HAL
open science

Hemodynamically informed parcellation of cerebral FMRI data

Aina Frau-Pascual, Thomas Vincent, Florence Forbes, Philippe Ciuciu

To cite this version:

Aina Frau-Pascual, Thomas Vincent, Florence Forbes, Philippe Ciuciu. Hemodynamically informed parcellation of cerebral FMRI data. ICASSP 2014 - 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, May 2014, Florence, Italy. pp.2079-2083, 10.1109/ICASSP.2014.6853965 . hal-01100186v2

HAL Id: hal-01100186

https://inria.hal.science/hal-01100186v2

Submitted on 6 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HEMODYNAMICALLY INFORMED PARCELLATION OF CEREBRAL FMRI DATA

Aina Frau-Pascual[1,3], *Thomas Vincent*[1], *Florence Forbes*[1], *Philippe Ciuciu*[2,3]

(1) INRIA, MISTIS, Grenoble University, LJK, Grenoble, France
(2) CEA/DSV/I^{2}BM NeuroSpin center, Bât. 145, F-91191 Gif-sur-Yvette, France
(3) INRIA, Parietal, F-91893 Orsay, France.

ABSTRACT

Standard detection of evoked brain activity in functional MRI (fMRI) relies on a fixed and known shape of the impulse response of the neurovascular coupling, namely the hemodynamic response function (HRF). To cope with this issue, the joint detection-estimation (JDE) framework has been proposed. This formalism enables to estimate a HRF per region but for doing so, it assumes a prior brain partition (or parcellation) regarding hemodynamic territories. This partition has to be accurate enough to recover accurate HRF shapes but has also to overcome the detection-estimation issue: the lack of hemodynamics information in the non-active positions. An hemodynamically-based parcellation method is proposed, consisting first of a feature extraction step, followed by a Gaussian Mixture-based parcellation, which considers the injection of the activation levels in the parcellation process, in order to overcome the detection-estimation issue and find the underlying hemodynamics.

***Index Terms*—** joint detection-estimation, hemodynamics, Gaussian mixtures, parcellation, brain

1. INTRODUCTION

Functional MRI (fMRI) is an imaging technique that indirectly measures neural activity through the Blood-oxygen-level-dependent (BOLD) signal [1], which captures the variation in blood oxygenation arising from an external stimulation. This variation also allows the estimation of the underlying dynamics, namely the characterization of the so-called hemodynamic response function (HRF). The hemodynamic characteristics are likely to spatially vary, but can be considered constant up to a certain spatial extent. Hence, it makes sense to estimate a single HRF shape for any given area of the brain. To this end, parcel-based approaches that segment fMRI data into functionally homogeneous regions and perform parcelwise fMRI data analysis provide an appealing framework [2].

In [3–5], a joint detection-estimation (JDE) approach has been proposed for simultaneously localizing evoked brain activity and estimating HRF shapes at a parcel-level. This spatial scale allows one to make a spatial compromise between hemodynamics reproducibility and signal aggregation, the latter operation enhancing the inherent low signal-to-noise ratio (SNR) of fMRI data. This aggregation has to be well spatially controlled because of the known fluctuation of hemodynamics across brain regions [6, 7]. Hence, the JDE approach operates on a prior partitioning of the brain into functionally homogeneous parcels, where the hemodynamics is assumed constant. A robust parcellation is needed to ensure a good JDE performance. A few attempts have been proposed to cope with this issue [2, 4, 8–11] but most of them are either too computationally demanding [4, 10, 11] or do not account for hemodynamics variability [2, 8, 9]. Here, our goal is to provide a *fast* hemodynamics parcellation procedure that can be used in daily applications prior to JDE inference.

