Adaptive Mean Shift Based Hemodynamic Brain Parcellation in fMRI

Mohanad Albughdadi[1], Lotfi Chaari[1] and Jean-Yves Tourneret[1] ⋆

[1] University of Toulouse, IRIT,INP-ENSEEIHT, France
firstname.lastname@enseeiht.fr

Abstract. One of the remaining challenges in event-related fMRI is to discriminate between the vascular response and the neural activity in the BOLD signal. This discrimination is done by identifying the hemodynamic territories which differ in their underlying dynamics. In the literature, many approaches have been proposed to estimate these underlying dynamics, which is also known as Hemodynamic Response Function (HRF). However, most of the proposed approaches depend on a prior information regarding the shape of the parcels (territories) and their number. In this paper, we propose a novel approach which relies on the adaptive mean shift algorithm for the parcellation of the brain. A variational inference is used to estimate the unknown variables while the mean shift is embedded within a variational expectation maximization (VEM) framework to allow for estimating the parcellation and the HRF profiles without having any prior information about the number of the parcels or their shape. Results on synthetic data confirms the ability of the proposed approach to estimate accurate HRF estimates and number of parcels. It also manages to discriminate between voxels in different parcels especially at the borders between these parcels. In real data experiment, the proposed approach manages to recover HRF estimates close to the canonical shape in the bilateral occipital cortex.

1 Introduction

Functional magnetic resonance imaging (fMRI) is a powerful non-invasive imagining technique to indirectly measure neural activity from the blood-oxygen-level dependent (BOLD) signal [1]. fMRI data analysis relies on two main task; the detection of activation in brain areas after a given stimulus and the estimation of the underlying dynamics of the brain which is also called as the hemodynamic response function (HRF). Many attempts to describe the link between stimuli and the induced BOLD signal have been proposed. The basic model is the so-called general linear model (GLM). In this model, the link between the stimuli and the induced BOLD signal is modelled as a convolution between the HRF and the binary stimulus sequence. However, this model considers a fixed HRF shape [2,3]. Many extensions have been proposed to account for the variability of

⋆ The authors would like to thank Dr. Philippe Ciuciu for providing them with the real data used for validation.

the HRF by using more regressors [4–7] which leads to less reliability in the detection task. Other approaches depending on physiologically-informed non-linear models (like the Balloon model) have been used to recover the hemodynamics in brain areas where activation has already been detected [5, 8–10]. However, these approaches are computationally intensive for a whole brain analysis and the presence of noise causes some identifiability issues.

The detection of the evoked activity and the estimation of the dynamics have been mainly addressed as two separate tasks while each of them depends on the other. 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 [11,12]. In this context, the joint detection estimation (JDE) model performs both tasks simultaneously [13–15]. In the JDE model, a single HRF shape is considered for a specific parcel (group of voxels). Although the JDE model jointly detects the evoked activity within the brain and estimates the HRF, it still requires a prior parcellation of the brain into functionally homogeneous regions. This challenge motivated the development of the joint parcellation detection estimation (JPDE) model [16,17] that performs online parcellation along with the detection and estimation tasks by setting voxels that share the same HRF pattern in the same HRF group (parcel). The JPDE model can be inferred using the VEM algorithm. However, this model still requires manual settings of the number of parcels. To overcome this issue, a model selection procedure was proposed in [18] to select the optimum number of parcels. This procedure depends mainly on free energy calculations where the model that maximizes the free energy is the best fit for the data. The limitation of this procedure arises from the fact that it needs to be run for each candidate model which can be time consuming especially if no prior information exists about the number of parcels. The standard JPDE model has been adopted in a Bayesian non-parametric approach [19] by making use of the Dirichlet process mixtures model combined with a hidden Markov random field to automatically infer the number of parcels and their shapes simultaneously with the estimation and detection tasks.

In this paper, a new approach is proposed to estimate the number of parcels from the fMRI BOLD signal. More precisely, we propose to embed the adaptive mean shift algorithm (which is a common clustering algorithm) within the variational inference framework associated with the JPDE model to estimate the parcels and their corresponding HRF profiles.

The rest of the paper is organized as follows. The JPDE model is summarized in Section 2. The adaptive mean shift (AMS) algorithm is illustrated in Section 3. The inference strategy of the AMS-JPDE (for adaptive mean shift with JPDE model) model is described in Section 4 along with the use of the AMS algorithm within the VEM framework. Experimental validation on synthetic and real fMRI data is presented in Section 5. Finally, conclusion and future work are drawn in Section 6.

