

Mélanges spatiaux gaussiens multivariés pour la classification de paramètres hémodynamiques en IRMf

Anne-Laure FOUQUE¹, Philippe CIUCIU¹, Laurent RISSER¹ et Thomas VINCENT¹,

¹NeuroSpin/CEA, 91191 Gif-sur-Yvette cedex, France
annelaure.fouque@cea.fr, philippe.ciuciu@cea.fr
laurent.risser,thomas.vincent@cea.fr

Résumé – Nous présentons un nouvel algorithme de classification statistique de paramètres hémodynamiques. A partir de données 4D d’IRM fonctionnelle (IRMf), il fournit une parcellisation du cerveau en zones fonctionnellement homogènes. Nous procédons en deux étapes: *i.*) une estimation non-paramétrique voxel à voxel du filtre hémodynamique qui permet d’extraire des caractéristiques telles que l’instant du pic, son amplitude, la largeur du filtre, suivie *ii.*) d’une classification de ces paramètres extraits à l’aide d’un modèle de mélange gaussien multivarié spatialement régularisé. En simulation et sur données réelles, notre approche fournit une parcellisation cérébrale robuste des paramètres hémodynamiques.

Abstract – A novel statistical clustering algorithm of haemodynamic parameters is proposed. From 4D fMRI data, it exhibits a brain parcellation into functionally homogenous regions of interest. Our algorithm proceeds in two steps: *i.*) a non-parametric voxel-based estimation of the haemodynamic filter function is performed before extracting haemodynamic features (time-to-peak, peak magnitude, width, ...). *ii.*) A subsequent classification of these parameters is performed using a multivariate spatial Gaussian mixture model. On artificial and real fMRI data, it is shown that our approach provides a reliable brain parcellation of haemodynamics parameters.

1 Introduction

En IRM fonctionnelle (IRMf), la localisation des activations cérébrales est conduite habituellement dans le cadre du modèle linéaire général (MLG), qui postule une forme unique de Fonction de Réponse Hémodynamique (FRH). Dans ce cadre, la modélisation du signal BOLD (Blood Oxygen Level Dependent) est invariante spatialement. Afin de prendre en compte la variabilité inter-régionale évidente de la FRH [1, 2], il est possible d’adapter au niveau de chaque voxel la définition d’un MLG en considérant plusieurs régresseurs par stimulus au lieu d’un seul : l’ajout de la dérivée temporelle permet par exemple de modéliser une fluctuation du délai de la FRH pourvu qu’il soit court [3]. Bien qu’élégante, cette modélisation induit une perte de sensibilité de détection. Pour éviter ce problème, l’introduction de MLGs spatialement adaptatifs à régresseur unique semble plus judicieuse comme illustré dans [4, 5]. Le signal BOLD attendu dérive d’une *unique* régresseur qui s’écrit comme la convolution du paradigme et d’une estimée non-paramétrique de la FRH locale. Les méthodes de détection-estimation conjointe de l’activité cérébrale permettent de définir de tels MLGs sur une parcellation du cerveau en régions fonctionnellement homogènes [5, 6]. La question sensible demeure la dépendance des résultats vis-à-vis du choix de la parcellisation comme illustré dans l’analyse de sensibilité réalisée dans [7]. Les méthodes sta-

tistiques ségrègent le cerveau en régions connexes homogènes fonctionnellement [8]. Toutefois, la régularité fonctionnelle est le plus souvent estimée à partir d’un MLG.

L’objet ici est de définir une parcellisation à partir des données 4D. Dans ce but, nous proposons une procédure en deux étapes. La première consiste à extraire voxel à voxel des paramètres hémodynamiques caractéristiques (*cf.* Section 2). La seconde est une étape de classification statistique des paramètres extraits (*cf.* Section 3). Pour ce faire, nous utilisons un modèle de Mélange gaussien multivarié (MGM) spatialement régularisé. Un échantillonneur de Gibbs permet d’en identifier les paramètres et d’en déduire les parcelles. Cet algorithme étend au cadre multivarié les travaux antérieurs de Fernandez et Green [9]. En Section 4, nous validons notre approche sur des données synthétiques et réelles avant de conclure en Section 5.

