Hemodynamic Brain Parcellation Using A Non-Parametric Bayesian Approach

Mohanad Albughdadi, Lotfi Chaari, Jean-Yves Tourneret, Florence Forbes, Philippe Ciuciu

To cite this version:

Mohanad Albughdadi, Lotfi Chaari, Jean-Yves Tourneret, Florence Forbes, Philippe Ciuciu. Hemodynamic Brain Parcellation Using A Non-Parametric Bayesian Approach. 2016. hal-01275622

HAL Id: hal-01275622

https://inria.hal.science/hal-01275622

Preprint submitted on 17 Feb 2016

Hemodynamic Brain Parcellation Using A Non-Parametric Bayesian Approach

M. Albughdadi[a,*], L. Chaari[a], J.-Y. Tourneret[a], F. Forbes[b], P. Ciuciu[c]

[a]*University of Toulouse, IRIT, INP-ENSEEIHT, France*
[b]*INRIA, MISTIS, Grenoble University, LJK, Grenoble, France*
[c]*CEA/NeuroSpin and INRIA Saclay, Parietal, France*

Abstract

One of the most challenging issues in task-related fMRI data analysis consists of deriving a meaningful functional brain parcellation. The joint parcellation detection estimation (JPDE) model addresses this issue through an automatic inference of the parcels directly from fMRI data. However, for doing so, the number of parcels needs to be fixed a priori and an appropriate initialization for the mask parcellation must be provided too. Hence, this difficult task generally depends on the subject. In this paper, an automatic model selection approach is proposed to overcome this limitation at the subject-level. Our approach relies on a non-parametric Bayesian approach that estimates the number of parcels online using a Dirichlet process mixture model combined with a hidden Markov random field. The inference is carried out using a variational expectation maximization strategy. As compared to a standard model selection approach in the original JPDE framework, our non-parametric extension appears more efficient in terms of computational time and does not require finely tuned initialization. Our method is first validated on synthetic data to demonstrate its robustness in selecting the right model order and providing accurate estimates for the parcellation, the hemodynamic response function (HRF) shapes and the activation maps. The method is then validated on real fMRI data in two regions of interest (ROIs): right motor and bilateral occipital ROIs. The results show the ability of the proposed method to aggregate parcels with similar behaviour from a hemodynamic point of view, while discriminating them from other parcels having

---

[*]Corresponding author
Email address: `mohanad.albughdadi@enseeiht.fr.` (M. Albughdadi)

different hemodynamic properties. The HRF estimates of the different hemodynamic territories obtained with our approach are close the the canonical HRF shape in both the right motor and the bilateral occipital cortices. The discrimination power of the proposed approach is increased compared to its ancestors where the results on real data show its ability to discriminate HRF profiles with different Full Width at Half Maximum (FWHM). The robust performance of detecting the elicited task-related activity is confirmed by comparing the neural response level estimates obtained using our approach with those obtained using the joint detection estimation (JDE) model.

Keywords: fMRI, hemodynamic parcellation, VEM, Dirichlet process mixture model, Non-parametric Bayesian, HMRF

---

1. Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive imaging technique that indirectly measures neural activity from the blood-oxygen-level dependent (BOLD) signal (Ogawa et al., 1990). This signal reflects the variations in the blood oxygenation level induced by oxygen consumption of neural population involved during task performance. Task-related fMRI data analysis generally focuses on two main issues: *i)* detecting the activated brain areas in response to a given stimulus, and *ii)* estimating the underlying dynamics associated with such an activation through the estimation of the so called hemodynamic response function (HRF).

So far, many approaches have been proposed to characterize the link between stimuli and the induced BOLD signal through the brain, the simplest relying on a general linear model (GLM) where the link between the stimulus onset and the BOLD effect is actually modelled through a convolution between the HRF and a binary stimulus sequence. The GLM has been primarily used for detecting task-related brain activity in a massive univariate manner (Friston et al., 1995), considering a constant and fixed canonical HRF shape (Boynton et al., 1996). Then, it has been progressively extended to account for the HRF variability using more regressors and hence more flexible design matrices (Glover, 1999; Friston et al., 2000; Henson et al., 2001; Lindquist et al., 2009). Nonetheless, due to the increase of regressors the main difficulty that comes up in this context is the decrease of statistical sensitivity in the subsequent tests, making the detection task less reliable. Besides, other approaches that rely on physiologically-informed non-linear models (e.g., the Balloon model) have been pushed forward for recovering hemodynamics but most often they are deployed in brain regions where evoked activity has already been detected (Buxton and Frank, 1997; Friston et al., 2000; Riera et al., 2004; Deneux and Faugeras, 2006). Their computational cost is actually prohibitive for whole brain analysis and some identifiability issues (different pairs of state variables and parameters give the same goodness-of-fit) arise because of the presence of noise.

The above mentioned approaches mainly address detection of evoked activity and HRF recovery as a two-step procedure whereas both tasks are strongly linked. A precise localization of activations depends on a reliable HRF estimate, while a robust HRF shape is only achievable in brain regions eliciting task-related activity (Kershaw et al., 1999; Ciuciu et al., 2003). Moreover, most of linear and non-linear models are designed for univariate inference whereas it is known that the BOLD signal is spatially smooth and thus the HRF shapes remain similar over a certain spatial distance (Handwerker et al., 2004; Ciuciu et al., 2004; Handwerker et al., 2012; Badillo et al., 2013b).

One of the approaches that accounts for this interdependence is the joint detection-estimation (JDE) framework, where both tasks are performed jointly (Makni et al., 2008; Vincent et al., 2010; Chaari et al., 2013). To improve robustness in the estimation task and account for spatial correlation of the BOLD signal, a single HRF shape model was assumed for a specific group of voxels, also referred to as a *parcel*. Within this JDE formalism, two approaches for posterior inference have been developed, the first one relying on computationally intensive stochastic sampling (Makni et al., 2008; Vincent et al., 2010) and the second one based on the variational expectation maximization (VEM) algorithm (Chaari et al., 2013) to achieve numerical convergence at lower cost.

However, whatever the numerical algorithm deployed, the JDE formalism requires a prior parcellation of the brain into functionally homogeneous regions. These parcels should achieve a fair compromise between *homogeneity* and *reliability* (Thirion et al., 2014). Homogeneity means that the parcels should be small enough to meet the assumption of HRF shape invariance within each parcel, whereas reliability should guarantee that parcels are large enough to ensure reliable HRF estimation and detection performance. This issue has motivated a number of recent developments that try to cope with the identification of relevant brain parcellation of the brain (Flandin et al., 2002; Thirion et al., 2006; Golland et al., 2007; Lashkari et al., 2010, 2012;

Eickhoff et al., 2011). In (Lashkari et al., 2012), a non-parametric Bayesian approach, relying on a Dirichlet process mixture model, is considered for the activation classes in a multi-subject framework but they assume that the HRF is fixed for a given region of interest. However, among the latter works, none tries to uncover functional regions that appear homogeneous with respect to their hemodynamic profile. This issue has been specifically addressed in (Badillo et al., 2013a) using either random parcellation and consensus clustering, or multivariate Gaussian probabilistic modelling (Fouque et al., 2009) or joint parcellation within the JDE framework (Chaari et al., 2012, 2015; Frau-Pascual et al., 2014), giving rise to the joint parcellation detection estimation (JPDE) appraoch. This strategy performs online parcellation during the detection estimation step through the selection of hemodynamic territories, *i.e*, sets of voxels that share the same HRF pattern.

Although automated inference of parcellation is performed in the JPDE methodology, the algorithm still requires the manual setting of the number of parcels. In a previous work (Albughdadi et al., 2014), we have proposed to finely tune this parameter using an off-line model selection strategy. This procedure was based on the computation of the free energy associated with models of increasing complexity, (*i.e*, with an increasing number of parcels) in the VEM framework. The best model was then selected as the one maximizing the free energy. This technique was however of limited interest since it requires to run the JPDE algorithm for many candidate models, which is quite time-consuming especially when no prior information is available on the approximate number of parcels.

This paper proposes a more original technique to perform on-line model selection by adopting a non-parametric Bayesian (NPB) model. A Dirichlet process (DP) prior combined with a hidden Markov random field is specifically used to estimate the number of parcels from the data itself without any prior knowledge on the initial parcellation. Injected within the JPDE formulation, we end up with an algorithm that needs to be run only once for getting an estimate of the number of parcels and the corresponding HRF territories, with their own hemodynamic signature and evoked responses.

The rest of the paper is organized as follows. The JPDE model is summarized in Section 2. The proposed BNP model selection algorithm is introduced in Section 3. Experimental validation on synthetic and real fMRI data is presented in Section 4. Finally, discussions and conclusions are drawn in Section 5.

2. Joint parcellation detection estimation model

2.1. Notation

In this paper, a vector is by convention a column vector. The transpose is denoted by ^{t}. Matrices and vectors are denoted with bold capital and lowercase letters (e.g., $\boldsymbol{X}$ and $\boldsymbol{z}$). We use letters j, m as indexes that run over voxels and experimental conditions, respectively. Tab. 1 details the variables and parameters defining the JPDE model while other acronyms used in this paper are listed in Tab. 2.

2.2. Observed and missing variables

In this paper, we consider the JPDE model proposed in (Chaari et al., 2012, 2015). This model is the extension of the parcel-based JDE model developed in (Makni et al., 2008; Vincent et al., 2010) to a whole-brain or a large brain area. For the ease of exposure, the JPDE model is summarized in what follows. Let $\boldsymbol{\mathcal{P}}$ be the set of voxels of interest within the brain mask or the mask of the region of interest (ROI) under study. 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 number of different stimulus types or experimental conditions is M. At each voxel j, a voxel dependent HRF $\boldsymbol{h}_j \in \mathbb{R}^D$ is assumed with D is the HRF size and $\boldsymbol{H} = \{\boldsymbol{h}_j, j \in \boldsymbol{\mathcal{P}}\}$ the set of all possible HRF shapes. Each $\boldsymbol{h}_j$ is associated with an HRF group among K possible groups. These groups are specified by a set of hidden labels $\boldsymbol{z} = \{z_j, j \in \boldsymbol{\mathcal{P}}\}$ where $z_j \in \{1, ..., K\}$ and $z_j = k$ means that the voxel j belongs to the k-th group. An estimation of $\boldsymbol{z}$ corresponds to a partition of the domain into K hemodynamic territories whose connected components define a parcellation of the brain or of the considered region of interest (ROI). The link to the observed BOLD data reads

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

where $\boldsymbol{X}_m = \{x_m^{n-d\Delta t}, n = 1, \ldots, N, d = 0, \ldots, D-1\}$ is a known binary matrix which provides information on the stimulus occurrences for the m-th experimental condition, $N \times D$ is the dimension of this matrix,

$\Delta t \leq TR$ being the sampling period of the unknown HRFs. The scalar weights a_j^m are the unknown response amplitudes. They are generally referred to as neural response levels (NRL). Denote as $\boldsymbol{A} = \{\boldsymbol{a}^m, m = 1, \ldots, M\}$

with $\boldsymbol{a}^m = \left\{a_j^m, j \in \boldsymbol{\mathcal{P}}\right\}$, a_j^m being the amplitude at voxel j for the m-th experimental condition. Similarly to the HRF, each NRL is assumed to be in one of I groups specified by activation class assignment variables $\boldsymbol{Q} = \{\boldsymbol{q}^m, m = 1, \ldots, M\}$ where $\boldsymbol{q}^m = \left\{q_j^m, j \in \boldsymbol{\mathcal{P}}\right\}$ and $q_j^m \in \{1, ..., I\}$ represents the activation class at voxel j for the m-th experimental condition. Two classes are considered here ($I = 2$) where $i = 0$ and $i = 1$ refer to non-activated and activated voxels, respectively. Finally, the rest of the signal is made of the vector $\boldsymbol{P}\boldsymbol{\ell}_j$, which corresponds to low frequency drifts where $\boldsymbol{P}$ is an $N \times O$ matrix, $\boldsymbol{\ell}_j \in \mathbb{R}^O$ is a vector to be estimated and $\boldsymbol{L} = \{\boldsymbol{\ell}_j, j \in \boldsymbol{\mathcal{P}}\}$. Regarding the observation noise, the $\boldsymbol{b}_j$'s are assumed to be independent, zero-mean Gaussian vectors with covariance matrix $\boldsymbol{\Gamma}_j^{-1}$, *i.e*, $\boldsymbol{b}_j \sim \mathcal{N}(0, \boldsymbol{\Gamma}_j^{-1})$. The set of all unknown precision matrices is denoted by $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_j, j \in \boldsymbol{\mathcal{P}}\}$.

Table 1: Notation for variables and parameters used in the JPDE model.

	Notation	**Definition**	**Dimension**
Variables	$\boldsymbol{y}_j \in \mathbb{R}^N$	Observed BOLD signal at voxel j	N
	$\boldsymbol{b}_j \in \mathbb{R}^N$	Acquisition noise vector at voxel j	N
	$\boldsymbol{h}_j = \{h_{d\Delta t}, d = 0, \ldots, D-1\} \in \mathbb{R}^D$	HRF sampled at Δ_t	D
	$a_j^m \in \mathbb{R}$	NRL at voxel j for condition m	1
	$\boldsymbol{a}^m = \{a_j^m, j \in \mathcal{P}\} \in \mathbb{R}^J$	NRLs for condition m	J
	$q_j^m \in \{0, 1\}$	Activation class for voxel j and condition m	1
	$\boldsymbol{q}^m = \{q_j^m, j \in \mathcal{P}\} \in \{0, 1\}^J$	Activation classes for condition m	J
	$z_j \in \{0, \ldots, K\}$	HRF group for voxel j	1
Unknown Parameters	$\boldsymbol{\ell}_j \in \mathbb{R}^O$	Low frequency drifts for voxel j	O
	$\boldsymbol{\Gamma}_j \in \mathbb{R}^{N \times N}$	Noise precision matrix for voxel j	$N \times N$
	$\boldsymbol{\mu}_m = \{\mu_{m0}, \mu_{m1}\} \in \mathbb{R}^2$	Mixture model means for NRLs in condition m	2
	$\boldsymbol{v}_m = \{v_{m0}, v_{m1}\} \in \mathbb{R}_+^2$	Mixture model variances for NRLs in condition m	2
	$\beta_m \in \mathbb{R}_+$	Potts regularization parameter for condition m	1
	$\theta_m = \{\boldsymbol{\mu}_m, \boldsymbol{v}_m, \boldsymbol{\beta}_m\}$	Parameters of the condition m	1
	$\beta_z \in \mathbb{R}_+$	Potts regularization parameter for HRF groups	1
	$\bar{\boldsymbol{h}} = \{\bar{\boldsymbol{h}}_k, k = 0, \ldots, K\} \in \mathbb{R}^{D \times K}$	HRF patterns	$D \times K$
	$\bar{\boldsymbol{\Sigma}} = \{\bar{\boldsymbol{\Sigma}}_k, k = 0, \ldots, K\}$	HRF covariance matrices	D
Fixed	$\boldsymbol{X}_m \in \{0, 1\}^{N \times D}$	Binary stimulus occurrence matrix for condition m	$N \times D$
	$\boldsymbol{P} \in \mathbb{R}^{N \times O}$	Low frequency orthonormal function basis	$N \times O$

Table 2: Acronyms used through the paper.

