

MULTIVARIATE SPATIAL GAUSSIAN MIXTURE MODELING FOR STATISTICAL CLUSTERING OF HEMODYNAMIC PARAMETERS IN FUNCTIONAL MRI

Anne-Laure Fouque^{1,2}, Philippe Ciuciu^{1,2} and Laurent Risser^{1,2}

¹ NeuroSpin/CEA, F-91191 Gif-sur-Yvette, France

² IFR 49, Institut d'Imagerie Neurofonctionnelle, Paris, France

¹ firstname.lastname@cea.fr

ABSTRACT

In this paper, a novel statistical parcellation of intra-subject functional MRI (fMRI) data is proposed. The key idea is to identify functionally homogenous regions of interest from their hemodynamic parameters. To this end, a non-parametric voxel-based estimation of hemodynamic response function is performed as a prerequisite. Then, the extracted hemodynamic features are entered as the input data of a Multivariate Spatial Gaussian Mixture Model (MSGMM) to be fitted. The goal of the spatial aspect is to favor the recovery of connected components in the mixture. Our statistical clustering approach is original in the sense that it extends existing works done on univariate spatially regularized Gaussian mixtures [1]. A specific Gibbs sampler is derived to account for different covariance structures in the feature space. On realistic artificial fMRI datasets, it is shown that our algorithm is helpful for identifying a parsimonious functional parcellation required in the context of joint detection-estimation of brain activity [2]. This allows us to overcome the classical assumption of spatial stationarity of the BOLD signal model.

Index Terms— multivariate Gaussian mixture model, spatial regularization, functional MRI, statistical clustering.

1. INTRODUCTION

Intra-subject analysis in fMRI is usually addressed using a hypothesis-driven approach that actually postulates a model for the Hemodynamic Response Function (HRF) and enables voxelwise inference in the General Linear Model (GLM) framework. In this formulation, the modeling of the BOLD response is crucial. In its simplest form, this modeling relies on a spatially invariant canonical HRF reflecting the BOLD signal best in the visual cortex [3]. However, intra-individual differences in the characteristics of the HRF have been exhibited between cortical areas in [4, 5]. This regional variability is large enough to be regarded with care. To account for these spatial fluctuations at the voxel level, one usually resorts to a hemodynamic function basis. For instance, the canonical HRF can be supplemented with its first and second derivatives to model differences in time [6]. Although powerful and elegant, the price to be paid for such a flexible modeling lies in a loss of sensitivity of detection. To facilitate cognitive interpretations, spatially adaptive GLMs in which a local *non-parametric* estimation of the HRF is performed, have been proposed in [2, 7]. The expected BOLD response is thus factorized with a *single* regressor making direct statistical comparisons between response magnitudes easier. Early investigations have shown that this joint detection-estimation framework of brain activity over functionally homogeneous Regions-Of-Interest (ROIs) pro-

vides more reliable results [2]. The critical issue is to exhibit such ROIs over the whole brain. To that end, several parcellation algorithms have been proposed [8, 9], which segregate the brain into connected and functionally homogeneous regions by minimizing a criterion reflecting both the spatial and functional structures of the dataset. The functional part of this criterion is usually computed from regression coefficients estimated by fitting a GLM.

The goal of the present paper is to avoid fitting a GLM for the definition of the parcellation. To this end, we propose a two-step procedure. As shown in Section 2, it first consists in extracting Hemodynamic Parameters (HP) from non-parametric HRF estimates computed as temporally regularized impulse responses [10]. Second, these voxel-based hemodynamic parameters (*e.g.*, time-to-peak (TTP), time-to-undershoot (TTU), peak magnitude (PM), undershoot magnitude (UM), ...) are entered as input multivariate datasets in a statistical clustering algorithm based upon a MSGMM, as explained in Section 3. A Gibbs sampler is developed for estimating the mixture parameters and identifying the clusters. It extends previous works performed on spatially correlated univariate mixtures [1]. Our approach is validated on synthetic fMRI dataset in Section 4 and conclusions are drawn in Section 5.

