

# ANALYSE PARCIMONIEUSE DES DONNÉES D'IRM FONCTIONNELLE DANS UN CADRE BAYÉSIEN VARIATIONNEL

C. Bakhous<sup>1,3</sup>, F. Forbes<sup>1</sup>, F. Enikeeva<sup>1</sup>, T. Vincent<sup>1</sup>, M. Dojat<sup>3</sup>, P. Ciuciu<sup>2</sup>

<sup>1</sup> INRIA Rhône-Alpes, 655 avenue de l'Europe, Montbonnot 38334 Saint Ismier Cedex, France firstname.lastname@inria.fr	<sup>2</sup> CEA/Neurospin & INRIA, Parietal, CEA Saclay, Bbt. 145, Point Courrier 156, 91191 Gif-sur-Yvette cedex, France firstname.lastname@cea.fr	<sup>3</sup> Centre de Recherche Inserm U836/UJF/CEA/CHU Chemin Fortuné Ferrini, Bâtiment : Edmond J. Safra, 38706 La Tronche Cedex firstname.lastname@ujf- grenoble.fr
--	--	---

**Résumé.** L'analyse des données d'Imagerie par Résonance Magnétique fonctionnelle (IRMf) est principalement effectuée à travers le modèle linéaire général (GLM) dans lequel l'activité d'une région cérébrale est supposée dépendre des différents types de stimuli (moteur, visuel, etc.) or la spécialisation fonctionnelle cérébrale indique que l'activation d'une région donnée n'est induite que par certains de ces stimuli. Inclure des conditions non pertinentes peut dégrader les résultats, en particulier quand la fonction de réponse hémodynamique (FRH) est conjointement estimée. De plus la sélection *a priori* des conditions pertinentes pour chaque région cérébrale n'est pas toujours possible (e.g. pathologie). Afin de faire face à ces difficultés, nous proposons une procédure variationnelle efficace permettant la sélection automatique des conditions selon l'activité cérébrale qu'elles suscitent. Une amélioration de la détection d'activation ainsi que de l'estimation de la FRH sont illustrées sur données réelles.

**Mots-clés.** Sélection de variable, détection-estimation conjointe, inférence bayésienne, imagerie par résonance magnétique fonctionnelle.

**Abstract.** Analysing functional Magnetic Resonance Imaging (fMRI) data is mainly done using the general linear model (GLM) in which the activation of a brain area is supposed to depend on all delivered stimuli (e.g. motor, visual, etc.) although activation is likely to be induced by only some of them in specific brain areas. Inclusion of irrelevant events may degrade the results, particularly when the Hemodynamic Response Function (HRF) is jointly estimated. In addition, a prior selection of relevant condition for each brain region is not always possible (e.g. pathology). To face this issue, we propose an efficient variational procedure that automatically selects the conditions according to the brain activity they elicit. It follows an improved activation detection and local HRF estimation that we illustrate on real fMRI data.

**Keywords.** Variable selection, joint detection-estimation, bayesian inference, functional magnetic resonance imaging.

# 1 Introduction

Les études en IRM fonctionnelle ont pour but principal la détection de l'activité cérébrale correspondante aux différents types de stimulus ainsi que l'estimation de la fonction de réponse hémodynamique (FRH) exprimant le niveau d'oxygénation sanguine. Le deuxième point est généralement ignoré et la FRH est supposée avoir une forme canonique constante à travers le cerveau entier. Ceci ne reflète pas toujours l'état physiologique surtout dans les cas pathologiques et peut donc dégrader les résultats de traitement.

L'analyse pertinente de ce type de donnée est basée sur le fait que l'activité de chaque région cérébrale est suscitée par des tâches cognitives spécifiques. Ce qui se résume donc à la prise en compte ou pas de chaque type de stimulus durant l'analyse de cette région.