Acronym	Definition
JDE	Joint Detection-Estimation
JPDE	Joint Parcellation Detection-Estimation
NP-JPDE	Non-Parametric Joint Parcellation Detection-Estimation
VEM	Variational Expectation Maximization
MCMC	Markov Chain Monte Carlo
NPB	Non-Parametric Bayesian
DPMM	Dirichlet Process Mixture Model
HMRF	Hidden Markov Random Field
HRF	Hemodynamic Response Function
NRL	Neural Response Level
PV	HRF Peak Value: $\max \{h_{d\Delta t}\}_{d=0:D-1}$
TTP	HRF Time-to-Peak: $\Delta t \times \arg\max_{d} \{h_{d\Delta t}\}_{d=0:D-1}$
FWHM	Full Width at Half Maximum: $\Delta t \times (d_2 - d_1)$ such that $h_{d_1\Delta t} = h_{d_2\Delta t} = \mathrm{PV}/2, d_1 < d_2$
TTU	HRF Time-to-Undershoot: $\Delta t \times \arg\min_{d>d_2} \{h_{d\Delta t}\}_{d=0:D-1}$

2.3. Hierarchical model

Adopting a Bayesian formulation for the above JPDE model, the joint distribution of $\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}$ is defined as follows

$$\begin{aligned} p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}; \boldsymbol{\Theta}) \; &= p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{\Theta})\, p(\boldsymbol{A} \mid \boldsymbol{Q}; \boldsymbol{\Theta})\, p(\boldsymbol{Q}; \boldsymbol{\Theta}) \\ & \quad p(\boldsymbol{H} \mid \boldsymbol{z}; \boldsymbol{\Theta})\, p(\boldsymbol{z}; \boldsymbol{\Theta}) \end{aligned} \tag{2}$$

where $\boldsymbol{\Theta}$ is the set of all parameters which will be defined later. More details about the right-hand side term of Eq. (2) are provided below.

(a) **Likelihood**

To account for serial correlation in fMRI time series, an autoregressive (AR) noise model has been adopted (Makni et al., 2008; Woolrich et al., 2001; Chaari et al., 2011, 2012, 2015). It follows that the covariance matrix reads $\boldsymbol{\Gamma}_j = \sigma_j^{-2}\boldsymbol{\Lambda}_j$ where $\boldsymbol{\Lambda}_j$ is a tridiagonal symmetric matrix whose components depend on the AR(1) parameter ρ_j (Makni et al., 2008): $(\boldsymbol{\Lambda}_j)_{1,1} = (\boldsymbol{\Lambda}_j)_{N,N} = 1$, $(\boldsymbol{\Lambda}_j)_{n,n} = 1 + \rho_j^2$ for $n = 2, \ldots, N-1$ and $(\boldsymbol{\Lambda}_j)_{n+1,n} = (\boldsymbol{\Lambda}_j)_{n,n+1} = -\rho_j$ for $n = 1, \ldots, N-1$. Using the notation $\boldsymbol{\theta}_0 = (\sigma_j^2, \rho_j)_{1 \le j \le J}$ and $\overline{\boldsymbol{y}}_j = \boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j - \boldsymbol{S}_j\boldsymbol{h}_j$ with $\boldsymbol{S}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}_m$, the likelihood factorizes over voxels as follows

$$p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{\theta}_0) \propto \prod_{j=1}^{J} \left[\frac{\sqrt{\det \boldsymbol{\Lambda}_j}}{\sigma_j^N} \right] \exp\left(-\frac{\overline{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Lambda}_j \overline{\boldsymbol{y}}_j}{2\sigma_j^2} \right). \tag{3}$$

(b) **Neural response levels**
The NRLs are assumed to be statistically independent across conditions, *i.e*,

$$p(\boldsymbol{A}; \boldsymbol{\theta}_{\boldsymbol{a}}) = \prod_{m=1}^{M} p(\boldsymbol{a}^m; \boldsymbol{\theta}_m) \tag{4}$$

where $\boldsymbol{\theta}_{\boldsymbol{a}} = \{\boldsymbol{\theta}_m, m = 1, \ldots, M\}$ 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, *i.e*,

$$p(\boldsymbol{a}^m \mid \boldsymbol{q}^m; \boldsymbol{\theta}_m) = \prod_{j \in \boldsymbol{\mathcal{P}}} p(a_j^m \mid q_j^m; \boldsymbol{\mu}_m, \boldsymbol{v}_m) \tag{5}$$

with $p(a_j^m \mid q_j^m = i; \boldsymbol{\theta}_m) \sim \mathcal{N}(\mu_{mi}, v_{mi})$. All the means and variances of the response amplitudes are gathered in the two unknown vectors $\boldsymbol{\mu} = \{\boldsymbol{\mu}_{mi}, m = 1, \ldots, M, i = 0, 1\}$ and $\boldsymbol{v} = \{\boldsymbol{v}_{mi}, m = 1, \ldots, M,\ i = 0, 1\}$, respectively. Note that for non-activating voxels ($i = 0$) we have set $\mu_{m0} = 0$ for all $m = 1, \ldots, M$.

(c) **Activation classes**
As in (Vincent et al., 2010), the M experimental conditions are assumed to be independent a priori regarding the activation class assignments, *i.e*, $p(\boldsymbol{Q}) = \prod_{m=1}^{M} p(\boldsymbol{q}^m; \beta_m)$ with $p(\boldsymbol{q}^m; \beta_m)$ a Markov random field prior, namely a Potts model with interaction parameter β_m,

$$p(\boldsymbol{q}^m; \beta_m) = W(\beta_m)^{-1} \exp\big(\beta_m U(\boldsymbol{q}^m)\big) \quad \text{with} \quad U(\boldsymbol{q}^m) = \sum_{j \sim l} I(q_j^m = q_l^m) \tag{6}$$

where $W(\beta_m)$ is a normalizing constant and I is an indicator function such that $I(a = b) = 1$ if $a = b$ and 0 otherwise. The notation $j \sim l$

means that the sum ranges over all neighboring voxels. Moreover, the neighboring system is a 6-connexity 3D scheme.

(d) **HRF patterns**
A unique HRF shape was considered for the whole parcel in (Makni et al., 2008; Vincent et al., 2010; Chaari et al., 2011). The JPDE model considered here assumes that the HRF is voxel-dependent and that the distribution of $\boldsymbol{h}_j$ is expressed conditionally to the HRF group variable z_j

$$p(\boldsymbol{H}|\boldsymbol{z}) = \prod_{j \in \boldsymbol{\mathcal{P}}} p(\boldsymbol{h}_j \,|\, z_j) \tag{7}$$

with $p(\boldsymbol{h}_j \,|\, z_j = k) \sim \mathcal{N}(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)$ where $\bar{\boldsymbol{h}}_k$ denotes the mean HRF pattern of group $\#k$, while $\bar{\boldsymbol{\Sigma}}_k = \nu_k \boldsymbol{I}_{D+1}$ adjusts the stochastic perturbations around $\bar{\boldsymbol{h}}_k$ via the value of the hyperparameter ν_k. The smoothness of the HRF pattern is ensured by assigning the prior distribution $\bar{\boldsymbol{h}}_k \sim \mathcal{N}(\boldsymbol{0}, \sigma_h^2 \boldsymbol{R})$, 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 parameter to be estimated or fixed. Moreover, $\bar{\boldsymbol{h}}_{k0} = \bar{\boldsymbol{h}}_{kD\Delta t} = 0$ as in (Makni et al., 2008; Vincent et al., 2010; Chaari et al., 2011). Hence, $\bar{\boldsymbol{h}}_k \in \mathbb{R}^{D-1}$.

(e) **HRF groups**
A spatial K-class Potts model with interaction parameter β_z is used to promote parcellation connexity

$$p(\boldsymbol{z}; \beta_z) = W(\beta_z)^{-1} \exp\big(\beta_z U(\boldsymbol{z})\big) \quad \text{with} \quad U(\boldsymbol{z}) = \sum_{j \sim l} I(z_j = z_l) \tag{8}$$

where $W(\beta_z)$ is a normalizing constant. Using this kind of prior forces the neighbouring voxels to belong to the same HRF group and thus to share the same HRF pattern.

The set of all parameters is denoted by $\boldsymbol{\Theta} = \big\{\boldsymbol{\Gamma}, \boldsymbol{L}, \boldsymbol{\theta}_a, \beta_z, \sigma_h^2, (\bar{\boldsymbol{h}}_k, \nu_k)_{1 \le k \le K}\big\}$.

Computing closed form expressions of the Bayesian estimators associated with the posterior distribution $p(\ \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z} \ | \ \boldsymbol{Y}; \boldsymbol{\Theta})$ is difficult and intractable for the JPDE model. A variational expectation maximization (VEM) approach is used to estimate the unknown parameters $\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}$ and the vector $\boldsymbol{\Theta}$ from its posterior distribution. The interested reader can refer to (Chaari et al., 2012, 2015, 2013) for more details about the variational expectation maximization estimation.

3. Non-parametric Bayesian model selection for Hemodynamic Brain Parcellation

One of the remaining issues for fitting the JPDE model to fMRI data is to determine the number of parcels. Even though the algorithm presented in (Chaari et al., 2012, 2015) automatically estimates the parcels from the data in parallel to the joint detection-estimation task, it still requires to manually set the number of parcels which limits its usefulness. Determining the optimal number of parcels K, which yields the best fits to the data is a model selection problem. In this case, the likelihood cannot be used directly as a model score since it does not account for the model complexity. Some alternatives are available in the literature and usually based on a penalized likelihood such as the Bayesian information criterion (BIC) and the minimum description length (MDL) criterion (Chickering and Heckerman, 1997).

We present here a novel approach to perform model selection. This approach relies on a non-parametric approach to estimate the number of parcels from the data. In this approach, the above mentioned JPDE model is reformulated to incorporate a non-parametric model selection using Dirichlet processes combined with a hidden Markov random field prior.

This section first recalls the basic principles of Dirichlet processes. In a second step, we show how Dirichlet processes can be included in the JPDE framework. The resulting inference scheme still based on VEM is finally presented in detail.

3.1. Dirichlet processes

Dirichlet processes were first proposed in (Ferguson, 1973) as distributions placed over distributions. A Dirichlet process (DP), denoted by $DP(G_0, \alpha)$, is characterized by a base distribution G_0 and a positive scaling parameter α. More precisely, a random distribution G is distributed according to a Dirichlet Process (Ferguson, 1973) with scaling parameter α and base distribution G_0, if for all natural numbers k and for all k-partitions $\{B_1, ..., B_k\}$

$$(G(B_1), G\left(B_2\right), ..., G\left(B_k\right)) \sim Dir\left(\alpha G_0\left(B_1\right), \alpha G_0\left(B_2\right), ..., \alpha G_0\left(B_k\right)\right) \quad (9)$$

where $Dir\left(\alpha G_0\left(B_1\right), \alpha G_0\left(B_2\right), ..., \alpha G_0\left(B_k\right)\right)$ is the Dirichlet distribution with parameter $\left(\alpha G_0\left(B_1\right), \ldots, \alpha G_0\left(B_k\right)\right)$.

A Dirichlet process mixture model (DPMM) uses the DP as a non-parametric prior in a hierarchical Bayesian model. Let us consider a mixture model where

η_n is the parameter associated with the n-th data point x_n, η_n is not observed and the DP is used to induce a prior on the η_n's. If G is a measure generated according to a DP, G is discrete with probability one. As a consequence, the following hierarchical representation can be seen as a countably infinite mixture model

$$\begin{aligned} x_n \mid \eta_n &\sim p(x_n \mid \eta_n) \\ \eta_n \mid G &\sim G \\ G \mid \{\alpha, G_0\} &\sim DP(\alpha, G_0) \end{aligned} \tag{10}$$

where $n = 1, \ldots, N$. Among the generated parameter values η_n, a number of them are equal. These unique values are used to partition the generated $x_1, \ldots, x_N$ into clusters. Thus, the DP mixture is a flexible mixture model with a random number of clusters which grows with new observed data. An explicit DP characterization, which will be useful hereafter, is provided in terms of stick-breaking construction (Blei et al., 2006). Consider two infinite collections of independent random variables $\boldsymbol{\tau}_i \sim Be(1, \alpha)$,where $Be(1, \alpha)$ is a beta distribution with parameters 1 and α, and $\eta_i^* \sim G_0$, for $i = 1, 2, \ldots$. With $\boldsymbol{\tau} = \tau_1, \tau_2, \ldots$, the stick-breaking representation of G is

$$\begin{aligned} \pi_i(\boldsymbol{\tau}) =& \tau_i \prod_{j=1}^{i-1} (1 - \tau_j) \\ G =& \sum_{i=1}^{\infty} \pi_i(\boldsymbol{\tau}) \delta_{\eta_i^*}. \end{aligned} \tag{11}$$

It is clear that G is a discrete distribution whose mixing proportions $\pi_i(\boldsymbol{\tau})$ are given by successively breaking a unit length stick into an infinite number of pieces. The size of each successive piece is proportional to the rest of the stick and is given by an independent draw from a beta distribution $Be(1, \alpha)$. Let z_n be the cluster assignment variable of the n-th data point. The hierarchical model of a Dirichlet process mixture model can be represented as follows

(i) $\tau_i \mid \alpha \sim Be(1, \alpha), i = 1, 2, ...$

(ii) $\eta_i^* \mid G_0 \sim G_0, i = 1, 2, ...$

(iii) for the n-th data point

(a) $z_n \mid \boldsymbol{\tau}$ is distributed according to a multinomial distribution, *i.e*, $z_n \mid \boldsymbol{\tau} \sim Mult(\pi(\boldsymbol{\tau}))$ with $\boldsymbol{\tau} = \tau_1, \tau_2, \ldots$

(b) $x_n \mid z_n \sim p(x_n \mid \eta^*_{z_n})$

3.2. Non-parametric Bayesian JPDE model

Following the line of DPMM, we address the issue of automatically selecting the number of parcels by considering a countably infinite number of parcels. This requires the extension of the standard finite state space Potts model to a countably infinite number of states in which we use a DPM prior on the $\boldsymbol{z}$ variable in the JPDE formulation. Our proposal differs from that in (Chatzis and Tsechpenakis, 2010) in that it is not a mean field approximation by a set of independent variables but a direct generalization of the Potts model that uses a stick breaking representation. The stick breaking representation is used to allow for the representation of an infinite number of states. For such a generalization, we need to consider the Potts model with an external field defined over $\boldsymbol{z} = \{z_1, \ldots, z_J\}$ with for all $j = 1, \ldots, J$, $z_j \in \{1, \ldots, K\}$

$$p(\boldsymbol{z}; \beta_z, \boldsymbol{\alpha}) \propto \exp\left(\sum_{j=1}^{J} \alpha_{z_j} + \beta_z \sum_{i \sim j} I(z_i = z_j)\right) , \tag{12}$$

where β_z is an interaction parameter as in Eq. (8) and $\boldsymbol{\alpha}$ is a parameter vector such that $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_K\}$ represents an additional external field parameter where each α_k is scalar. Such a Potts model is defined up to a multiplicative constant depending on $\boldsymbol{\alpha}$, meaning that the distribution (12) can be also obtained when adding the same constant value to all the α_k's. To avoid such an identifiability issue, it is common to consider additional constraints on the α_k's. One way to make the parameter vector $\boldsymbol{\alpha}$ unique is to asssume $\alpha_k = \log \pi_k$ with $\sum_{k=1}^{K} \pi_k = 1$. The Potts model in (12) then rereads

$$\begin{aligned} p(\boldsymbol{z}; \beta_z, \boldsymbol{\pi}) &\propto \exp\left(\sum_{j=1}^{J} \log \pi_{z_j} + \beta_z \sum_{i \sim j} I(z_i = z_j)\right) \\ &\propto \left(\prod_{j=1}^{J} \pi_{z_j}\right) \exp\left(\beta_z \sum_{i \sim j} I(z_i = z_j)\right) . \end{aligned} \tag{13}$$

Define $V(\boldsymbol{z};\boldsymbol{\pi},\beta_z)=\sum_{j=1}^J \log \pi_{z_j}+\beta_z \sum_{i\sim j} I(z_i=z_j)$, which is called the energy function, where the first and the second sum respectively represents the first and the second order potentials. In the finite state space case, such a representation is equivalent, via the Hammersley-Clifford theorem (Besag, 1974), to assume that the distribution in (12) is a Markov random field.