2. EXTRACTION OF HEMODYNAMIC PARAMETERS

Let (V_1, \dots, V_J) be the set of voxels covering the brain at the functional spatial resolution. To voxel V_j is associated a fMRI signal $\mathbf{y}_j \in \mathbb{R}^N$ where N is the scan number. The following linear model is used to extract the voxel-based HRFs:

$$\mathbf{y}_j = \mathbf{X} \mathbf{h}_j + \mathbf{P} \boldsymbol{\ell}_j + \mathbf{b}_j, \quad (1)$$

where $\mathbf{h}_j \in \mathbb{R}^{K+1}$ denotes the HRF shape, $\mathbf{X} \in \mathbb{R}^N \times \mathbb{R}^{K+1}$ is the binary onset matrix coding the stimulus occurrences and $\mathbf{P} \in \mathbb{R}^N \times \mathbb{R}^Q$ a orthogonal basis against which the low frequency trend $\boldsymbol{\ell}_j \in \mathbb{R}^Q$ is fitted. Noise $\mathbf{b}_j \in \mathbb{R}^N$ is assumed to be $\mathcal{N}(\mathbf{0}, \sigma_j^2 \mathbf{I}_N)$ -distributed. The session likelihood thus reads:

$$p(\mathbf{y}_j | \mathbf{h}_j, \boldsymbol{\ell}_j, \sigma_j^2) \propto \sigma_j^{-N} \exp(-\|\mathbf{y}_j - \mathbf{X} \mathbf{h}_j - \mathbf{P} \boldsymbol{\ell}_j\|^2 / 2\sigma_j^2).$$

Akin to [10], prior information is introduced on \mathbf{h}_j by setting the first and last parameters of \mathbf{h}_j to zero and considering an a priori Gaussian probability density function (pdf) $\mathcal{N}(0, \sigma_h^2 \mathbf{R})$ with $\mathbf{R} = (\mathbf{D}_2^t \mathbf{D}_2)^{-1}$ and \mathbf{D}_2 stands for a discrete approximation of the second order derivative. The Maximum a posteriori estimate is then derived from the combination of the likelihood with the prior density using Bayes' rule: $p(\mathbf{h}_j | \mathbf{y}_j, \boldsymbol{\ell}_j, \theta_j) \sim \mathcal{N}(\hat{\mathbf{h}}_j^{\text{MAP}}, \boldsymbol{\Omega}_j)$, where $\boldsymbol{\Omega}_j^{-1} = \sigma_j^{-2} \mathbf{X}^t \mathbf{X} + \sigma_h^{-2} \mathbf{R}^{-1}$ and $\hat{\mathbf{h}}_j^{\text{MAP}} = \boldsymbol{\Omega}_j \mathbf{X}^t (\mathbf{y}_j - \mathbf{P} \boldsymbol{\ell}_j) / \sigma_j^2$. An Expectation-Maximization algorithm gives us the simultaneous

The authors thank the ANR for funding the OPTIMED Project.

estimates of \mathbf{h}_j , ℓ_j and $\theta_j = (\sigma_j^2, \sigma_h^2)$.

3. MULTIVARIATE GAUSSIAN MIXTURE MODELS

3.1. The mixture sampling model

Let $\mathbf{y}_j \in \mathbb{R}^n$ be now the feature vector to be clustered such that $n < N$ and n typically describes hemodynamic parameters: TPP, TTO, PM, PU. These parameters have been extracted from the voxel-based HRF estimates computed in Section 2. We consider *conditionally independent* but not identically distributed multivariate observations $\mathbf{y} = (\mathbf{y}_j)_j$ from a GMM:

$$p(\mathbf{y}_j | \mathbf{w}_j, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{k=1}^K w_{jk} \mathcal{N}_n(\mathbf{y}_j; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (2)$$