Dans le contexte du modèle linéaire général utilisant une forme prédéfinie de la FRH, une étude pertinente est effectuée à travers des tests de Fisher entre les modèles réduits et complets [1, 2]. Une méthode alternative de détection-estimation conjointe bayésienne (DEC) proposée par [3], prend en compte la variabilité de la FRH à travers le cerveau en traitant la détection et l'estimation simultanément. Dans ce contexte, [4] propose une approche de sélection de modèle basée sur la comparaison de différents modèles afin de sélectionner le meilleur. Cette procédure nécessite l'exécution de tous les modèles possibles, de plus la famille de stimuli pertinents change entre les régions cérébrales. Dans ce papier, nous proposons une procédure adaptative permettant la sélection automatique des conditions expérimentales nécessaires grâce à une variable binaire supplémentaire exprimant la pertinence (en terme d'activité évoquée) de chaque type de stimuli ce qui nous ramène à une procédure de sélection de variables [5] déjà utilisée dans [6, 1] afin d'exprimer parcimonieusement l'activité cérébrale. Dans l'approche DEC la détection d'activité est traitée en modélisant les voxels activés et non-activés par un mélange gaussien à deux classes au lieu d'un *a priori* Bernoulli-Gaussien [6]. L'utilisation des variables binaires dans notre cas concerne plutôt la sélection des types de stimuli effectuée à travers tous les voxels (Section 2). L'approche proposée est mise en place dans un cadre EM variationnel (Section 3) plus efficace en temps de calcul comparé à la procédure stochastique intense utilisée dans [7]. Des résultats sur données réelles montrent l'intérêt de notre modèle appelé *modèle parcimonieux* par rapport au *modèle complet* qui prend en compte tous les types de stimulus.

## 2 Détection-estimation conjointe parcimonieuse

Un vecteur est par convention un vecteur colonne. On note la transposée par  $^t$ . Sauf cas contraire,  $j$ ,  $m$  et  $i$  représentent successivement les indices à travers les voxels, types de stimuli et composantes du modèle de mélange (classes d'activation). La distribution gaussienne de moyenne  $\mu$  et variance  $\Sigma$  est écrite sous la forme  $\mathcal{N}(\mu, \Sigma)$ .

Selon [3, 8], pour une parcelle  $\gamma$ , les données observées sont  $y = \{y_j, j \in \gamma\}$  où  $y_j$  est un vecteur de dimension  $N$  représentant le signal BOLD mesuré au voxel  $j \in \gamma$ . Les

états d'activation sont représentés par les étiquettes  $q = \{q^m, m=1:M\}$  avec  $q^m = \{q_j^m, j \in \gamma\}$  où  $q_j^m = i$  signifie que le voxel  $j$  appartient à la classe d'activation  $i$  pour la condition expérimentale  $m$ . Généralement, le nombre de classe est de 2 pour les voxels activés ( $i=1$ ) et inactivés ( $i=0$ ). Les amplitudes d'activation sont modélisées par les niveaux de réponses neuronales (NRNs)  $a = \{a^m, m=1:M\}$  avec  $a^m = \{a_j^m, j \in \gamma\}$  et  $a_j = \{a_j^m, m=1:M\}$ . Le but du modèle parcimonieux est de détecter la sous-famille des  $M$  conditions expérimentales provoquant le signal BOLD dans une parcelle donnée.

Un type de stimuli est identifié comme non pertinent si la moyenne de l'activité cérébrale qu'il suscite est très faible. Cette activité est donc considérée comme un artefact et le stimulus correspondant exclu de l'analyse. Ceci est différent de [7] dans lequel une condition est considérée non pertinente si le nombre de voxels activés est très faible. Comme le confirme les résultats ci-dessous, le choix du seuil de pertinence est plus facile dans le cas d'un critère basé sur la moyenne d'activation. Afin de mettre en place une telle sélection, nous considérons un groupe de  $M$  variables binaires  $w = \{w^m, m=1:M\}$  où  $w^m=1$  (resp.  $w^m=0$ ) implique que le  $m$ ème stimulus est pertinent (resp. non pertinent). La FHR est représentée par  $h = [h_0, h_{\Delta t}, \dots, h_{D\Delta t}]^t$ , un vecteur de dimension  $(D+1)$  avec  $\Delta t$  le pas d'échantillonnage. Finalement les artefacts physiologiques sont modélisés par  $\mathbf{P}\ell$  où  $\mathbf{P}$  est une base orthogonale de fonctions bases fréquences de dimension  $N \times L$  et  $\ell = \{\ell_j, j \in \gamma\}$  les coefficients des dérives bases fréquences.