2 Estimation de l’hémodynamique

Soit (V_1, \dots, V_J) l’ensemble des voxels couvrant le cerveau à la résolution spatiale de l’IRMf. Le voxel V_j est associé au signal $\mathbf{y}_j \in \mathbb{R}^N$ où N est le nombre de scans acquis. L’estimation voxelique de la FRH obéit au modèle suivant : $\mathbf{y}_j = \mathbf{X}\mathbf{h}_j + \mathbf{P}\mathbf{l}_j + \mathbf{b}_j$ où $\mathbf{h}_j \in \mathbb{R}^{K+1}$ définit la réponse impulsionnelle, $\mathbf{X} \in \mathbb{R}^N \times \mathbb{R}^{K+1}$ la matrice des occurrences des stimuli et $\mathbf{P} \in \mathbb{R}^N \times \mathbb{R}^Q$ une base

orthogonale modélisant les basses fréquences sur laquelle $\ell_j \in \mathbb{R}^Q$ est estimé. Le bruit $\mathbf{b}_j \in \mathbb{R}^N$ est supposé blanc gaussien *i.e.*, $\mathcal{N}(\mathbf{0}, \sigma_j^2 \mathbf{I}_N)$ conduisant à une vraisemblance gaussienne. Comme dans [10], l'*a priori* gaussien introduit sur \mathbf{h}_j , $\mathcal{N}(0, \sigma_h^2 \mathbf{R})$ permet de conserver un estimateur du maximum *a posteriori* (MAP) à structure linéaire $\hat{\mathbf{h}}_j^{\text{MAP}} = \mathbf{\Omega}_j \mathbf{X}^t (\mathbf{y}_j - \mathbf{P} \hat{\ell}_j) / \sigma_j^2$ où $\mathbf{\Omega}_j^{-1} = \sigma_j^{-2} \mathbf{X}^t \mathbf{X} + \sigma_h^{-2} \mathbf{R}^{-1}$. Un algorithme EM fournit l'estimé au sens du maximum de vraisemblance de ℓ_j et $\theta_j = (\sigma_j^2, \sigma_h^2)$.

3 Clustering multivarié

3.1 Modèle de mélange gaussien

Dans la suite on note $\mathbf{z}_j \in \mathbb{R}^n$, le vecteur de paramètres hémodynamiques extraits en V_j ; Par exemple, on prendra l'instant et l'amplitude du maximum (pic) et du minimum (*undershoot*) de $\hat{\mathbf{h}}_j^{\text{MAP}}$. Les observations $\mathbf{z} = (\mathbf{z}_j)_j$ sont supposées conditionnellement indépendantes mais non identiquement distribuées selon un MGM :

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

où K définit le nombre de classes. La loi sur la k^{e} composante du mélange est une distribution gaussienne de dimension n de moyenne $\boldsymbol{\mu}_k \in \mathbb{R}^n$ et de covariance $\boldsymbol{\Sigma}_k \in \mathbb{R}^n \times \mathbb{R}^n$. On note $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K)$ et $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K)$ l'ensemble de ces vecteurs et matrices, respectivement. D'après [9], les poids $\mathbf{w} = (w_{jk})_{j=1:J}^{k=1:K}$ varient spatialement et vérifient : $w_{jk} > 0$ et $\sum_{k=1}^K w_{j,k} = 1$. Ici, nous supposons que $\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ sont inconnus alors que K est fixé à l'avance. Une généralisation pour inférer une valeur de K est néanmoins possible via un schéma d'inférence à sauts réversibles (voir [9, 11] pour les détails). Soit $\mathbf{q} = (q_j)_{j=1:J}$ des variables latentes indépendantes de distribution multinomiale $p(q_j = k | \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = w_{jk}, \forall k = 1 : K$. Nous supposons que \mathbf{z} sont indépendantes étant donné $\mathbf{q} : p(\mathbf{z}_j | \mathbf{q}, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}_n(\boldsymbol{\mu}_{q_j}, \boldsymbol{\Sigma}_{q_j})$. Ainsi, nous obtenons (1) en intégrant q_j hors du problème. Les covariances $(\boldsymbol{\Sigma}_k)_k$ sont définies selon des hypothèses que l'on fait sur les n dimensions caractéristiques :

- *hyp. homoscédastique* : la covariance s'écrit $\boldsymbol{\Sigma}_k = \sigma_k^2 \mathbf{I}_n$ où \mathbf{I}_n est la matrice identité de taille $n \times n$ et une seule variance est à estimer par classe.
- *hyp. hétéroscédastique* : $\boldsymbol{\Sigma}_k = \text{diag}[\sigma_{1k}^2, \sigma_{2k}^2, \dots, \sigma_{nk}^2]$ et n variances doivent être estimées au sein de chaque classe k .