where K denotes the set of components. The distribution under the k th component ($k = 1 : K$) is a n -dimensional Gaussian distribution with mean $\boldsymbol{\mu}_k \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma}_k \in \mathbb{R}^n \times \mathbb{R}^n$. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K)$ and $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K)$ denote the sets of mean vectors and covariance matrices, respectively. Following [1], the weights $\mathbf{w} = (w_{jk})_{j=1:J}^{k=1:K}$ vary with space index j and satisfy: $w_{jk} > 0$ and $\sum_{k=1}^K w_{j,k} = 1$. In this paper, we shall assume that \mathbf{w} , $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ are unknown and subject to inference while K is supposed to be fixed and known. A generalization to an unknown value of K is nonetheless feasible (see [1, 11] for details). Let $\mathbf{q} = (q_j)_{j=1:J}$ be J independent allocation variables with the multinomial distribution $p(q_j = k | \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = w_{jk}, \forall k = 1 : K$ and assume that \mathbf{y} are independent given \mathbf{q} with $p(\mathbf{y}_j | \mathbf{q}, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}_n(\boldsymbol{\mu}_{q_j}, \boldsymbol{\Sigma}_{q_j})$. Then, we obtain Eq. (2) by integrating out q_j .

The covariance matrices $(\boldsymbol{\Sigma}_k)_k$ can be modeled specifically according to the assumptions that seem tenable between the n features:

- *iid features*: The covariance matrix is given by $\boldsymbol{\Sigma}_k = \sigma_k^2 \mathbf{I}_n$ where \mathbf{I}_n is the n -dimensional identity matrix. Here, a single variance component has to be estimated for each class k .
- *independent features*: $\boldsymbol{\Sigma}_k = \text{diag}[\sigma_{1k}^2, \sigma_{2k}^2, \dots, \sigma_{nk}^2]$. Here, n variance components have to be estimated for each class k .

Full matrices $\boldsymbol{\Sigma}$ are beyond the scope of this paper. Interestingly, in case of correlated features, this phenomenon can be handled using a parsimonious correlation structure like a first order autoregressive model that only depends on $(\sigma_k^2, \rho_k), \forall k$: $\boldsymbol{\Sigma}_k = \sigma_k^2 \boldsymbol{\Lambda}_k^{-1}$ where $\boldsymbol{\Lambda}_k$ is tri-diagonal and built up from ρ_k (cf. [2] for details).

3.2. Proper priors on mixture parameters

The following priors are introduced on the mixture parameters. The retained values for the hyper-parameters allow us to define proper but less informative prior densities.

- Mean vectors: $\boldsymbol{\mu}_k | \boldsymbol{\kappa}, \boldsymbol{\xi} \sim \mathcal{N}_n(\boldsymbol{\xi}, \boldsymbol{\kappa}^{-1}), \forall k = 1 : K$ where vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ and ξ_l corresponds to the half of the range of the l th dimension of $\mathbf{y}^l = (y_1^l, \dots, y_J^l)$ and matrix $\boldsymbol{\kappa}$ is given by: $\boldsymbol{\kappa} = \text{diag}[R_1^{-2}, \dots, R_n^{-2}]$, with R_l the length of the range of the l th dimension of \mathbf{y}^l .
- Covariance matrices: For the iid model, we consider $\forall k = 1 : K, \sigma_k^{-2} | \alpha, \beta \sim \mathcal{G}(\alpha, \beta)$ as prior density while for the independent model, we need the following extension: $\forall, k = 1 : K, l = 1 : n, \sigma_{kl}^{-2} | \alpha, \beta_l \sim \mathcal{G}(\alpha, \beta_l)$. Parameter α is typically fixed ($\alpha = 3$) while scalar β or vector $\boldsymbol{\beta} = (\beta_l)_{l=1:n}$ are unknown and distributed according to $\beta \sim \mathcal{G}(g, s)$ and $\beta_l \sim \mathcal{G}(g, s_l)$, respectively. Again, g is fixed ($g = 0.3$) while

scalar s or vector \mathbf{s} are respectively given by $s = 100g/(\alpha R^2)$ with $R = \sum_{l=1}^n R_l/n$ and $s_l = 100g/(\alpha R_l^2)$ depending on the covariance model.