Using the stick breaking construction, we can then consider a countably infinite number of probabilities π_k that sum to 1, *i.e*, $\sum_{k=1}^\infty \pi_k = 1$. From this, we can define the same energy function V as before but consider it over an infinite countable set (homogeneous to the set of positive integers),

$$V\left(\boldsymbol{z};\boldsymbol{\pi},\beta_z\right)=\sum_{j=1}^J \log \pi_{z_j}+\beta_z \sum_{i\sim j} I(z_i=z_j)$$

for $z_j\in\{1,2,\ldots\}$. Next, using the Gibbs representation $p(\boldsymbol{z})\propto \exp(V(\boldsymbol{z};\boldsymbol{\pi},\beta_z))$, the Hammersley-Clifford theorem still holds if $\sum_{\boldsymbol{z}} \exp(V(\boldsymbol{z};\boldsymbol{\pi},\beta_z)) < \infty$. Our choice of $\boldsymbol{\pi}$ ensures this property. Indeed,

$$\begin{aligned}\sum_{\boldsymbol{z}} \exp(V(\boldsymbol{z};\boldsymbol{\pi},\beta_z)) &= \sum_{\boldsymbol{z}} \left(\prod_{j=1}^J \pi_{z_j}\right) \exp\left(\beta_z \sum_{i\sim j} I(z_i=z_j)\right)\\ &< \exp\left(\beta_z J(J-1)\right) \sum_{\boldsymbol{z}} \prod_{j=1}^J \pi_{z_j}\\ &< \exp\left(\beta_z J(J-1)\right) < \infty\end{aligned}$$

where $J(J-1)$ is the maximum number of neighbors among J sites. We also used that for all $j=1,\ldots,J$, $\sum_{z_j}\pi_{z_j}=\sum_{k=1}^\infty \pi_k = 1$. It follows that $p(\boldsymbol{z};\boldsymbol{\pi},\beta_z)$, in the infinite state space case, is still a valid probability distribution and is a Markov field by the Hammersley-Clifford theorem. Note that such a generalization of the Potts model is possible because of the presence of the external field parameters π_k that satisfy $\sum_{k=1}^\infty \pi_k = 1$. A Potts model with equal external field parameters cannot be as simply extended to an infinite countable state space. For a Potts model with no external field, such an extension is not possible because in the K-state case this Potts model is equivalent to $\pi_k = 1/K$ for all k where their sum does not tend to 1 when K tends to infinity.

In the stick breaking setting, we then consider $\pi_k(\boldsymbol{\tau}) = \tau_k \prod_{l=1}^{k-1}(1-\tau_l)$

and

$$p(\boldsymbol{z};\beta_z,\boldsymbol{\tau}) \propto \left(\prod_{j=1}^{J}\pi_{z_j}(\boldsymbol{\tau})\right)\ \exp\left(\beta_z\sum_{i\sim j}I(z_i=z_j)\right)\ . \tag{14}$$

Such a construction is valid for any set of parameters $\boldsymbol{\tau}=\{\tau_k\}_{k=1}^{\infty}$ with each $\tau_k\in[0,1]$. However we would be left with an infinite number of parameters τ_k to estimate. The Bayesian point of view solves this problem by assuming that all τ_k's are i.i.d. variables following the same $Be(1,\alpha)$ distribution so that the number of parameters to estimate is now reduced to a single parameter α.

To reformulate the JPDE model in a non-parametric Bayesian framework coupled with HMRF, the original model of (Chaari et al., 2012, 2015) is kept except for the HRF groups $\boldsymbol{z}$. The extension of JPDE to an infinite number of parcels therefore consists of augmenting the original JPDE formulation with additional variables $\{\tau_k\}_{k=1}^{\infty}$ and of considering the following hierarchical construction which we call the NP-JPDE model

(i) $\tau_k \mid \alpha \sim Be(1,\alpha), k=1,2,...$

(ii) $(\Theta_k^*=(\bar{\boldsymbol{h}}_k,\bar{\boldsymbol{\Sigma}}_k)\mid G_0)\sim G_0, k=1,2,...$ where $G_0=\mathcal{N}(0,\sigma_h^2\boldsymbol{R})\otimes\delta_{\nu\boldsymbol{I}}$

(iii) $p(\boldsymbol{z}|\boldsymbol{\tau};\beta_z)\propto(\prod_{j=1}^{J}\pi_{z_j}(\boldsymbol{\tau}))\ \exp(\beta_z\sum_{i\sim j}I(z_i=z_j))$

(iv) $\boldsymbol{h}_j \mid z_j \sim p(\boldsymbol{h}_j|\Theta_{z_j}^*)$, where $p(\boldsymbol{h}_j|\Theta_k^*)=\mathcal{N}(\boldsymbol{h}_j;\bar{\boldsymbol{h}}_k,\bar{\boldsymbol{\Sigma}}_k)$ is a Gaussian distribution whose parameters $\bar{\boldsymbol{h}}_k,\bar{\boldsymbol{\Sigma}}_k$ are associated with the k-th parcel[1].

Fig. 1 illustrates the new hierarchical model which differs from the model of (Chaari et al., 2012, 2015) by the green circled variables required for model selection. The resulting joint distribution of the $\boldsymbol{Y},\boldsymbol{A},\boldsymbol{H},\boldsymbol{Q},\boldsymbol{z}$ and the additional variable $\boldsymbol{\tau}$ reads

$$\begin{aligned} p(\boldsymbol{Y},\boldsymbol{A},\boldsymbol{H},\boldsymbol{Q},\boldsymbol{z},\boldsymbol{\tau};\boldsymbol{\Theta}) =& p(\boldsymbol{Y}\mid\boldsymbol{A},\boldsymbol{H};\boldsymbol{\Theta})\ p(\boldsymbol{A}\mid\boldsymbol{Q};\boldsymbol{\Theta})\ p(\boldsymbol{Q};\boldsymbol{\Theta})\\ & p(\boldsymbol{H}\mid\boldsymbol{z};\boldsymbol{\Theta})\ p(\boldsymbol{z}|\boldsymbol{\tau};\boldsymbol{\Theta})\ p(\boldsymbol{\tau};\boldsymbol{\Theta}) \end{aligned} \tag{15}$$

where $\boldsymbol{\Theta}=\left\{\boldsymbol{\Gamma},\boldsymbol{L},\boldsymbol{\theta}_a,\beta_z,\sigma_h^2,(\bar{\boldsymbol{h}}_k)_{1\le k\le K},\nu,\alpha\right\}$ and $p(\boldsymbol{\tau})=\prod_{k=1}^{\infty}p(\tau_k)$, $p(\boldsymbol{z}|\boldsymbol{\tau})$ are defined in steps (i) and (iii), respectively.

---

[1] The other distributions defining the model remain the same as in the standard JPDE model. Note that in the extended version above we assume $\nu_k=\nu$ for all k to define G_0.

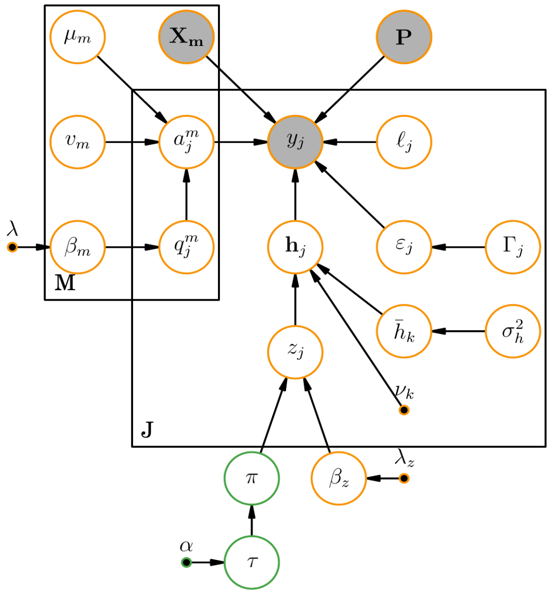


Figure 1: Graphical model describing dependencies between observed and missing variables involved in the non-parametric JPDE generative model for a given parcel $\boldsymbol{\mathcal{P}}$ with J voxels. Orange circled variables are identical to the original model, whereas the green circled variables are the new ones.

3.3. VEM algorithm

Different inference strategies can be used to estimate the missing variables $\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}, \boldsymbol{\tau}$ and the parameters $\boldsymbol{\Theta}$ from the posterior $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}, \boldsymbol{\tau} \mid \boldsymbol{Y}; \boldsymbol{\Theta})$ associated with Eq. (15). Due to the computational complexity of MCMC methods, we here use a VEM algorithm to derive an approximation of the true posterior distribution $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}, \boldsymbol{\tau} \mid \boldsymbol{Y}; \boldsymbol{\Theta})$ of the form

$$\widetilde{p}(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}, \boldsymbol{\tau}; \boldsymbol{\Theta}) = \widetilde{p}_A(\boldsymbol{A})\ \widetilde{p}_H(\boldsymbol{H})\ \prod_{j=1}^{J} \widetilde{p}_{Q_j}(Q_j)\ \prod_{j=1}^{J} \widetilde{p}_{z_j}(z_j)\ \widetilde{p}_\tau(\boldsymbol{\tau}). \quad (16)$$

In the variational distribution above, the approximations $\prod_{j=1}^{J} \widetilde{p}_{Q_j}(Q_j)$ and $\prod_{j=1}^{J} \widetilde{p}_{z_j}(z_j)$ are sought in a form that factorizes over voxels (mean field) to handle intractability due to the spatial neighborhood. The infinite state space for $\boldsymbol{z}$ is dealt with by considering a truncation to a number K which consists of assuming that the variational distribution satisfies $\widetilde{p}_{z_j}(k) = 0$ for $k > K$ and $\widetilde{p}_\tau(\boldsymbol{\tau}) = \prod_{k=1}^{K-1} \widetilde{p}_{\tau_k}(\tau_k)$. This amounts to setting $\tau_k = 1$ for $k \geq K$ or $\widetilde{p}_{\tau_k}(\tau_k) = \delta_1(\tau_k)$.

The VEM approach requires five steps associated with five expectations referred to as: VE-A, VE-H, VE-Q, VE-Z and VE-$\boldsymbol{\tau}$. Compared to the original version of JPDE, the new steps are the VE-Z and VE-$\boldsymbol{\tau}$ steps which are detailed below. (See (Chaari et al., 2012, 2015) for more details about the other expectation steps.)

- **VE-$\boldsymbol{\tau}$ step**

  The VE-$\boldsymbol{\tau}$ step results straightforwardly from results on variational approximation in the exponential family. Given Eq. (11) and for $k = 1, \ldots, K-1$,

$$\widetilde{p}_{\tau_k}(\tau_k) \propto p(\tau_k)\ \exp\left(\sum_{j=1}^{J} \mathrm{E}_{\widetilde{p}_{z_j}\widetilde{p}_{\tau\setminus\{k\}}}\left[\log \pi_{z_j}(\boldsymbol{\tau})\right]\right) \quad (17)$$

$$\propto p(\tau_k) \exp\left(\sum_{j=1}^{J} \sum_{l=k+1}^{K} \widetilde{p}_{z_j}(l) \log(1-\tau_k) + \sum_{j=1}^{J} \widetilde{p}_{z_j}(k) \log(\tau_k)\right) \quad (18)$$

$$\propto Be(\gamma_{k,1}, \gamma_{k,2}) \quad (19)$$

with

$$\gamma_{k,1} = 1 + \sum_{j=1}^{J} \widetilde{p}_{z_j}(k) \tag{20}$$

$$\gamma_{k,2} = \alpha + \sum_{j=1}^{J} \sum_{l=k+1}^{K} \widetilde{p}_{z_j}(l). \tag{21}$$

A gamma prior is placed over the scaling parameter α following (Escobar and West, 1995) with parameters $(\hat{s}_1, \hat{s}_2)$. The gamma distribution is conjugate to the stick lengths and the parameters $\hat{s}_1$ and $\hat{s}_2$ are given by

$$\hat{s}_1 = s_1 + K - 1 \tag{22}$$

$$\hat{s}_2 = s_2 - \sum_{k=1}^{K-1} \mathrm{E}_{\widetilde{p}_{\tau_k}}\big[\log(1-\tau_k)\big]. \tag{23}$$

After computing these parameters, we replace α in Eq. (21) with its expectation $\mathrm{E}_q\big[\alpha\big] = \frac{\hat{s}_1}{\hat{s}_2}$.

- **VE-Z step**
  The VE-Z step is divided into J VE-Z_j steps. Since we assume $\widetilde{p}_{z_j}(z_j) = 0$ for $z_j > K$, we only need to compute the distributions for $z_j \leq K$,

$$\widetilde{p}_{z_j}(z_j) \propto \exp\left(\mathrm{E}_{\widetilde{p}_{H_j}}\big[\log p(\boldsymbol{h}_j|z_j)\big] + \mathrm{E}_{\widetilde{p}_{\tau}}\big[\log \pi_{z_j}(\boldsymbol{\tau})\big] + \beta_z \sum_{i\sim j} \widetilde{p}_{z_i}(z_j) \right) \tag{24}$$

where

$$\mathrm{E}_{\widetilde{p}_{\tau}}\big[\log \pi_k(\boldsymbol{\tau})\big] = \mathrm{E}_{\widetilde{p}_{\tau_k}}\big[\log \tau_k\big] + \sum_{l=1}^{k-1} \mathrm{E}_{\widetilde{p}_{\tau_l}}\big[\log(1-\tau_l)\big] . \tag{25}$$

The expectations above can be computed using the fact that $\widetilde{p}_{\tau_k}$ is a beta distribution, *i.e*,$Be(\gamma_{k,1}, \gamma_{k,2})$ defined by (20) and (21)

$$\mathrm{E}_{\widetilde{p}_{\tau_k}}\big[\log(\tau_k)\big] = \Psi(\gamma_{k,1}) - \Psi(\gamma_{k,1} + \gamma_{k,2}) \tag{26}$$

$$\mathrm{E}_{\widetilde{p}_{\tau_k}}\big[\log(1-\tau_k)\big] = \Psi(\gamma_{k,2}) - \Psi(\gamma_{k,1} + \gamma_{k,2}) \tag{27}$$

where $\Psi(.)$ is the digamma function defined by

$$\Psi(z) = \frac{d}{dz} \log \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}.$$

The term $\mathrm{E}_{\widetilde{p}_{H_j}}\big[\log p(H_j|z_j)\big]$ is computed as in the JPDE model as shown in the VE-H step of (Chaari et al., 2015).

- **VM step**

  The maximization step in this extended JPDE is the same as in (Chaari et al., 2015). As a consequence, it can be rewritten as

$$\begin{aligned}\boldsymbol{\Theta}^{(r)} = \arg\max_{\boldsymbol{\Theta}}\{\ & \mathrm{E}_{\widetilde{p}_A^{(r)}\widetilde{p}_H^{(r)}}\big[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{h}; \boldsymbol{L}, \boldsymbol{\Gamma})\big] + \mathrm{E}_{\widetilde{p}_A^{(r)}\widetilde{p}_Q^{(r)}}\big[\log p(\boldsymbol{A} \mid \boldsymbol{Q}; \boldsymbol{\mu}, \boldsymbol{v})\big] \\ & + \mathrm{E}_{\widetilde{p}_H^{(r)}\widetilde{p}_z^{(r)}}\big[\log p(\boldsymbol{h}|\boldsymbol{z}; \{\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k\}_{k=1:K}\big] + \mathrm{E}_{\widetilde{p}_Q^{(r)}}\big[\log p(\boldsymbol{Q}; \boldsymbol{\beta})\big] \\ & + \mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_\tau^{(r)}}\big[\log p(\boldsymbol{z} \mid \boldsymbol{\tau}; \beta_z)\big]\}. \end{aligned} \tag{28}$$

  The only step that differs from the original JPDE model is the maximization of Eq. (28) with respect to β_z which leads to

$$\beta_z^{(r)} = \arg\max_{\beta_z} \mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_\tau^{(r)}}\big[\log p(\boldsymbol{z}|\boldsymbol{\tau}; \beta_z)\big]. \tag{29}$$

  More details about this step are given in Appendix B. The solution proposed for estimating β_z involves Monte Carlo runs that are computationally expensive. For this reason, the experiments considered in this paper have been conducted with a fixed value of β_z.