In this work, we propose a two-step approach consisting first of hemodynamics feature extraction, in which a general linear model (GLM) is used to discriminate hemodynamics information, followed by a parcellation of these features. The goal here is finding features that are able to catch most of the hemodynamic information, without the need of perfectly estimating the HRF function. Afterwards, an agglomerative clustering algorithm is used to segment the features. Our main contribution is the consideration of the detection-estimation effect within the parcellation step: there is a lack of hemodynamics information in the non- or slightly-activating voxels. The idea is then to enforce grouping these uncertain voxels with neighboring activating voxels. This is done through a spatial constraint in the agglomerative step of the parcellation procedure. Moreover, the uncertainty in a given voxel can be quantified by a statistics linked to its activation level, namely a p-value obtained in the GLM feature extraction step. This statistics is hence injected within the agglomeration criterion.

This paper is organized as follows. First, we present the artificial data generation, the feature extraction step in the GLM framework, and the proposed parcellation algorithm, which relies on a model-based agglomerative spatial clustering. In the results, the proposed parcellation is compared with a classical Ward parcellation refined with a spatial constraint. Quantitative results are assessed in a Monte Carlo experiment where the variability against several random data sets is assessed. Finally, we investigate the impact of the input parcellation on the HRF estimation provided by the JDE approach.

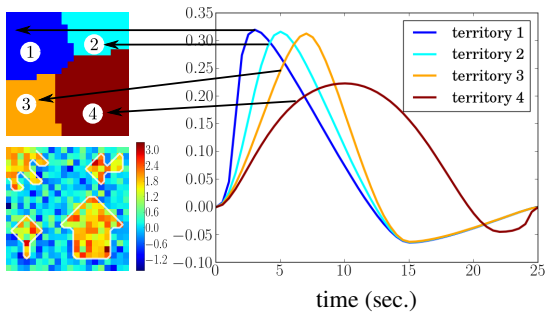

Fig. 1. Artificial fMRI data sets. Top left: hemodynamic parcels. Bottom left: response levels. Right: HRFs associated with parcels.

2. HEMODYNAMICALLY BASED PARCELLATION

2.1. Generation of artificial fMRI data sets

Let us define the set of all parcels as $\mathbb{P} = \{\mathcal{P}_1, ..., \mathcal{P}_\gamma, ..., \mathcal{P}_\Gamma\}$ where $\mathcal{P}_\gamma$ is the set of position indexes belonging to parcel γ and $\mathbb{J}_\gamma$ denotes the set of positions in parcel γ. Artificial BOLD data sets are generated using the following regional BOLD model, for a given voxel $j \in \mathbb{J}_\gamma$, and a given parcel γ:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h}_\gamma + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j, \quad (1)$$

where a_j^m is the response amplitude at voxel j for a certain condition m, $\boldsymbol{X}^m = (x^m_{n-d\Delta t})_{n=1:N,d=0:D}$ is the binary matrix encoding the stimulus for each condition m, $\boldsymbol{h}_\gamma = (h_{d\Delta t,\gamma})_{d=0:D}$ is the HRF corresponding to parcel $\mathcal{P}_\gamma$, Δt being the HRF sampling period, $\boldsymbol{P}$ the orthogonal function basis multiplied by the drift $\boldsymbol{\ell}_j$, and $\boldsymbol{b}_j$ the noise vector. Note that $\boldsymbol{y}_j = [y_j(t_1), ..., y_j(t_N)]^t$, where $t_k = k\,TR$ and $TR \gg \Delta t$. Typical values are $\Delta t = 0.6$ and $TR = 1.8$ sec.