D'autres modèles non-diagonaux de $\boldsymbol{\Sigma}$ invoquant des corrélations entre composantes peuvent être envisagés.

Outre ces hypothèses, nous introduisons des lois *a priori* propres sur les paramètres du mélange pour éviter les dégénérescences. Pour les vecteurs moyennes, nous choisissons $\boldsymbol{\mu}_k | \boldsymbol{\kappa}, \boldsymbol{\xi} \sim \mathcal{N}_n(\boldsymbol{\xi}, \boldsymbol{\kappa}^{-1}), \forall k = 1 : K$ où $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ et ξ_l correspond à la moitié de l'échelle de variation selon

la l^{e} dimension de $\mathbf{z}^l = (z_1^l, \dots, z_J^l)$ et la covariance est donnée par $\boldsymbol{\kappa} = \text{diag}[R_1^{-2}, \dots, R_n^{-2}]$, avec R_l la gamme de variation de la l^{e} dimension de \mathbf{z}^l . Pour le modèle *homoscédastique*, la loi *a priori* considérée est $\forall k = 1 : K, \sigma_k^{-2} | \alpha, \beta \sim \mathcal{G}(\alpha, \beta)$. Pour l'extension *hétéroscédastique*, on a $\forall k = 1 : K, l = 1 : n, \sigma_{kl}^{-2} | \alpha, \beta_l \sim \mathcal{G}(\alpha, \beta_l)$. Le paramètre α est fixé ($\alpha = 3$) alors que le scalaire β ou le vecteur $\boldsymbol{\beta} = (\beta_l)_{l=1:n}$ sont inconnus et distribués selon $\beta \sim \mathcal{G}(g, s)$ et $\beta_l \sim \mathcal{G}(g, s_l)$, respectivement. A nouveau, g est fixé ($g = 0.3$) et s ou \mathbf{s} sont donnés par $s = 100g/(\alpha R^2)$ avec $R = \sum_{l=1}^n R_l/n$ et $s_l = 100g/(\alpha R_l^2)$ fonction de la covariance.

3.2 Modélisation de la corrélation spatiale

Une différence fondamentale entre (1) et le cas d'un mélange iid [11] réside dans les poids w_{jk} variables spatialement. Au lieu d'introduire une structure d'interactions locales sur les variables latentes q_j comme dans [12, 5], nous définissons un modèle spatial portant sur la distribution *a priori* des poids $\boldsymbol{\omega}$. Nous modélisons le fait que les observations correspondant à des voxels voisins ont une probabilité plus élevée d'avoir des poids plus proches que des observations éloignées. Ainsi, nous introduisons des champs de Markov gaussiens 3D aux plus proches voisins sur chacune des K classes :

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

$$= \mathcal{N}_n(0, \mathbf{Q}^{-1}) \quad \text{avec} \quad \mathbf{x} = (x_1, \dots, x_J), u > 0.$$

L'inverse de covariance s'écrit $\mathbf{Q} = \mathbf{I}_J + u\mathbf{A}$ où $\mathbf{A} = (a_{jj'})$ définit les adjacences avec $a_{jj} = \nu_j$ le nombre de voisins de V_j , et $a_{jj'} = -1$ pour $V_j \sim V_{j'}$ voisins et nul sinon. Le paramètre $u \geq 0$ règle le niveau de régularisation spatiale et $c(u)$ normalise le champ gaussien (2) : $c(u) = (2\pi)^{-J/2} \prod_{j=1}^J (1 + u g_j)^{-1/2}$. Les valeurs propres $(g_j)_j$ de \mathbf{A} ne sont calculées qu'une fois pour une topologie donnée. De l'Eq. (2), on déduit que des voxels voisins ont des composantes \mathbf{x} d'autant plus proches que u est grand. A l'inverse, $u = 0$ correspond à la situation d'indépendance. Le terme séparable $\sum_j x_j^2$ dans le modèle (2) rend la distribution *a priori* intégrable, condition suffisante pour obtenir une loi *a posteriori* non dégénérée dans le cas des mélanges. D'après [9], nous établissons maintenant le lien entre (2) et les poids $\boldsymbol{\omega}$ dans (1).

