a_j}^{(r)}$, where $\widetilde{p}_{h_j}^{(r)} \sim \mathcal{N}(\boldsymbol{m}_{h_j}^{(r)}, \boldsymbol{\Sigma}_{h_j}^{(r)})$ and $\widetilde{p}_{a_j}^{(r)} \sim \mathcal{N}(\boldsymbol{m}_{a_j}^{(r)}, \boldsymbol{\Sigma}_{a_j}^{(r)})$. More specifically, the VE-steps write as follow:

- **VE-H** step: Developing the right hand side of (9), it comes that the Gaussian $\widetilde{p}_{h_j}^{(r)}$ is defined by its parameters $\boldsymbol{\Sigma}_{h_j}^{(r)} = (\boldsymbol{V}_1 + \boldsymbol{V}_2)^{-1}$ and $\boldsymbol{m}_{h_j}^{(r)} = \boldsymbol{\Sigma}_{h_j}^{(r)}(\boldsymbol{m}_1 + \boldsymbol{m}_2)$, where

$$\boldsymbol{V}_1 = \sum_{m,m'} \boldsymbol{\Sigma}_{a_j(m,m')}^{(r-1)} \boldsymbol{X}_m^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \boldsymbol{X}_{m'} + \widetilde{\boldsymbol{S}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \widetilde{\boldsymbol{S}}_j,$$

$$\boldsymbol{V}_2 = \sum_{k=1}^{K} \widetilde{p}_{z_j}(k)^{(r-1)} \bar{\boldsymbol{\Sigma}}_k^{(r-1)-1},$$

$$\boldsymbol{m}_1 = \widetilde{\boldsymbol{S}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j^{(r-1)}),\ \boldsymbol{m}_2 = \sum_{k=1}^{K} \bar{\boldsymbol{\Sigma}}_k^{(r-1)-1} \widetilde{p}_{z_j}(k)^{(r-1)} \bar{\boldsymbol{h}}_k^{(r-1)}.$$

  Above, $\widetilde{\boldsymbol{S}}_j = \sum_{m=1}^{M} m_{a_j^m}^{(r-1)} \boldsymbol{X}_m$ and $m_{a_j^m}^{(r-1)}$, $\boldsymbol{\Sigma}_{a_j(m,m')}^{(r-1)}$ denote respectively the m and (m,m') entries of $\boldsymbol{m}_{a_j}^{(r-1)}$ and $\boldsymbol{\Sigma}_{a_j}^{(r-1)}$.

- **VE-A** step : This step is similar to the one of the JDE model developed in [13] if we consider a different HRF per voxel j. Using Eq. (10), standard algebra rules allow to identify the Gaussian distribution $\widetilde{p}_{a_j}^{(r)} \sim \mathcal{N}(\boldsymbol{m}_{a_j}^{(r)}, \boldsymbol{\Sigma}_{a_j}^{(r)})$, with

$$\boldsymbol{\Sigma}_{a_j}^{(r)} = \left(\sum_{i\in\{0,1\}} \boldsymbol{\Delta}_{ij} + \widetilde{\boldsymbol{H}_j} \right)^{-1}, \quad \boldsymbol{m}_{a_j}^{(r)} = \boldsymbol{\Sigma}_{a_j}^{(r)} \left(\sum_{i\in\{0,1\}} \boldsymbol{\Delta}_{ij} \boldsymbol{\mu}_i^{(r-1)} + \widetilde{\boldsymbol{G}}^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \big(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j^{(r-1)}\big) \right), \tag{14}$$

  where a number of intermediate quantities need to be specified. First, $\boldsymbol{\mu}_i^{(r-1)} = \left[\mu_{1i}^{(r-1)}, \ldots, \mu_{Mi}^{(r-1)}\right]^{\mathrm{t}}$ and $\widetilde{\boldsymbol{G}} = \mathrm{E}_{\widetilde{p}_H^{(r)}}\left[\boldsymbol{G}\right]$ where $\boldsymbol{G}$ is the matrix $\boldsymbol{G} = [\boldsymbol{g}_1 \mid \ldots \mid \boldsymbol{g}_M]$ made of columns $\boldsymbol{g}_m = \boldsymbol{X}_m \boldsymbol{h}_j$. The m-th column of $\widetilde{\boldsymbol{G}}$ is then also denoted by $\widetilde{\boldsymbol{g}}_m = \boldsymbol{X}_m \boldsymbol{m}_{h_j}^{(r)} \in \mathbb{R}^N$. Then, $\boldsymbol{\Delta}_{ij} = \mathrm{diag}_M \left[\widetilde{p}_{q_j^m}^{(r-1)}(i)/v_{mi}^{(r-1)}\right]$

and $\widetilde{\boldsymbol{H}}_j = \mathrm{E}_{\widetilde{p}^{(r)}_{h_j}}\big[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\boldsymbol{G}\big]$ is an $M \times M$ matrix whose element (m, m') is given by:

$$\begin{aligned}\mathrm{E}_{\widetilde{p}^{(r)}_{h_j}}\big[\boldsymbol{g}_m^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\boldsymbol{g}_{m'}\big] &= \mathrm{E}_{\widetilde{p}^{(r)}_{h_j}}\big[\boldsymbol{g}_m\big]^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\mathrm{E}_{\widetilde{p}^{(r)}_{h_j}}\big[\boldsymbol{g}_{m'}\big] + \mathrm{trace}\big(\boldsymbol{\Gamma}_j^{(r-1)} cov_{\widetilde{p}^{(r)}_{h_j}}(\boldsymbol{g}_m, \boldsymbol{g}_{m'})\big)\\ &= \widetilde{\boldsymbol{g}}_m^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\widetilde{\boldsymbol{g}}_{m'} + \mathrm{trace}\big(\boldsymbol{\Gamma}_j^{(r-1)}\boldsymbol{X}_m\boldsymbol{\Sigma}_{h_j}^{(r)}\boldsymbol{X}_{m'}^{\mathrm{t}}\big).\end{aligned}$$

- **VE-Q** step: This step remains conceptually the same as for the JDE model developed in [13] where no parcellation is estimated online. However, we provide below a slightly different and simpler presentation. Note also that we have corrected here previous typos on p.6 and Appendix D of [13]. In contrast to [13], we assume from the start that for all $m = 1 : M$, $\widetilde{p}^{(r)}_{\boldsymbol{q}^m}(\boldsymbol{q}^m) = \prod_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}^{(r)}_{q_j^m}(q_j^m)$ so that the VE-Q step (11) actually divides into $M \times J$ successive sub-steps. More specifically, for $m = 1 : M$, each VE-$\boldsymbol{q}^m$ step is decomposed into J successive VE-q_j^m step. It comes, for $j \in \boldsymbol{\mathcal{V}}$:

$$\widetilde{p}^{(r)}_{q_j^m}(q_j^m) \propto \exp(\mathrm{E}_{\widetilde{p}^{(r)}_A \widetilde{p}^{(r)}_{q^{\setminus m}} \widetilde{p}^{(r)}_{q^m_{\setminus j}}}\big[\log p(q_j^m|\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{q}^m_{\setminus j}, \boldsymbol{q}^{\setminus m}; \boldsymbol{\Theta}^{(r-1)}\big]) \tag{15}$$

where $\boldsymbol{q}^m_{\setminus j} = \{q^m_{j'}, j' \neq j\}$ and $\boldsymbol{q}^{\setminus m} = \{\boldsymbol{q}^{m'}, m' \neq m\}$. Keeping only the terms that involve q_j^m, it follows,

$$\widetilde{p}^{(r)}_{q_j^m}(q_j^m) \propto \exp(\mathrm{E}_{\widetilde{p}^{(r)}_{a_j^m}}\big[\log p(a_j^m|q_j^m; \mu_m^{(r-1)}, v_m^{(r-1)}\big] + \mathrm{E}_{\widetilde{p}^{(r-1)}_{\boldsymbol{q}^m_{\setminus j}}}\big[\log p(q_j^m|\boldsymbol{q}^m_{\setminus j}; \beta_m^{(r-1)}\big]) \tag{16}$$

$$\propto \exp\left(-\frac{1}{2}\log v^{(r-1)}_{mq_j^m} - \frac{1}{2}\frac{(\boldsymbol{m}^{(r)}_{a_j^m} - \mu^{(r-1)}_{mq_j^m})^2 + \boldsymbol{\Sigma}^{(r)}_{a_j(m,m)}}{v^{(r-1)}_{mq_j^m}} + \beta_m^{(r-1)} \sum_{j'\in\boldsymbol{\mathcal{V}}, j'\sim j} \widetilde{p}^{(r-1)}_{q^m_{j'}}(q_j^m)\right) \tag{17}$$

For $q_j^m = i$, it follows

$$\widetilde{p}^{(r)}_{q_j^m}(i) \propto \mathcal{N}(\boldsymbol{m}^{(r)}_{a_j^m}; \mu^{(r-1)}_{mi}, v^{(r-1)}_{mi})\ \exp\left(-\frac{1}{2}\frac{\boldsymbol{\Sigma}^{(r)}_{a_j(m,m)}}{v^{(r-1)}_{mi}} + \beta_m^{(r-1)} \sum_{j'\in\boldsymbol{\mathcal{V}}, j'\sim j} \widetilde{p}^{(r-1)}_{q^m_{j'}}(i)\right)\ , \tag{18}$$

where the Gaussian distribution with mean μ_{mi} and variance v_{mi} is denoted by $\mathcal{N}(\ .\ ; \mu_{mi}, v_{mi})$. Note that the above expression corresponds to a synchrone updating of each $\widetilde{p}_{q_j^m}$ in turn from the previous $\widetilde{p}^{(r-1)}_{q_j^m}$'s. However in practice, as soon as $\widetilde{p}_{q_j^m}$ is updated for some j, it will be taken into account in the following updates of the subsequent $\widetilde{p}_{q^m_{j'}}$ for $j' \neq j$.