4. Validation

To validate the NP-JPDE model, we performed numerical experiments on both synthetic and real data [2]. The proposed non-parametric Bayesian algorithm is compared with the strategy adopted in (Albughdadi et al., 2014) which consists of selecting the model that provides the highest free energy. The free energy calculations must be done for all the candidate models which is time consuming specially when no prior information is available about the optimum number of parcels. The interested reader can refer to Appendix A for more details on the free energy-based model selection.

---

[2] These experiments were implemented in Python within the framework offered by the Pyhrf software (Vincent et al., 2014), see also `http://pyhrf.org`.

4.1. Synthetic fMRI time series

To validate the NP-JPDE model, three different synthetic experiments referred to as Exps. 1-3 have been conducted. Different parcellation masks have been used in each experiment to generate BOLD signal according to Eq. (1). Two experimental conditions ($M = 2$) have been considered with 30 trials for each of them. The reference activation labels are shown in Fig. 2.

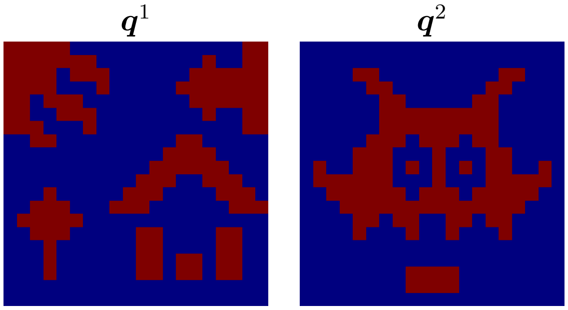


Figure 2: Reference activation labels for the two experimental conditions (grid size = 20×20).

Using Pyhrf, the NRLs were drawn according to their prior distribution conditionally to the activation labels $\boldsymbol{Q}$ of Fig. 2. Given these 20×20 binary labels, the NRLs were simulated as follows, for $m = 0, 1$: $a_j^m \mid q_j^m = 0 \sim \mathcal{N}(0, 0.5)$ and $a_j^m \mid q_j^m = 1 \sim \mathcal{N}(3.2, 0.5)$. The representative NRLs are displayed in Fig. 3. The onsets of these trials were randomly generated with a mean inter stimuli interval of 3 s and a variance of 5 s. The fMRI time series $\boldsymbol{y}_j$ were then generated according to Eq. (1) using $\Delta t = 0.5$ and $TR = 1$ s.

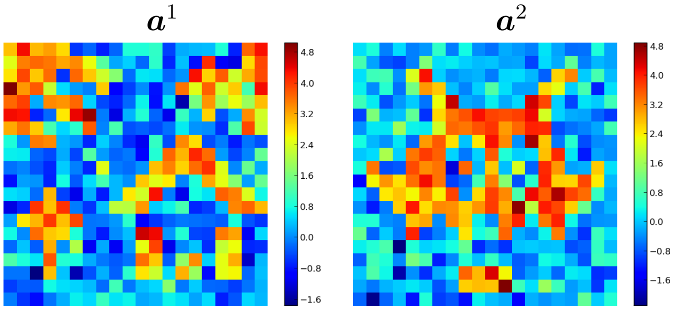


Figure 3: Reference NRLs for the two experimental conditions (grid size = 20×20).

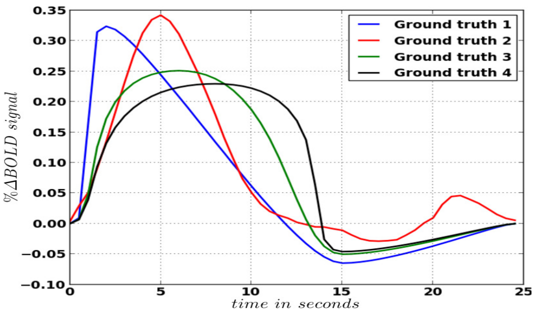


Figure 4: Ground truth HRF shapes ($\bar{h}_k, k = 1, \ldots, K^\omega$ with $\omega = \{1, \ldots, 3\}$) used for generating synthetic fMRI time series.

As a ground truth for the parcellation, different HRFs groups were considered, each with $K^\omega = \omega + 1$ parcels where $\omega \in \{1, \ldots, 3\}$. The HRFs associated with these groups were selected from the ground truth HRFs $(\bar{\boldsymbol{h}}_k)_{k=1}^{K^\omega}$ shown in Fig. 4. Reference parcellations for the three experiments are displayed in Fig. 5. These reference parcellations were chosen with different cardinalities and overlap with activation areas in order to investigate the robustness of the NP-JPDE model to the total amount of evoked activity in each parcel. Indeed, from a statistical point of view, the estimation of parcels involving a large amount of activated voxels should be more accurate than the estimation of parcels overlapping only a few activated voxels. Importantly, to mimic a real scenario in all experiments, we set the percentage of the activated voxels to be approximately 53% of the total number of voxels (this percentage was calculated by performing a bitwise OR between the reference activation binary labels of the two experimental conditions Fig. 2). Tab. 3 reports for each experiment the percentage of activated voxels in each parcel of the ground truth.

These synthetic fMRI time series were then processed by the JPDE and NP-JPDE models. Results obtained with the two models were compared especially in terms of model selection. When using the original JPDE, three competing models $K^\omega = \omega + 1$ where $\omega \in \{1, \ldots, 3\}$ were run and their corresponding free energy was computed according to Eq. (A.1). As regards the NP-JPDE, it is worth noting that we do not need to specify any specific initialization. Hence, the latter was done randomly in contrast to the shown initializations for the original JPDE reported in Fig. 5[bottom]. The NP-JPDE model only requires to set the maximum number of parcels K (truncation level) for the variational approximation. This number was set to

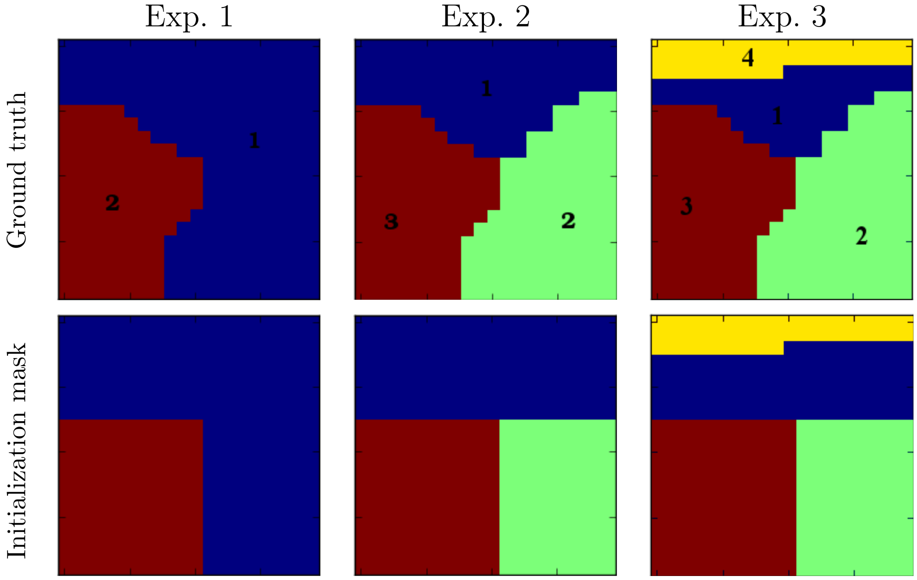


Figure 5: Ground truth parcellations used for the 3 experiments and corresponding initialization masks (only used for the original version of the JPDE approach) (grid size = 20×20).

Table 3: Percentage of activated voxels in each parcel of the ground truth parcellations for the three experiments. The parcels indexes are shown in Fig. 5

Experiment	# Parcel	% of activated voxels
Exp. 1	1	66.7%
	2	33.3%
Exp. 2	1	22.2%
	2	44.5%
	3	33.3%
Exp. 3	1	19.5%
	2	44.5%
	3	33.3%
	4	2.7%

$K = 20$ for the three experiments, while the Potts parameter β_z was fixed to 1.2 for the spatial regularity of the parcellation [3]. The parameter β_m for activation classes which corresponds to the m-th experimental condition is estimated in the maximization step as in (Chaari et al., 2013, 2015). The prior values over the scaling parameter α of the DPMM were set to $\hat{s}_1 = 20, \hat{s}_2 = 5$. The estimated parcellations obtained by the two JPDE versions are shown in Fig. 6. This figure shows accurate parcellation estimates from a visual point of view. A comparison with the ground truth allows one to conclude that the proposed NP-JPDE algorithm recovers accurate parcels especially for activated parcels. Quantitative evaluation of the parcellation estimates is provided in Tab. 4 where the error rate with respect to the ground truth is given. First, one can notice the small error probabilities for both models in all experiments. Furthermore, the NP-JPDE outperforms the original JPDE seen in the error reported for Exps. 2-3. This remark corroborates the better visual performance of the proposed NP-JPDE model.

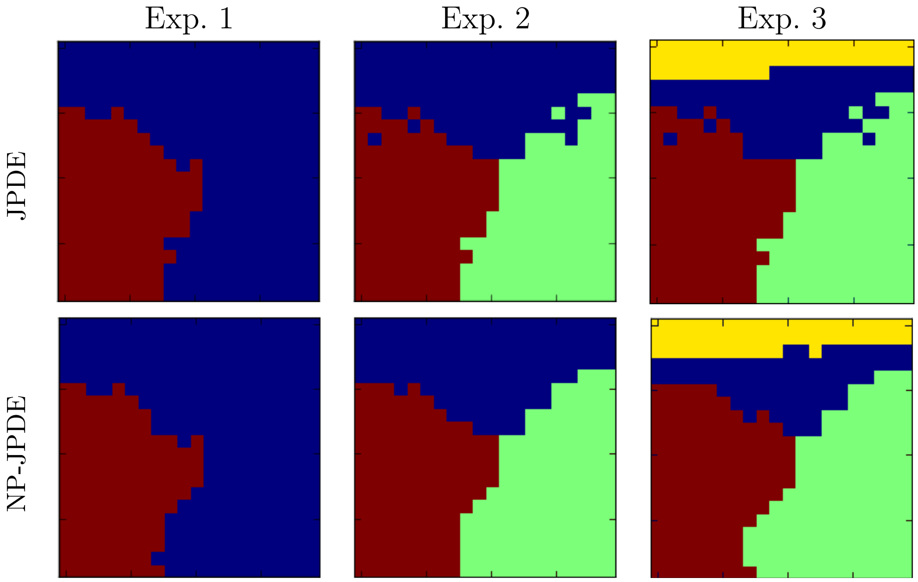


Figure 6: Parcellation estimates for the three experiments using the original JPDE and NP-JPDE (grid size = 20 × 20).

To investigate more deeply the robustness of the parcellation estimation using

---

[3] This value of β_z was adjusted by cross validation.

Table 4: Error probabilities on the parcellation estimates using the original JPDE and the NP-JPDE algorithms.

Model	Exp. 1	Exp. 2	Exp. 3
NP-JPDE	1.5%	**0.25%**	**1.5%**
JPDE	1.5%	2.75%	3.25%

the NP-JPDE model, the confusion matrix for each of the three experiments was computed and shown in Tabs. 5-7. We observed that the proposed NP-JPDE is highly accurate regarding the parcellation estimation step as the overlap between the reference and estimate for each parcel is larger than 95% in all experiments.

In order to further investigate the robustness of the proposed model, Tab. 8 provides the mean square errors (MSEs) for the NRLs and activation labels associated with the JPDE and NP-JPDE models. These results corroborate the fact that the NP-JPDE model ensures precise estimation of the NRLs for both experimental conditions and outperforms the classical JPDE version. The construction of the parcellation for the NP-JPDE model has therefore very little impact on the NRL estimates and the detection task. Next, we investigated the accuracy of the estimation task by looking at the HRF estimates using the NP-JPDE model as reported in Fig. 7. A comparison between the reference and estimated HRF shapes shows that the NP-JPDE model is able to recover precise hemodynamics profiles and they are close to the HRF estimates of the original JPDE version (shown in the same figure).

Table 5: Confusion matrix for Exp. 1. (NP-JPDE model). RP and EP refer to the reference and the estimated parcellations, respectively.

RP EP	Parcel. 1	Parcel. 2
Parcel. 1	**1.0**	0.046
Parcel. 2	0.0	**0.954**

Last, we studied the convergence of the number of parcels over iterations within the NP-JPDE. To this end, we present in Fig. 8 the parcellation estimate for Exp. 2 along different iterations until convergence. Starting with a

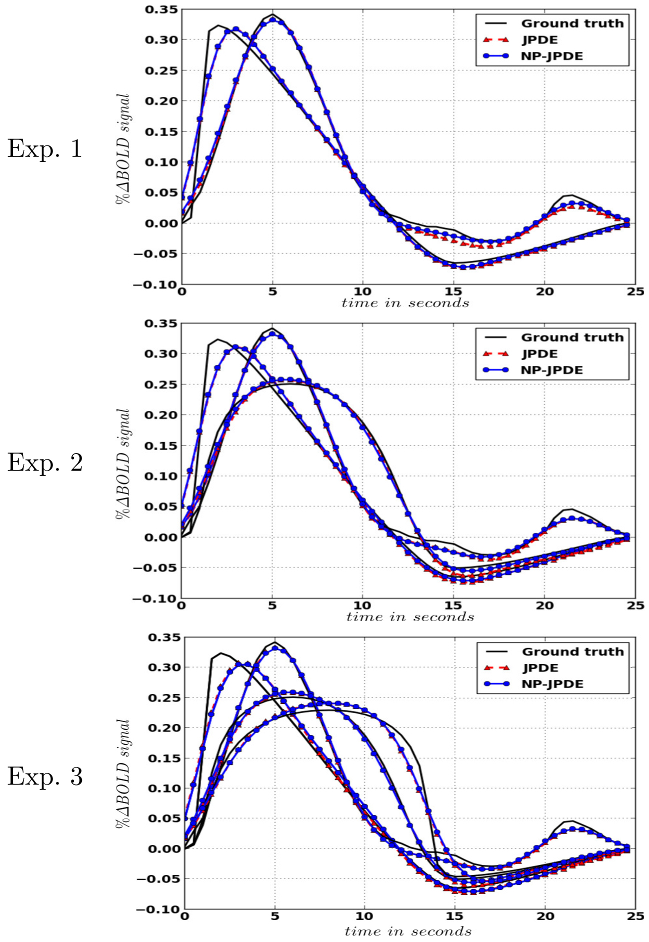


Figure 7: HRF estimates for the three experiments using JPDE and NP-JPDE models.

Table 6: Confusion matrix for Exp. 2. (NP-JPDE model). RP and EP refer to the reference and the estimated parcellations, respectively.

RP EP	Parcel. 1	Parcel. 2	Parcel. 3
Parcel. 1	**1.0**	0.0	0.008
Parcel. 2	0.0	**1.0**	0.0
Parcel. 3	0.0	0.0	**0.992**

Table 7: Confusion matrix for Exp. 3. (NP-JPDE model). RP and EP refer to the reference and the estimated parcellations, respectively.