3.3 Modèle logistique gaussien

Pour un mélange à K classes, il suffit d'introduire K vecteurs indépendants $\mathbf{x}_k \equiv (x_{1k}, \dots, x_{Jk}), \forall k = 1 : K$, chacun étant distribué selon le modèle (2). Chaque \mathbf{x}_k modélise une dépendance spatiale entre ses J éléments selon une dimension parmi n . En faisant interagir les n dimensions, on définit ainsi les poids à l'aide d'une *transformer*

mation logistique : $w_{jk} = \frac{\exp(x_{jk}/\phi)}{\sum_{l=1}^K \exp(x_{jl}/\phi)}$, $\forall k = 1 : K$, où $\phi > 0$. Ainsi, les poids pour V_j dépendent du j^e élément de chaque vecteur $\mathbf{x}_1, \dots, \mathbf{x}_K$. Quand u augmente dans (2), les \mathbf{x}_k deviennent plus homogènes et ont tendance à se concentrer vers zéro mais le paramètre ϕ peut atténuer cet effet. Remarquons que $w_{jk}/w_{jl} = \exp((x_{jk} - x_{jl})/\phi)$ est une fonction monotone de ϕ qui tend vers 1 quand $\phi \rightarrow \infty$ et vers 0 ou ∞ quand $\phi \rightarrow 0$. Une faible valeur de ϕ peut donc éliminer cet écrasement vers zéro des valeurs de \mathbf{x} . Le cas $\phi = 0$ correspond à $w_{jk} = 1$ si $x_{jk} = \max_{l=1:K} x_{jl}$ et à $w_{jk} = 0$ sinon. Cela implique que les variables latentes \mathbf{q} sont des fonctions déterministes de \mathbf{x} . L'autre cas limite ($\phi \rightarrow \infty$) conduit à $w_{jk} = 1/K, \forall (j, k)$, empêchant ainsi toute structuration spatiale des poids. Ici, nous restreignons ϕ à des valeurs suffisamment faibles pour éviter ces effets indésirables.

Pour tenir compte de l'incertitude sur les paramètres scalaires u, ϕ , nous introduisons des lois *a priori* indépendantes : $u \sim \mathcal{U}([0, u_{\max}])$ et $\phi \sim \mathcal{U}([0, \phi_{\max}])$ pour des seuils u_{\max} et ϕ_{\max} positifs. Les supports sont choisis suffisamment grands ($u_{\max} = \phi_{\max} = 10$) pour contenir des modes de la distribution *a posteriori* mais pas trop grands afin d'éviter l'apparition de caractéristiques indésirables dans \mathbf{w} quand ϕ augmente.

3.4 Inférence bayésienne

Un échantillonneur de Gibbs est utilisé comme algorithme pour inférer les différents paramètres à K fixé. Pour simplifier l'échantillonnage, les variables latentes et auxiliaires (\mathbf{q} et \mathbf{x} respectivement) sont introduites dans le schéma de simulation de la distribution conjointe *a posteriori*. Deux propriétés saillantes de cet algorithme sont à mentionner : d'abord, comme les poids \mathbf{w} dépendent de façon déterministe de (\mathbf{x}, ϕ) ils ne sont pas simulés. Ensuite, dans la plupart des lois conditionnelles à simuler, les variables \mathbf{q} sont marginalisées afin d'éviter un faible taux de mélange de la chaîne de Markov. Ce schéma génère néanmoins une chaîne irréductible tout en conservant la loi *a posteriori* $p(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, u, \phi | \mathbf{z})$ comme distribution invariante. L'ergodicité est donc garantie et les moyennes d'échantillons convergent donc vers les espérances *a posteriori* marginales, nos estimateurs.

4 Validation expérimentale

4.1 Simulations

Des signaux synthétiques ont été générés selon le modèle linéaire présenté en Section (2). Vingt répétitions d'un seul stimulus ont été utilisées pour générer un paradigme événementiel rapide de même que la FRH canonique pour simuler les décours temporels. Nous avons joué sur sa forme paramétrique pour faire varier spatialement sa forme selon

les trois paramètres décrits en haut de la Fig. 1. Du bruit blanc gaussien \mathbf{b}_j et des dérives basse fréquence $\mathbf{P}\ell_j$ ont été ajoutées à $\mathbf{X}\mathbf{h}_j$ dans chaque pixel d'une grille 2D. Le rapport signal sur bruit varie dans $[-10, 12]$ dB. L'extraction des paramètres hémodynamiques à partir des FRHs voxeliques estimées fournit les cartes du bas en Fig. 1.

