Journal de la Société Française de Statistique
Vol. 151 No. 1 (2010)

A joint detection-estimation framework for analysing within-subject fMRI data

Titre: Un cadre de détection-estimation conjointe pour analyser les données individuelles d'IRMf

Philippe Ciuciu[1], Thomas Vincent[1], Laurent Risser[2] and Sophie Donnet[3]

Abstract: In this paper, we review classical and advanced methodologies for analysing within-subject functional Magnetic Resonance Imaging (fMRI) data. Such data are usually acquired during sensory or cognitive experiments that aims at stimulating the subject in the scanner and eliciting *evoked* brain activity. From such four-dimensional datasets (three in space, one in time), the goal is twofold: first, *detecting* brain regions involved in the sensory or cognitives processes that the experimental design manipulates; second, *estimating* the underlying activation dynamics. The first issue is usually addressed in the context of the General Linear Model (GLM), which a priori assumes a *functional* form for the impulse response of the hemodynamic filter. The second question aims at estimating this shape which makes sense in activating regions only. In the last five years, a novel Joint Detection-Estimation (JDE) framework addressing these two questions simultaneously has been proposed in [59, 60, 102]. We show to which extent this methodology outperforms the GLM approach in terms of statistical sensitivity and specificity, which additional questions it allows us to investigate theoretically and how it provides us with a well-adapted framework to treat spatially unsmoothed real fMRI data both in the 3D acquisition volume and on the cortical surface.

Résumé : Dans cet article, nous synthétisons la méthodologie usuelle et des variantes plus élaborées pour analyser des données individuelles d'Imagerie par Résonance Magnétique fonctionnelle (IRMf). De telles données sont acquises au cours d'expériences sensorielles ou cognitives dont le but est de stimuler le sujet dans le scanner afin d'évoquer une activité cérébrale typique. À partir de données quadri-dimensionnelles (trois dans l'espace et le temps), le but est double : d'abord, détecter les régions impliquées dans les processus sensoriels ou cognitifs que le protocole expérimental manipule ; ensuite, estimer la dynamique cérébrale sous-jacente. La première question est généralement abordée dans le contexte du Modèle Linéaire Général (MLG) qui suppose *a priori* connue la réponse impulsionnelle du filtre hémodynamique. La seconde question traite de l'estimation de la forme de cette réponse qui fait sens uniquement dans les régions activées. Durant ces cinq dernières années, un nouveau cadre de détection-estimation conjointe traitant ces deux questions simultanément a été proposé successivement dans [59, 60, 102]. Nous montrons ici jusqu'à quel point cette méthodologie étend les approches à base de MLG, les améliore en termes de sensibilité et spécificité statistique, quelles questions supplémentaires ce cadre permet d'investiguer théoriquement et comment il fournit un moyen aux utilisateurs de traiter de façon adaptée leur données non-lissées spatialement, aussi bien dans le répère 3D de l'acquisition que sur la surface corticale.

Keywords: fMRI, neuroimaging, nonparametric hemodynamics, Bayesian inference, MCMC, model selection, Markov random fields, partition function, Potts fields
Mots-clés : IRMf, neuro-imagerie, hémodynamique non-paramétrique, inférence bayésienne, MCMC, sélection de modèles, champs de Markov, fonction de partition, champs de Pots
AMS 2000 subject classifications: 62F15, 62H30, 62M10, 62M40, 62P10

[1] NeuroSpin, CEA Saclay, Bât 145 - Point Courrier 156, 91191 Gif-sur-Yvette cedex.
E-mail: philippe.ciuciu@cea.fr and and E-mail: thomas.vincent@cea.fr
[2] Institute for Mathematical Science, Imperial College, SW7, London UK.
E-mail: laurent.risser@gmail.com
[3] CEREMADE UMR CNRS 7534, Université de PARIS IX- DAUPHINE, Pl. du Maréchal DeLattre De Tassigny 75775 Paris Cedex 16
E-mail: donnet@ceremade.dauphine.fr

 ISSN: 2102-6238

1. Introduction

Since the first report of the Blood Oxygen Level-Dependent (BOLD) effect in human [70], fMRI has represented a powerful tool to non-invasively study the relation between cognitive tasks and the hemodynamic (BOLD) response reflecting evoked neuronal activity indirectly. To this end, several subjects (typically fifteen) are sampled from a population of interest and then scanned during an fMRI experiment consisting of the presentation of several stimulus types usually in a random order to avoid anticipation effects.