- **VE-Z** step: As for the VE-Q step, we look for a variational solution of the form $\widetilde{p}^{(r)}_Z(\boldsymbol{Z}) = \prod_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}^{(r)}_{z_j}(z_j)$, which leads to J successive sub-steps performed in turn. For each $j \in \boldsymbol{\mathcal{V}}$, and

$k = 1 : K$,

$$\widetilde{p}_{z_j}^{(r)}(k) \propto \mathcal{N}(\boldsymbol{m}_{h_j}^{(r)}; \bar{\boldsymbol{h}}_k^{(r-1)}, \bar{\boldsymbol{\Sigma}}_k^{(r-1)}) \ \exp\left(-\frac{1}{2}\text{trace}(\boldsymbol{\Sigma}_{h_j}^{(r)} \bar{\boldsymbol{\Sigma}}_k^{(r-1)-1}) + \beta_z^{(r-1)} \sum_{j' \in \boldsymbol{\mathcal{V}}, j' \sim j} \widetilde{p}_{z_{j'}}^{(r-1)}(k) \right). \tag{19}$$

- **VM** step: For this maximization step, we can first rewrite Eq. (13) as

$$\boldsymbol{\Theta}^{(r)} = \arg\max_{\boldsymbol{\Theta}} \ \Big[\mathrm{E}_{\widetilde{p}_A^{(r)} \widetilde{p}_H^{(r)}} \big[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}, \boldsymbol{\Gamma}) \big] + \mathrm{E}_{\widetilde{p}_A^{(r)} \widetilde{p}_Q^{(r)}} \big[\log p(\boldsymbol{A} \mid \boldsymbol{Q}; \boldsymbol{\mu}, \boldsymbol{v}) \big] + \mathrm{E}_{\widetilde{p}_Q^{(r)}} \big[\log p(\boldsymbol{Q}; \boldsymbol{\beta}) \big]$$
$$+ \mathrm{E}_{\widetilde{p}_Z^{(r)}} \big[\log p(\boldsymbol{Z}; \beta_z) \big] + \mathrm{E}_{\widetilde{p}_H^{(r)} \widetilde{p}_Z^{(r)}} \Big[\log p(\boldsymbol{H} \mid \boldsymbol{Z}; \boldsymbol{\nu}, \bar{\boldsymbol{h}}) \Big] + \log p(\bar{\boldsymbol{h}}; \sigma_h^2) \big]. \tag{20}$$

  Note that the prior on $\bar{\boldsymbol{h}}$ is taken into account in the last term. The maximization step can therefore be divided into five sub-steps (two additional ones compared to [13]) involving separately $(\boldsymbol{\mu}, \boldsymbol{v})$, $\boldsymbol{\beta}$, β_z, $(\mathbf{L}, \boldsymbol{\Gamma})$, σ_h^2 and $(\boldsymbol{\nu}, \bar{\boldsymbol{h}})$. For some of the above sub-steps, a closed form can be derived for the related parameters. For some others, numerical iterative procedures are requested. See [13] and Appendix A for details.

IV. Validation of the proposed approach

A. Validation on synthetic data

1) Case of study: Experiments have been carried out on artificial fMRI data generated according to Eq. (1). We simulated a random mixed sequence of indexes coding for $M = 2$ different stimuli composed of 30 trials each with a time of repetition set to $TR = 1s$. The resulting sequence was then multiplied by stimulus-dependent and space-varying NRLs, which were drawn from the prior distribution $p(\boldsymbol{A}; \boldsymbol{\theta}_a)$. To this end, 2D slices composed of 20 x 20 binary labels $\boldsymbol{q}^m$ (activating and non-activating voxels) were constructed for each stimulus type m (see Fig. 2[top]). Given these labels, the NRLs were simulated as follows, for $m = 1, 2$: $(a_j^m \mid q_j^m = 0) \sim \mathcal{N}(0, 0.5)$ and $(a_j^m \mid q_j^m = 1) \sim \mathcal{N}(3.2, 0.5)$ (see Fig. 5[top]). As regards HRFs, three groups ($K = 3$) were considered and spatially organized in three parcels of similar size (labels $\boldsymbol{Z}$) as shown in Fig. 3. Within each parcel, all voxels share the same HRF prior parameters $(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)$. The mean HRF shapes $\{\bar{\boldsymbol{h}}_k, k = 1 : K\}$ are depicted in Fig. 4 (ground truth) and show strong fluctuations across parcels in terms of peak positions and widths. Isotropic prior covariance matrices $\{\bar{\boldsymbol{\Sigma}}_k = \nu_k I_{D+1}, k = 1 : K\}$ were considered to draw voxel-specific HRFs according to $p(\boldsymbol{h}_j | z_j = k)$ where ν_k has been set to 0.02.

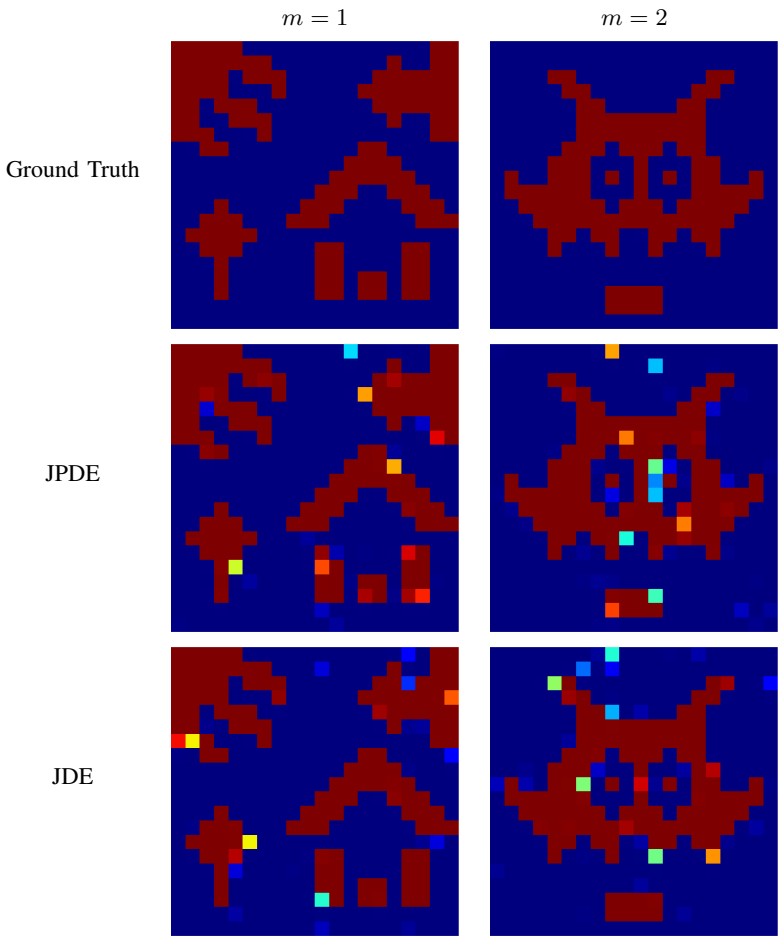

Fig. 2. Reference activation labels and Posterior Probability Maps (PPM) for JPDE and JDE (a single parcel is assumed for JDE).

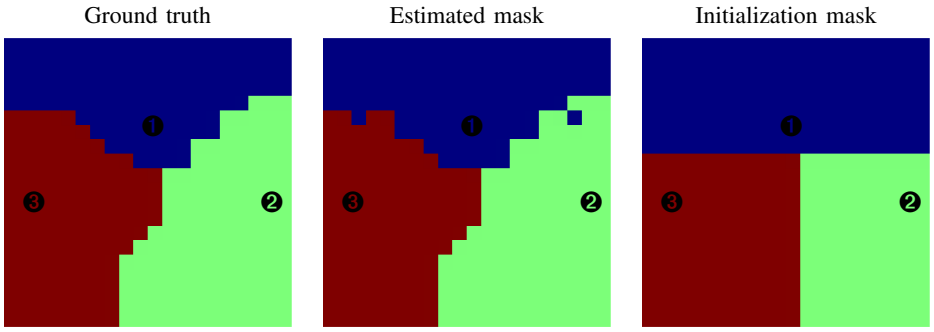

Fig. 3. Ground truth (left), JPDE estimated (middle) and the initial parcellation (right) masks.

As regards parcellation, Fig. 3[middle] shows the ability of JPDE to recover the spatial support of hemodynamic territories with high accuracy (1% of misclassified voxels and a DICE index [25] of 0.993)

from an arbitrary initialization (Fig. 3[right]). The HRF variability does not seem to affect the activation maps which are equally well estimated in the JPDE and JDE cases (Fig. 2[middle] and Fig. 2[bottom]). However, a clear difference is seen on the estimated HRFs, which are depicted in Fig. 4 together with the ground truth: the three parcel-specific HRF estimates using JPDE are plotted in addition to the JDE-based HRF time course obtained by merging all parcels of the arbitrary initial parcellation. Three JDE experiments have therefore been conducted, where each of the three ground truth HRFs in Fig. 4 has been used as global reference HRF in a separate experiment. The estimated HRFs are then compared to the ground truth and the one estimated with JPDE for the corresponding parcel. It turns out that the JPDE estimation is accurate for all parcels although the parcels cover different proportions of activation areas (*i.e.* the amount of useful signal varies from one parcel to another). In contrast, JDE provides an intermediate HRF shape that lies between those of the three parcels. This explains the observed differences between the two models in terms of NRL estimates and points out the JDE sensitivity to the choice of the *a priori* parcellation. When the parcellation is imperfect, JDE enforces fitting a wrong HRF model to the real data, and therefore a wrong activation dynamics. In contrast, JPDE is able to automatically refine an initial candidate for the parcellation and provides reliable detection and estimation results.