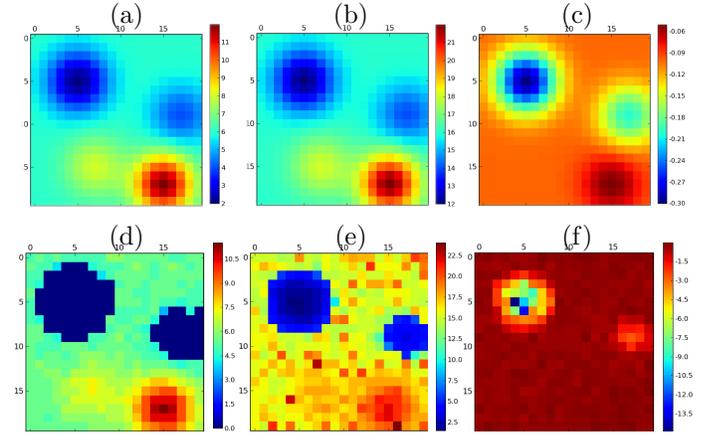


FIG. 1 – **En haut**, cartes des vrais paramètres de FRHs simulées : (a) : instants du pic (b) : instants de l'undershoot et (c) : amplitude de l'undershoot. **En bas** : Paramètres correspondants extraits de $\hat{\mathbf{h}}_j^{\text{MAP}}$

Ici, on compare les modèles homo et hétéroscélastiques de $\boldsymbol{\Sigma}_k$ sur un mélange à $K = 3$ classes (une couleur par classe) identifié sur les paramètres hémodynamiques estimés en Fig.1(d)-(f). La carte du modèle homoscélastique Fig. 2(a) apparaît beaucoup plus bruitée que celle du modèle hétéroscélastique en Fig. 2(b). La raison tient au manque de souplesse du modèle homoscélastique, comme l'illustre la Table 1 où pour chaque classe, la variance estimée selon chaque dimension est très différente. L'estimation de certains paramètres comme l'amplitude de l'undershoot ($n = 3$, Fig 1(f)) étant plus robuste, le modèle hétéroscélastique adapte son degré de précision à chacun des paramètres et conserve ainsi l'information de cohérence spatiale des données.

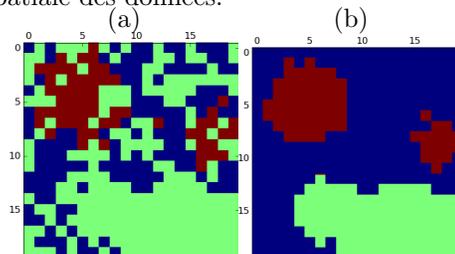


FIG. 2 – Classifications par le MGM homoscélastique (a) et hétéroscélastique (b) avec $K = 3$ et une couleur par classe.

4.2 Données réelles

Les données réelles ont été acquises au cours d'un paradigme événementiel rapide (constitué de 60 stimuli répartis en dix conditions) visant à localiser les cortex sensoriels (auditif, visuel, moteur) et d'autres aires cognitives

TAB. 1 – Variances des classes des modèles de mélange homo et hétéroscédastiques.

	Homoscédastique	Hétéroscédastique		
	σ_k^2	σ_{1k}^2	σ_{2k}^2	σ_{3k}^2
Classe 1	0.28	0.33	4.34	0.15
Classe 2	1.47	0.22	6.02	0.52
Classe 3	0.88	2.06	2.77	0.17

de plus haut niveau impliquées dans le calcul et la lecture au cours d’une seule session constituée de $N = 125$ volumes définis par les paramètres $TR/TE = 2400/30$ ms à la résolution spatiale $3 \times 3 \times 3\text{mm}^3$.