2 The JPDE model

The adopted JPDE model is the one proposed in [16, 17]. Let $\boldsymbol{\mathcal{P}}$ be the set of voxels (J voxels) and $\boldsymbol{y}_j \in \mathbb{R}^N$ the fMRI time series for the voxel j at times $\{t_n,\ n = 1, \ldots, N\}$, where $t_n = nTR$, N is the number of scans and TR is the repetition time. The BOLD time series is denoted as $\boldsymbol{Y} = \{\boldsymbol{y}_j, j \in \boldsymbol{\mathcal{P}}\}$. M different experimental conditions are considered. The model assumes that the HRFs are voxel-dependent and the whole set is denoted as $\boldsymbol{H} = \{\boldsymbol{h}_j, j \in \boldsymbol{\mathcal{P}}\}$ with $\boldsymbol{h}_j \in \mathbb{R}^D$. Each $\boldsymbol{h}_j$ is associated with one among K considered HRF groups (parcels). A set of hidden variables $\boldsymbol{z} = \{z_j, j \in \boldsymbol{\mathcal{P}}\}$ is used to encode these groups where $z_j \in \{1, \ldots, K\}$ and $\boldsymbol{z}$ follows a K-class Potts model with interaction parameter β_z to account for spatial connexity. In the group $\#k$ and voxel $\#j$, the HRF $\boldsymbol{h}_j$ is a stochastic perturbation of an HRF pattern $\bar{\boldsymbol{h}}_k$ such that $\boldsymbol{h}_j \sim \mathcal{N}(\bar{\boldsymbol{h}}_k, \nu_k \boldsymbol{I}_D)$, where ν_k is a parameter which controls the stochastic perturbations around $\bar{\boldsymbol{h}}_k$ and $\boldsymbol{I}_D$ is the identity matrix of size $D \times D$. Smooth HRF patterns are forced by assigning them the prior $\bar{\boldsymbol{h}}_k \sim \mathcal{N}(0, \sigma_h^2 \boldsymbol{R})$ where $\boldsymbol{R} = (\Delta t)^4 \left(\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2\right)^{-1}$, $\Delta t < TR$ is the sampling period of the unknown HRFs and $\boldsymbol{D}_2$ is the second order finite difference matrix. The following observation model is considered

$$\forall j \in \boldsymbol{\mathcal{P}}, \quad \boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}_m \boldsymbol{h}_j + \boldsymbol{P}\ell_j + \boldsymbol{\varepsilon}_j, \tag{1}$$

where the low frequency drifts are denoted by $\boldsymbol{P}\ell_j$ and $\boldsymbol{X}_m = \{x_m^{n-d\Delta t}, n = 1, \ldots, N, d = 0, \ldots, D-1\}$ is a binary matrix that provides information on the stimulus occurrences for the m^{th} experimental condition. The neural response levels (NRL) are denoted by $\boldsymbol{A} = \{\boldsymbol{a}^m, m = 1, \ldots, M\}$ with $\boldsymbol{a}^m = \left\{a_j^m, j \in \boldsymbol{\mathcal{P}}\right\}$. The amplitudes $\boldsymbol{a}^m$'s follow spatial Gaussian mixtures defined by a set of parameters $\boldsymbol{\theta_a}$ and governed by binary Markov fields. More specifically, each NRL is assigned to one of the activation classes encoded by the variables $\boldsymbol{Q} = \{\boldsymbol{q}^m, m = 1, \ldots, M\}$ where $\boldsymbol{q}^m = \left\{q_j^m, j \in \boldsymbol{\mathcal{P}}\right\}$ is a binary Markov field with interaction parameter β_m distributed according to an exponential distribution with parameter λ_m. Two classes are considered; $(q_j^m = 1)$ if voxel j is activated by condition $\#m$ and $(q_j^m = 0)$ otherwise. An additive Gaussian noise $\boldsymbol{\varepsilon}_j$ is considered with covariance matrix $\boldsymbol{\Gamma}_j^{-1}$. The set of all parameters is denoted as $\boldsymbol{\Theta} = \{\boldsymbol{L}, \boldsymbol{\Gamma}, \boldsymbol{\theta_a}, \boldsymbol{\nu}, \boldsymbol{\lambda}, \sigma_h^2, \bar{\boldsymbol{h}}, \boldsymbol{\beta}, \beta_z\ \}$ where $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_j, j \in \boldsymbol{\mathcal{P}}\}, \boldsymbol{\nu} = \{\nu_k, k = 1, \ldots, K\}$, $\boldsymbol{\beta} = \{\beta_m, m = 1, \ldots, M\}$ and $\bar{\boldsymbol{h}} = \left(\bar{\boldsymbol{h}}_k\right)_{1 \leq k \leq K}$. More details about the meaning of these parameters are available in [16, 17].