At the subject-level, fMRI data analysis essentially addresses two problems. The first one is about the *detection* of activated brain areas in response to given stimulus types or to behavioral tasks, while the second one concerns the *estimation* of the underlying temporal dynamics, usually referenced as the characterisation of the Hemodynamic Response Function (HRF). Detection has been extensively treated in the context of General Linear Models (GLM) [47, 106, 76]. In this framework, detection consists of a mass univariate inference of evoked brain activity at the voxel level. This univariate inference rests on the modelling of the temporal BOLD response i.e., on the definition of a *design matrix*, which usually relies on a spatially invariant temporal model of the BOLD signal throughout the brain. The precise localisation of evoked brain activity thus coincides with performing detection in every voxel of the brain.

The second question has been investigated in a parametric [29, 57, 18, 80, 56], semi-parametric [37, 40, 106, 107] or nonparametric [69, 41, 63, 14, 64, 65, 8] setting. The main differences between these approaches lie in the underlying assumptions regarding the evoked responses: parametric methods treat the HRF as a deterministic analytic function which depends on a small number of parameters. Semi-parametric approaches are more flexible since they make use of function basis, sometimes operate a selection [107] or introduce control time points in the HRF time course [37, 106]. Nonparametric approaches do not assume a functional form for the hemodynamic filter and implicitely rely on stick function basis to define it. Their nonparametric feature lies in their large number of parameters compared to alternative methods that may induce ill-posedness of the inverse problem aiming at identifying the filter, especially at the voxel level [14, 64, 65, 8]. Hence, to get a robust nonparametric HRF shape, the hemodynamic filter is usually transformed into a stochastic object, which enables the introduction of prior information through a probability distribution function (pdf).

Of course, detection and estimation are intrinsically connected to one another. On the one hand, the detection of brain activation requires the specification of a HRF shape throughout the brain. On the other hand, a robust and accurate estimate of the hemodynamic response is only available in regions eliciting signal fluctuations correlated with the paradigm because the SNR is too poor elsewhere. In the GLM and massively univariate context, the extra sum of squares or likelihood ratio principle has been applied to select the best HRF shape in terms of brain activation recovery. In particular, it has been shown that the *voxelwise* Finite Impulse Response (FIR) model is very attractive to capture HRF shape fluctuations such as differences in time-to-peak [48, 47], while enabling localisation of evoked activity. Nonetheless, the gain achieved in flexibility using FIR models has a direct cost in terms of statistical sensitivity, because the larger the number of HRF coefficients to be estimated the smaller the degree of freedoms in subsequent statistical tests. Hence, voxelwise FIR modelling remains efficient in *synchronous* experimental designs [14] but fails to provide a reliable answer to hemodynamics fluctuations

 ISSN: 2102-6238

in asynchronous designs in which the number of HRF coefficients becomes too large. The key point is therefore to set up a formulation in which *detection* and *estimation* enter naturally and simultaneously. This setting cannot be the classical hypothesis testing framework. Indeed, the sequential procedure which first consists in estimating the HRF on a given dataset and then building a specific GLM upon this estimate for detecting activations in the same dataset, entails statistical problems in terms of sensitivity and specificity: the control of the false positive rate actually becomes hazardous due to the use of an erroneous number of degrees of freedom. To overcome this problem, a Joint Detection-Estimation (JDE) formalism has been developed in the a Bayesian framework [59, 60, 86, 102]. where detection coincides with finding the voxels that elicit an evoked response while estimation makes reference to the inference on the underlying hemodynamics.