Interestingly, the NRL differences in Fig. 5 (see the JPDE-JDE plots in Fig. 5[bottom]) show that NRL estimates with JPDE have higher values, which means that JPDE allows retrieving stronger activation dynamics closer to the ground truth.

The most significant NRL differences lie in parcels 2 and 3 where the JDE HRF estimate differs the most from the ground truth. In terms of Mean Square Error (MSE), reported values confirm the improved performance of JPDE over JDE: $MSE^{m=1}_{\text{JDE}} = 0.0182$ *vs* $MSE^{m=1}_{\text{JPDE}} = 0.0107$ and $MSE^{m=2}_{\text{JDE}} = 0.0183$ *vs* $MSE^{m=2}_{\text{JPDE}} = 0.0141$.

2) Robustness study: The aim of this part consists of assessing the robustness of the JPDE model to the initial conditions. To this end, we first investigated the robustness of JPDE to the input SNR in terms of estimation error on the output NRLs. We then investigated the robustness of JPDE to the initial parcellation in terms of DICE coefficient for the estimated output parcellation. These first two experiments were conducted on 50 Monte Carlo (MC) runs in order to compute mean estimate and standard deviations values. A third set of experiments was designed to measure the impact of the overlapping percentage between the parcellation and activation maps in order to study how the presence of activations within each parcel helps achieving good parcellation results.

a) Robustness to input SNR: In this experiment, 50 MC simulations have been used to evaluate the output MSE on estimated NRLs for five different input SNR values. Fig. 6 shows the evolution of the

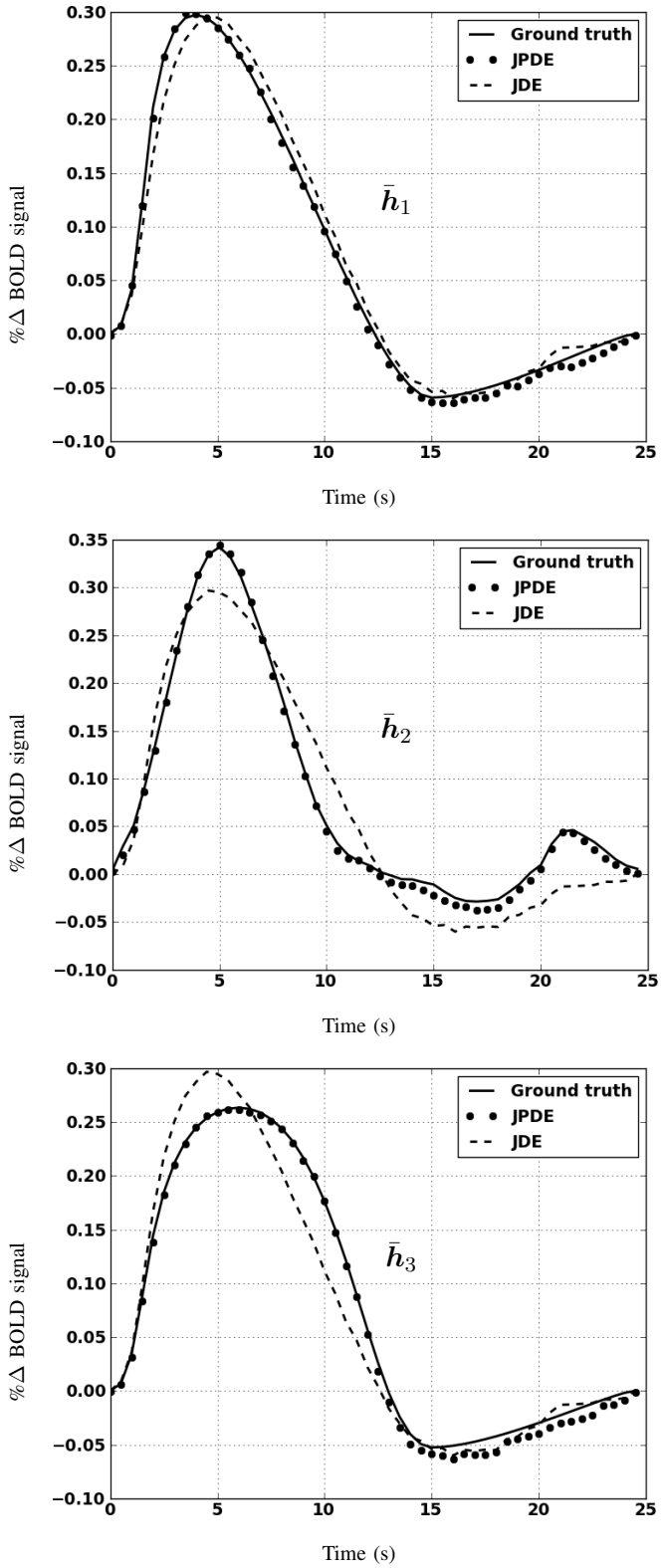

Fig. 4. Reference and estimated HRF patterns $\{\bar{h}_k, k = 1 : K\}$ for $K = 3$ parcels using JPDE and JDE.

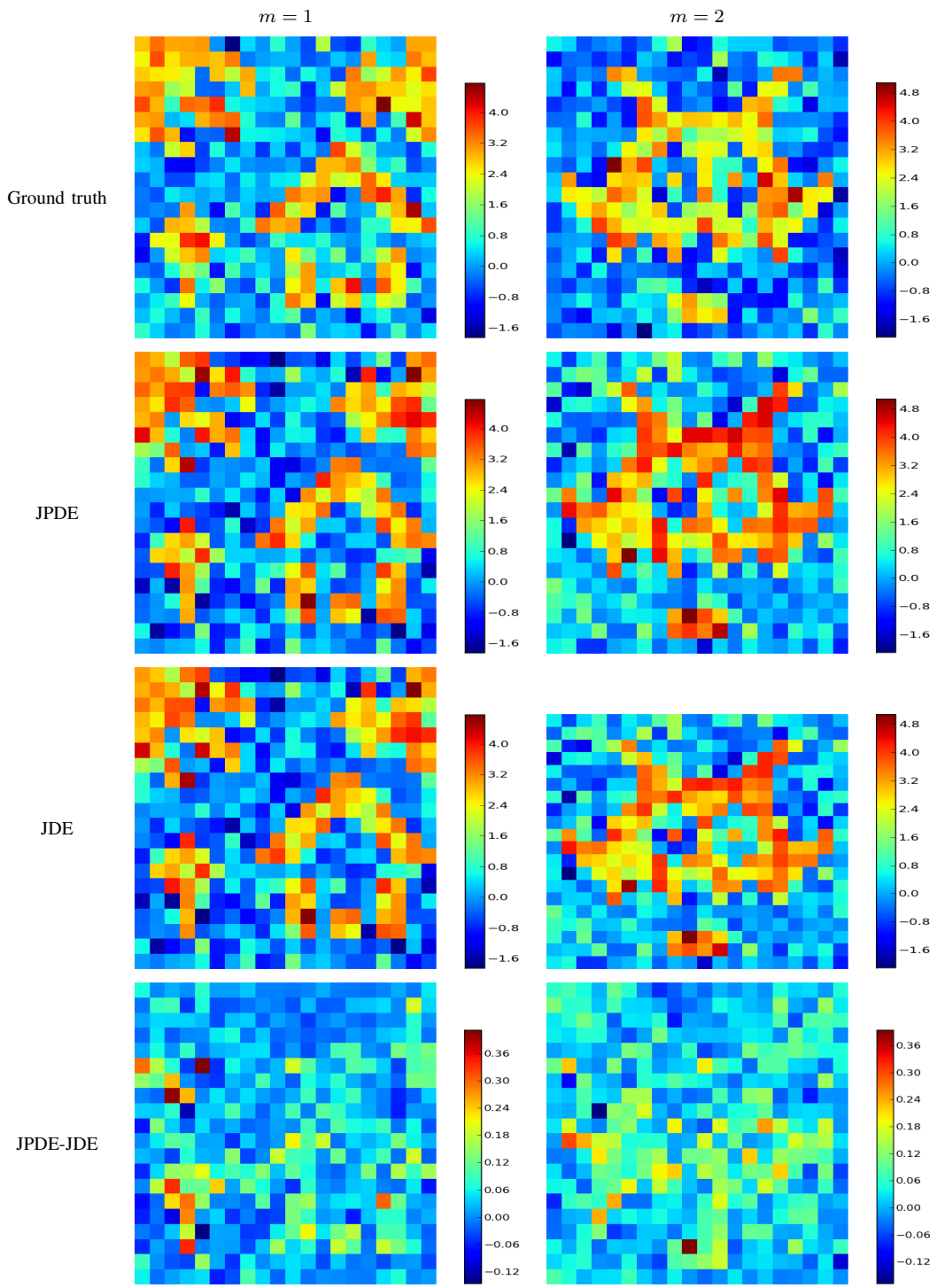

Fig. 5. Reference and estimated NRLs using JPDE (3 parcels) and JDE (1 parcel).

mean MSE on estimated NRLs in addition to the standard deviation over this mean value with respect to the input SNR. Mean MSE and standard deviations are displayed separately for each experimental condition m. Our results indicate low error level on estimated NRLs. The obtained curves also show that the standard deviation around the mean estimated value decreases as well as this mean value as soon as the input SNR increases. These results are consistent with those observed in out previous work on the JDE model [13]. Note that for this experiment, a 3-parcel model has been used with the same ground truth parcellation and activation masks in Fig 3 and Fig 2, respectively.