In this work, we considered artificial data at low SNR, with one experimental condition represented with a 20×20-voxel binary activation labels $\boldsymbol{q} = [q_1...q_J]$, and levels of activation $\boldsymbol{a} = [a_1...a_J]$, with $(a_j|q_j = 1) \sim \mathcal{N}(1.8, 0.25)$. We simulated a map of hemodynamics parcels, with different HRF shapes $\boldsymbol{h}_\gamma$ (duration 25 sec., $TR = 1$ sec., $\Delta t = 0.5$ sec.) in each parcel $\mathcal{P}_\gamma$ (see Fig. 1), by using the combination of 3 Bezier's curves, each being controlled by 4 points, to describe the curves until the peak, from the peak to the undershoot, and from the undershoot to the end, given specific peak and undershoot widths. We considered a Discrete Cosine Transform for $\boldsymbol{P}$, a drift $\boldsymbol{\ell}_j \sim \mathcal{N}(0, 11\boldsymbol{I}_4)$, and white Gaussian noise with variance $\epsilon_j^2 = 1.5$.

2.2. Feature extraction

As regards hemodynamic feature extraction, several approaches are available. Here, we only focus on GLM-based ones involving either canonical HRF and its derivative(s) [12]. We chose not to rely in Finite Impulse Response (FIR) models [13] since they may be to sensitive to noise. Although more flexible regularized FIR (RFIR) approaches such as [14] are also theoretically able to recover any HRF shape accurately in contrast to canonical GLM, RFIR inference is pretty difficult when the SNR is too low since it proceeds voxelwise. Moreover, it is time consuming because it performs unsupervised estimation (cf [10]). Here, we are more interested in a quick feature extraction step that also allows us to disentangle true activated voxels from non-activated ones.

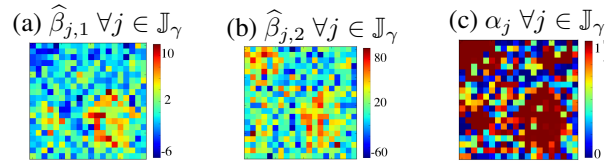

Fig. 2. Hemodynamics features extracted $\boldsymbol{\phi}_j = [\widehat{\beta}_{j,1}, \widehat{\beta}_{j,2}]^t$ $\forall j \in \mathbb{J}$, and activation levels $\alpha_j\ \forall j \in \mathbb{J}_\gamma$.

As regards canonical GLMs, our feature extraction step consists in fitting the following linear model: $\boldsymbol{y}_j = \mathcal{X}\boldsymbol{\beta}_j + \boldsymbol{b}_j$, where $\boldsymbol{y}_j$ is the BOLD signal, $\mathcal{X}$ the design matrix, $\boldsymbol{\beta}_j$ the parameter estimates and $\boldsymbol{b}_j$ the noise at voxel j. Let us denote $\beta_{j,0}$, $\beta_{j,1}$ and $\beta_{j,2}$ the parameters associated with the regressors in $\mathcal{X}$, which derives from the convolution of the experimental paradigm with the canonical HRF $\boldsymbol{h}$, its temporal $\boldsymbol{h}'$ and dispersion derivatives $\boldsymbol{h}''$, respectively. We assume that $\beta_{j,0}$ rather contains information about the HRF magnitude, whereas $\beta_{j,1}$ and $\beta_{j,2}$ provide information about the HRF delay and dispersion respectively, and, hence, are useful to differentiate hemodynamic territories. Maximum likelihood (ML) inference enables to get the parameter estimates $\widehat{\boldsymbol{\beta}}_j$ in each voxel among which we only retain $\boldsymbol{\phi}_j = [\widehat{\beta}_{j,1}, \widehat{\beta}_{j,2}]^t$, as input features to the parcellation method.

To quantify the activation level, we consider the p-value $p_{j,0}$ associated with testing $H_0 : \beta_{j,0} = 0$ and we use the notation $\alpha_j = 1 - p_{j,0} \in (0, 1)$ for these statistics in voxel j: The higher the α_j value, the larger our confidence in the presence of evoked activity in voxel j. Importantly, the statistics α_j does not enter in the parcellation along with the previously defined features $\boldsymbol{\phi}_j$, it is rather used as weights in the agglomeration criterion.