Les signaux observés sont donc exprimés par le modèle génératif bilinéaire parcimonieux suivant:

$$\forall j \in \gamma, \quad y_j = \sum_{m=1}^M w^m a_j^m \mathbf{X}^m h + \mathbf{P}\ell_j + \varepsilon_j \quad (1)$$

Où  $\mathbf{X}^m$  est une matrice binaire de dimension  $N \times (D+1)$  codant les temps d'arrivée du  $m$ ème stimulus échantillonné sur une grille de résolution  $\Delta t$ ,  $\varepsilon_j$  représente un bruit gaussien indépendamment distribué  $\varepsilon_j \sim \mathcal{N}(\mathbf{0}, \sigma_j^2 \mathbf{I}_N)$  ( $\mathbf{I}_N$  est la matrice identité de dimension  $N \times N$ ). (Pour plus de détails voir [3, 8]).

La densité jointe est donc  $p(y, w, a, h, q) = p(y | w, a, h) p(a | w, q) p(w) p(h) p(q)$  Dans laquelle on distingue termes suivant:

**La vraisemblance**  $p(y | w, a, h)$  définie par  $p(y | w, a, h) = \prod_{j \in \gamma} p(y_j | w, a_j, h; \sigma_j^2)$  avec  $(y_j | w, a_j, h; \sigma_j^2) \sim \mathcal{N}(\sum_{m=1}^M w^m a_j^m \mathbf{X}^m h + \mathbf{P}\ell_j, \sigma_j^2 \mathbf{I}_N)$ .

**Le terme**  $p(a | w, q)$  dans lequel les étiquettes  $q_j^m$  permettent de différencier les voxels actifs des non-actifs. Les NRNs sont supposés indépendants conditionnellement aux étiquettes  $q_j^m$  et aux variables de pertinence  $w^m$ :  $p(a | w, q) = \prod_m \prod_j p(a_j^m | w^m, q_j^m)$ . Pour  $i \in \{0, 1\}$ ,  $(a_j^m | w^m=1, q_j^m=i) \sim \mathcal{N}(\mu_i^m, v_i^m)$  et  $(a_j^m | w^m=0, q_j^m=i) \sim \mathcal{N}(\mu_0^m, v_0^m)$  d'où le rôle de la variable de pertinence  $w^m$ . Les paramètres gaussiens  $\theta_a = \{\mu_1^m, v_1^m, v_0^m, m=1:M\}$  doivent être estimés mais on fixe  $\mu_0^m=0$  pour tout  $m$ . L'idée est que pour un stimulus pertinent ( $w^m=1$ ) la distribution de  $a_j^m$  dépend de l'étiquette  $q_j^m$  au voxel  $j$ . Alors que pour un stimulus non pertinent ( $w^m=0$ ),  $q_j^m$  n'a aucune influence sur  $a_j^m$ , qui est distribué autour de 0 exprimant l'absence d'une réponse significative correspondante au stimulus  $m$ .

**Le terme**  $p(w)$  reliant les variables de pertinence aux moyennes d'activation  $\mu_1^m$ . Ces variables sont indépendantes à travers les différents stimuli  $p(w) = \prod_{m=1}^M p(w^m)$  et pour chaque  $m$ ,  $w^m$  suit une distribution de Bernoulli dont la probabilité de succès est une sigmoïde de paramètre  $\mu_1^m$ ,  $p(w^m = 1) = (1 + \exp(-\tau_1[(\mu_1^m)^2 - \tau_2]))^{-1}$  où  $\tau_1$  contrôle la pente de la sigmoïde et  $\tau_2$  son point d'inflexion considéré comme le seuil de pertinence au-dessus duquel le stimulus a une grande probabilité d'être considéré comme pertinent. Une condition  $m$  est considérée comme non pertinente quand l'estimation de  $\mu_1^m$  est proche de 0 reflétant l'absence d'une activité significative pour le stimulus  $m$ .