RP EP	Parcel. 1	Parcel. 2	Parcel. 3	Parcel. 4
Parcel. 1	**1.0**	0.013	0.0	0.0
Parcel. 2	0.00	**0.961**	0.0	0.0
Parcel. 3	0.0	0.0	**1.0**	0.023
Parcel. 4	0.0	0.026	0.0	**0.977**

Table 8: MSEs of NRL estimates and activation labels for the JPDE and NP-JPDE models.

		Exp. 1		Exp. 2		Exp. 3	
		JPDE	NP-JPDE	JPDE	NP-JPDE	JPDE	NP-JPDE
NRLs	$m = 1$	0.016	**0.007**	0.017	**0.008**	0.017	**0.008**
	$m = 2$	0.012	**0.006**	0.012	**0.006**	0.012	**0.006**
Labels	$m = 1$	**0.003**	0.004	0.011	**0.003**	0.011	**0.003**
	$m = 2$	0.003	0.003	0.003	**0.002**	0.003	0.003

random initialization, this figure shows that after about 7 iterations all the main parcels are well established. Furthermore, for the same experiment, fifty runs of the VEM algorithm using different random initializations were performed and the subsequent box plot graph was drawn to investigate the sensitivity of the NP-JPDE model to this setting. Fig. 9 shows the evolution of the number of parcels over iterations for the fifty runs. It appears first that all the parcels were present after the first few iterations. Second, this number decreased through the iterations. Finally, we investigated the computational load. For doing so, we computed the running time for the standard JPDE framework by accumlating all elapsed times required for assessing the free energy associated with each candidate model, as done in (Albughdadi et al., 2014). Using a machine with 8 cores, each corresponding to an Intel® Xeon(R) CPU E3-1240 v3 chipset clocking at 3.40GHz processor and 16 GB of RAM, the four investigated models in the classical JPDE framework run in about 35 mins whereas for the NP-JPDE model it takes less than 9 mins. Thus, the computational cost of the NP-JPDE model is reduced when compared to free energy calculations of many candidate models.

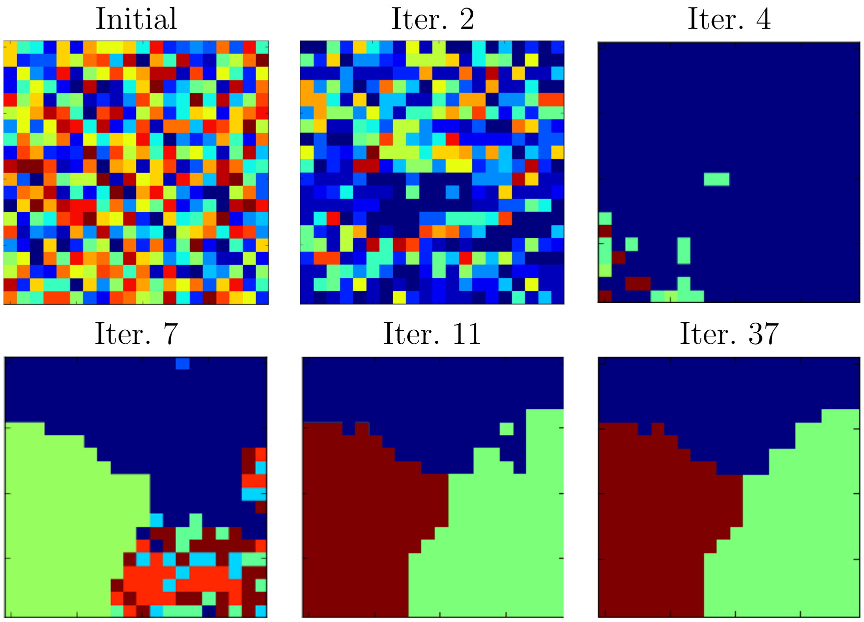


Figure 8: Parcellation estimates for Exp. 2 using the NP-JPDE model along successive iterations (grid size $= 20 \times 20$).

4.2. Real data

Two experiments were conducted on real fMRI data to validate the proposed NP-JPDE model. The two experiments differ by the regions of interest (ROI) under consideration. Exp. 1 and Exp. 2 focused on the right motor and bilateral occipital ROIs, respectively. These ROIs are shown in Fig. 10 and have been defined from the statistical results of a standard subject-level GLM analysis of fMRI data. More precisely, Student$-t$ maps associated with the two contrasts of interest, namely `(Left Click - Right Click)` and `(Visual stimuli - Auditory stimuli)`, have been thresholded at $p = 0.05$, corrected for multiple comparisons according to the FWER criterion, see (Badillo et al., 2013b; Chaari et al., 2014) for details. The fMRI data were collected using a gradient-echo EPI sequence (TE $=$ 30 ms/TR $=$ 2.4s/thickness $=$ 3 mm/FOV $=$ 192$\times$192 mm^2, matrix size: 96$\times$96) at a 3 Tesla during a localizer experiment (Pinel et al., 2007). Sixty auditory, visual and motor stimuli were involved in the paradigm and defined in ten experimental conditions ($M = 10$) (see (Badillo et al., 2013b; Chaari et al., 2014) for details). During this paradigm, $N = 128$ scans were

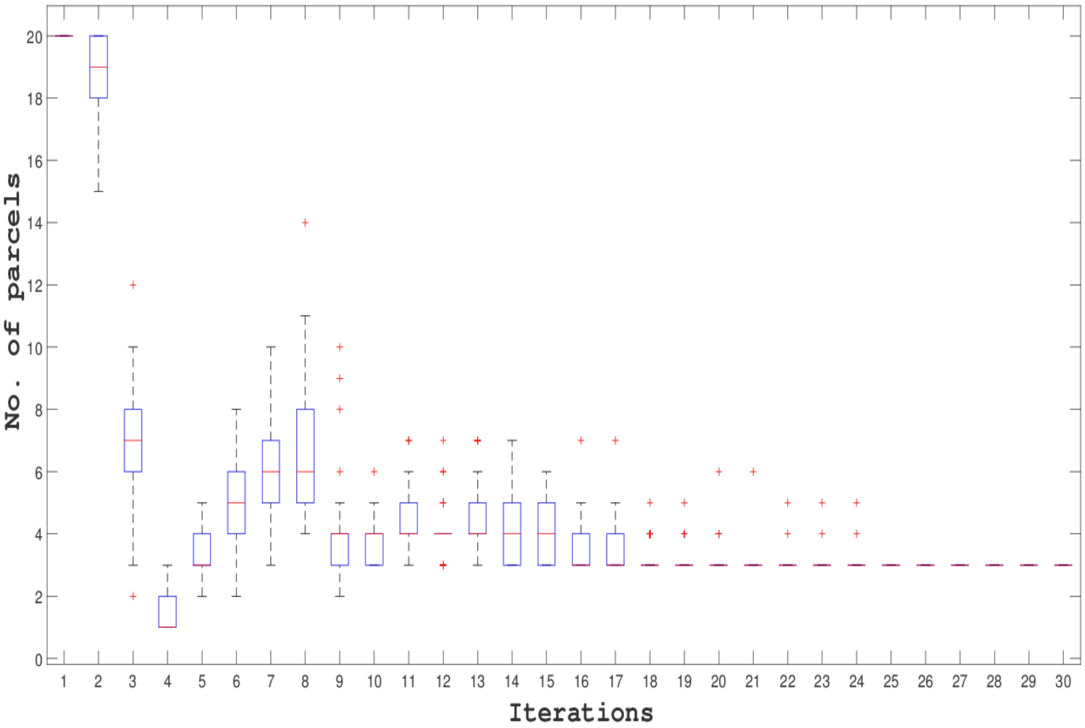
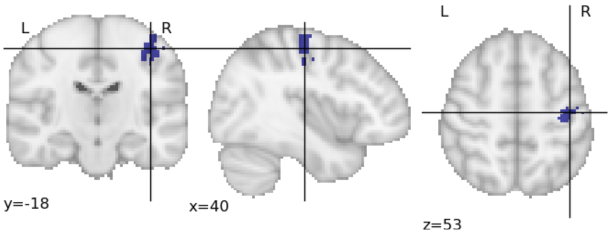
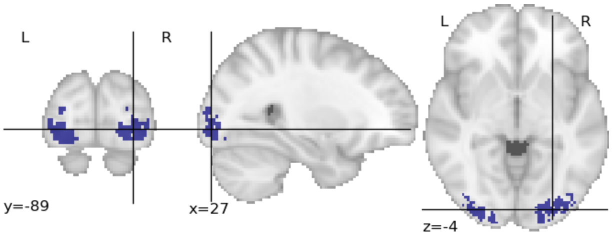


Figure 9: Boxplot for fifty different runs of Exp. 2 using the NP-JPDE model showing the convergence of the parcellation up to 30 iterations. The convergence is achieved from iteration 16.



(a) Right motor ROI



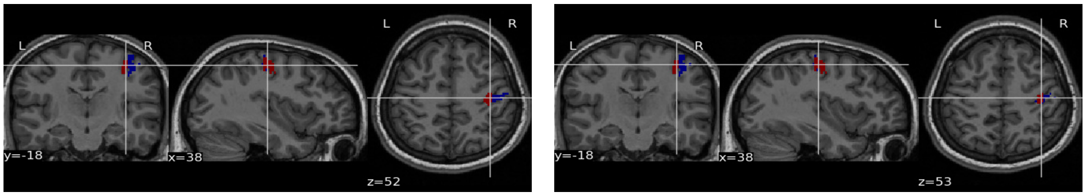
(b) Bilateral occipital ROI

Figure 10: Anatomical localization of brain regions. On top, the ROI is located in the right motor cortex and consists of a single connected component. At the bottom, the ROI is located in the primary visual cortex and made up of two connected components, one in each hemisphere.

acquired. For both experiments, we considered the truncation level $K = 20$, the parameter of the HMRF β_z was empirically set to 1.8 and the parameters of the gamma prior for the scaling parameter α were set to $\hat{s}_1 = 20$, $\hat{s}_2 = 5$.[4] In Exp. 1, two parcels were estimated in the right motor cortex. Different slices of the estimated parcellation are shown in Fig. 11. The HRF shape estimates are shown in Fig. 12 along with the canonical HRF and the HRF estimated with the JDE model. These HRF estimates have the same value of the time to peak (TTP) and the full width at half maximum (FWHM): TTP = 4.8 s and FWHM = 4.2 s. As regards the HRF obtained with JDE, the TTP and FWHM values are 4.8 s and 3.6 s, respectively. We notice that both models recover the same TTP whereas the JDE yields a slightly narrower HRF (lower FWHM). The Euclidean distances were calculated between the HRF estimates and the canonical HRF. These values are reported in Tab. 9 in addition to the distance between the individual NP-JPDE HRF estimates. The reported distances indicate that the NP-JPDE model pro-

---

[4] These parameters were determined empirically by cross validation.



(a) Slice. 1 (b) Slice. 2

Figure 11: Consecutive slices of the estimated parcellation located in the right motor cortex.

vides closer HRF estimates to the canonical one (average Euclidean distance of 0.4) compared to the JDE model (average Euclidean distance of 0.43). In this sense, the NP-JPDE model provides more coherent results than the JDE one in terms of closeness of the HRF estimates to the canonical shape in the motor cortex as it has already been shown in the literature (Badillo et al., 2013b).

As regards the NRL estimates, we focused on the left and right click visual and auditory experimental conditions which are expected to elicit evoked activity in the right motor cortex. Fig 13 shows the NRL estimates using the NP-JPDE and JDE models (with respect to the left and right auditory experimental conditions) and the computed contrast (auditory left click-auditory right click). Fig 14 shows the NRL estimates for both models with respect to the left and right click visual experimental conditions and the computed contrast (visual left click-visual right click). These results confirm the coherence between the NRL estimates obtained with the JDE and NP-JPDE models, especially in terms of maximum activation location and amplitude values.

Table 9: Euclidean distance between the HRF estimates in the right motor cortex and the canonical HRF. Distance between the individual NP-JPDE HRF estimates are also provided.

	HRF. 1	HRF. 2	JDE
Canonical HRF	0.37	0.43	0.43
HRF. 2	0.30	–	–

The NP-JPDE was also run for Exp. 2 on the bilateral occipital cortex. Four parcels were detected as shown in Fig. 15. The corresponding HRF shape estimates for these parcels are shown in Fig. 16. These HRF estimates are

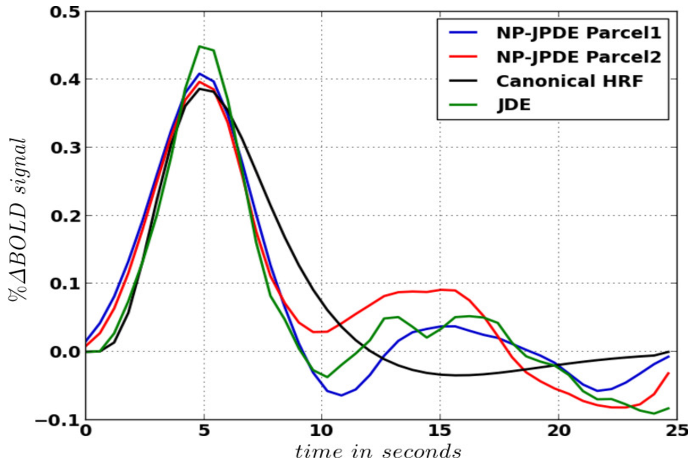
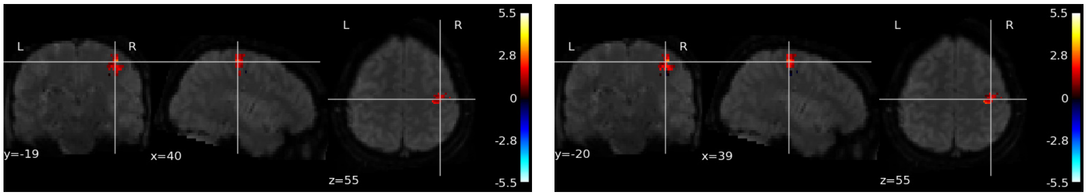
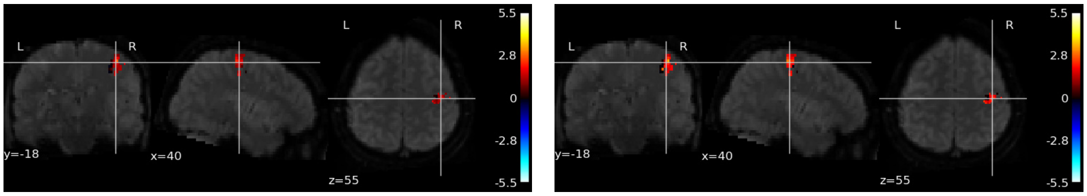


Figure 12: HRF shape estimates using the NP-JPDE and JDE models in the right motor cortex and the canonical HRF.