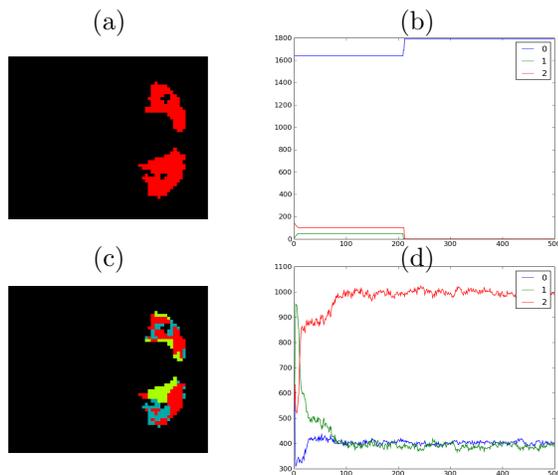


FIG. 3 – (a)-(b) **Homoscédastique** : classification à $K = 3$ classes du cortex visuel (a) et nombre de voxels par classe (b) ; (c)-(d) **Hétéroscédastique** : même vue axiale (c) et cardinal des classes (d).

La Fig. 3 illustre sur une coupe axiale du cerveau la comparaison entre les modèles homo et hétéroscédastique d’un MGM à $K = 3$ classes. La dégénérescence observée dans le cas homoscédastique (une seule classe en rouge) est illustrée en Fig. 3(a)-(b) à travers l’évolution du cardinal de chaque classe au cours des $K = 500$ itérations de l’échantillonneur de Gibbs. A nouveau, le modèle hétéroscédastique régularise ce problème et produit une carte assez homogène et réaliste par rapport à l’organisation rétinotopique du cortex visuel : la classe 2 en rouge correspond au cortex primaire et la classe en vert à une partie du cortex secondaire. Le cortex primaire étant de taille plus grande, on retrouve logiquement que la classe en rouge est de cardinalité maximale.

5 Conclusion

Nous avons proposé dans ce papier une approche de classification multivariée à base de MGM spatialement régularisé. La méthode se décline selon différents modèles en fonction de la structure de covariance des classes. Tant sur simulation que sur données réelles d’IRMf, le modèle

hétéroscédastique apparaît comme le plus robuste. Cette approche a été validée dans un cadre « mono-condition » et s’étend sans difficulté au cadre multi conditions. A l’heure actuelle, la classification reste supervisée au sens où le nombre de classes est choisi à l’avance. L’ajout d’une étape de sélection de modèle *a posteriori* nous permettra dans le futur de rendre la méthode non supervisée.

Références

- [1] G. K. Aguirre, E. Zarahn et M. D’Esposito, « The variability of human BOLD hemodynamic responses », *Neuroimage*, vol. 7, pp. 574, 1998.
- [2] D. A. Handwerker, J. M. Ollinger, et 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, 2004.
- [3] R. Henson, C. Price, M. Rugg, R. Turner et K. Friston, « Detecting latency differences in event-related BOLD responses : application to words vs. nonwords and initial versus repeated face presentations », *Neuroimage*, vol. 15, pp. 83–97, 2002.
- [4] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz et P. Ciuciu, « A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI », *Neuroimage*, vol. 41, 3, pp. 941–969, juil. 2008.
- [5] T. Vincent, P. Ciuciu et J. Idier, « Spatial mixture modeling for the joint detection-estimation of brain activity in fMRI », in *32th Proc. IEEE ICASSP*, Honolulu, Hawaii, USA, avr. 2007, vol. I, pp. 325–328.
- [6] L. Risser, T. Vincent, P. Ciuciu et J. Idier, « Robust extrapolation scheme for fast estimation of 3D Ising field partition functions. application to within-subject fMRI data analysis. », *MICCAI’09*, London, UK, sep. 2009.
- [7] T. Vincent, P. Ciuciu et B. Thirion, « Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI », in *5th Proc. IEEE ISBI*, Paris, France, mai 2008, pp. 568–571.
- [8] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu et J.-B. Poline, « Dealing with the shortcomings of spatial normalization : Multi-subject parcellation of fMRI datasets », *Hum. Brain Mapp.*, vol. 27, 8, pp. 678–693, 2006.
- [9] S. Fernández et P. J. Green, « Modelling spatially correlated data via mixtures : a Bayesian approach », *J. R. Statist. Soc. B*, vol. 64, 4, pp. 805–826, 2002.
- [10] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, C. Pallier et H. Benali, « Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment », *IEEE Trans. Med. Imag.*, vol. 22, 10, pp. 1235–1251, oct. 2003.
- [11] S. Richardson et P. J. Green, « On Bayesian analysis of mixtures with an unknown number of components », *J. R. Statist. Soc. B*, vol. 59, 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, 442, pp. 585–595, juin 1998.