Fig. 2 shows the features $\boldsymbol{\phi}_j$ extracted $\widehat{\beta}_{j,1}$ (Fig. 2(a)) and $\widehat{\beta}_{j,2}$ (Fig. 2(b)), which contain information about the hemodynamics territories, as we can see in the activated regions of $\widehat{\beta}_{j,1}$ with different values in the different territories; and the activation levels (Fig. 2(c)) α_j, with values from 0 to 1.

2.3. Informed Gaussian mixture based parcellation

2.3.1. Agglomerative clustering algorithms

The model based interpretation of agglomerative clustering algorithms [15] makes the assumption that features have been generated by probability distributions that vary across parcels. In the context of model-based hard clustering, which aims at assigning classes to the input samples instead of weights, the

goal is to maximize the classification likelihood with respect to both $\boldsymbol{\Theta}$ and $\boldsymbol{z}$ given a set of features $\boldsymbol{\Phi}$:

$$\mathcal{L}(\boldsymbol{\Theta}, \boldsymbol{z} \mid \boldsymbol{\Phi}) = \prod_{j=1}^{J} p(\boldsymbol{\phi}_j | \boldsymbol{\theta}_{z_j}) \tag{2}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_\gamma\}_{\gamma=1:\Gamma}$ is the set of parcel-specific model parameters and $\boldsymbol{z} = \{z_j\}_{j=1:J}$ denotes the set of parcel labels associated with each voxel.

In an agglomerative approach, this function is *approximately* optimized by successive merge operations, starting from an initial clustering guess or singleton clusters. Hence, at each step, i.e. when merging two clusters $\mathcal{P}_\gamma$ and $\mathcal{P}_\tau$ of the current parcellation $\boldsymbol{z}$ into the parcel $\mathcal{P}_{\gamma'} = \mathcal{P}_\gamma \cup \mathcal{P}_\tau$ of the resulting parcellation $\boldsymbol{z}'$, the relative increase of the log-likelihood has to be maximized:

$$\log \Delta\mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{z}, \boldsymbol{z}' | \boldsymbol{\Phi}) = \log \left(\frac{\max_{\Theta} \mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{z}' | \boldsymbol{\Phi})}{\max_{\Theta} \mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{z} | \boldsymbol{\Phi})} \right) \tag{3}$$

A given merging step thus involves several likelihood maximization over parameters $\boldsymbol{\Theta}$.

2.3.2. *Gaussian-mixture model*

To account for the activation level α_j associated with each voxel j, we rely on a independent two-class Gaussian mixture in the agglomerative step. The rationale is that features $\boldsymbol{\phi}_j$ are distributed differently within a given parcel depending on the corresponding activation levels α_j. Hence, the two-class mixture is expressed on every $\boldsymbol{\phi}_j$ in parcel γ as a way to model parameter differences related to activation levels:

$$\begin{aligned} p(\boldsymbol{\phi}_j | \boldsymbol{\theta}_\gamma) &= \sum_{i=0}^{1} \Pr(q_j = i) f(\boldsymbol{\phi}_j \mid q_j = i; \boldsymbol{\theta}_\gamma) \\ &= \sum_{i=0}^{1} \lambda_{\gamma,i} \mathcal{N}(\mu_{\gamma,i}, \Sigma_{\gamma,i}) \end{aligned} \tag{4}$$

The latent variable q_j encodes the activation state of voxel j and $\Pr(q_j = 1) = \lambda_{\gamma,1}$ reflects the probability of activation. This latent variable can be directly linked to the activation statistics α_j obtained at the feature extraction step: $\hat{\lambda}_{\gamma,1} = \sum_{j \in \mathbb{J}_\gamma} \alpha_j / J_\gamma = 1 - \hat{\lambda}_{\gamma,0}$ where $\mathbb{J}_\gamma$ denotes the set of voxels in parcel γ and J_γ their cardinality. Then, straightforward calculations give the following ML estimators for the parcel-level mixture moments:

$$\hat{\mu}_{\gamma,i} = \sum_{j \in \mathbb{J}_\gamma} \delta_{i,j} \boldsymbol{\phi}_j / \Delta, \tag{5}$$

$$\hat{\Sigma}_{\gamma,i} = \sum_{j \in \mathbb{J}_\gamma} \delta_{i,j} (\boldsymbol{\phi}_j - \hat{\mu}_{\gamma,i})(\boldsymbol{\phi}_j - \hat{\mu}_{\gamma,i})^{\mathrm{t}} / \Delta, \tag{6}$$

where $\delta_{i,j} = (1 - i - (-1)^i \alpha_j)$, $\Delta = \sum_{j \in \mathbb{J}_\gamma} \delta_{i,j}$ and $\hat{\mu}_{\gamma,i}$ and $\hat{\Sigma}_{\gamma,i}$ define the empirical weighted mean and covariance of features in parcel γ. Once the parameters $\boldsymbol{\Theta}$ have been estimated, the two parcels $\mathcal{P}_\gamma$ and $\mathcal{P}_\tau$ that are selected to be merged into $\mathcal{P}_{\gamma'} = \mathcal{P}_\gamma \cup \mathcal{P}_\tau$ are those which maximizes:

$$\begin{aligned} \log \Delta\mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{z}, \boldsymbol{z}' | \boldsymbol{\Phi}) = & \sum_{j \in \mathbb{J}_\gamma} \log\Big(\sum_{i=0}^{1} \log \hat{\lambda}_{\gamma,i} \mathcal{N}(\hat{\mu}_{\gamma,i}, \hat{\Sigma}_{\gamma,i})\Big) \\ & + \sum_{j \in \mathbb{J}_\tau} \log\Big(\sum_{i=0}^{1} \log \hat{\lambda}_{\tau,i} \mathcal{N}(\hat{\mu}_{\tau,i}, \hat{\Sigma}_{\tau,i})\Big) \\ & - \sum_{j \in \mathbb{J}_{\gamma'}} \log\Big(\sum_{i=0}^{1} \log \hat{\lambda}_{\gamma',i} \mathcal{N}(\hat{\mu}_{\gamma',i}, \hat{\Sigma}_{\gamma',i})\Big) \end{aligned}$$

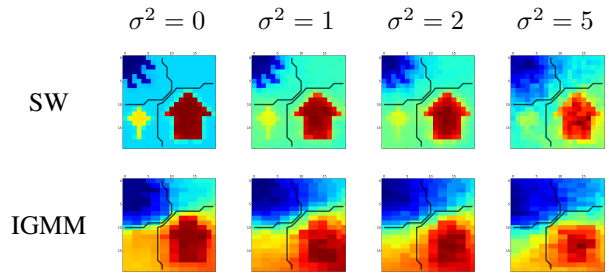

Fig. 3. Parcellations obtained by SW and IGMM methods, considering an averaged 100 iterations Monte Carlo experiment. From left to right: noise variance $0, 1, 2$ and 5, respectively.

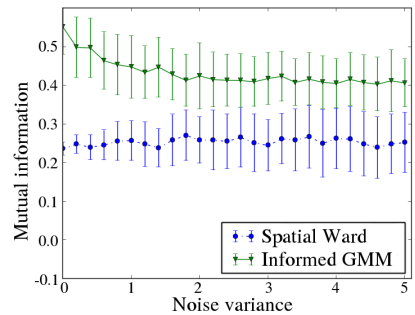

Fig. 4. MI-based quantitative comparison of SW and IGMM parcellation results with the ground truth territories for noise variances ranging from 0.0 to 5.0.

3. SIMULATION RESULTS

3.1. Informed Gaussian Mixture based Parcellation

In this section, the proposed Informed GMM (IGMM) method is compared with the Ward [16] algorithm with connectivity constraints, that here we name Spatial Ward (SW). A Monte Carlo experiment is used, where the variability against several random data sets is assessed, to quantify the results of both parcellation methods. Fig. 3 shows results averaged across 100 runs and for different noise variance levels. SW makes a big parcel for all the non-activated positions, while IGMM overcomes this issue and partitions the positions independently of the activation level.