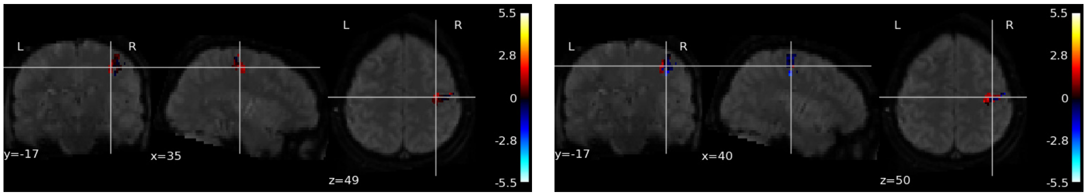
(a) Auditory left click (NP-JPDE)

(b) Auditory left click (JDE)



(c) Auditory right click (NP-JPDE)

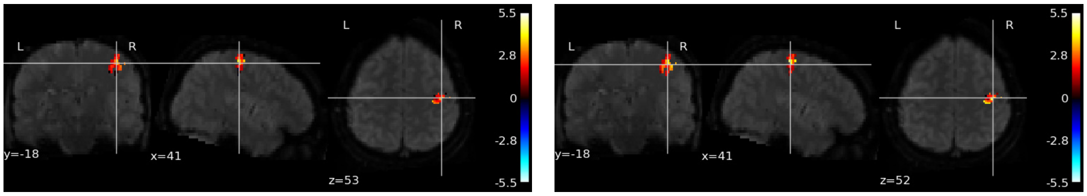
(d) Auditory right click (JDE)



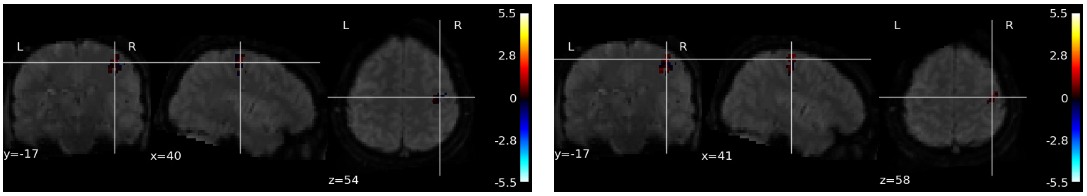
(e) Left click-Right click contrast (NP-JPDE)

(f) Left click-Right click contrast (JDE)

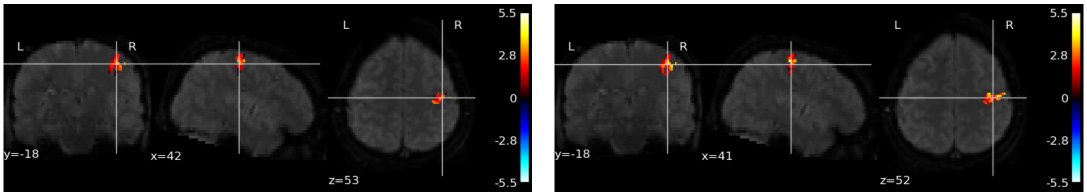
Figure 13: NRL estimates for the auditory left and right click experimental conditions and their computed contrast (left click-right click) using NP-JPDE and JDE models.



(a) Visual left click (NP-JPDE) (b) Visual left click (JDE)



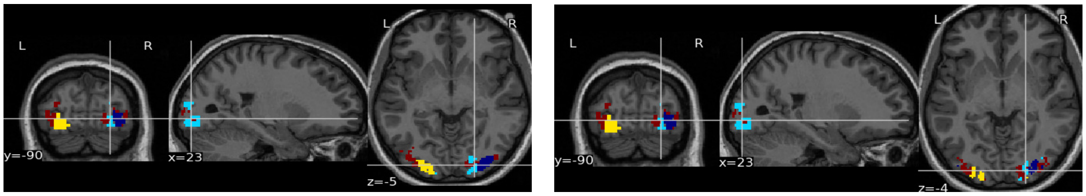
(c) Visual right click (NP-JPDE) (d) Visual right click (JDE)



(e) Left click-Right click contrast (NP-JPDE) (f) Left click-Right click contrast (JDE)

Figure 14: NRL estimates for the visual left and right click experimental conditions and their computed contrast (left click-right click) using NP-JPDE and JDE models.

displayed along with the canonical HRF and the one estimated using the JDE model. The computed TTP for the HRF profiles of parcels 1, 2 and 4 is TTP = 5.4 s, while for parcel 3 we have TTP = 6.0 s. The FWHM was also computed and is equal to 4.2 s for parcels 1 and 4, and to 4.8 s for parcels 2 and 3. As regards the HRF estimated using the JDE model, we have TTP = 5.4 s and FWHM = 4.2 s. Moreover, Tab. 10 reports the computed Euclidean distances between the different HRF estimates and the canonical HRF. It also reports the same distance between the individual NP-JPDE HRF estimates. The reported distances indicate that the NP-JPDE model provides closer HRF estimates to the canonical shape with average Euclidean distance of 0.42. More interestingly, it is clear that the NP-JPDE model is able to discriminate between parcels that have very close HRFs in terms of Euclidean distance, namely those of parcels 1 and 2. Indeed, these two parcels have similar TTPs, but different FWHM values. They are therefore detected as different parcels by the NP-JPDE model. Fig. 17 shows the NRL estimates for some of the experimental conditions which are supposed to induce evoked activity in the bilateral occipital cortex (namely, video calculations, video sentences, horizontal checkerboard and vertical checkerboard). The obtained NRL estimates with the NP-JPDE and the JDE are similar in terms of amplitude values and the location of the highest activation.



(a) Slice.1 (b) Slice.2

Figure 15: Consecutive slices of the estimated parcellation located in the occipital cortex.

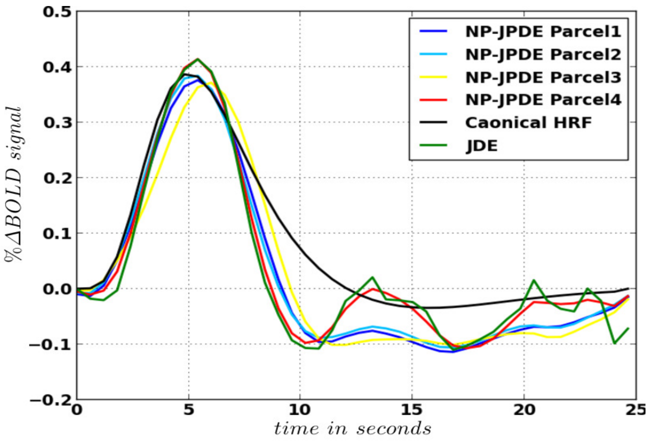
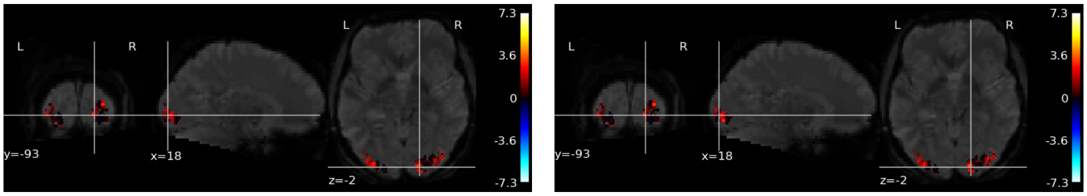


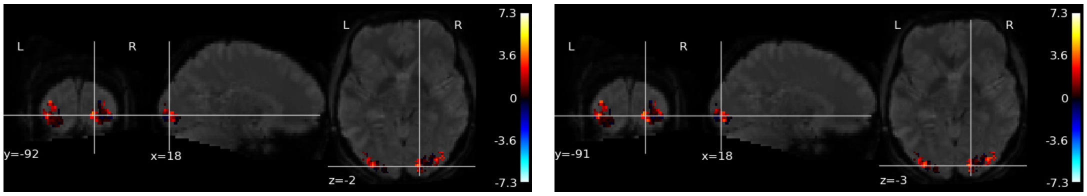
Figure 16: HRF shape estimates using the NP-JPDE and JDE models in the bilateral occipital cortex and the canonical HRF shape.

Table 10: Euclidean distance between the HRF estimates in the bilateral occipital cortex and the canonical HRF. Dinstance between the individual NP-JPDE HRF estimates are also provided.

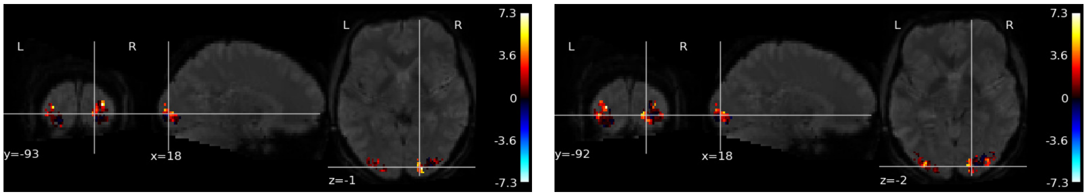
	HRF. 1	HRF. 2	HRF. 3	HRF. 4	JDE
Canonical HRF	0.42	0.41	0.43	0.41	0.47
HRF. 2	0.06	–	0.22	0.20	–
HRF. 3	0.17	–	–	0.35	–
HRF. 4	0.23	–	–	–	–



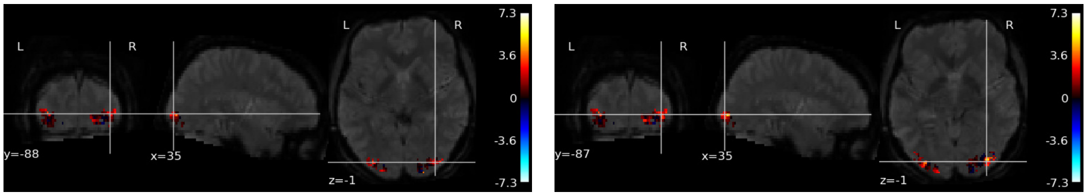
(a) Video calculations (NP-JPDE) (b) Video calculations (JDE)



(c) Video sentences (NP-JPDE) (d) Video sentences (JDE)



(e) Horizontal checkerboard (NP-JPDE) (f) Horizontal checkerboard (JDE)



(g) Vertical checkerboard (NP-JPDE) (h) Vertical checkerboard (JDE)

Figure 17: NRL estimates for the visual sentences and calculation experimental conditions using NP-JPDE and JDE models.

5. Discussion and conclusion

In this paper, we proposed a new approach to estimate the number of hemodynamic parcels in fMRI data analysis where model selection was formulated as a clustering issue. This approach is based on a Dirichlet process mixture model combined with a hidden Markov random field. A direct generalization of the Potts model that uses a stick breaking representation allows for the representation of an infinite number of states. The advantage of the non-parametric framework was to allow an automatic estimation of number of parcels from the fMRI data. The use of a hidden Markov random field accounted for the spatial constraints on the connexity of the estimated parcels. The JPDE model, proposed in (Chaari et al., 2012, 2015), was extended using this non-parametric Bayesian formulation yielding the so called NP-JPDE model. Following its ancestor (the JPDE), the NP-JPDE relies on the VEM as an inference strategy. However, the new layers in the hierarchy of the NP-JPDE model result in two new expectations steps (namely, VE$-Z$ and VE$-\boldsymbol{\tau}$ steps) while the others remain the same as in the classical JPDE model. Also, the maximization of the interaction parameter of the Potts model (β_z) over the parcellation labels is not straightforward and has been changed to account for the dependency of the parcellation labels variable $\boldsymbol{z}$ on the stick breaking length $\boldsymbol{\tau}$ of the DPMM. The proposed model is extended also by injecting a new prior on the concentration parameter α of the DPMM which allows for its automatic estimation through the VEM iterations.

Synthetic and real data experiments were used to validate the proposed approach. Using synthetic data experiments, we studied the performance of our approach in estimating an accurate parcellation and HRFs and detecting the task-related activity. Experiments with different scenarios were conducted using synthetic data generated for a different number of parcels. Our results were consistent with the classical JPDE model that relies on the model selection procedure (Albughdadi et al., 2014) in terms of HRF recovery and evoked activation detection. However, the proposed NP-JPDE extension outperformed the original JPDE formulation in terms of parcellation inference. We also investigated the performance of the NP-JPDE in terms of convergence speed and computational time, and we showed again its superiority over its ancestor. As regards the NRL estimates, the NP-JPDE managed to preserve the robust performance of the classical JPDE in detecting the task-related activity. On real fMRI data, we used two ROIs to validate the proposed approach, the right motor cortex and the bilateral occipital area embodying the primary visual cortices. In the right motor cortex, two different parcels were estimated with HRF estimates close to the canonical HRF. These results came consistent with the HRF estimate of the JDE model and with the conclusion in (Badillo et al., 2013b). In the bilateral occipital cortex, the left and the right parcels showed similar hemodynamic territories. The HRF estimates with the NP-JPDE were close to the canonical HRF especially in terms of TTP and they were better recovered than using the JDE model. For both experiments, the NRL estimates using the NP-JPDE model were coherent with those yielded by analysing the same fMRI data using the JDE model. Future work will focus on finding a more efficient approach to estimate the parameter β_z of the Potts model to avoid the computational complexity of the theoretical estimation provided in this paper. A further extension of this model will be applied to multi-subject studies to derive a meaningful group-level parcellation and HRF estimates in a non-parametric framework.

Appendix A. Free energy-based model selection

This section summarizes the technique proposed in (Albughdadi et al., 2014) relying on the calculation of the free energy for model selection (Attias, 2000; Beal, 2003). This technique is based on the JPDE model (Chaari et al., 2012, 2015). It is worth noting that when the sample size is large enough, the free energy approximates the well known BIC and MDL criteria. If we consider Ω different candidate models, the number of parcels in each of them will be denoted by K^ω where $\omega \in \{1, ..., \Omega\}$. Here, we use different notations for the parameters $\boldsymbol{\Psi}$ and the hyperparameters $\boldsymbol{\Phi}$ where $\Psi = \{\bar{\boldsymbol{h}}, \beta_z, \boldsymbol{\beta}\}$ and $\boldsymbol{\Phi} = \{\boldsymbol{L}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{\upsilon}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \gamma_z, \sigma_h^2\}$. After running the adopted VEM algorithm, for each model order we end up with estimated posteriors denoted as $\tilde{p}_A^{(\infty)}, \tilde{p}_H^{(\infty)}, \tilde{p}_Q^{(\infty)}, \tilde{p}_z^{(\infty)}$ and pointwise estimates $\boldsymbol{\Psi}^{(\infty)}$, $\boldsymbol{\Phi}^{(\infty)}$. Following (Albughdadi et al., 2014), we use ω with the previous quantities to refer to the model ω , *i.e*, we write p_A^ω instead of $\tilde{p}_A^{(\infty)}$. The free energy reads

$$\mathcal{F}(p^\omega, \boldsymbol{\Psi}^\omega, \boldsymbol{\Phi}^\omega) = \mathrm{E}_{p^\omega}\left[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{H}, \boldsymbol{z}, \boldsymbol{\Psi}^\omega; \boldsymbol{\Phi}^\omega)\right] + \mathcal{G}(p^\omega) \tag{A.1}$$

where p^ω is the factorized posterior approximation:
$p^\omega(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{H}, \boldsymbol{z}) = p_A^\omega(\boldsymbol{A}) p_Q^\omega(\boldsymbol{Q})\ p_H^\omega(\boldsymbol{H})\ p_z^\omega(\boldsymbol{z})$. The free energy in Eq. (A.1) can therefore be rewritten as

$$\begin{aligned}\mathcal{F}(p^\omega, \boldsymbol{\Psi}^\omega, \boldsymbol{\Phi}^\omega) = {} & \mathrm{E}_{p_A^\omega p_H^\omega}\left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{\Phi}^\omega)\right] + \mathrm{E}_{p_A^\omega p_Q^\omega}\left[\log p(\boldsymbol{A} \mid \boldsymbol{Q}; \boldsymbol{\Phi}^\omega)\right] \\ & + \mathrm{E}_{p_Q^\omega}\left[\log p(\boldsymbol{Q} | \boldsymbol{\beta}^\omega)\right] + \mathrm{E}_{p_H^\omega p_z^\omega}\left[\log p(\boldsymbol{H} \mid \boldsymbol{z}, \bar{\boldsymbol{h}}^\omega; \boldsymbol{\Phi}^\omega)\right] \\ & + \mathrm{E}_{p_z^\omega}\left[\log p(\boldsymbol{z} | \beta_z^\omega)\right] + \log p(\bar{\boldsymbol{h}}^\omega; \boldsymbol{\Phi}^\omega) + \log p(\boldsymbol{\beta}^\omega; \boldsymbol{\Phi}^\omega) \\ & + \log p(\beta_z^\omega; \boldsymbol{\Phi}^\omega) + \mathcal{G}(p_A^\omega) + \mathcal{G}(p_Q^\omega) + \mathcal{G}(p_H^\omega) + \mathcal{G}(p_z^\omega).\end{aligned} \tag{A.2}$$

Each of the above terms can be calculated from the outputs of the VEM algorithm based on the estimated posteriors and hyperparameters. For more details about these calculations, the interested reader can refer to Appendix A[5]. The terms of the free energy appearing in Eq. (A.2) are computed after convergence of the JPDE VEM model. The exact mathematical calculations are detailed below.