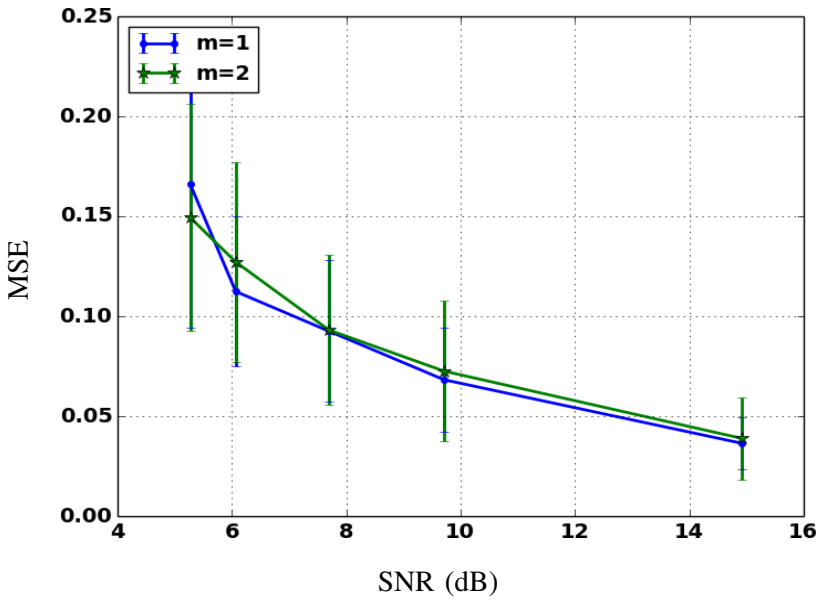

Fig. 6. Evolution of the mean MSE and standard deviation on estimated NRLs with JPDE, for 5 different input SNR values and for the two experimental conditions.

b) Robustness to initial parcellation: Here, we investigated the sensitivity of the proposed JPDE model to the initial parcellation mask. The accuracy of the parcellation mask, which was given as input to JPDE, may actually have a strong impact on the final estimate of the parcellation. The accuracy was measured in terms of DICE coefficients for each estimated mask. More specifically, different DICE levels between the ground truth and the initial parcellation have been tested to cover a large number of situations. Over 50 MC runs, for each DICE level, different initial parcellation masks were given to the JPDE algorithm. These masks were generated in a pseudo-random way so as to guarantee a random border between two parcels in a 2-parcel model, while approximately ensuring connexity and the same DICE coefficient value for each such generated mask. Fig. 7 shows a reference parcellation mask and three examples of generated masks with two parcels, all sharing approximately the same value of DICE

coefficient. More details about this pseudo-random synthesis are given in Appendix B.

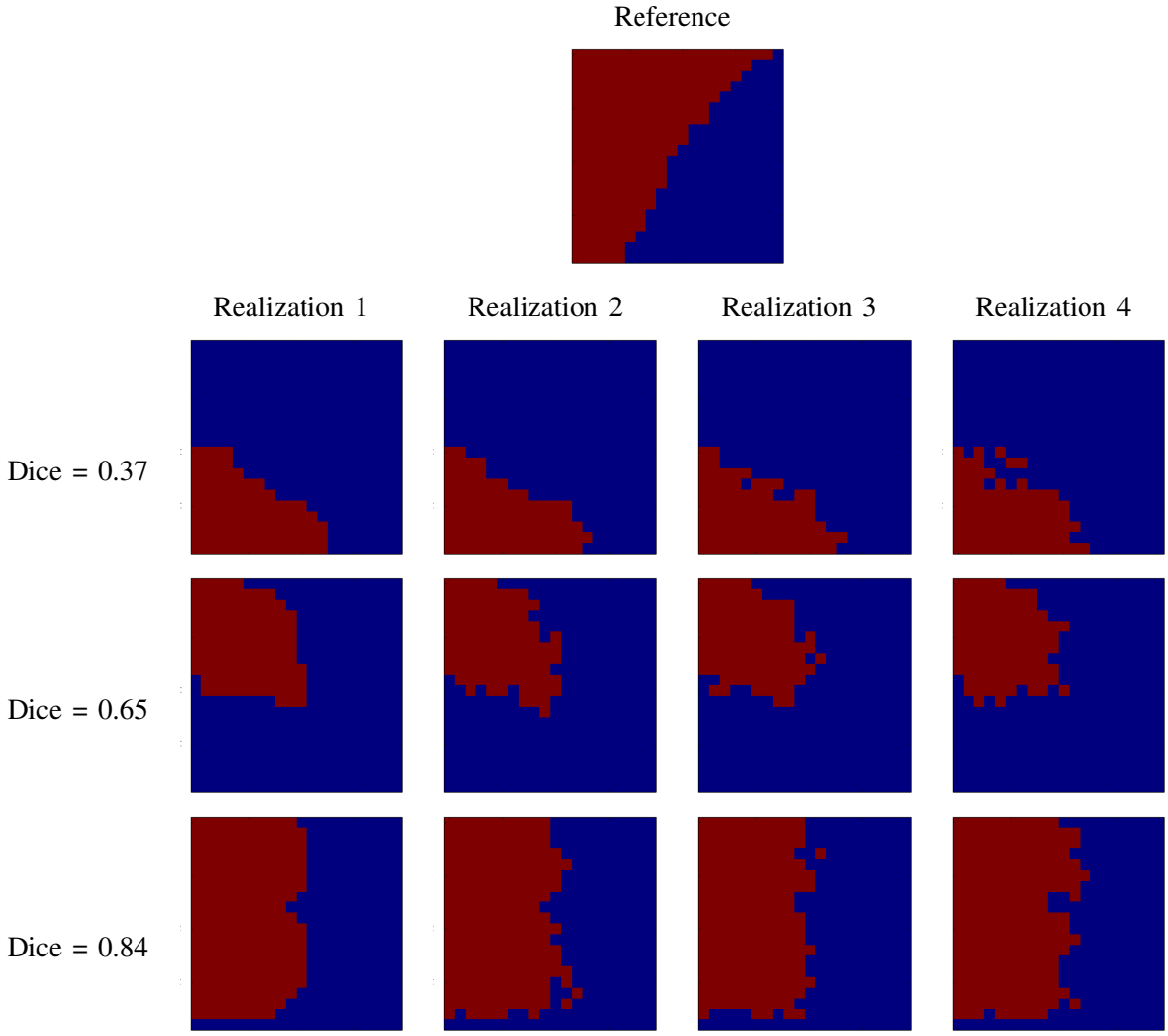

Fig. 7. Reference parcellation mask (top row) and pseudo-randomly generated parcellations with a 2-parcel model. Displayed masks for three DICE coefficient groups. Each row illustrates different generated masks sharing the same DICE coefficient.

Fig. 8 shows the DICE coefficient (mean value and standard deviation) for the estimated parcellation with respect to the DICE level for the initial parcellation. Through the illustrated curve, one can easily notice that the output mean DICE value increases with the input DICE, which was expected since an accurate initialization naturally leads to an accurate final estimate. As regards standard deviations around the mean DICE values, they remain almost the same. One can notice a slightly higher standard deviation value for the second mask (DICE=0.65). This could be explained by the fact that the red parcel of this example mainly covers an activated area from the first experimental condition ($m = 1$). The pseudo-

random generation for this example can then involve more or less activated voxels from the second experimental condition ($m = 2$). This coverage effect is addressed in the following section.

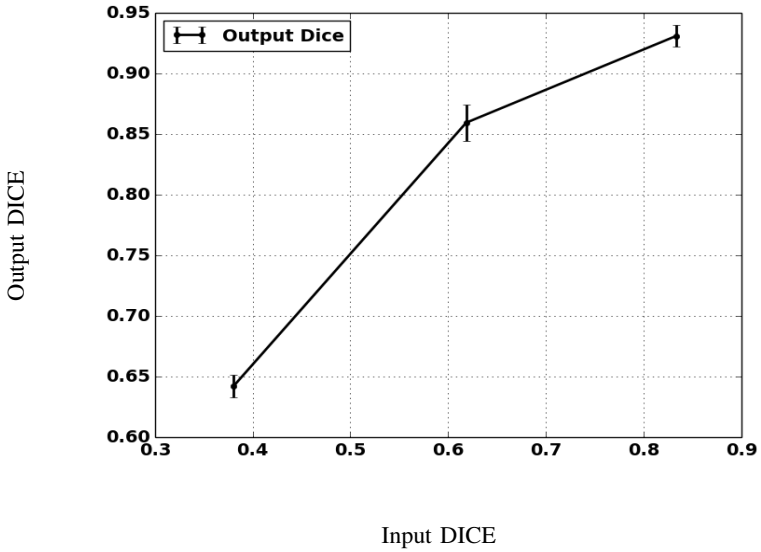

Fig. 8. Output DICE coefficient for the estimated parcellation with respect to the input DICE over 50 MC runs for the three initial parcellation masks in Fig. 7 .

c) Activation coverage: Next, we investigated the impact of the amount of activations inside target parcels. Owing to the hierarchical model adopted for JPDE, parcels are estimated (parcellation task) from the HRF patterns (estimation task), which are strongly linked to the accuracy of detected activations (detection task). These three tasks are thus strongly linked one to the other within the JPDE formulation. For this reason, and as already shown in previous works on the standard JDE model [9, 10, 13], precise HRF estimation is mainly expected in activated brain areas, and vice versa, improved activation detection should be achieved from more accurate HRF estimates. In the same vein, precise delineation of parcels is mainly expected if the latter embed evoked brain activity. The presence of activations within each target parcel is therefore an important concern that is now investigated. To this end, several simulations have been conducted using different configurations for the reference parcellation masks. A 2-parcel model has been used with the same activation maps as in Fig.2. The overlapping percentage between the two parcels and the activation maps changes over simulations. Fig. 9 shows the three parcellation masks we used. They are superimposed to the global activation maps ($m = 1$ and $m = 2$). Note that this global maps have been obtained with a logical "OR" applied to the two activation label maps so that a voxel is said activated if it induces evoked activity for at least one experimental condition ($m = 1$ or $m = 2$).