A variational approach was proposed in [16, 17] to approximate the posterior of the JPDE model as the product of simple distributions. More precisely, the posterior of the JPDE model was approximated as

$$\tilde{p}(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z} \,|\, \boldsymbol{Y}) = \tilde{p}_A(\boldsymbol{A}) \tilde{p}_H(\boldsymbol{H}) \tilde{p}_Q(\boldsymbol{Q}) \tilde{p}_z(\boldsymbol{z}). \tag{2}$$

The inference was carried out in two parts: the expectation step which was divided into four main steps to compute approximate posteriors of the variables

$\{\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}\}$, and the maximization step to estimate the unknown parameters. The interested reader can refer to [16, 17] for further details.

3 Adaptive Mean Shift algorithm

Mean shift is a feature space analysis algorithm which has been widely used for computer vision tasks. Feature space analysis is used to reduce the data to sets of significant features which is also known as clustering or classification. This algorithm is a robust clustering technique which does not require setting the number of clusters. It is an iterative algorithm that estimates the modes of a multivariate distribution underlying the feature space. The number of clusters is obtained automatically by estimating centres of these clusters [20]. Dense regions presented in the feature space correspond to the modes (local maxima) of the probability density function (pdf) of the observed data. Each data point is associated with the nearby peak of the pdf. The mean shift defines a window (kernel) around each data point and then computes its mean. The center of the window is shifted to the mean in an iterative procedure until convergence. The mean shift algorithm relies on kernel density estimation which is a non-parametric approach to estimate the pdf of a random variable. Akin to [21], with a kernel K, each $\boldsymbol{h}_j \in \mathbb{R}^D$ is associated with a bandwidth value w_j that defines the radius of the kernel. Since the set $\boldsymbol{H} = \{\boldsymbol{h}_j, j \in \boldsymbol{\mathcal{P}}\}$ is not available, we will use the set of HRF estimates $\boldsymbol{m}_H = \{\boldsymbol{m}_{H_j}, j = 1, \ldots, J\}$ as in [16, 17]. The kernel density estimator at point $\boldsymbol{m}_{H_x}$ can be defined as

$$\hat{f}_K(\boldsymbol{m}_{H_x}) = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{w_j^D} k\left(\left\| \frac{\boldsymbol{m}_{H_x} - \boldsymbol{m}_{H_j}}{w_j} \right\|_2^2 \right), \tag{3}$$

where $k(\mathbf{x})$ is the profile of the spherically symmetric kernel K satisfying

$$K(\mathbf{x}) = c_{k,D} k(\|\mathbf{x}\|_2^2) > 0 \quad ; \quad \|\mathbf{x}\|_2 \leq 1. \tag{4}$$

The normalization constant $c_{k,D}$ ensures that the kernel $K(\mathbf{x})$ integrates to one. Whenever the derivative of the kernel profile $k(\mathbf{x})$ exists, we can define a function $g(\mathbf{x}) = -k'(\mathbf{x})$. Using $g(\mathbf{x})$ as a profile, a kernel $G(\mathbf{x})$ can be defined as $G(\mathbf{x}) = c_{g,D} g(\|\mathbf{x}\|_2^2)$. Applying the gradient to (3), we can obtain the following result (see [20] for more details)

$$\boldsymbol{S}_G(\boldsymbol{m}_{H_x}) = C \frac{\hat{\nabla} f_K(\boldsymbol{m}_{H_x})}{\hat{f}_G(\boldsymbol{m}_{H_x})}, \tag{5}$$

where $\hat{\nabla} f_K(\mathbf{x})$ is the gradient density estimator. In (5), C is a positive constant and $\boldsymbol{S}_G(\boldsymbol{m}_{H_x})$ is the mean shift vector which can be rewritten as

$$\boldsymbol{S}_G(\boldsymbol{m}_{H_x}) = \frac{\sum_{j=1}^{J} \frac{1}{w_j^{D+2}} \boldsymbol{m}_{H_j} g\left(\left\| \frac{\boldsymbol{m}_{H_x} - \boldsymbol{m}_{H_j}}{w_j} \right\|_2^2 \right)}{\sum_{j=1}^{J} \frac{1}{w_j^{D+2}} g\left(\left\| \frac{\boldsymbol{m}_{H_x} - \boldsymbol{m}_{H_j}}{w_j} \right\|_2^2 \right)} - \boldsymbol{m}_{H_x}. \tag{6}$$

The mean shift vector will always move toward the maximum increase of the density [20]. This is due to the fact that at location $\boldsymbol{m}_{H_x}$, the weighted mean of the data points with kernel G is proportional to the normalized density gradient estimate with kernel K (see (5)). The stationary points obtained via a gradient ascent method represent the modes of the density function and all the points associated with the same stationary point belong to the same cluster. Let us define $\{\boldsymbol{B}_l\}_{l=1,2,\ldots}$ as the sequence of successive locations of kernel G. Using (6), we can write

$$\boldsymbol{B}_{l+1} = \frac{\sum_{j=1}^{J} \frac{1}{w_j^{D+2}} \boldsymbol{m}_{H_j} g\left(\left\| \frac{\boldsymbol{m}_{H_l} - \boldsymbol{m}_{H_j}}{w_j} \right\|_2^2\right)}{\sum_{j=1}^{J} \frac{1}{w_j^{D+2}} g\left(\left\| \frac{\boldsymbol{m}_{H_l} - \boldsymbol{m}_{H_j}}{w_j} \right\|_2^2\right)} \quad l = 1, 2, \ldots. \tag{7}$$