---

[5] This free energy has to be calculated once the convergence of the VEM algorithm has been reached for each model. The best model is therefore the one associated with the highest free energy value.

(1) $\mathrm{E}_{p_A^\omega p_H^\omega}\left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{\Phi}^\omega)\right]$
This term corresponds to $E_{p_A^\omega p_H^\omega}\left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}^\omega, \boldsymbol{\Gamma}^\omega)\right]$ and it can be computed as follows

$$
\begin{aligned}
E_{p_A^\omega p_H^\omega}\left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}^\omega, \boldsymbol{\Gamma}^\omega)\right] &= -\frac{NJ}{2}\log 2\pi + \frac{J}{2}\log|\boldsymbol{\Lambda}_j| - N\sum_{j=1}^{J}\log\sigma_j \\
&+ \sum_{j}^{J}\boldsymbol{m}_{H_j}^{\mathrm{t}}\Big(\sum_{m=1}^{M} m_{A_j^m}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j^\omega(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\Big) - \frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j^\omega(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \\
&- \frac{1}{2}\sum_{j=1}^{J}\sum_{m,m'}\Big\{\Big(m_{A_j^m} m_{A_j^{m'}} + v_{A_j^m A_j^{m'}}\Big) \\
&\times\Big[\boldsymbol{m}_{H_j}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j^\omega\boldsymbol{X}^{m'}\boldsymbol{m}_{H_j} + \mathit{trace}\,\Big(\boldsymbol{\Sigma}_{H_j}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j^\omega\boldsymbol{X}^{m'}\Big)\Big]\Big\} \qquad \text{(A.3)}
\end{aligned}
$$

(2) $\mathrm{E}_{p_A^\omega p_Q^\omega}\left[\log p(\boldsymbol{A} \mid \boldsymbol{Q}; \boldsymbol{\Phi}^\omega)\right]$
This term is equivalent to $E_{p_A^\omega p_Q^\omega}\left[\log p(\boldsymbol{A} \mid \boldsymbol{Q}, \boldsymbol{\mu}^\omega, \boldsymbol{v}^\omega)\right]$. Straightforward calculations lead to the following

$$
\begin{aligned}
&E_{p_A^\omega p_Q^\omega}\left[\log p(\boldsymbol{A} \mid \boldsymbol{Q}, \boldsymbol{\mu}, \boldsymbol{v})\right] \\
&= \sum_{m=1}^{M}\sum_{j=1}^{J}\left\{\left[1 - p_{q_j^m}^\omega(1)\right]\left[\log\frac{1}{\sqrt{2\pi v_{m0}^\omega}} - \frac{(m_{A_j^m} - \mu_{m0}^\omega)^2 + v_{A_j^m A_j^{m'}}}{2v_{m0}^\omega}\right]\right. \\
&\quad\left. + p_{q_j^m}^\omega(1)\left[\log\frac{1}{\sqrt{2\pi v_{m1}^\omega}} - \frac{(m_{A_j^m} - \mu_{m1}^\omega)^2 + v_{A_j^m A_j^{m'}}}{2v_{m1}^\omega}\right]\right\} \qquad \text{(A.4)}
\end{aligned}
$$

(3) $\mathrm{E}_{p_Q^\omega}\left[\log p(\boldsymbol{Q}|\boldsymbol{\beta}^\omega)\right]$
Using straightforward calculations we obtain

$$
E_{p_Q^\omega}\left[\log p(\boldsymbol{Q} \mid \boldsymbol{\beta}^\omega)\right] = \sum_{m=1}^{M}\left\{-\log W(\beta_m^\omega) + \beta_m^\omega E_{p_{\boldsymbol{q}^m}}\left[U(\boldsymbol{q}^m)\right]\right\} \qquad \text{(A.5)}
$$

with

$$
E_{p_{Q^m}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_{j=1}^{J}\sum_{l\sim j}E_{p_{Q_j^m}p_{Q_l^m}}\left[I(q_j^m = q_l^m)\right] = \frac{1}{2}\sum_{j}\sum_{l\sim j}\sum_{i=0}^{1}p_{Q_j^m}^\omega(i)\,p_{Q_l^m}^\omega(i) \qquad \text{(A.6)}
$$

- **Computing $W(\beta_m^\omega)$:** Using a mean field approximation, we obtain

$$W(\beta_m^\omega) \simeq W_{mf}(\beta_m^\omega) \exp\left(\beta_m^\omega E_{p_{mf}}\left[U(\boldsymbol{q}^m) - U_{mf}(\boldsymbol{q}^m)\right]\right) \tag{A.7}$$

where $U_{mf}(\boldsymbol{q}^m)$,$W_{mf}(\beta_m^\omega)$ and p_{mf} have been derived in (Bakhous, 2013)

$$\begin{aligned} W(\beta_m^\omega) \;\simeq\; & \prod_{j=1}^{J} \sum_{i=0}^{1} \exp\left(\beta_m^\omega \sum_{l\sim j} p^\omega_{Q_l^m}(i)\right) \\ & \times \exp\left(\beta_m^\omega \sum_{j} \sum_{l\sim j} \sum_{i=0}^{1} \left[p_{mfj}(i)\left(\frac{p_{mfl}(i)}{2} - p^\omega_{Q_l^m}(i)\right)\right]\right) \end{aligned} \tag{A.8}$$

with $P_{mfj}(i) = \dfrac{\exp\left(\beta_m^\omega \sum_{l\sim j} p^\omega_{Q_l^m}(i)\right)}{\sum_{i\in\{0,1\}} \exp\left(\beta_m^\omega \sum_{l\sim j} p^\omega_{Q_l^m}(i)\right)}$.

(4) $\mathrm{E}_{p_H^\omega p_z^\omega}\left[\log p(\boldsymbol{H} \mid \boldsymbol{z}, \bar{\boldsymbol{h}}^\omega; \boldsymbol{\Phi}^\omega)\right]$
This term is equivalent to $\mathrm{E}_{p_H^\omega p_z^\omega}\left[\log p(\boldsymbol{H} \mid \boldsymbol{z}, \bar{\boldsymbol{h}}^\omega; \boldsymbol{\nu}^\omega)\right]$ and can be computed as follows

$$\begin{aligned} & \mathrm{E}_{p_H^\omega p_z^\omega}\left[\log p(\boldsymbol{H} \mid \boldsymbol{z}, \bar{\boldsymbol{h}}^\omega; \boldsymbol{\nu}^\omega)\right] \\ & = \sum_{j=1}^{J} \sum_{k=1}^{K} p^\omega_{z_j}(k)\left(-\frac{D+1}{2}\log 2\pi - \frac{D+1}{2}\log \nu_k^\omega\right) \\ & - \sum_{j=1}^{J} \sum_{k=1}^{K} p^\omega_{z_j}(k)\left(\frac{1}{2\nu_k^\omega} E_{p_{H_j}}[(H_j - \bar{\boldsymbol{h}}_k^\omega)^T (H_j - \bar{\boldsymbol{h}}_k^\omega)]\right) \\ & = \sum_{j=1}^{J} \sum_{k=1}^{K} p^\omega_{z_j}(k)\left(-\frac{D+1}{2}\log 2\pi - \frac{D+1}{2}\log \nu_k^\omega\right) \\ & - \sum_{j=1}^{J} \sum_{k=1}^{K} p^\omega_{z_j}(k)\left(\frac{1}{2\nu_k^\omega}[(m_{H_j} - \bar{\boldsymbol{h}}_k^\omega)^T (m_{H_j} - \bar{\boldsymbol{h}}_k^\omega) + \mathrm{trace}(\boldsymbol{\Sigma}_{H_j})\right) \end{aligned} \tag{A.9}$$

(5) $\mathrm{E}_{p_z^\omega}\left[\log p(\boldsymbol{z}|\beta_z^\omega)\right]$

$$E_{p_z^\omega}\left[\log p(\boldsymbol{z} \mid \beta_z^\omega)\right] = -\log W^z(\beta_z^\omega) + \beta_z^\omega E_{p_z}\left[U(\boldsymbol{z})\right] \tag{A.10}$$

with

$$E_{p_z^\omega}[U(\boldsymbol{z})] = \frac{1}{2}\sum_{j=1}^{J}\sum_{l\sim j} E_{p_{z_j}p_{z_l}}[I(z_j = z_l)] = \frac{1}{2}\sum_{j=1}^{J}\sum_{l\sim j}\sum_{i=1}^{K} p_{z_j}^{\omega}(i)p_{z_l}^{\omega}(i)$$

- **Computing $W^z(\beta_z^\omega)$:** Using a mean field approximation yields

$$W^z(\beta_z^\omega) \simeq W_{mf}^z(\beta_z^\omega)\exp\left(\beta_z^\omega E_{p_{mf}}\left[U(\boldsymbol{z}) - U_{mf}(\boldsymbol{z})\right]\right) \tag{A.11}$$

where $U_{mf}(\boldsymbol{z})$, $W_{mf}^z(\beta_z^\omega)$ and p_{mf} have been derived in (Bakhous, 2013)

$$\begin{aligned} W^z(\beta_z^\omega) \;\simeq\; & \prod_{j=1}^{J}\sum_{i=1}^{K}\exp\left(\beta_z^\omega\sum_{l\sim j}p_{z_l}^{\omega}(i)\right) \\ & \times\exp\left(\beta_z^\omega\sum_{j=1}^{J}\sum_{l\sim j}\sum_{i=1}^{K}\left[p_{mfj}(i)\left(\frac{p_{mfl}(i)}{2} - p_{z_l}^{\omega}(i)\right)\right]\right) \end{aligned} \tag{A.12}$$

with $P_{mfj}(i) = \dfrac{\exp\left(\beta_z^\omega\sum_{l\sim j}p_{z_l}^{\omega}(i)\right)}{\sum_{i\in\{1,\dots,K\}}\exp\left(\beta_z^\omega\sum_{l\sim j}q_{z_l}(i)\right)}$.

(6) $\log p(\bar{\boldsymbol{h}}^\omega;\boldsymbol{\Phi}^\omega)$

This term corresponds to $\log p(\bar{\boldsymbol{h}}^\omega;\sigma_h^{2(\omega)})$ whose computation is summarized below

$\log p(\bar{\boldsymbol{h}}^\omega;\sigma_h^{2(\omega)}) = \sum_{k=1}^{K}\log p(\bar{\boldsymbol{h}}_k^\omega;\sigma_h^{2(\omega)})$ where $p(\bar{\boldsymbol{h}}_k^\omega;\sigma_h^{2(\omega)})$ is a zero mean Gaussian distribution with covariance matrix $\sigma_h^{2(\omega)}\boldsymbol{R}$

$$\log p(\bar{\boldsymbol{h}}^\omega;\sigma_h^2) = \sum_{k=1}^{K}\log\mathcal{N}(\bar{\boldsymbol{h}}_k^\omega;0,\sigma_h^2\boldsymbol{R}) \tag{A.13}$$

(7) $\log p(\boldsymbol{\beta}^\omega;\boldsymbol{\Phi}^\omega)$

This term is equivalent to $\log p(\boldsymbol{\beta}^\omega;\lambda^\omega)$ which is an exponential distribution with parameter λ^ω

$$\log p(\boldsymbol{\beta}^\omega;\lambda^\omega) = \lambda^\omega\exp(-\lambda^\omega\boldsymbol{\beta}^\omega) \tag{A.14}$$

where $\boldsymbol{\beta}^\omega \geqslant 0$.

(8) $\log p(\beta_z^\omega; \mathbf{\Phi}^\omega)$
Following the same steps as above, this term is equivalent to $\log p(\beta_z^\omega; \lambda^\omega)$, and it is also an exponential distribution with parameter λ_z^ω.

$$\log p(\hat{\beta}_z; \lambda_z^\omega) = \lambda_z^\omega \exp(-\lambda_z^\omega \beta_z^\omega) \tag{A.15}$$

where $\beta_z^\omega \geqslant 0$.

(9) $\mathcal{G}(p_A^\omega) = \sum_{j=1}^{J} \sum_{m=1}^{M} \mathcal{G}(p_{A_j^m}^\omega)$
To calculate this term,we use the expression of the Gaussian entropy

$$\mathcal{G}(p_A^\omega) = \sum_{j=1}^{J} \sum_{m=1}^{M} \mathcal{G}(p_{a_j^m}^\omega) = \frac{1}{2} \sum_{j=1}^{J} \sum_{m=1}^{M} \log \left(\left| \Sigma_{a_j^m} \right| (2\pi e)^D \right) \tag{A.16}$$

(10) $\mathcal{G}(p_H^\omega) = \sum_{j=1}^{J} \mathcal{G}(p_{H_j}^\omega)$
The Gaussian entropy is used as well to calculate this term

$$\mathcal{G}(p_H^\omega) = \sum_{j=1}^{J} \mathcal{G}(p_{H_j}^\omega) = \frac{1}{2} \sum_{j=1}^{J} \log \left(\left| \Sigma_{H_j} \right| (2\pi e)^D \right) \tag{A.17}$$

(11) $\mathcal{G}(p_Q^\omega) = \sum_{j=1}^{J} \sum_{m=1}^{M} \mathcal{G}(p_{Q_j^m}^\omega)$
Using the definition of the entropy for a discrete random variable, the following results is obtained

$$\mathcal{G}(p_Q^\omega) = -\sum_{j=1}^{J} \sum_{m=1}^{M} p_{Q_j^m}^\omega \log_2 \left(p_{Q_j^m}^\omega \right) \tag{A.18}$$

(12) $\mathcal{G}(p_z^\omega) = \sum_{j=1}^{J} \mathcal{G}(p_{z_j}^\omega)$
As $\mathcal{G}(p_Q^\omega)$, we obtain

$$\mathcal{G}(p_z^\omega) = -\sum_{j} p_{z_j}^\omega \log_2 \left(p_{z_j}^\omega \right) \tag{A.19}$$

Appendix B. Estimating the parameter $\boldsymbol{\beta_z}$ in the NP-JPDE model

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 (29) with respect to β_z admits a unique solution. Indeed, it is equivalent to solve

$$\beta_z^{(r)} \;=\; \arg\max_{\beta_z} \mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\right] - \mathrm{E}_{\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\log\mathcal{K}(\beta_z,\boldsymbol{\tau})\right] \tag{B.1}$$

where $\mathcal{K}$ denotes the normalizing constant that depends on $\boldsymbol{\tau}$ and β_z. Denoting the gradient vector and Hessian matrix respectively by ∇_{β_z} and $\nabla^2_{\beta_z}$, it comes

$$\begin{aligned}&\nabla_{\beta_z}\mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\log p(\boldsymbol{z}|\boldsymbol{\tau};\beta_z)\right]\\ &= \mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\nabla_{\beta_z}V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\right] - \mathrm{E}_{p(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\nabla_{\beta_z}V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\right]\end{aligned} \tag{B.2}$$

$$\begin{aligned}&\nabla^2_{\beta_z}\mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\log p(\boldsymbol{z}|\boldsymbol{\tau};\beta_z)\right]\\ &= \mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\nabla^2_{\beta_z}V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\right] - \mathrm{E}_{p(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\nabla^2_{\beta_z}V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\right]\\ &- \mathrm{E}_{\widetilde{p}_{\boldsymbol{\tau}}^{(r)}}\left[\mathrm{var}_{p(\boldsymbol{z};\boldsymbol{\tau},\beta_z)}[\nabla_{\beta_z}V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)]\right]\,.\end{aligned} \tag{B.3}$$