3.3. Modeling spatial dependence

A crucial difference between (2) and the case of independent and identically distributed mixture previously examined in [11] is that the weights in model (2) are indexed by j so that they are allowed to vary from voxel to voxel where observations take place. There are actually two strategies for encoding spatial dependencies in a mixture model: the first one consists in introducing a spatial structure connecting voxels through a hidden Markov model on allocation variables (e.g. Potts model) [7, 12]. The alternative retained here consists in encoding spatial dependence through the prior distribution of the weights. In simple terms, the basic behaviour that we try to capture is that observations that correspond to nearby locations are more likely to have similar values of the weights (i.e. similar allocation probabilities) than observations from locations that are far apart. Our analysis will therefore be conditional on a given graph with the locations as vertices, and certain designated pairs of nearby locations as edges. This graph will be the conditional independence graph of a component of our models. Thus, we are working with Markov random fields indexed by j . Pairs of locations connected by an edge will be called neighbors.

Here, the proposed model to allow for spatially correlated weights is introduced by means of a Gaussian Markov random field with pdf:

$$p(\mathbf{x}|u) = c(u) \exp\left(-\frac{1}{2}\left(u \sum_{j \sim j'} (x_j - x_{j'})^2 + \sum_{j=1}^J x_j^2\right)\right) \quad (3)$$

$$= \mathcal{N}_n(0, \mathbf{Q}) \quad \text{with} \quad \mathbf{x} = (x_1, \dots, x_J), u > 0.$$

Here, matrix $\mathbf{Q} = \mathbf{I}_J + u\mathbf{A}$ where $\mathbf{A} = (a_{jj'})$ encodes the adjacencies with $a_{jj} = \nu_j$ the number of neighbors of V_j and off-diagonal elements $a_{jj'} = -1$ if voxels V_j and $V_{j'}$ are neighbors and $a_{jj'} = 0$ otherwise. The parameter u is non-negative and $c(u)$ is the appropriate partition function of the MRF (3): $c(u) = (2\pi)^{-J/2} \prod_{j=1}^J (1 + ug_j)^{\frac{1}{2}}$ where g_1, \dots, g_J denote the eigenvalues of \mathbf{A} . For a given graph, these vectors only need to be computed once and can then be stored for any further analyses using the same graph. From Eq. (3), it is clear that neighboring locations are induced to have similar values as the value of the corresponding elements in \mathbf{x} , and this effect is more pronounced as parameter u increases. The limiting case $u = 0$ corresponds to independence across locations, whereas as $u \rightarrow \infty$ the distribution in Eq. (3) tends to a degenerate distribution where neighboring locations are forced to have exactly the same value of the corresponding elements in \mathbf{x} . Note that in model (3) the term $\sum_j x_j^2$ makes model (3) a proper distribution and is necessary to obtain a proper posterior distribution in the mixture setting. In the next subsection, following [1], we explain how Eq. (3) is used to obtain a spatially correlated prior for the weights \mathbf{w} in the mixture model (2).

3.4. The Logistic Normal (LN) model

For a mixture with K components, our approach requires the introduction of K independent vectors $\mathbf{x}_k \equiv (x_{1k}, \dots, x_{Jk}), \forall k = 1 : K$, each distributed according to model (3). Although the vectors are independent of each other, each of them incorporates spatial dependence among its J elements. Next, the weights are defined by using

the logistic transform:

$$w_{jk} = \frac{\exp(x_{jk}/\phi)}{\sum_{l=1}^K \exp(x_{jl}/\phi)}, \quad k = 1 : K, \quad (4)$$

with $\phi > 0$. Thus, the weights for V_j depend on the j th element of each of $\mathbf{x}_1, \dots, \mathbf{x}_K$. The dependence structure of (3) induces spatial dependence among the weights. As u increases in pdf (3), realizations of the \mathbf{x}_k -processes become smoother, and there is also stronger shrinkage of the mean towards zero but the parameter ϕ can compensate for this. Note that $w_{jk}/w_{jl} = \exp((x_{jk} - x_{jl})/\phi)$ is a monotonic function of ϕ converging to 1 as $\phi \rightarrow \infty$ and to either 0 or ∞ when $\phi \rightarrow 0$. A smaller value of ϕ can thus alleviate the effects of increasing shrinkage in the \mathbf{x} -values. The case $\phi = 0$ corresponds to $w_{jk} = 1$ if $x_{jk} = \max_{l=1:K} x_{jl}$ and to $w_{jk} = 0$ otherwise. This implies that the allocation variables \mathbf{q} are deterministic functions of \mathbf{x} . The other limiting case ($\phi \rightarrow \infty$) leads to $w_{jk} = 1/K, \forall (j, k)$, thus precluding any spatial patterns in the weights. We shall restrict ϕ to be sufficiently small to avoid this undesirable feature.