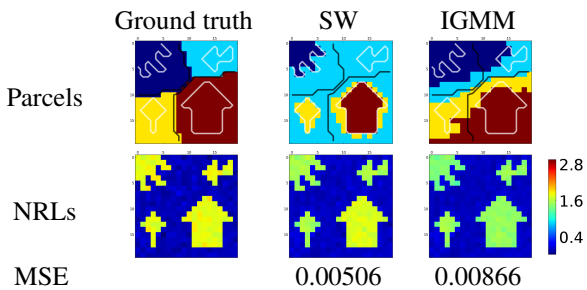

Fig. 5. JDE detection results. From top to bottom: parcellation used, averaged estimated NRLs over 100 JDE iterations and detection MSE $= \frac{\|\boldsymbol{e}^d\|^2}{\|\boldsymbol{a}^{true}\|^2}$, being $\boldsymbol{e}^d = [e_1^d ... e_J^d]$, $\boldsymbol{a}^{true} = [a_1^{true} ... a_J^{true}]$, and $e_j^d = \hat{a}_j - a_j^{true}$. From left to right: ground truth, SW and IGMM.

For a quantitative evaluation, mutual information (MI) [17] was used to compare our parcellation results with the ground truth territories. Fig. 4 shows the evolution of both parcellation techniques with respect to increasing noise variance. IGMM outperforms SW and has a decreasing mean value until $MI \approx 0.4$, whereas SW has an almost constant mean $MI \approx 0.25$. Note that both methods are sensitive to noise and have a high variance.

3.2. JDE analysis

Here, we study the impact of the prior parcellation as input knowledge to the JDE procedure. As input parcellations, we considered the hard clustering resulting from either the SW- or IGMM-based average parcels computed over the 100 individual results of our previous Monte Carlo experiment, i.e., the ones shown in Fig. 3 for $\sigma^2 = 1.5$.

Fig. 5 compares the averaged detection results over 100 JDE iterations. First, we can see a slightly lower mean detection MSE for SW-based hard parcellation. Also, a lower activation level can be observed for voxels considered in the wrong IGMM-based parcel. Fig. 6 shows the averaged estimated HRF profiles over 100 JDE iterations whether it is based on SW or IGMM parcellation methods, compared with the ground-truth HRFs in Fig. 1. Overall, HRFs are well recovered by both SW-based or IGMM-based JDE analyses. In region 2 (cyan), IGMM-based better fits the ground-truth since it mixes less voxels with different hemodynamics than SW which includes all non-activating voxels. In region 3 (yellow), SW seems to yield a MC-averaged HRF estimate slightly closer to the ground-truth than IGMM which produces a parcel that also spans region 4 (red). However, the MC variability is higher in this region than in the others for both parcellation methods as shown by the error bars.

Regarding to computation time, IGMM takes 130 times more than SW, but 1800 times less than CC-JDE [11].