**les termes**  $p(h)$  et  $p(q)$  ne sont pas détaillés ici (voir [8]). Ils impliquent respectivement un paramètre de lissage de la FRH  $\sigma_h^2$  et un paramètre de régularisation spatiale pour chaque condition  $\{\beta^m, m=1:M\}$ .

**Les paramètres**  $\theta = \{\sigma_j^2, \mu_1^m, v_1^m, v_0^m, \tau_1, \tau_2, \sigma_h^2, \beta^m, j \in \gamma, m = 1:M\}$  sont estimés à travers la procédure EM de manière semblable à [8]. En ce qui concerne les paramètres de la sigmoïde,  $\tau_1 = \frac{1}{\tau_2} \log\left(\frac{1-p_0}{p_0}\right)$  où  $p_0 = 0.001$  est l'ordonnée à l'origine  $p(w^m = 1 | \mu_1^m = 0, \tau_1, \tau_2)$ . Le deuxième paramètre  $\tau_2$  est estimé avec un *a priori* gamma dont les paramètres sont choisis de manière à éviter les très petites valeurs de  $\tau_2$  causant des grandes pentes ( $\tau_1$  élevé) et induisant donc des problèmes numériques et les grandes valeurs pouvant influencer négativement les conditions pertinentes.

### 3 Estimation maximisation variationnelle

Les calculs analytiques directs étant complexes, on se réfère à une procédure itérative d'Estimation-Maximisation (EM) variationnelle. Par rapport à [8],  $M$  étapes-E additionnelles prennent en compte les  $M$  variables binaires  $w$  en plus de la considération de leur impact sur les autres variables et paramètres du modèle. On s'intéresse plus particulièrement à l'estimation des probabilités *a posteriori*  $p(w^m | y)$ . A l'itération ( $r$ ), notons l'estimation courante de ces probabilités  $\tilde{p}_{w^m}^{(r)}$ , celle des paramètres  $\theta^{(r)}$  et  $E_{\tilde{p}}[\cdot]$  la moyenne par rapport à une densité de probabilité  $\tilde{p}$ . Pour chaque  $m=1:M$ ,  $\tilde{p}_{w^m}^{(r)}(w^m)$  est proportionnel à  $\exp\left(E_{\tilde{p}_{a,q,h,w^{m' \neq m}}^{(r-1)}}[\log p(w^m | y, a, q, h, w^{m' \neq m}, \theta^{(r)})]\right)$  où  $\tilde{p}_{a,q,h,w^{m' \neq m}}^{(r-1)} = \tilde{p}_a^{(r-1)} \tilde{p}_q^{(r-1)} \tilde{p}_h^{(r-1)} \tilde{p}_{w^{m' \neq m}}^{(r-1)}$  représentent l'estimation variationnelle courante de la densité de probabilité des variables  $a, q, h$  et  $w^{m' \neq m}$ . Il s'ensuit que  $\tilde{p}_{w^m}^{(r)}(w^m)$  est proportionnel au produit des trois termes suivants:  $\prod_{j \in \gamma} \exp(E_{\tilde{p}_{h,a_j,w^{m' \neq m}}^{(r-1)}}[\log p(y_j | w, a_j, h; \sigma_j^{2(r)})])$ ,  $\prod_{j \in \gamma} \exp(E_{\tilde{p}_{a_j^m, q_j^m}^{(r-1)}}[\log p(a_j^m | w^m, q_j^m; \theta_a^{(r)})])$  et  $p(w^m; \mu_1^{m(r)}, \tau_1, \tau_2)$ . Les calculs détaillés montrent que la probabilité de  $w^m$  est influencée par la valeur estimée de  $\mu_1^m$  à travers le troisième terme. Quand  $\mu_1^{m2} \geq \tau_2$ , ce terme favorise  $w^m = 1$  vs  $w^m = 0$ . De façon similaire, le second terme dépend de  $\tilde{p}_{a_j^m}^{(r-1)}$  pour  $j \in \gamma$ . Quand la moyenne de ces densités de probabilités tend vers 0,  $w^m = 0$  est favorisé. Enfin,  $w^m = 0$  est favorisé quand les  $y_j$  dans le premier terme sont bien expliqués par le modèle après suppression des stimulus non pertinents. Autrement dit, quand le bruit est mieux estimé. Les sous-étapes-E des autres variables restent approximatives à celles de [8] après la prise