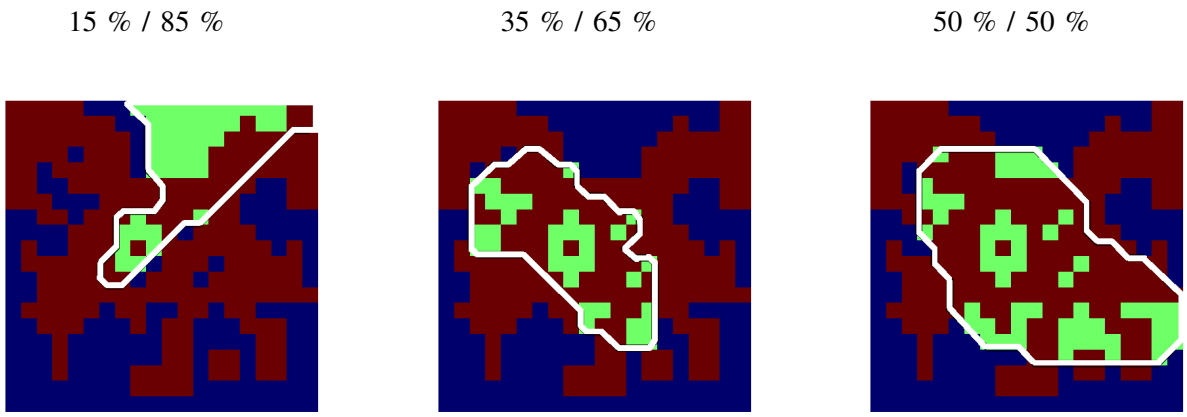

Fig. 9. Overlaps between the parcellation masks (green parcel delimited by the white line) and the global activation map (red). Solid black line delineates the boundary between the two parcels. The overlapping percentage is reported for each mask.

We then focused on the parcellation results and evaluated them in terms of output DICE coefficients. Fig. 10 provides the output DICE curve with respect to the overlapping percentage between the parcels and the activation labels. Masks with different overlapping levels were used for simulating the data. Slightly different masks were given as initialization to the JPDE algorithm based on the pseudo-random generation detailed in Section IV-A2b. One can easily notice that a fair partition (50%/50%) of activations across parcels helps us achieving better parcellation performance compared to a non-uniformly distributed case (15%/85% for example). This results is thus coherent with our expectations: the presence of activations helps us uncovering accurate brain parcellations based on hemodynamic properties.

B. Validation on real data

This section is devoted to the validation of the proposed JPDE model on real fMRI data.

We analyzed fMRI data acquired on a 6-year old child in the context of a study designed to investigate the large-scale organization of spoken language areas in dyslexic children [26]. The paradigm consisted of a slow event-related design comprising 40 short sentences in French (native language) and Japanese (a foreign language that none of the children understood), randomly presented every 12 s. Each sentence was repeated once in a row to study repetition effect [27] and the sentences were produced by different native woman speakers. The mean sentence duration was similar in both languages (2707 ms vs 2724 ms). The paradigm comprised a single run which consisted of 4 different trials in each language, a trial being composed of two consecutive instances of the same sentence. Hence, the total duration of the run was 3 min 12 s during which 16 sentences were delivered to the child. In this period of time, $N = 80$

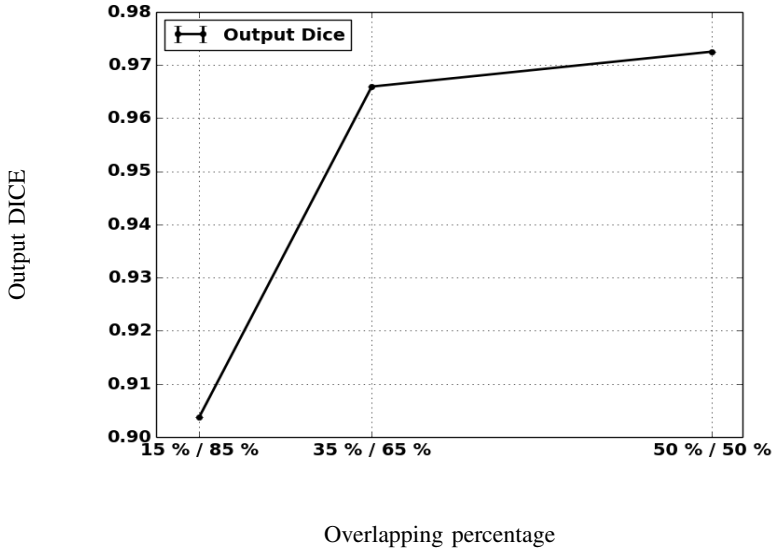

Fig. 10. Output DICE with respect to the overlapping percentage of activations by the two parcels of a 2-parcel model. Used parcellation masks are those in Fig. 9.

volumes were acquired. MRI structural (TR = 2300 ms, TE = 4.18 ms, matrix 256 x 256 x 176, voxel size = 1x1x1 mm) and functional (TR = 2400 ms, TE = 30 ms, matrix, 64 x 64 x 40, voxel size =3x3x3mm) whole brain data were acquired on a 3.0 Tesla Siemens Tim Trio scanner.

We compared the hemodynamic results we obtained using both the JDE and JPDE models in order to investigate the improvements brought by the parcellation inference. Our comparison is based on two key properties: (i) the brain regions known to be involved in language processing and (ii) the known specificity of activation dynamics. As regards (i), the regions of interest typically lie along the superior temporal sulcus (STS) as illustrated in Fig. 11, where they are superimposed on an anatomical slice of the six-year old child. Six regions are emphasized, based on anatomical criteria:

- A central region located in the Heschl's gyrus, and termed `Heschl`;
- A region located in the middle part of the STS, corresponding to the primary auditory cortex, termed middle STS and denoted `mSTS`;
- Two regions in the posterior part of the STS, the most posterior being denoted `pSTS` and the other one `intermediate pSTS`;
- Two regions in the anterior part of the STS, the most anterior one being denoted `aSTS` and the other one `intermediate aSTS`.

As regards (ii), it has been shown in the literature [27] that infants present a gradient in the dynamics

of activations along the superior temporal sulcus, meaning that some regions respond earlier than others or that the time-to-peak (TTP) of the corresponding HRF profile is shorter[1]. This gradient has also been replicated in children [28] and adults [29]. In each case, the fastest regions were found around Heschl's gyrus, with a slow-down along the dorso-ventral and rostral-caudal axes. The more anterior and posterior parts of the STS are known to be the slowest ones. Our results will thus be interpreted with respect to this state-of-the-art.

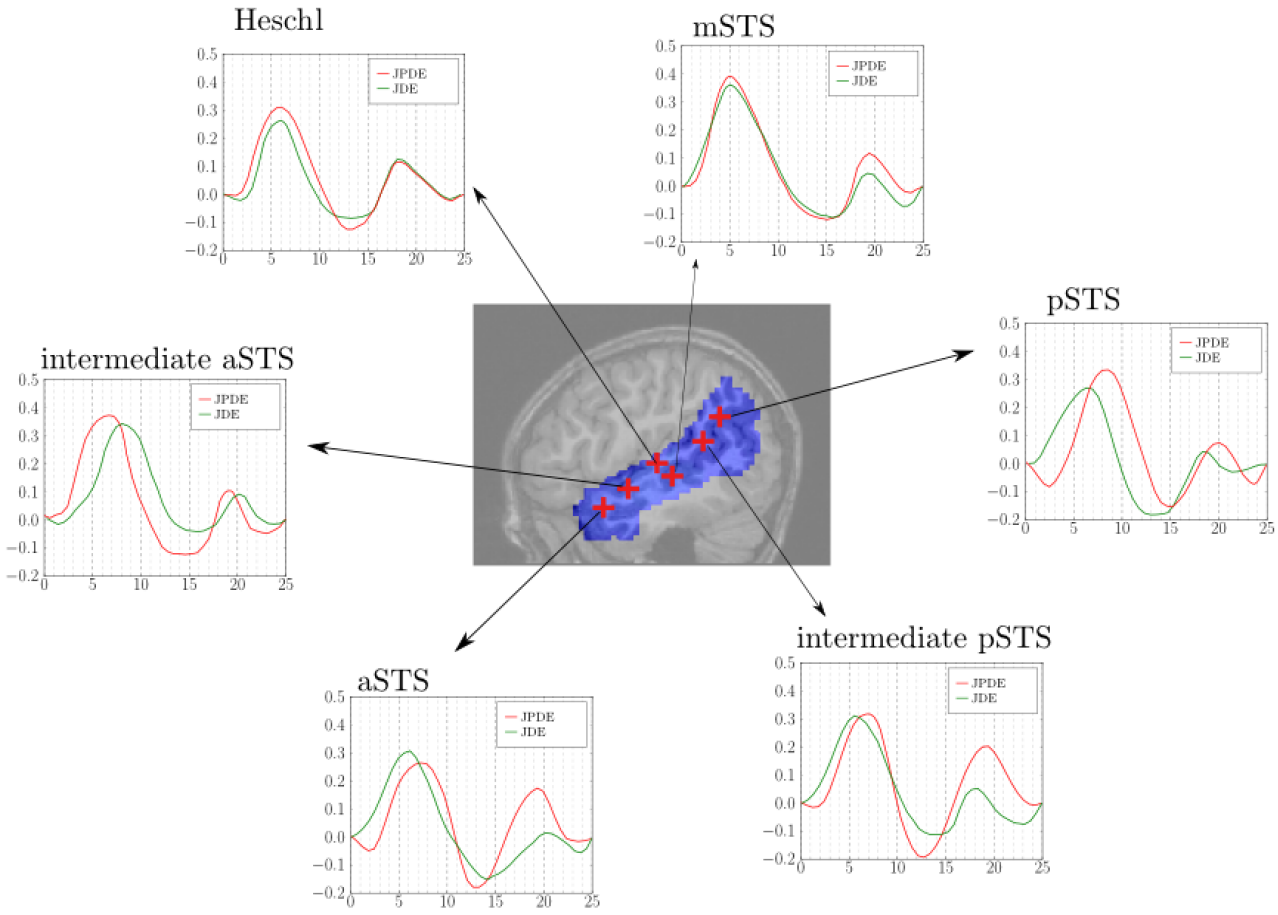

Fig. 11. The HRF profiles estimated in six regions along the Superior Temporal Sulcus using the JPDE (red) and JDE (green) models. In the center, the STS mask is superimposed on an anatomical slice of the six-year old child (sagittal view, x=48). The region centres are indicated by red marks.