Eq. (7) is one of the properties of the mean shift: it is nothing but a hill climbing iterative procedure until the density gradient vanishes. After the convergence of this iterative procedure, we will obtain the local maxima of the density (modes). To overcome the problem of setting an optimal global bandwidth, w_j is estimated using an ℓ_1 norm as in [21]. Assuming that $\boldsymbol{m}_{H_{j,k}}$ is the k-nearest neighbour of the point $\boldsymbol{m}_{H_j}$, the bandwidth associated with $\boldsymbol{m}_{H_j}$ can be computed as follows

$$w_j = \left\| \boldsymbol{m}_{H_j} - \boldsymbol{m}_{H_{j,k}} \right\|_1. \tag{8}$$

After convergence of this procedure, the estimated modes are the HRF estimates of the parcels (see [20] for more details). The practical algorithm for mode detection can be summarized in two main steps; 1) run the AMS algorithm to find stationary points of $\hat{f}_K$; 2) keep only the local maxima of these points.

4 AMS within the VEM framework of the JPDE model

The AMS-JPDE model relies on the VEM algorithm for inference as in the standard JPDE model. However, modifications have to be carried out to embed the AMS algorithm within the VEM framework. The hierarchy of the classical JPDE model is slightly modified: no spatial prior is imposed over the HRF group assignment labels $\boldsymbol{z}$. The posterior distribution in (2) is factorized as a product of pdfs of the missing variables yielding four different expectation steps (VE-A, VE-H, VE-Z and VE-Q). In the context of the AMS-JPDE model, the expectation over $\boldsymbol{z}$ is different compared to the standard JPDE model while the rest of the expectation steps (VE-A, VE-Q and VE-H) remains the same (see [16, 17]).

– **VE-A step:** This step is exactly the same as in [16, 17]. It reads as follows

$$\widetilde{p}_A^{(r)}(\boldsymbol{A}) \propto \exp\left(\mathrm{E}_{\widetilde{p}_H^{(r)} \widetilde{p}_Q^{(r-1)}}\left[\log p(\boldsymbol{A} \mid \boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{Q}; \boldsymbol{\Theta}^{(r-1)})\right]\right). \tag{9}$$

– **VE-Q step:** It corresponds to the same VE-Q step in [16, 17]

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

– **VE-H step:** As in the JPDE model, the VE-H step is

$$\widetilde{p}_H^{(r)}(\boldsymbol{H}) \propto \exp\left(\mathrm{E}_{\widetilde{p}_A^{(r-1)}\widetilde{p}_z^{(r-1)}}\left[\log p(\boldsymbol{H} \mid \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{z}; \boldsymbol{\Theta}^{(r-1)}\right]\right). \tag{11}$$

– **VE-Z step:** This step is similar to the VE-Z step in the standard JPDE model [16, 17]. Neglecting the term that comes from the spatial prior over the labels $\boldsymbol{z}$, $\widetilde{p}_{z_j}$ can be rewritten as follows

$$\begin{aligned}\widetilde{p}_{z_j}^{(r)}(k) &\propto \mathcal{N}\left(\boldsymbol{m}_{H_j}^{(r)}; \bar{\boldsymbol{h}}_k^{(r-1)}, \bar{\boldsymbol{\Sigma}}_k^{(r-1)}\right), \\ &\propto \exp\left(-\left(\boldsymbol{m}_{H_j}^{(r)} - \bar{\boldsymbol{h}}_k^{(r-1)}\right)^{\mathrm{t}} \bar{\boldsymbol{\Sigma}}_k^{-1}\left(\boldsymbol{m}_{H_j}^{(r)} - \bar{\boldsymbol{h}}_k^{(r-1)}\right)\right),\end{aligned} \tag{12}$$

where $\bar{\boldsymbol{h}} = \left(\bar{\boldsymbol{h}}_k^{(r-1)}\right)_{1\leq k\leq K}$ are the modes of the parcels (HRF patterns) obtained by the AMS algorithm in the maximization step at iteration $(r-1)$.