The rest of this paper is organized as follows. In Section 2, we review the recent literature that motivates the development of the JDE methodology. In short, we focus on the within- and between-subject variability sources in hemodynamics. Also, we summarize the salient features of the most advanced version of the JDE approach. In Section 3, we overview the observation model involved in the stationary parcelwise JDE formalism, explain how it extends any GLM-based approach and how it can deal with trial-varying BOLD response. In Section 4, we focus on the most relevant priors that give the best results in terms of HRF estimation accuracy and optimal sensitivity/specificity trade-off for detection purposes. In Section 5, our parcelwise Bayesian inference scheme is exposed with a particular attention to *unsupervised* estimation of Spatial Mixture Models (SMMs). Next, the extension to whole brain analysis is presented in Section 6. Section 7 is finally devoted to within-subject results on real fMRI data both in the three-dimensional acquisition volume and also on the cortical surface. In Section 8, special attention is then paid to model comparison and selection in the JDE context. Conclusions are drawn in Section 9.

2. State of the art

Intra-individual differences in the characteristics of the HRF have been exhibited between cortical areas in [1, 67, 45]. Although smaller than inter-individual fluctuations, this regional variability is large enough to be treated with care. To account for these spatial fluctuations at the voxel level one usually resorts to hemodynamic function basis. For instance, the canonical HRF can be supplemented with its first and second derivatives to model differences in time [47]. To make the basis spatially adaptive, more flexible *semi-parametric* approaches have been proposed later to capture these variations [37, 40, 107]. For instance, Woolrich et al. have proposed a half-cosine parameterisation in combination with the selection of the best basis set [106, 107], while other have resorted to FIR models [1].

Due to the low Signal-to-Noise Ratio (SNR) in fMRI time series, prior information has to be introduced in order to get robust HRF estimates, whatever the model in use. Then, most of these contributions take place in the Bayesian setting and constrain the HRF. In a semi-parametric framework, the HRF time course is typically decomposed into different periods (initial dip, attack, rise, decay, fall, ...), each of them being described by specific parameters over which prior boundary

[1] this number corresponds to the number of HRF time points to be estimated in the FIR model case.

 ISSN: 2102-6238

constraints are involved. In the *non-parametric* approaches or FIR models that have emerged in the fMRI literature as a powerful tool to infer on the HRF shape [69, 41, 63, 14, 65], prior information is usually embedded in a Gaussian process prior that induces temporal smoothness constraints on the HRF shape.

On the other hand, several works segregating neuro- and hemodynamics events from fMRI time series have been proposed. Most contributions rely on a stochastic modelling of hemo- and neurodynamics in a state space formulation: the simplest approach called bilinear Dynamic Systems (BDS) [75, 61] embodies a nonparametric HRF modelling as in the JDE framework but proceeds voxelwise. The most sophisticated formalism is known as Dynamic Causal Modeling (DCM) and enables the study of experimentally induced changes in functional integration among a small number of brain regions [30, 97, 96, 95, 78]. DCM is a general Bayesian framework for inferring hidden neuronal states from measurements of brain activity (fMRI, MEG/EEG, ...) [31, 55, 62]: every DCM instance rests on biophysically plausible and physiologically interpretable models of neuronal network dynamics that can predict distributed responses to experimental stimuli. Hence, DCM could be regarded as a generalisation of our JDE framework, since the HRF convolution kernels are replaced by bio-physically informed differential equations mediating the hemodynamic convolution [6, 28, 7, 84, 93, 94]. Because of the strong complexity behind most of existing physiological models, several works have considered cruder but numerically simpler models based on less informed evolution schemes (e.g. ARX, ARMAX models) [83, 2, 53]. Since DCM is able to infer effective connectivity only between an a priori graph of regions but not over the whole brain, this constitutes a strong limitation of DCM that prevents us from using it to study hemodynamics fluctuations at a larger scale. However, the principle of modelling neuronal and hemodynamic convolutions under biologically motivated constraints using Bayesian model inversion are exactly the same than those behind the JDE framewok: DCM attempts to optimise both the neuronal and hemodynamic convolution kernels simultaneously, but resorts to Variational Bayes (VB) approximations to enable the computation of posterior estimators in a reasonable amount of time.