en compte de l'influence de  $w$ . L'étape M est décomposée en 5 sous-étapes traitant séparément  $\sigma_h^2$ ,  $\{\beta^m, m=1:M\}$ ,  $\{\sigma_j^2, j \in \gamma\}$ ,  $\{\mu_i^m, v_i^m, i=0,1\}$  et  $\tau_2$ . Les deux premières restent inchangées par rapport à [8], les deux autres montrent une dépendance supplémentaire à  $\hat{p}_{w^m}^{(r)}(w^m=1)$ , la quatrième donne des expressions assez proches pour  $v_i^m$  et en fin un algorithme de dichotomie est nécessaire pour l'estimation de  $\mu_1^m$  et  $\tau_2$  dans les deux dernières sous-étapes.

## 4 Résultats et conclusion:

On considère un jeu de données réelles non lissées issues d'un paradigme événementiel rapide [9] comprenant 80 stimuli de type auditif, visuel et moteur ainsi que des tâches cognitives (calcul et compréhension de phrases). Les données comprennent une seule session de  $N=128$  fonctionnels avec un temps de répétition de  $2.4s$ , chaque scan représente un volume 3D de dimension  $96 \times 96 \times 40$  voxels. L'intervalle inter stimuli (ISI) moyen est de  $3.76s$  avec un écart-type de  $1.99s$ . Afin d'obtenir un fort contraste, les conditions expérimentales ont été regroupées en quatre catégories: calcul, phrase, damier et clic (i.e., motor response). On s'intéresse à la région motrice droite (162 voxels) dans laquelle il est attendu que les conditions: damier, phrase et calcul soient considérées comme non pertinentes. Le *modèle parcimonieux* détecte ces trois conditions comme non pertinentes avec des probabilités *a posteriori*  $p(w^m=0|y)$  proches de 1 et la condition clic comme pertinente avec une probabilité 1. De plus ce modèle estime des NRNs et NRNs normalisés supérieurs au modèle complet pour la condition motrice pertinente (clic). Comme alternative à (SPMs) utilisés dans l'inférence classique les cartes de probabilité *a posteriori* ( $PPM^\alpha$ ) expriment la probabilité  $p(a_j^m > \alpha|y)$  qu'un NRN  $a_j^m$  dépasse un certain seuil  $\alpha$ , sachant les données [10]. Le seuil  $\alpha$  définit le sens de l'activation et est fixé à l'abscisse du point d'intersection entre les gaussiennes active et inactive. On considère qu'en dessous de cette valeur le NRN n'est pas suffisamment élevé pour être considéré comme activé. Etant donnée que cette valeur est probablement estimée différemment entre les deux modèles, on a fixé  $\alpha=6.3$  (la moyenne des deux). Les histogrammes superposés des PPMs ( $PPM^\alpha$ ) supérieurs à  $\rho=0.75$  pour les deux modèles sont représentés dans la Fig.1-(a). Les Fig.1-(b,c) montrent (en rouge) les voxels  $j$  pour lesquels  $p(a_j^m > \alpha|y) > 0.95$  pour les deux modèles parcimonieux et complet (56 vs 76 voxels pour le modèle parcimonieux). Finalement, la Fig.1-(d) nous montre la différence entre les FRHs estimées et la forme canonique ce qui confirme l'importance de l'estimation conjointe de cette fonction.

Les résultats précédents valident l'intérêt et l'efficacité de notre approche par rapport à celle définie dans [3, 8]. En plus de la mise en place de l'approche variationnelle, les avantages par rapport à [7] sont l'estimation du paramètre  $\tau_2$  de la sigmoïde et la possibilité de détection de petits groupes de voxels de forte activation. Finalement, il est nécessaire de valider l'approche sur d'autres données réelles en plus d'une adaptation de l'analyse au niveau du groupe [11].