– **M-Step:** In the maximization step, the HRF profiles corresponding to the estimated parcels are obtained using the AMS algorithm while the maximization of the rest of the parameters remains the same as in the standard JPDE model. The corresponding M-step for the AMS-JPDE model reads

$$\boldsymbol{\Theta}^{(r)} = \arg\max_{\boldsymbol{\Theta}}\left[\mathrm{E}_{\tilde{p}_A^{(r)}\tilde{p}_H^{(r)}\tilde{p}_Q^{(r)}\tilde{p}_z^{(r)}}\left[\log p\left(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{z}; \boldsymbol{\Theta}\right)\right]\right]. \tag{13}$$

Eq. (13) can be rewritten as

$$\begin{aligned}\boldsymbol{\Theta}^{(r)} = \arg\max_{\boldsymbol{\Theta}}\Big[&\mathrm{E}_{\tilde{p}_A^{(r)}\tilde{p}_H^{(r)}}\left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}, \boldsymbol{\Gamma})\right] + \mathrm{E}_{\tilde{p}_A^{(r)}\tilde{p}_Q^{(r)}}\left[\log p(\boldsymbol{A} \mid \boldsymbol{Q}; \boldsymbol{\mu}, \boldsymbol{v})\right] \\ &+\mathrm{E}_{\tilde{p}_Q^{(r)}}\left[\log p(\boldsymbol{Q}; \boldsymbol{\beta})\right] + \mathrm{E}_{\tilde{p}_H^{(r)}\tilde{p}_z^{(r)}}\left[\log p(\boldsymbol{H} \mid \boldsymbol{z}; \bar{\boldsymbol{h}}, \boldsymbol{\nu})\right]\Big].\end{aligned} \tag{14}$$

The term $\mathrm{E}_{\tilde{p}_H^{(r)}\tilde{p}_z^{(r)}}\left[\log p(\boldsymbol{H} \mid \boldsymbol{z}; \bar{\boldsymbol{h}}, \boldsymbol{\nu})\right]$ is associated with the maximization of $\bar{\boldsymbol{h}}$ and is replaced by the AMS algorithm (see Section 3 for more details).
In the standard JPDE model, the smoothness of $\left(\bar{\boldsymbol{h}}_k\right)_{1\leq k\leq K}$ is favoured by controlling their second order derivatives with the following prior: $\bar{\boldsymbol{h}}_k \sim \mathcal{N}\left(0, \sigma_h^2 \boldsymbol{R}\right)$. In the AMS-JPDE model, we rely on a weighted least squares regularization for smoothness. A smooth $\bar{\boldsymbol{h}}_k$ that approximates the non-smooth one $\bar{\boldsymbol{h}}_k^0$ (the output of the AMS algorithm) can be obtained by solving the following problem

$$\arg\min_{\bar{\boldsymbol{h}}_k} ||\bar{\boldsymbol{h}}_k^0 - \bar{\boldsymbol{h}}_k||_2^2 + \lambda_h ||\boldsymbol{D}_2 \bar{\boldsymbol{h}}_k||_2^2, \tag{15}$$

where λ_h is a parameter to be fixed by the user and $\boldsymbol{D}_2$ is the second order finite difference matrix. In the above expression, minimizing $||\bar{\boldsymbol{h}}_k^0 - \bar{\boldsymbol{h}}_k||_2^2$ forces the smooth $\bar{\boldsymbol{h}}_k$ to be close to the non-smooth one $\bar{\boldsymbol{h}}_k^0$. On the other hand, minimizing the term $||\boldsymbol{D}_2 \bar{\boldsymbol{h}}_k||_2^2$ favours the smoothness of $\bar{\boldsymbol{h}}_k$. Straightforward computations lead to the following expression of $\bar{\boldsymbol{h}}_k$ minimizing (15)

$$\bar{\boldsymbol{h}}_k = \left(\boldsymbol{I}_D + \lambda_h \boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2\right)^{-1} \bar{\boldsymbol{h}}_k^0. \tag{16}$$

Note that such a quadratic regularization is equivalent to fixing a Gaussian prior on $\bar{\boldsymbol{h}}_k$ in the hierarchical Bayesian model.

5 Experimental validation

To validate the AMS based parcellation with the JPDE model, we have performed numerical experiments on both synthetic and real data [1].

5.1 Synthetic fMRI time series

The proposed model was validated on four different experiments denoted as Exps. 1-4 and defined by different parcellation masks (see Fig. 1)[top row] to generate the BOLD signal according to (1). As regards the experimental condi-

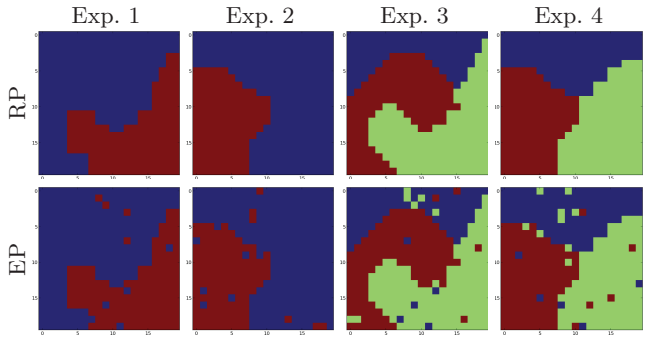


Fig. 1: Reference parcellations (RP) used for the 4 experiments and corresponding estimated parcellation (EP) (grid size = 20×20).