The last expectations in (B.2) and (B.3) are taken over the Potts prior (14). It follows that whenever $V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)$ is linear in β_z, $\nabla^2_{\beta_z}V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)$ is zero, the Hessian matrix is a semi-definite negative matrix and the function to optimize is concave. Unfortunately, due to the intractable normalizing constant $\mathcal{K}$, expressions (B.2) and (B.3) are not directly available. It is necessary to approximate the terms involving the true MRF prior $p(\boldsymbol{z};\boldsymbol{\tau},\beta_z)$ using an approximation. A natural approach is to use

$$\widetilde{p}_z^{prior}(\boldsymbol{z};\boldsymbol{\tau},\beta_z) = \prod_{j=1}^{J}\widetilde{p}_{z_j}^{prior}(z_j;\boldsymbol{\tau},\beta_z) \tag{B.4}$$

with $\widetilde{p}_{z_j}^{prior}(z_j;\boldsymbol{\tau},\beta_z)$ defined by

$$\widetilde{p}_{z_j}^{prior}(z_j=k;\boldsymbol{\tau},\beta_z) = \frac{\exp(\log\pi_k + \beta_z\sum\limits_{i\sim j}\widetilde{p}_{z_i}^{(r-1)}(k))}{\sum\limits_{l=1}^{K}\exp(\log\pi_l + \beta_z\sum\limits_{i\sim j}\widetilde{p}_{z_i}^{(r-1)}(l))}\,. \tag{B.5}$$

This MRF prior approximation induced by the posterior variational approximation has been proposed in (Celeux et al., 2003) and also exploited in (Chaari et al., 2013). Similarly, we obtain

$$\mathrm{E}_{\widetilde{p}_z^{(r)}\widetilde{p}_\tau^{(r)}}\left[\nabla_{\beta_z} V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\right] = \textstyle\sum_{k=1}^{K}\sum_{i\sim j}\widetilde{p}_{z_j}(k)\widetilde{p}_{z_i}(k) \tag{B.6}$$

$$\mathrm{E}_{p(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\widetilde{p}_\tau^{(r)}}\left[\nabla_{\beta_z} V(\boldsymbol{z};\boldsymbol{\tau},\beta_z)\right] = \sum_{k=1}^{K}\sum_{i\sim j}\mathrm{E}_{\widetilde{p}_\tau^{(r)}}\left[\widetilde{p}_{z_j}^{prior}(k;\boldsymbol{\tau},\beta_z)\widetilde{p}_{z_i}^{prior}(k;\boldsymbol{\tau},\beta_z)\right]. \tag{B.7}$$

The additional difficulty is that we have to compute an additional expectation with regards to $\widetilde{p}_\tau^{(r)}$. The last expectation with respect to $\widetilde{p}_\tau^{(r)}$ can be approximated using Monte Carlo sums. Since $\widetilde{p}_\tau^{(r)}$ is the product of $K-1$ independent beta distributions defined in Eq. (20) and Eq. (21), it is straightforward to draw T realizations of $(\tau_1,\ldots,\tau_{K-1})$. Denoting these realizations as $\boldsymbol{\tau}^{(t)}$ for $t=1,\ldots,T$, the following approximation is obtained

$$\begin{aligned}&\mathrm{E}_{\widetilde{p}_\tau^{(r)}}\left[\sum_{k=1}^{K}\sum_{i\sim j}\widetilde{p}_{z_j}^{prior}(k;\boldsymbol{\tau},\beta_z)\widetilde{p}_{z_i}^{prior}(k;\boldsymbol{\tau},\beta_z)\right]\\&=\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{k=1}^{K}\sum_{i\sim j}\widetilde{p}_{z_j}^{prior}(k;\boldsymbol{\tau}^{(t)},\beta_z)\widetilde{p}_{z_i}^{prior}(k;\boldsymbol{\tau}^{(t)},\beta_z)\right].\end{aligned}$$

An approximation of the gradient in (Eq. (B.2))can then be computed. Setting this approximate gradient to zero leads to an equation that can be solved to provide β_z.

References

Albughdadi, M., Chaari, L., Forbes, F., Tourneret, J.-Y., Ciuciu, P., Sept 2014. Model selection for hemodynamic brain parcellation in fMRI. In: Proc. EUSIPCO. Lisbon, Portugal, pp. 31–35.

Attias, H., 2000. A variational Bayesian framework for graphical models. In: Proc. Advances in Neural Information Processing Systems 12. Denver, Colorado, United States.

Badillo, S., Varoquaux, G., Ciuciu, P., June 2013a. Hemodynamic estimation based on consensus clustering. In: Proc. PRNI. Philadelphia, USA, pp. 211–215.

Badillo, S., Vincent, T., Ciuciu, P., 15 Nov. 2013b. Group-level impacts of within- and between-subject hemodynamic variability in fMRI. Neuroimage 82, 433–448.

Bakhous, C., 2013. Modèles d'encodage parcimonieux de l'activité cérébrale mesurée par irm fonctionnelle. Ph.D. thesis, University of Grenoble.

Beal, M. J., 2003. Variational algorithms for approximate bayesian inference. Ph.D. thesis, University of London.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B (Methodological), 192–236.

Blei, D. M., Jordan, M. I., et al., 2006. Variational inference for Dirichlet process mixtures. Bayesian Analysis 1 (1), 121–143.

Boynton, G. M., Engel, S. A., Glover, G. H., Heeger, D. J., 1996. Linear systems analysis of functional magnetic resonance imaging in human v1. The Journal of Neuroscience 16 (13), 4207–4221.

Buxton, R. B., Frank, L., 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. Journal of Cerebral Blood Flow & Metabolism 17 (1), 64–72.

Celeux, G., Forbes, F., Peyrard, N., 2003. EM procedures using mean field-like approximations for Markov model-based image segmentation. Pattern Recognition 36, 131–144.

Chaari, L., Badillo, S., Vincent, T., Dehaene-Lambertz, G., Forbes, F., Ciuciu, P., Nov. 2015. Hemodynamic-Informed Parcellation of fMRI Data in a Joint Detection Estimation Framework, submitted to IEEE Transactions on Medical Imaging.

Chaari, L., Ciuciu, P., Mériaux, S., Pesquet, J.-C., 2014. Spatio-temporal wavelet regularization for parallel mri reconstruction: application to functional mri. Magnetic Resonance Materials in Physics, Biology and Medicine (MAGMA)DOI: 10.1007/s10334-014-0436-5.

Chaari, L., Forbes, F., Vincent, T., Ciuciu, P., 2012. Hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework. In: et al., N. A. (Ed.), Medical Image Computing and Computer-Assisted Intervention. Vol. 7512. Springer, pp. 180–188.

Chaari, L., Forbes, F., Vincent, T., Dojat, M., Ciuciu, P., 2011. Variational solution to the joint detection estimation of brain activity in fMRI. In: Fichtinger, G., Martel, A., Peters, T. (Eds.), Medical Image Computing and Computer-Assisted Intervention. Vol. 6892. Springer, pp. 260–268.

Chaari, L., Vincent, T., Forbes, F., Dojat, M., Ciuciu, P., 2013. Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach. IEEE Trans. Med. Imag. 32 (5), 821–837.

Chatzis, S. P., Tsechpenakis, G., 2010. The infinite hidden Markov random field model. IEEE Trans. Neural Networks 21 (6), 1004–1014.

Chickering, D. M., Heckerman, D., 1997. Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. Machine Learning 29 (2-3), 181–212.

Ciuciu, P., Idier, J., Roche, A., Pallier, C., Apr. 2004. Outlier detection for robust region-based estimation of the hemodynamic response function in event-related fMRI. In: IEEE International Symposium on Biomedical Imaging. Arlington, VA USA, pp. 392–395.

Ciuciu, P., Poline, J., Marrelec, G., Idier, J., Pallier, C., Benali, H., Oct 2003. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. IEEE Transactions on Medical Imaging 22 (10), 1235–1251.

Deneux, T., Faugeras, O., 2006. Using nonlinear models in fMRI data analysis: model selection and activation detection. Neuroimage 32 (4), 1669–1689.

Eickhoff, S., Bzdok, D., Laird, A., Roski, C., Caspers, S., Zilles, K., Fox, P., 2011. Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. Neuroimage 57 (3), 938–949.

Escobar, M. D., West, M., 1995. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association 90 (430), 577–588.

Ferguson, T. S., 1973. A Bayesian analysis of some nonparametric problems. The Annals of Statistics, 209–230.

Flandin, G., Kherif, F., Pennec, X., Rivière, D., Ayache, N., Poline, J.-B., 2002. A new representation of fMRI data using anatomo-functional constraints. In: Proc. Neuroimage (HBM'02). Sendai, Japan.

Fouque, A.-L., Ciuciu, P., Risser, L., April 2009. Multivariate spatial Gaussian mixture modeling for statistical clustering of hemodynamic parameters in functional MRI. In: Proc. ICASSP. Taipei, Taiwan, pp. 445–448.

Frau-Pascual, A., Vincent, T., Forbes, F., Ciuciu, P., May 2014. Hemodynamically informed parcellation of cerebral fMRI data. In: Proc. ICASSP. Florence, Italy, pp. 2079–2083.

Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., Turner, R., 1995. Analysis of fMRI time-series revisited. Neuroimage 2 (1), 45–53.

Friston, K. J., Mechelli, A., Turner, R., Price, C. J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. Neuroimage 12 (4), 466–477.

Glover, G. H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9 (4), 416–429.

Golland, P., Golland, Y., Malach, R., 2007. Detection of spatial activation patterns as unsupervised segmentation of fMRI data. In: Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 110–118.

Handwerker, D., Gonzalez-Castillo, J., D'Esposito, M., Bandettini, P., 2012. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. Neuroimage 62 (2), 1017–1023.

Handwerker, D. A., Ollinger, J., D'Esposito, M., Apr. 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses 21, 1639–1651.

Henson, R. N. A., Rugg, M. D., Friston, K. J., 2001. The choice of basis functions in event-related fMRI. Neuroimage 13 (6), 149.

Kershaw, J., Ardekani, B. A., Kanno, I., 1999. Application of Bayesian inference to fMRI data analysis. IEEE Trans. Med. Imag. 18 (12), 1138–1153.

Lashkari, D., Sridharan, R., Vul, E., Hsieh, P.-J., Kanwisher, N., Golland, P., 2012. Search for patterns of functional specificity in the brain: a nonparametric hierarchical Bayesian model for group fMRI data. Neuroimage 59 (2), 1348–1368.

Lashkari, D., Vul, E., Kanwisher, N., Golland, P., 2010. Discovering structure in the space of fMRI selectivity profiles. Neuroimage 50 (3), 1085–1098.

Lindquist, M. A., Loh, J. M., Atlas, L. Y., Wager, T. D., 2009. Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. Neuroimage 45 (1), S187–S198.

Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., Ciuciu, P., 2008. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. Neuroimage 41 (3), 941–969.

Ogawa, S., Lee, T.-M., Kay, A. R., Tank, D. W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. 87 (24), 9868–72.

Pinel, P., Thirion, B., Mériaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.-B., Dehaene, S., 2007. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neuroscience 8 (1), 91.

Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., Kawashima, R., 2004. A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. Neuroimage 21 (2), 547–567.

Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B., 2006. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Human Brain Mapping 27 (8), 678–693.

Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.-B., 2014. Which fMRI clustering gives good brain parcellations? Frontiers in neuroscience 8.

Vincent, T., Badillo, S., Risser, L., Chaari, L., Bakhous, C., Forbes, F., Ciuciu, P., 2014. Flexible multivariate hemodynamics fMRI data analyses and simulations with PyHRF. Frontiers in Neuroscience 8 (67).

Vincent, T., Risser, L., Ciuciu, P., 2010. Spatially adaptive mixture modeling for analysis of fMRI time series. IEEE Trans. Med. Imag. 29 (4), 1059–1074.

Woolrich, M. W., Ripley, B. D., Brady, M., Smith, S. M., 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. Neuroimage 14 (6), 1370–1386.

Vitae

- Mohanad ALBUGHDADI received his B.Sc. in Computer Engineering from IUG in Palestine in 2010. In 2011, he attained his M.Sc. in Electronic Systems Design Engineering from the University of Science in Malaysia where he was awarded as the Best Student of the year. After obtaining the French government scholarship to pursue his doctoral degree, he is currently enrolled as a Ph.D. student at (ENSEEIHT) the National Polytechnic Institute of Toulouse, and the IRIT laboratory in France. His research focus is on fMRI analysis, image and signal processing, image clustering and Bayesian inference.

- Lotfi CHAARI (IEEE M'08) received the engineering and master degrees from the High School of Telecommunication in Tunis (SUP'COM) in 2007. In 2010 he received the PhD degree from the university of Paris-Est. He prepared his PhD with the Signal and Communication group of the Laboratoire d'Informatique Gaspard Monge (LIGM), an associated lab of the Centre national de la Recherche Scientifique (UMR 8049 - CNRS). He has been working during his PhD on medical image reconstruction for Magnetic Resonance Imaging, in collaboration with NeuroSpin-CEA, the first center in Europe dedicated to ultra-high field MRI and applications in cognitive neurosciences. He then moved to Grenoble as a postdoctoral fellow at INRIA with the Mistis team where he worked on functional MRI data analysis and detection-estimation approaches. He is assistant professor at INP-Toulouse since 2012 and member of the TCI team of the IRIT laboratory. Prof. Chaari's research field covers image reconstruction and restoration techniques under sparsity constraints, especially in application to medical imaging. He is also interested in variational, Bayesian and variational Bayesian approaches for image processing and analysis. He has been involved in different technical committees of the signal processing society and co-organized several special sessions on medical imaging.

- Jean-Yves TOURNERET (SM'08) received the ingénieur degree in electrical engineering from in 1989 and the Ph.D. degree in 1992 from (ENSEEIHT) the National Polytechnic Institute of Toulouse. He is currently a professor in the university of Toulouse (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo (MCMC) methods. He has been involved in the organization of several conferences and has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001-2007, 2010-present). He has been serving as an associate editor for the IEEE Transactions on Signal Processing (2008-2011, 2015-present) and for the EURASIP journal on Signal Processing (2013-present).

- Florence FORBES is the director of research at INRIA and head of the MISTIS team since 2003, she has 20 years' experience on probabilistic graphical models, clustering methods, spatial data analysis and statistical image analysis. Her publications in international journals and conferences relate to 4 main fields: Statistics and Probability, Machine learning and Shape recognition, Image Processing and Signal Processing, Biology and Medicine. Florence Forbes has coordinated several national projects and has contributed to 2 European STREP projects as main co-Investigator. She supervised 9 doctoral students and 9 post-doctoral research fellows between 2006 and 2014.

- Philippe CIUCIU received his engineering and M.Sc. degrees in automatic control and signal processing both in 1996 from ESIEA Paris. He received his Ph.D. in signal processing in 2000 from the Université Paris-Sud, France. He has been the CEA Expert Senior rank in the field of biomedical research since 2014 for his contributions to signal/image processing for functional brain imaging. His research interests are interdisciplinary ranging from signal and image processing (Compressed Sensing for MR Imaging) to functional brain imaging (fMRI, MEG) for applications to cognitive neuroscience (multi-perceptual learning, plasticity) and clinical trials in Alzheimer's disease and neurological disorders (Stroke).