For JDE-based inference, we used a parcellation of the STS (see Fig. 12(a)) generated by a spatially

[1]The TTP corresponds to the time to reach the maximum of the HRF curve

constrained hierarchical clustering (Ward) algorithm of functional features (effects maps), the latter being extracted from a classical GLM analysis [14]. The number of parcels was set to 10. After fitting the JDE model to the fMRI data, parcel-specific HRF shapes were computed (see Fig. 11). For JPDE-based inference, the initial parcellation was set up using the same ascendant hierarchical clustering except that some noise was randomly introduced (cf. Fig. 13(a)): the number of parcels K in this initial parcellation was varied from 10 to 20 corresponding to a slight over-segmentation compared to the JDE setup. Our motivation for changing K was to assess the robustness of JPDE for retrieving a coherent parcellation from this noisy initial guess and to see to what extent the corresponding HRF shapes were well recovered (Fig. 11). In what follows, we report the best results we got, namely for $K = 13$ and discuss hereafter how to cope with this issue.

First, it is worth noting that the JPDE model was able to recover a regular parcellation, as shown in Fig. 13(b)). Hence, the parcellation yielded by the JPDE model not only differs from its noisy initialization (Fig. 13 (a)) but also from the fixed parcellation used in JDE (Fig. 12 (a)).

Second, significant differences between the two models can be seen on the HRF profiles, whose discrepancies are reported in Fig. 11. Such differences are actually related to the parcel support changes between the two competing approaches. In Fig. 11, for the sake of clarity, we show HRF profiles for only six parcels along the STS. We observed similar time courses in the center of the STS, namely in Heschl's gyrus and middle STS, whereas we found medium to large shape differences in its anterior and posterior parts. We noticed the presence of a second HRF peak about 18 to 20 s after the first sentence onset. This peaks occurs because of the sentence repetition effect 12 s after the presentation of the first stimulus. In what follows, we will restrict our TTP analysis to the study of the first peak although a specific extension of the JDE formulation has been proposed in the past to deal with repetition suppression effect [30].

Next, to assess the recovery of the temporal gradient along the STS, we compared the parcel-specific TTPs between HRF estimates delivered by the two approaches. Fig. 12(c) and 13(c) report the TTP maps in a given slice (sagittal view) for the JDE and JPDE models, respectively. The colors closer to red (resp. purple) correspond to smaller (resp. larger) TTPs (i.e. to faster and slower responses, respectively). The faster TTP is close to 5 s where the slower one is around 8.5 s. The JPDE model recovers a clear gradient of response, from the middle STS regions (faster responses) to the anterior and posterior parts (slower responses). In the posterior and anterior regions (pSTS and aSTS), the corresponding hemodynamic responses are significantly slower than the one estimated by JDE inference ($\text{TTP}_{\text{JPDE}} = 8.5$ s vs $\text{TTP}_{\text{JDE}} = 6.5$ s in pSTS, $\text{TTP}_{\text{JPDE}} = 7.5$ s vs $\text{TTP}_{\text{JDE}} = 6$ s in aSTS). The same observation is replicated in the intermediate posterior STS region. Furthermore, a slower HRF was in the more anterior

region ($\text{TTP}_{\text{JPDE}} = 8$ s) than in the intermediate aSTS ($\text{TTP}_{\text{JPDE}} = 7$ s). Hence, in these regions, the JPDE model provides more coherent results with the existing literature. In contrast, this gradient assumption is less tenable when analyzing the results delivered by the JDE inference. In particular, JDE-based hemodynamic time courses revealed a TTP decrease down to 6 s in the anterior part of the STS. Still in the same vein, the HRF is slower in the intermediate aSTS ($\text{TTP}_{\text{JDE}} = 8$ s) than in the aSTS ($\text{TTP}_{\text{JDE}} = 6$ s).

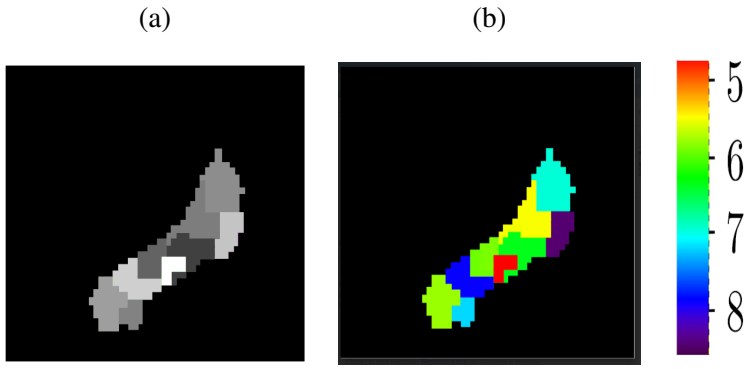

Fig. 12. Results obtained using the JDE model. (a): Parcellation provided as input mask to the JDE algorithm. (b): TTP map indicating the time-to-peak in seconds for each parcel. The color bar encodes the shortest and longest TTPs in red and purple, respectively.

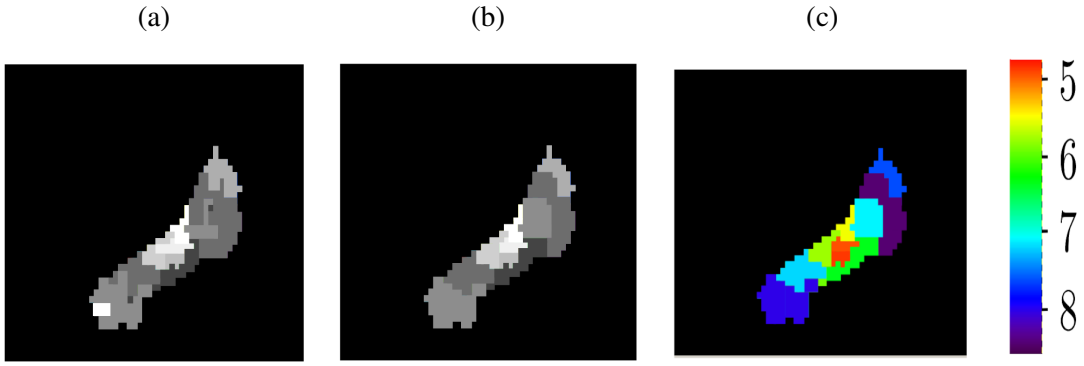

Fig. 13. Results obtained using the JPDE model. (a): Initial parcellation provided as input to the JPDE algorithm. (b): Final parcellation computed by the algorithm. (c): TTP map indicating the time-to-peak in each parcel of the final parcellation.

Globally, for $K = 13$ the JPDE approach recovers a more coherent final parcellation of fMRI data, with corresponding HRF shapes more accurately estimated than the ones provided by the JDE model. Indeed, using a small number of parcels could lead to an "under-parcellation" effect where similar parcels are

aggregated in a single one. On the other hand, if K is too large some parcels can be left empty. Indeed, each voxel is allocated to a given parcel based on the maximum value of the $\widetilde{p}_{z_j}^{(r)}(k)$ for $k = 1 \ldots, K$. If K is too large, some classes will be left empty even if their HRF pattern $\bar{\boldsymbol{h}}_k^{(r)}$ can be estimated. This could be explained by resorting to (36) in Appendix A where it is clear that each voxel has its own contribution in the estimation of all the HRF pattenrs $\bar{\boldsymbol{h}}_k$, but with different proportions depending on the value of $\widetilde{p}_{z_j}^{(r)}(k)$ for $k = 1 \ldots, K$. For this reason, the algorithm can end up with a number of non-empty parcels smaller than K.

This raises the issue of automatically selecting the number of parcels, which can be considered in an extension of the JPDE model in future work.

V. Conclusion

We proposed a JPDE framework that provides an automatic parcellation of the brain into homogeneous hemodynamic territories directly from the fMRI data. The quality and reliability of such a parcellation is at the core of robust neural activity detection and brain hemodynamics estimation. By enabling a fully adaptive data-dependent identification of the parcels, the JPDE framework greatly extends the possibilities of detection-estimation approaches by allowing interaction between detection and estimation, but also the parcellation tasks. The gain in removing the commitment to a priori fixed territories has been confirmed through experiments that showed that JPDE achieved better results than its JDE ancestor both on synthetic and real data. Our real fMRI data experiments also showed that the proposed JPDE framework does not only perform automatic parcellation, but also adds more regularity and accuracy in the estimation of the hemodynamic profiles related to each parcel. However, the initial parcellation (in particular the number of parcels) impacts the estimation of the final parcellation and the corresponding HRF curves. Thus, an important remaining question raised by this new framework is related to the issue of choosing the right number of HRF groups at best *i.e.* in a sparse manner so as to capture the spatial variability in hemodynamic territories while enabling the reproducibility of parcel identification across fMRI datasets. Another perspective to the present work would be to include anatomical priors in the model that could be more sophisticated than the Potts one in order to mix functional and anatomical information in the parcellation task.

APPENDIX

A. VM step:

This appendix is devoted to the M-step of our VEM procedure. Specifically, this step involves different sub M-steps related to $(\boldsymbol{\mu}, \boldsymbol{v})$, $(\mathbf{L}, \boldsymbol{\Gamma})$, $\boldsymbol{\beta}$, β_z, σ_h^2 and $(\bar{\boldsymbol{h}}, \boldsymbol{\nu})$, respectively. The different sub-M steps are summarized below.