tions, we consider here two of them ($M = 2$) with 30 trials for each. The reference binary labels are shown in Fig. 2(left column). The NRLs are simulated from their prior conditionally to the activation labels $\boldsymbol{Q}$ as shown in Fig. 2(right column). Given these 20×20 binary labels, the NRLs were simulated as follows, for $m = 0, 1$: $a_j^m \,|\, q_j^m = 0 \sim \mathcal{N}(0, 0.5)$ and $a_j^m \,|\, q_j^m = 1 \sim \mathcal{N}(3.2, 0.5)$. The inter stimuli interval and variance to generate the onsets of the trials were 3 s and 5 s, respectively. Finally, the fMRI time series $\boldsymbol{y}_j$ were generated according to (1) with $\Delta t = 0.5$ and $TR = 1$ s.

We analyzed the generated fMRI time series for the four experiments using the AMS-JPDE model. The parcellation estimates for each experiment is shown in Fig. 1[bottom row]. It is worth noticing that for the AMS-JPDE, no prior initialization for the parcellation or truncation level for the maximum number of parcels is needed. The number of K-nearest neighbours (K_{NN}) is the only parameter that needs to be manually set. For the synthetic data experiments, we set $K_{NN} = 50$. The computed parcellation errors between the reference and the estimated parcellation is $2.25\%, 3.25\%, 4.5\%$ and 4.75% for Exps. 1 to 4, respectively. These results show a good ability of the AMS-JPDE to recover the hemodynamic territories with low error probability. Moreover, for each experiment, we

[1] These experiments were implemented in Python within the framework offered by the Pyhrf software [22].

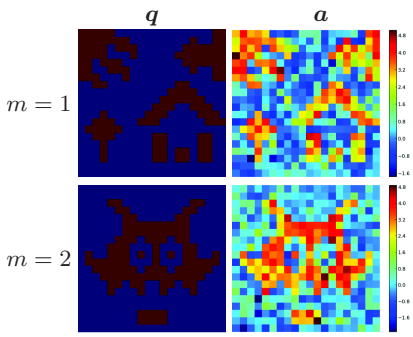


Fig. 2: Reference activation labels (left column) and reference NRLs (right column) for the two experimental conditions (grid size = 20×20).

computed the confusion matrix between the reference and the estimated parcellation. The results displayed in Tables. 1 and 2 show a major intersection between them. Although some voxels were misclassified (since no spatial constraints are imposed over the parcellation step), the AMS-JPDE model managed to establish a good parcellation especially for those voxels located on the borders between parcels. The results of the AMS-JPDE model were coherent with the results of the model selection procedure in [18] that calculates the free energy of different competing models each with K^{ω} parcels and $K^{\omega} = \omega + 1$, $\omega \in \{1, \ldots, 3\}$. The models maximizing the free energy are the best fit for the data. These optimal models lead to two parcels for Exp. 1 and 2 and three parcels for Exp. 3 and 4. Regarding the running time of the algorithm and considering Exp. 4 as an example, using the model selection procedure in [18] the accumulated time required to run the 3 competing models is around 35 mins. On the other hand, the AMS-JPDE model takes less than 12 mins. Thus, the computational time of the AMS-JPDE is reduced compared to free energy calculations of the competing models. Fig. 3 shows the transformed voxel-dependent HRFs in 3-D representation using Principal Component Analysis (PCA) with the HRF groups labels for each experiment. We also explored the ability of the AMS-JPDE model to estimate the HRF profiles for the estimated parcels, as shown in Fig. 4. The modes of the parcels are the outputs of the AMS algorithm and here they represent the HRF estimate for each parcel. These results are close to the ground truth. The AMS-JPDE also managed to obtain a good performance in detecting the activation as in the JPDE model. The mean square error (MSE) was computed for each experimental condition in the four experiments. The average MSEs for the estimated NRLs and labels were 0.006 and 0.004, respectively.

5.2 Real data

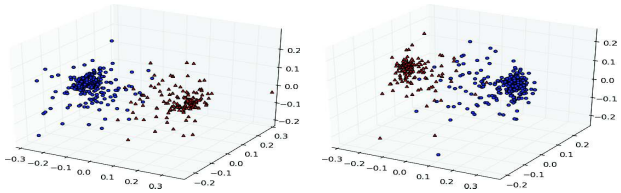
One experiment was conducted on real fMRI data to validate the AMS-JPDE model. The considered region of interest (ROI) is the bilateral occipital cortex.

Table 1: Confusion matrices for Exps. 1 and 2. RP and EP refer to the reference and the estimated parcellations, respectively.