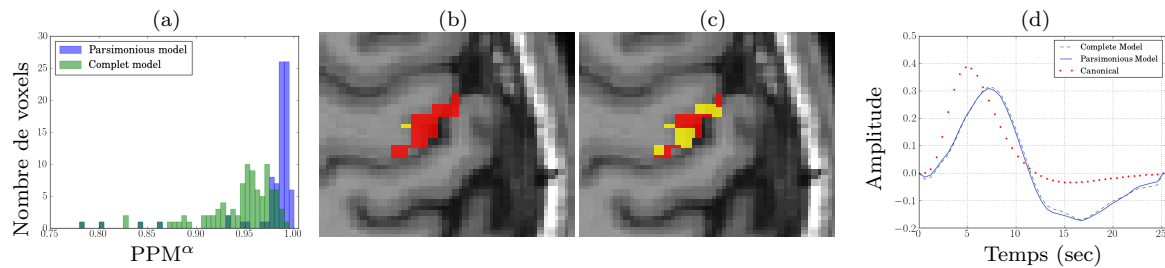


Figure 1: Histogrammes des  $PPM^\alpha$  de la condition pertinente (clic) pour la région motrice droite avec un seuil d’activation  $\alpha = 6.3$  (a) (la couleur foncée représente le chevauchement). Cartes des  $PPM^\alpha$  seuillées (1 en rouge et 0 en jaune) à un niveau de confiance de 95% pour le modèle parcimonieux (b) et complet (c). FRH canonique comparée à ceux estimées par le modèle *complet* et *parcimonieux* pour la région motrice droite (d).

## References

- [1] V.P. Oikonomou, K. Blekas, and L. Astrakas, “A Sparse and Spatially Constrained Generative Regression Model for fMRI Data Analysis,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 58–67, 2012.
- [2] H. Luo and S. Puthusserypady, “A Sparse Bayesian Method for Determination of Flexible Design Matrix for fMRI Data Analysis,” *IEEE Trans. on Circuits and systems-I: regular papers*, vol. 52, no. 12, pp. 2699–2706, 2005.
- [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, pp. 1059–1074, 2010.
- [4] S. Donnet, M. Lavielle, P. Ciuciu, and J.-B. Poline, “Selection of temporal models for event-related fMRI,” in *2th Proc. ISBI*, Arlington, VA, 2004, pp. 992–995.
- [5] R. B. O. Hara and M. J. Sillanpaa, “A review of Bayesian variable selection methods: What, how and which,” *Bayesian Analysis*, vol. 4, no. 1, pp. 85–118, 2009.
- [6] M. Smith and L. Fahrmeir, “Spatial Bayesian variable selection with application to functional MRI,” *J. Stat. Amer. Assoc.*, vol. 102, no. 478, pp. 417–431, 2007.
- [7] C. Bakhous, F. Forbes, T. Vincent, L. Chaari, M. Dojat, and P. Ciuciu, “Adaptive experimental condition selection in event-related fMRI,” in *8th Proc. ISBI*, Barcelona, Spain, 2012, pp. 1755–1758.
- [8] 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.*, 10.1109/TMI.2012.2225636, to appear.
- [9] P. Pinel, B. Thirion, S. Meriaux, A. Jobert, J. Serres, D. Le Bihan, J. B. Poline, and S. Dehaene, “Fast reproducible identification and large-scale databasing of individual functional cognitive networks,” *BMC Neurosci*, vol. 8, pp. 91, 2007.
- [10] K. J. Friston, W. Penny, C. Phillips, S. J. Kiebel, G. Hinton, and J. Ashburner, “Classical and Bayesian inference in neuroimaging,” *Neuroimage*, vol. 16, no. 2, pp. 465–483, Jun. 2002.
- [11] S. Badillo, T. Vincent, and P. Ciuciu, “Impact of the joint detection-estimation approach on random effects group studies in fMRI,” in *7th Proc. ISBI*, Chicago, IL, 2011, pp. 376–380.