3.5. Priors on spatial interaction parameters

The parameter u controls the dependence in the MRF (3). In particular, $u = 0$ leads to independent weights \mathbf{w} across locations. The scale parameter ϕ affects the size and general behavior of \mathbf{w} and therefore also influences the correlation between the allocation variables \mathbf{q} . To cope with uncertainty about these smoothing parameters, we assign independent prior pdfs to them: $u \sim \mathcal{U}([0, u_{\max}])$ and $\phi \sim \mathcal{U}([0, \phi_{\max}])$ for some positive numbers u_{\max} and ϕ_{\max} . The support intervals are chosen to be sufficiently large to allow for interior modes in the posterior law, but not so large that the unreasonable features displayed by \mathbf{w} as ϕ increases could emerge. Our basic settings are $u_{\max} = \phi_{\max} = 10$.

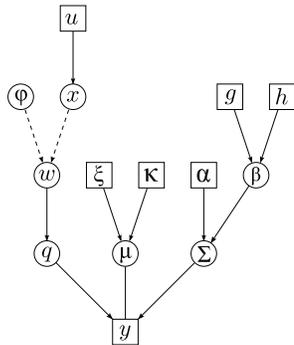


Fig. 1. Directed acyclic graph corresponding to the LN spatial mixture model (2)-(3). Square boxes represent fixed or observed quantities and circles the unknowns.

3.6. Computational implementation

The Bayesian model proposed in this paper is too complex to be amenable to analytical calculations. Hence, we resort to Gibbs sampling since K is assumed to be known. To facilitate sampling, the allocation and auxiliary variables (\mathbf{q} and \mathbf{x}) are introduced to compute the joint posterior pdf of all variables describes in Fig. 1. It is worth mentioning a couple of unusual features of our sampler. Firstly, since the weights \mathbf{w} are a deterministic function of (\mathbf{x}, ϕ) (cf. Eq. (4)) they cannot be included as additional random variables in the sampler. Secondly, in most steps of the sampler the allocation variables

\mathbf{q} will be integrated out to avoid slow mixing for values of ϕ which are small in relation to the values of x_{jk} . The resulting Markov chain is irreducible and has the required posterior law $p(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, u, \phi | \mathbf{y})$ as invariant distribution, so ergodic averages converge to the corresponding posterior expectations. In Table 1, we only give the main sampling steps.

Table 1. Gibbs sampler for fitting the MSGMM (iid features).

- Compute hyper-parameters $\boldsymbol{\kappa}, (R_l)_{l=1:n}, s$
- Initialize: $\forall k(\mathbf{x}_k, \sigma_k^2, \boldsymbol{\mu}_k), \phi, u, \beta$ according to the priors.
- Iteration t : draw samples from the full conditional posteriors:
 - $q : \forall j | \Pr(q_j = k) \propto w_{jk} \mathcal{N}(\mathbf{y}_j; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
 - $\beta \sim \mathcal{G}(g + K\alpha, s + \sum_k \sigma_k^{-2})$
 - Cpt $\forall k, n_k = \sum_{j=1}^J \mathbb{1}_{q_j=k}$ and $\mathbf{m}_k = \sum_{j=1}^J \mathbf{y}_j \mathbb{1}_{q_j=k}$
 - $\boldsymbol{\mu}_k \sim \mathcal{N}\left(\underbrace{(\boldsymbol{\kappa} + \sigma_k^{-2} n_k \mathbb{1}_n)^{-1}}_{=\boldsymbol{\Xi}_k} (\boldsymbol{\kappa} \boldsymbol{\xi} + \sigma_k^{-2} \mathbf{m}_k), \boldsymbol{\Xi}_k\right)$
 - Cpt $\forall j, \tilde{\mathbf{y}}_j = \mathbf{y}_j - \boldsymbol{\mu}_{q_j}$ and $\forall k, S_k^2 = \sum_{j=1}^J \|\tilde{\mathbf{y}}_j\|^2 \mathbb{1}_{q_j=k}$
 - $\forall j, \mathbf{x}_j \sim (\sum_k w_{jk} \mathcal{N}(\mathbf{y}_j; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) \times \dots$
 $\dots \times \prod_{k=1}^K \mathcal{N}(x_{jk}; \underbrace{(1 + w_{jk})^{-1}}_{\eta_j} \sum_{j' \sim j} x_{j'k}, \eta_j)$
 - $\phi \sim \prod_{j=1}^J \left(\sum_{k=1}^K w_{jk} \mathcal{N}(\mathbf{y}_j; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \mathbb{1}_{[0, \phi_{\max}]}(\phi)$
 - $u \sim c(u)^k \exp(-\frac{u}{2} \sum_{k=1}^K \sum_{j \sim j'} (x_{jk} - x_{j'k})^2) \mathbb{1}_{[0, u_{\max}]}(u)$
 - Update the weights \mathbf{w}
- Iterate until convergence is achieved.