The literature on Bayesian fMRI methods offers several approaches to adequately choose spatial priors for detection purpose that make the overall fitting procedure multivariate. To enhance the estimation of regression coefficients in the GLM context, global stationary and more recently local non-stationary spatial regularisation models have been introduced [106, 76, 25, 46]. These Bayesian spatial priors encode similarity between neighboring voxels and estimate one or several degrees of smoothness of the parameter images throughout the brain. In this regard, they do not require data to be smoothed prior to entering a statistical model. However, since these models take place in the family of edge-preserving regularization models [34, 35, 9, 13], posterior inference [2] is not numerically tractable at low computational cost without considering a mean-field approximation [26] and the VB formalism [4]. This kind of approximation reduces numerical complexity by assuming conditional posterior independence, which means independence across voxels in the fMRI context. It also provides closed form updating rules for iterative estimation of parameters of interest and hyper-parameters provided that conjugate priors have been retained. Nonetheless, the literature in spatial regularisation offers alternative choices that make the definition of nonstationary spatial models and their global optimization [5]

[2] *e.g.,* the computation of the maximum a posteriori or the posterior mean estimates do not admit a closed form expression.

 ISSN: 2102-6238

or simulation feasible [52, 49, 43, 91, 23, 105, 3], even if VB approximation can also be derived [104]. These models enter in the class of Spatial Mixture Models (SMM) in which the spatial structure is embedded on *hidden* allocation variables that specify the voxel states in the fMRI context (activating, deactivating or non-activating) through a *discrete* Markov Random Field (MRF) [98, 81, 91, 73, 105, 74] or a Conditional Random Field (CRF)[82]. In this regard, the detection problem becomes a segmentation issue that can be addressed using a region-based method [36, 52] instead of an edge-preserving restoration one [34, 35, 9, 13].

As introduced in [50, 19, 22, 101, 81] and further developed in [91, 105, 73, 25, 77], prior mixture models define an appropriate way to perform the classification or the segmentation of statistical parametric maps into activating, non-activating or deactivating brain regions. The pioneering contribution related to mixture modelling in neuroimaging (PET) [50] has represented the intensity distributions of Statistical Parametric Maps (SPMs) using a complex spatio-temporal MRF. Yet, the use of mixture modelling in a joint detection-estimation problem introduces specific concerns in comparison to the usual "hypothesis testing framework". Indeed, our data is *not* the voxelwise z-statistics but rather the raw fMRI time courses, which are required for the estimation step.

In [59, 60], we have investigated the choice of the best independent [3] mixture model (IMM) that serves as prior distribution on the evoked response magnitude, called the Neural Response Level (NRL) hereafter. More precisely, it has been shown that the use of inhomogenous gamma-Gaussian IMM permits to better disentangle activating from non-activating voxels at the expense of higher computational cost. Nonetheless, in case of very few activations, the IMM model overestimates the false positive rate (specificity). For this reason, homogeneous *supervised* SMM (SSMM) has been considered as a powerful alternative in [102]. It has been shown that SMM outperforms its IMM counterpart in terms of false positive control and activation cluster recovery. More recently, unsupervised SMM (USMM) has been developed to make spatial regularisation fully automatic [86, 102] at the whole brain scale.