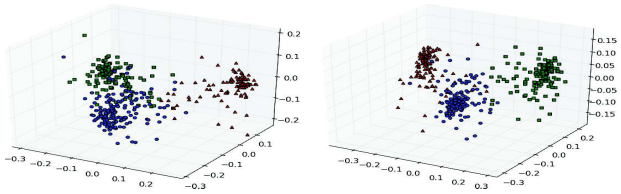
EP / RP	Parcel. 1		Parcel. 2	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2
Parcel. 1	**0.98**	**0.98**	0.03	0.05
Parcel. 2	0.02	0.02	**0.97**	**0.95**

Table 2: Confusion matrices for Exps. 3 and 4. RP and EP refer to the reference and the estimated parcellations, respectively.

EP / RP	Parcel. 1		Parcel. 2		Parcel. 3	
	Exp. 3	Exp. 4	Exp. 3	Exp. 4	Exp. 3	Exp. 4
Parcel. 1	**0.93**	**0.94**	0.02	0.01	0.01	0.02
Parcel. 2	0.05	0.01	**0.96**	**0.96**	0.02	0.02
Parcel. 3	0.02	0.05	0.02	0.03	**0.97**	**0.96**

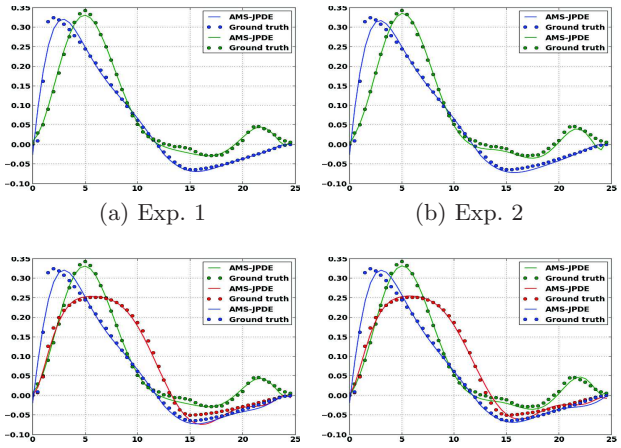


(a) Exp. 1 (b) Exp. 2



(c) Exp. 3 (d) Exp. 4

Fig. 3: Transformed voxel-dependent HRFs in 3-D representation using PCA with HRF groups labels.



(a) Exp. 1 (b) Exp. 2

(c) Exp. 3 (d) Exp. 4

Fig. 4: HRF estimates for the synthetic data experiments.

The fMRI data were collected using a gradient-echo EPI sequence (TE = 30 ms/TR = 2.4 s /thickness = 3 mm/ FOV = 192×192 mm^2, matrix size: 96×96) with a 3 Tesla magnetic field during a localizer experiment. Sixty auditory, visual and motor stimuli were involved in the paradigm and defined in ten experimental conditions ($M = 10$). During the experiment, $N = 128$ scans were acquired and $\Delta t = 0.6$ s. The number of K-nearest neighbours was set to $K_{NN} = 50$. Running this experiment using the AMS-JPDE model, 3 parcels were estimated as shown in Fig. 5. The corresponding HRF shape estimates are shown in Fig. 6. The computed time to peak (TTP) for the HRF estimates was 5.4 s for all of them while the full width at half maximum (FWHM) was 4.2 s for parcels 1 and 3 and 4.8 s for parcel 2. The obtained results are coherent with the conclusion that the HRF estimates in the bilateral occipital cortex should be consistent with the canonical shape [19, 23]. To verify these results, we also ran the JPDE model with the model selection procedure proposed in [18] on the same fMRI data using three candidate models with one, two and three parcels for initial parcellation. The candidate model maximizing the free energy was the one with three parcels and the reported value was -236604. This result is also coherent with our findings using the AMS-JPDE model.

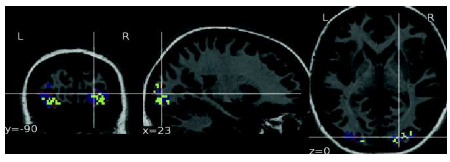


Fig. 5: The estimated parcellation in the bilateral occipital cortex.

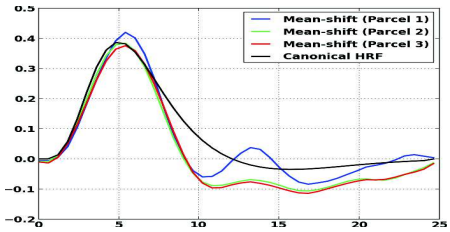


Fig. 6: HRF shape estimates using the AMS-JPDE model in the bilateral occipital cortex and the canonical HRF.