4. SIMULATION RESULTS

Artificial fMRI time series were generated according to the additive principle defined in Eq. (1). Twenty trials of a single experimental condition were involved in the experimental paradigm and thus in the definition of matrix \mathbf{X} . The true HRF shape followed the canonical hemodynamic filter defined in [3]. Its parametric definition allows us to specify true spatially-varying HPs as illustrated in Fig. 2. White Gaussian noise \mathbf{b}_j and low-frequency drift $\mathbf{P}\ell_j$ were then superimposed to $\mathbf{X}\mathbf{h}_j$ in every pixel of the 2D grid. The noise was varied in space according to a *signal-to-noise ratio* ranging from -10 to 12 dB.

As illustrated in Fig. 3, the first step of our approach consists in extracting HPs from the HRF estimates derived in Section 2. Clearly, these estimates are quite accurate, except for the TTU parameter whose estimate seems noisy (cf. Fig. 3(d)). Fig. 4 illustrates our MSGMM-based clustering results using $K = 4$. First, it is shown that the extracted HPs give clusters that are very close to those provided by the true HPs (compare Fig. 4(b)-(d) vs. Fig. 4(a)-(c)). Moreover, given that the current implementation of the MSGMM approach corresponds to iid features ($\boldsymbol{\Sigma}_k = \sigma_k^{-2} \mathbf{I}_n$), we studied the influence of varying the dimension n : it appears that the removal of the UM parameter, which is differently distributed than other HPs, induces stronger coherence in the clusters recovered from the true and estimated HPs. This probably indicates that diagonal but not iid covariance matrices $\boldsymbol{\Sigma}_k$ would be helpful in such Fig. 4(c)-(d) also demonstrate that the four mixtures components are well localized in space with respect to the true parameters. Finally, Fig. 4 shows that

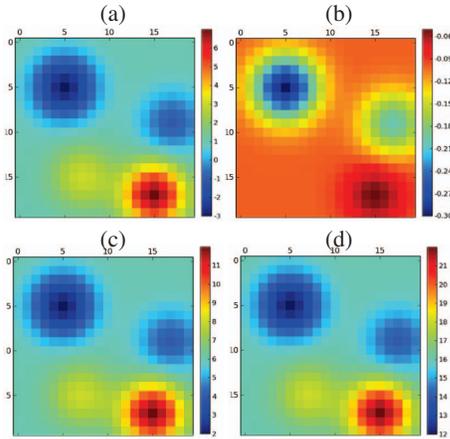


Fig. 2. 2D map of true HP. (a): peak magnitudes; (b): Undershoot magnitudes; (c): time-to-peak; (d): time-to-undershoot.