1) VM-$(\boldsymbol{\mu}, \boldsymbol{v})$ step:

By maximizing with respect to $(\boldsymbol{\mu}, \boldsymbol{v})$, Eq. (20) reads:

$$(\boldsymbol{\mu}^{(r)}, \boldsymbol{v}^{(r)}) = \underset{(\boldsymbol{\mu}, \boldsymbol{v})}{\arg\max}\ \mathrm{E}_{\widetilde{p}_A^{(r)} \widetilde{p}_Q^{(r)}} \big[\log p(\boldsymbol{A} \,|\, \boldsymbol{Q}; \boldsymbol{\mu}, \boldsymbol{v})\big]. \tag{21}$$

Estimates of $(\boldsymbol{\mu}, \boldsymbol{v})$ can be derived following the same procedure as in [13].

2) VM-$(\boldsymbol{L}, \boldsymbol{\Gamma})$ step:

For this maximization we need to compute:

$$(\boldsymbol{L}^{(r)}, \boldsymbol{\Gamma}^{(r)}) = \underset{(\boldsymbol{L}, \boldsymbol{\Gamma})}{\arg\max}\ \mathrm{E}_{\widetilde{p}_A^{(r)} \widetilde{p}_H^{(r)}} \big[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}, \boldsymbol{\Gamma})\big]. \tag{22}$$

By factorizing over voxels j, estimates of $\boldsymbol{L}$ and $\boldsymbol{\Gamma}$ can be derived following the same procedure as in [13].

3) VM-$\boldsymbol{\beta}$ step:

By maximizing with respect to $\boldsymbol{\beta}$, Eq. (20) reads:

$$\boldsymbol{\beta}^{(r)} = \underset{\boldsymbol{\beta}}{\arg\max}\ \mathrm{E}_{\widetilde{p}_Q^{(r)}} \big[\log p(\boldsymbol{Q}; \boldsymbol{\beta})\big]. \tag{23}$$

Using an additional exponential prior (with parameter λ_{β_m}) akin to [13] in order to avoid over-estimation of this parameter and hence having extra spatial regularization, this M-step can be reformulated as, for each $m = 1 : M$,

$$\beta_m^{(r)} = \underset{\beta_m}{\arg\max}\, E_{\widetilde{p}_{q^m}^{(r)}} [\log p(\boldsymbol{q}^m; \beta_m)] + \log p(\beta_m; \lambda_{\beta_m}). \tag{24}$$

The optimal value of $\boldsymbol{\beta}^{(r)}$ can be reached following the same iterative procedure as in [13]. However, we provide here the specific implementation used in this paper. This step does not admit an explicit closed-form expression but can be solved numerically using gradient ascent schemes. It is straightforward to show that the maximization of (24) in β_m admits a unique solution. Indeed, it is equivalent to solve,

$$\beta_m^{(r)} = \underset{\beta_m}{\arg\max}\, \beta_m (E_{\widetilde{p}_{q^m}^{(r)}} [U(\boldsymbol{q}^m)] - \lambda_{\beta_m}) - \log \mathcal{K}(\beta_m) \tag{25}$$

where $\mathcal{K}$ denotes the normalizing constant that depends on β_m. The first and second derivatives with regards to β_m can be easily computed as respectively:

$$\nabla_{\beta_m} = E_{\widetilde{p}^{(r)}_{q^m}}[U(\boldsymbol{q}^m)] - E_{p(q^m;\beta_m)}[U(\boldsymbol{q}^m)] - \lambda_{\beta_m} \tag{26}$$

$$\nabla^2_{\beta_m} = -var_{p(q^m;\beta_m)}[U(\boldsymbol{q}^m)] \leq 0 \tag{27}$$

where $p(q^m;\beta_m)$ is the Potts model defined in (3) and $var_{p(q^m;\beta_m)}$ denotes the variance w.r.t. $p(q^m;\beta_m)$. The function to optimize is thus concave. Unfortunately, due to the intractable normalizing constant $\mathcal{K}$, the two expressions above are not directly available. It is then necessary to approximate the terms involving the true MRF prior $p(\boldsymbol{q}^m;\beta_m)$. A simple solution is to use a Mean Field like approximation as presented in [31] in which we remove the spatial interactions between neighboring voxels by replacing the stochastic neighbors by fixed values:

$$p^{MF}_{q^m}(\boldsymbol{q}^m;\beta_m) = \prod_{j\in\boldsymbol{\mathcal{V}}} p^{MF}_{q^m_j}(q^m_j;\beta_m) \tag{28}$$

with $p^{MF}_{q^m_j}(q^m_j;\beta_m)$ defined by:

$$p^{MF}_{q^m_j}(q^m_j = i;\beta_m) = \frac{\exp(\beta_m \sum\limits_{j'\in\boldsymbol{\mathcal{V}},j'\sim j} \widetilde{p}^{(r)}_{q^m_{j'}}(i))}{\sum\limits_{l\in\{0,1\}} \exp(\beta_m \sum\limits_{j'\in\boldsymbol{\mathcal{V}},j'\sim j} \widetilde{p}^{(r)}_{q^m_{j'}}(l))} \tag{29}$$

It follows,

$$\begin{aligned}
\nabla_{\beta_m} &\approx \sum_{(j,j')\in\boldsymbol{\mathcal{V}}^2,j\sim j'} \sum_{i\in\{0,1\}} \left(\widetilde{p}^{(r)}_{q^m_j}(i)\widetilde{p}^{(r)}_{q^m_{j'}}(i) - p^{MF}_{q^m_j}(i;\beta_m)p^{MF}_{q^m_{j'}}(i;\beta_m)\right) - \lambda_{\beta_m} \qquad (30)\\
&\approx \frac{1}{2}\sum_{i\in\{0,1\}}\sum_{j\in\boldsymbol{\mathcal{V}}} \left(\widetilde{p}^{(r)}_{q^m_j}(i)\left(\sum_{j'\in\boldsymbol{\mathcal{V}},j'\sim j}\widetilde{p}^{(r)}_{q^m_{j'}}(i)\right) - p^{MF}_{q^m_j}(i;\beta_m)\left(\sum_{j'\in\boldsymbol{\mathcal{V}},j'\sim j} p^{MF}_{q^m_{j'}}(i;\beta_m)\right)\right) - \lambda_{\beta_m},
\end{aligned}$$

where we emphasize that the dependence in β^m is in the second term, the other terms being constant. In both $\widetilde{p}^{(r)}_{q^m_j}(i)$ and $p^{MF}_{q^m_j}(i;\beta_m)$, the Markov part is treated similarly using the *best* currently available posterior variational approximation $\widetilde{p}^{(r-1)}_{q^m}$ or $\widetilde{p}^{(r)}_{q^m}$. The main difference between $\widetilde{p}^{(r)}_{q^m_j}(i)$ and $p^{MF}_{q^m_j}(i;\beta_m)$ is that $p^{MF}_{q^m_j}(i;\beta_m)$ does not involve directly terms that depend on the observed data $\boldsymbol{Y}$ which is consistent with the fact that it represents an approximation of the prior MRF.

4) VM-β_z step:

By maximizing with respect to β_z, Eq. (20) reads:

$$\beta^{(r)}_z = \arg\max_{\beta_z} \ \mathrm{E}_{\widetilde{p}^{(r)}_Z}\left[\log p(\boldsymbol{Z};\beta_z)\right]. \tag{31}$$

If one wants to penalize high values of β_z to avoid over spatial regularization, an exponential prior can also be used as for the M-$\boldsymbol{\beta}$ step. The above optimization procedure can therefore be reformulated as

$$\beta_z^{(r)} = \arg\max_{\beta_z} E_{\widetilde{p}_Z^{(r)}}[\log p(\boldsymbol{Z};\beta_z)] + \log p(\beta_z;\lambda_z) \tag{32}$$

where λ_z is the parameter of the exponential distribution. The same iterative procedure as for the M-$\boldsymbol{\beta}$ step can be used to reach the optimal value of β_z making use of the possibility to approximate the gradient as

$$\nabla_{\beta_z} \approx \sum_{(j,j')\in\boldsymbol{\mathcal{V}}^2, j\sim j'} \sum_{k=1:K} \left(\widetilde{p}_{z_j}^{(r)}(k)\widetilde{p}_{z_{j'}}^{(r)}(k) - p_{z_j}^{MF}(k;\beta_z)p_{z_{j'}}^{MF}(k;\beta_z)\right) - \lambda_{\beta_z},$$

with

$$p_{z_j}^{MF}(z_j = k;\beta_z) = \frac{\exp(\beta_z \sum_{j'\in\boldsymbol{\mathcal{V}}, j'\sim j} \widetilde{p}_{z_{j'}}^{(r)}(k))}{\sum_{l=1:K} \exp(\beta_z \sum_{j'\in\boldsymbol{\mathcal{V}}, j'\sim j} \widetilde{p}_{z_{j'}}^{(r)}(l))} \tag{33}$$

5) VM-$(\sigma_h^2, \bar{\boldsymbol{h}}, \boldsymbol{\nu})$ step:

This maximization step can be reached by solving the following problem

$$\left(\sigma_h^{2(r)}, \bar{\boldsymbol{h}}^{(r)}, \boldsymbol{\nu}^{(r)}\right) = \arg\max_{\sigma_h^2, \bar{\boldsymbol{h}}, \boldsymbol{\nu}} \mathrm{E}_{\widetilde{p}_H^{(r)}\widetilde{p}_Z^{(r)}}\left[\log p(\boldsymbol{H} \mid \boldsymbol{Z}; \bar{\boldsymbol{h}}, \boldsymbol{\nu})\right] + \log p(\bar{\boldsymbol{h}}; \sigma_h^2). \tag{34}$$

Derivations with respect to the $\bar{\boldsymbol{h}}_k$'s and $\boldsymbol{\nu}_k$'s provide the following equations for a given k

$$\boldsymbol{\nu}_k = \frac{\sum_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)\left(\mathrm{trace}(\boldsymbol{\Sigma}_{h_j}^{(r)}) + (\boldsymbol{m}_{h_j}^{(r)} - \bar{\boldsymbol{h}}_k)^{\mathrm{t}}(\boldsymbol{m}_{h_j} - \bar{\boldsymbol{h}}_k)\right)}{(D+1)\sum_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)} \tag{35}$$

and

$$\bar{\boldsymbol{h}}_k = \left(I_{D+1} + \frac{\boldsymbol{\nu}_k \boldsymbol{R}^{-1}/\sigma_h^2}{\sum_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)}\right)^{-1} \frac{\sum_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)\boldsymbol{m}_{h_j}^{(r)}}{\sum_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)} \tag{36}$$

Then updated values $\boldsymbol{\nu}_k^{(r)}$ and $\bar{\boldsymbol{h}}_k^{(r)}$ can be obtained by solving equations (35) and (36). An estimation of σ_h^2 follows as

$$\sigma_h^{2(r)} = \frac{\sum_{k=1}^K \bar{\boldsymbol{h}}_k^{(r)t} \boldsymbol{R}^{-1} \bar{\boldsymbol{h}}_k^{(r)}}{K} \tag{37}$$

However, as σ_h^2 plays the role of a smoothing parameter for the HRF shapes, it will be held fixed in the present work.