6 Conclusion and future work

This paper proposed a new approach for automatic hemodynamic brain parcellation relying on the existing JPDE model and the adaptive mean shift algorithm yielding the so-called AMS-JPDE model. The AMS algorithm is used within the VEM framework of the JPDE model to formulate it as a non-parametric approach for model selection. In contrast with the standard JPDE model, the AMS-JPDE model requires no prior initialization for the parcellation as the HRF estimates are the modes obtained by the AMS algorithm of the underlying multivariate distribution of the voxel-dependent HRFs. Future work will focus on embedding spatial prior in the AMS algorithm to eliminate outlier voxels in the parcellation estimates.

References

1. S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *in Proc. Natl. Acad. Sci.*, vol. 87, no. 24, pp. 9868–72, 1990.
2. K. J. Friston, A. P. Holmes, J.-B. Poline, P. J. Grasby, S. C. R. Williams, R. S. J. Frackowiak, and R. Turner, "Analysis of fMRI time-series revisited," *Neuroimage*, vol. 2, no. 1, pp. 45–53, 1995.
3. G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger, "Linear systems analysis of functional magnetic resonance imaging in human v1," *The Journal of Neuroscience*, vol. 16, no. 13, pp. 4207–4221, 1996.
4. G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI," *Neuroimage*, vol. 9, no. 4, pp. 416–429, 1999.
5. K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, "Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics," *Neuroimage*, vol. 12, no. 4, pp. 466–477, 2000.
6. R. N. A. Henson, M. D. Rugg, and K. J. Friston, "The choice of basis functions in event-related fMRI," *Neuroimage*, vol. 13, no. 6, pp. 149, 2001.
7. M. A. Lindquist, J. M. Loh, L. Y. Atlas, and T. D. Wager, "Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling," *Neuroimage*, vol. 45, no. 1, pp. S187–S198, 2009.

8. R. B. Buxton and L. Frank, "A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation," *Journal of Cerebral Blood Flow & Metabolism*, vol. 17, no. 1, pp. 64–72, 1997.
9. J. J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and R. Kawashima, "A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals," *Neuroimage*, vol. 21, no. 2, pp. 547–567, 2004.
10. T. Deneux and O. Faugeras, "Using nonlinear models in fMRI data analysis: model selection and activation detection," *Neuroimage*, vol. 32, no. 4, pp. 1669–1689, 2006.
11. J. Kershaw, B. A. Ardekani, and I. Kanno, "Application of Bayesian inference to fMRI data analysis," *IEEE Trans. Med. Imag.*, vol. 18, no. 12, pp. 1138–1153, 1999.
12. P Ciuciu, J.B. Poline, G. Marrelec, J. Idier, Ch. Pallier, and H. Benali, "Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1235–1251, Oct 2003.
13. S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu, "A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI," *Neuroimage*, vol. 41, no. 3, pp. 941–969, 2008.
14. T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of fMRI time series," *IEEE Trans. Med. Imag.*, vol. 29, no. 4, pp. 1059–1074, 2010.
15. L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach," *IEEE Trans. Med. Imag.*, vol. 32, no. 5, pp. 821–837, 2013.
16. L. Chaari, F. Forbes, T. Vincent, and P. Ciuciu, "Hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework," in *Medical Image Computing and Computer-Assisted Intervention*, N. Ayache et al., Ed. 2012, vol. 7512, pp. 180–188, Springer.
17. L. Chaari, S. Badillo, T. Vincent, G. Dehaene-Lambertz, F. Forbes, and P. Ciuciu, "Subject-level Joint Parcellation-Detection-Estimation in fMRI," `https://hal.inria.fr/hal-01255465/file/JPDE_CHAARI_submitted06012016.pdf`, Jan. 2016.
18. M. Albughdadi, L. Chaari, F. Forbes, J.-Y. Tourneret, and P. Ciuciu, "Model selection for hemodynamic brain parcellation in fMRI," in *Proc. EUSIPCO*, Lisbon, Portugal, Sept 2014, pp. 31–35.
19. M. Albughdadi, L. Chaari, J.-Y. Tourneret, F. Forbes, and P. Ciuciu, "Hemodynamic Brain Parcellation Using A Non-Parametric Bayesian Approach," `https://hal.inria.fr/hal-01275622/file/albugdhadi_paper.pdf`, Feb. 2016.
20. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.
21. B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: A texture classification example," in *Proc. ICCV*, Nice, France, 2003, IEEE, pp. 456–463.
22. T. Vincent, S. Badillo, L. Risser, L. Chaari, C. Bakhous, F. Forbes, and P. Ciuciu, "Flexible multivariate hemodynamics fMRI data analyses and simulations with PyHRF," *Frontiers in Neuroscience*, vol. 8, no. 67, 2014.
23. S. Badillo, T. Vincent, and P. Ciuciu, "Group-level impacts of within- and between-subject hemodynamic variability in fMRI," *Neuroimage*, vol. 82, pp. 433–448, 15 Nov. 2013.