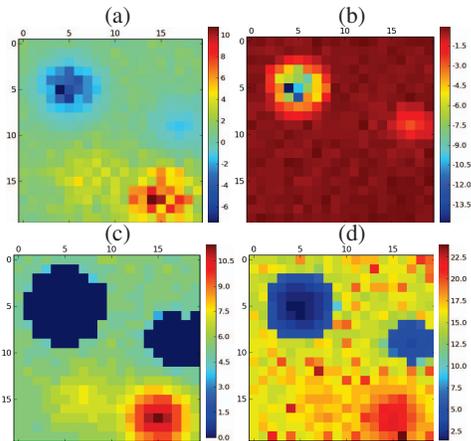


Fig. 3. 2D map of HP estimates extracted from voxel-based \hat{h}_j^{MAP} estimates. (a): PM; (b): UM; (c): TTP; (d): TTU.

each class is not necessarily a connected component suggesting that some post-processing for dividing each class into connected components is required to define a parcellation of the brain.

5. CONCLUSION

The present paper has achieved the goal of deriving a statistical clustering approach from HP estimates. This clustering does not guarantee the recovery of connected ROIs and thus makes necessary some post-processing to generate an input parcellation to the joint detection-estimation framework of brain activity developed in [2]. This approach has been presented in the context of a single stimulus type and could therefore be generalized to multiple conditions. In addition, our clustering is supervised: the number of classes has been chosen empirically while in practice, this number is not known making its unsupervised extension the topic for future works as well as the validation on real data.

6. REFERENCES

[1] Sylvia Fernández and Peter J. Green, “Modelling spatially correlated data via mixtures: a Bayesian approach,” *J. R. Statist. Soc. B*, vol. 64, no. 4, pp. 805–826, 2002.

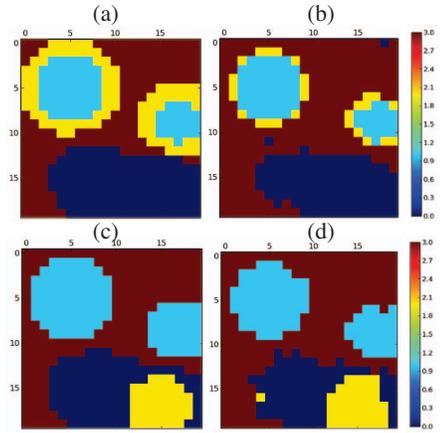


Fig. 4. Clustering results (2D \hat{q} -maps) based upon the MSGMM using $K = 4$; (a)–(c): from true HP; (b)–(d): from extracted HP. (a)–(b): $n = 4$; (c)–(d): $n = 3$ (UM removed).

[2] S. Makni, J. Idier, T. Vincent, B. Thirion, Ghislaine 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, July 2008.

[3] G. H. Glover, “Deconvolution of impulse response in event-related BOLD fMRI,” *Neuroimage*, vol. 9, pp. 416–429, 1999.

[4] G. K. Aguirre, E. Zarahn, and M. D’Esposito, “The variability of human BOLD hemodynamic responses,” *Neuroimage*, vol. 7, pp. 574, 1998.

[5] Daniel A. Handwerker, John M. Ollinger, , and Mark D’Esposito, “Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses,” *Neuroimage*, vol. 21, pp. 1639–1651, 2004.

[6] R. Henson, C. Price, M. Rugg, R. Turner, and K. Friston, “Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations,” *Neuroimage*, vol. 15, pp. 83-, 2002.

[7] T. Vincent, P. Ciuciu, and J. Idier, “Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI,” in *32th Proc. IEEE ICASSP*, Hawaii, Apr. 2007, vol. I, pp. 325-.

[8] 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.

[9] B. Thyreau, B. Thirion, G. Flandin, and J.-B. Poline, “Anatomo-functional description of the brain: a probabilistic approach,” in *Proc. 31th Proc. IEEE ICASSP*, Toulouse, France, May 2006, vol. V, pp. 1109–1112.

[10] 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, pp. 1235-, 2003.

[11] Sylvia Richardson and Peter J. Green, “On Bayesian analysis of mixtures with an unknown number of components (with discussion),” *JRSS B*, vol. 59, no. 4, pp. 731–792, 1997.

[12] D. M. Higdon, “Auxiliary variable methods for Markov chain Monte Carlo with applications,” *J. Amer. Statist. Assoc.*, vol. 93, no. 442, pp. 585–595, June 1998.