Note that in the more general case of non isotropic $\bar{\boldsymbol{\Sigma}}_k$, we would get

$$\bar{\boldsymbol{\Sigma}}_k = \frac{\sum\limits_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)\left(\boldsymbol{\Sigma}_{h_j}^{(r)} + (\boldsymbol{m}_{h_j}^{(r)} - \bar{\boldsymbol{h}}_k)(\boldsymbol{m}_{h_j}^{(r)} - \bar{\boldsymbol{h}}_k)^{\mathrm{t}}\right)}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)} \tag{38}$$

and

$$\bar{\boldsymbol{h}}_k = \left(I_{D+1} + \frac{\bar{\boldsymbol{\Sigma}}_k \boldsymbol{R}^{-1}/\sigma_h^2}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)}\right)^{-1} \frac{\sum\limits_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)\boldsymbol{m}_{h_j}}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(k)}. \tag{39}$$

The first case (35) can be obtained from the general one by setting $\boldsymbol{\nu}_k = \mathrm{trace}(\bar{\boldsymbol{\Sigma}}_k)/(D+1)$.

B. Pseudo-random generation of parcellation masks

This appendix details the technique used to pseudo-randomly generate parcellation masks approximately sharing the same DICE level. Starting from a reference initial parcellation mask such as the one in Fig. 7 (top), a 2×2 neighbourhood is selected for each voxel j. Over these neighbouring voxels, the histogram over the K classes is calculated. After normalization, we end up with a probability vector $Pr \in \mathbb{R}^K$ that sums to one. Each element $Pr(k)$ indicates the probability that the current voxel belongs to class k. For the target pseudo-random mask (Mask*), the value of the current voxel Mask$^*(j)$ is randomly drawn from a multinomial distribution using the probability vector Pr. When $K = 2$, this simply reduces to a random generation from a binomial distribution. This mask generation technique ensures the spatial coherence of the parcels code between different realisations. The random effect will mainly impact voxels at the parcel borders.

REFERENCES

[1] D.A. Handwerker, J.M. Ollinger, and D. Mark, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *Neuroim.*, vol. 21, pp. 1639–1651, 2004.

[2] J. Kershaw, B.A. Ardekani, and I. Kanno, "Application of Bayesian inference to fMRI data analysis," *IEEE Trans. Med. Imag.*, vol. 18, no. 12, pp. 1138–1152, 1999.

[3] G.M. Boynton, S.A. Engel, G.H. Glover, and D. J. Heeger, "Linear systems analysis of functional magnetic resonance imaging in human V1," *J. Neurosci.*, vol. 16, pp. 4207–4221, 1996.

[4] P. Ciuciu, J.B. Poline, G. Marrelec, J. Idier, Ch. Pallier, and H. Benali, "Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1235–1251, Oct 2003.

[5] T.D. Wager, A. Vazquez, L. Hernandez, and D.C Noll, "Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies.," *Neuroim.*, vol. 25, no. 1, pp. 206–18, Mar. 2005.

[6] G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI," *Neuroim.*, vol. 9, pp. 416–429, 1999.

[7] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith, "Fully Bayesian spatio-temporal modelling of fMRI data," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 213–231, Feb. 2004.

[8] C. Goutte, F. Nielsen, and L. K. Hansen, "Modeling the haemodynamic response in fMRI using smooth FIR filters," *IEEE Trans. Med. Imag.*, vol. 19, no. 12, pp. 1188–1201, 2000.

[9] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu, "A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI," *Neuroim.*, vol. 41, no. 3, pp. 941–969, 2008.

[10] T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of within-subject fMRI time series," *IEEE Trans. Med. Imag.*, vol. 29, pp. 1059–1074, 2010.

[11] J. Wang, H. Zhu, J. Fan, K. Giovanello, and W. Lin, "Adaptively and spatially estimating hemodynamic response functions in fMRI," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, Gabor Fichtinger, Anne Martel, and Terry Peters, Eds., vol. 6892 of *Lecture Notes in Computer Science*, pp. 269–276. Springer Berlin Heidelberg, Toronto, Canada, Sep. 2011.

[12] L. Chaari, F. Forbes, T. Vincent, M. Dojat, and P. Ciuciu, "Variational solution to the joint detection estimation of brain activity in fMRI," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, Gabor Fichtinger, Anne Martel, and Terry Peters, Eds., vol. 6892 of *Lecture Notes in Computer Science*, pp. 260–268. Springer Berlin Heidelberg, Sep. 2011.

[13] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach," *IEEE Trans. Med. Imag.*, vol. 32, no. 5, pp. 821–837, May 2013.

[14] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline, "Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets," *Hum. Brain Mapp.*, vol. 27, no. 8, pp. 678–693, Aug. 2006.

[15] G. Flandin, F. Kherif, X. Pennec, D. Rivière, N. Ayache, and J.-B. Poline, "A new representation of fMRI data using anatomo-functional constraints," in *Proc. 8th HBM*, Sendai, Japan, June 2002.

[16] T. Vincent, P. Ciuciu, and B. Thirion, "Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI," in *5th IEEE Int. Symp. on Biomed. Imag. (ISBI)*, Paris, France, May 2008, pp. 568–571.

[17] Y. Ji, P.-Y. Hervé, U. Aickelin, and A. Pitiot, "Parcellation of fMRI datasets with ICA and PLS–a data driven approach.," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, G.-Z. Yang, D. J. Hawkes, D. Rueckert, J. A. Noble, and C.-J. Taylor, Eds., vol. 5761 of *Lecture Notes in Computer Science*, pp. 984–991. Springer Berlin Heidelberg, Sep. 2009.

[18] S. Badillo, G. Varoquaux, and P. Ciuciu, "Hemodynamic estimation based on consensus clustering," in *PRNI 2013 – 3rd International Workshop on Pattern Recognition in NeuroImaging*, Philadelphia, United States, June 2013.

[19] A. Frau, T. Vincent, F. Forbes, and P. Ciuciu, "Hemodynamically informed parcellation of cerebral fMRI data," in *IEEE International Conference on Acoustics, Speech, and Signal (ICASSP)*, Florence, Italy, May 26 - 31 2014, pp. 2079 – 2083.

[20] M. Woolrich, B. Ripley, M. Brady, and S. Smith, "Temporal autocorrelation in univariate linear modelling of fMRI data," *Neuroimage*, vol. 14, no. 6, pp. 1370–1386, Dec. 2001.

[21] W.D. Penny, G. Flandin, and N. Trujillo-Bareto, "Bayesian comparison of spatially regularised general linear models," *Hum. Brain Mapp.*, vol. 28, no. 4, pp. 275–293, Apr. 2007.

[22] W. D. Penny, N. Trujillo-Barreto, and K. J. Friston, "Bayesian fMRI time series analysis with spatial priors," *Neuroimage*, vol. 23, no. 2, pp. 350–362, 2005.

[23] L. Chaari, F. Forbes, T. Vincent, and P. Ciuciu, "Hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*,

Nicholas Ayache, Hervé Delingette, Polina Golland, and Kensaku Mori, Eds., vol. 7512 of *Lecture Notes in Computer Science*, pp. 180–188. Springer Berlin Heidelberg, 2012.

[24] R.M. Neal and G.E. Hinton, "A view of the EM algorithm that justifies incremental, sparse and other variants," in *Lear. in Graph. Mod.*, Jordan, Ed., pp. 355–368. 1998.

[25] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content," *Kongelige Danske Videnskabernes Selskab. Biologiske Skrifter*, vol. 4, pp. 1–34, 1948.

[26] K. Monzalvo, J. Fluss, C. Billard, S. Dehaene, and G. Dehaene-Lambertz, "Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status," *Neuroim.*, vol. 61, no. 1, pp. 258–274, February 2012.

[27] G. Dehaene-Lambertz, L. Hertz-Pannier, J. Dubois, S. Mériaux, A. Roche, and M. Sigman, "Functional organization of perisylvian activation during presentation of sentences in preverbal infants," *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 14240–14245, 2006.

[28] J. Brauer, J. Neumann, and A. D. Friederici, "Temporal dynamics of perisylvian activation during language processing in children and adults," *Neuroim.*, vol. 41, no. 4, pp. 1484–1492, 2009.

[29] G. Dehaene-Lambertz, L. Hertz-Pannier, and J. Dubois, "Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants," *Trends in Neurosci.*, vol. 29, pp. 367–373, 2006.

[30] P. Ciuciu, S. Sockeel, T. Vincent, and J. Idier, "Modelling the neurovascular habituation effect on fMRI time series," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 433–436.

[31] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Patt. Rec.*, vol. 36, pp. 131–144, 2003.