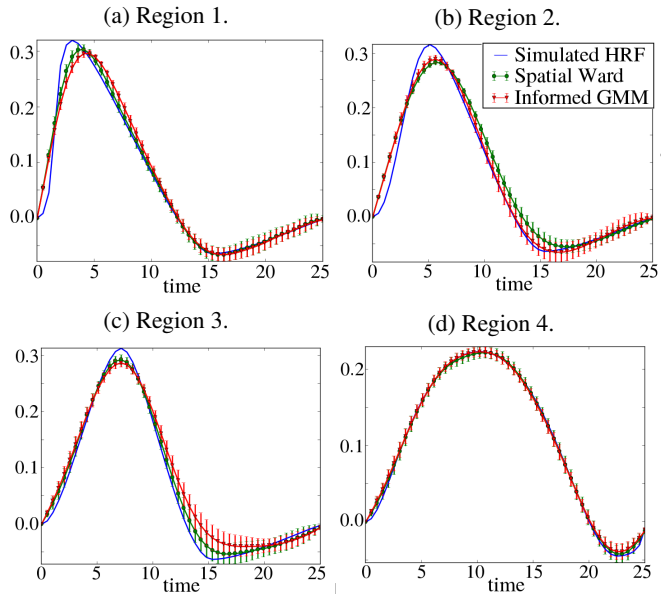

Fig. 6. JDE HRF estimation averaged results over 100 JDE iterations, for the different regions, labelled in Fig. 1. The errorbars correspond to the standard deviation of the estimation.

4. DISCUSSION AND CONCLUSION

A hierarchical parcellation method that takes into account the activation levels in the parcellation process has been developed, so as to be consistent with JDE assumption that one parcel contains active and inactive positions, and find the underlying hemodynamic territories, independently of the activation level. In terms of computation load, it is not a demanding method. However, hierarchical agglomerative algorithms need the selection in advance of the number of clusters, and this fact lead us to a definition problem. In the JDE framework, we want parcels big enough to be able to estimate accurately the HRF, although if parcels are too big, we will be losing information and averaging the hemodynamics. However, if we can have a good estimate of hemodynamics, we are interested in having as much as possible parcels to better recover the territories' singularities.

We quantitatively validated that the proposed IGMM approach enables a better recovery than a reference spatial Ward approach. Indeed, parcels obtained with IGMM are less influenced by highly activated positions and do not mix non-activation positions altogether. Finally, JDE results are quite comparable in terms of detection and HRF estimation whether the input parcellation comes from the SW method or the proposed IGMM approach. Still, the proposed approach yields more reliable parcellations than SW and may be more adequate to treat real data sets, which will be investigated in future works.

5. REFERENCES

[1] S. Ogawa, T. Lee, A. Kay, and D. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proc. Natl. Acad. Sci. USA*, vol. 87, no. 24, pp. 9868–9872, 1990.

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

[3] T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of within-subject fMRI time series," *IEEE Trans. Med. Imag.*, vol. 29, no. 4, pp. 1059–1074, Apr. 2010.

[4] L. Chaari, F. Forbes, T. Vincent, and P. Ciuciu, "Adaptive hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework," in *15th Proc. MICCAI*, Nice, France, Oct. 2012, LNCS 7512, (Part III), pp. 180–188, Springer Verlag.

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

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

[7] 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.

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

[9] T. Vincent, P. Ciuciu, and B. Thirion, "Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI," in *5th Proc. IEEE ISBI*, Paris, France, May 2008, pp. 568–571.

[10] A.-L. Fouque, P. Ciuciu, and L. Risser, "Multivariate spatial Gaussian mixture modeling for statistical clustering of hemodynamic parameters in functional MRI," in *34th Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 445–448.

[11] S. Badillo, G. Varoquaux, and P. Ciuciu, "Hemodynamic estimation based on consensus clustering," in *IEEE Pattern Recognition in Neuroimaging (PRNI)*, Philadelphia, USA, June 2013.

[12] K. Friston, P. Fletcher, O. Joseph, A. Holmes, M. Rugg, and R. Turner, "Event-related responses in fmri: characterising differential responses.," *Neuroimage*, vol. 7, pp. 30–40, 1998.

[13] R. Henson, M. Rugg, and K. J. Friston, "The choice of basis functions in event-related fmri," *Neuroimage*, vol. 13, no. 6, pp. 149–149, 2001.

[14] 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.

[15] Sepandar D. Kamvar, Dan Klein, and Christopher D. Manning, "Interpreting and extending classical agglomerative clustering algorithms using a model-based approach," in *ICML*, 2002.

[16] J.H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, pp. 236–244, 1963.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.