As in the GLM or DCM frameworks, the JDE formalism enables the analysis of parcelwise hemodynamics systems as well as the study of event-related fMRI data involving multiple experimental conditions. The trial by trial variability that is usually accounted for either by adding a specific regressor in the design matrix to parametrically modulate the effect size in time or by modelling successive stimulus trials as separate regressors for modelling the so-called *repetition suppression* effect in cognitive neuroscience is more efficienty handled in the JDE approach. It actually relies on a sparser representation [15]. Also, while deactivations or negative BOLD responses [89] are trivial to model within the standard framework by simply considering a negative NRLs, their spatial coherence is not garanteed. The JDE framework enables the recovery of deactivation clusters using a *three-class* Potts field, which is a generalization of the Ising model of interacting spins on a lattice. Although developed originally in statistical mechanics, the associated energy functions describing local interactions provide a nice statistical model for spatial correlations in images. We will use a Potts field, where the different states of neighboring voxels can be of three different sorts or "colors"; namely activated, not activated or deactivated. Furthermore, akin to [98, 91], the JDE methodology allows us to jointly estimate HRFs and perform activation detection at the parcel-level. Hence, it generalizes recent contributions [98, 91]

[3] not spatially correlated.

 ISSN: 2102-6238

following directions:

- Following [49, 105] but in contrast to [98, 91], spatial regularisation in the JDE formalism is *unsupervised* and makes the segmentation of brain activations fully automatic. Our unsupervised regularisation is also spatially adaptive meaning that the amount of regularisation varies across parcels. This requires a precise estimation of the *partition function* of the MRF defined over each parcel; see Section 5 for details.
- Spatial regularisation also varies across *experimental conditions* in every parcel since both the Signal-to-Noise Ratio (SNR) and the activation pattern may fluctuate from one condition to another according to the involvement of the given parcel in the experimental paradigm.

Hence, the proposed methodology introduces spatially adaptive levels of regularisation at a reasonable computational cost in adopting a fully *exact* Bayesian inference of brain activity. Parcel-based parameters of interest (HRF shape, NRLs) as well as hyper-parameters are estimated in the Posterior Mean (PM) sense from unsmoothed fMRI time series after convergence of a hybrid Metropolis within Gibbs sampling procedure [58].

3. The spatially regularised JDE approach

3.1. A parcelwise procedure

We denote vectors and matrices with bold lower and upper case letters, respectively (*e.g.,* $\boldsymbol{y}$ and $\boldsymbol{P}$). A vector is by convention a column vector. Scalars are denoted with non-bold lower case letters. The transpose is denoted by ^{t}. Unless stated otherwise, subscripts i, j, m and n are respectively indexes over mixture components, voxels, stimulus types and time points. The probability distribution functions (pdf) are denoted using calligraphic letters (eg, $\mathscr{N}$ and $\mathscr{G}$ for the Gaussian and gamma distributions).

The JDE framework proposed in [60] relies on a prior parcellation of the brain into $\mathscr{P} = (\mathscr{P}_\gamma)_{\gamma=1:\Gamma}$ functionally homogeneous and connected parcels [99], where typically $\Gamma \approx 500$ to cover the whole brain (see Subsection 6.1 and [100] for their computation and the assessment of Γ[4]); see also Figs. 4-5 for illustrations. Parcels are used to induce conditional dependencies among various parameters of our hemodynamic model. The most important dependency is that, within any parcel $\mathscr{P}_\gamma$ comprising voxels $(V_j)_{j=1:J_\gamma}$, the form of the HRF $\boldsymbol{h}_\gamma$ is the same. This does not mean that we treat each parcel as a single observation; we still optimise voxel-specific parameters relating to how neuronal activity excites a hemodynamic response. In this section, we describe this parcel-based model and the implicit assumptions that it entails. Then, we specify the priors involved over each parcel. In other words, our Bayesian inference of brain activity is independently repeated over the Γ different parcels. For notational simplicity, in what follows we drop the γ index that makes reference to the parcel except for $\boldsymbol{h}_\gamma$ and J_γ.

3.2. Forward parcel-based model of the BOLD signal

The forward bilinear model of the BOLD signal introduced in [59] and extended in [60] to account for serial correlation of fMRI time series is a time-invariant model that characterises

[4] The algorithm is available to the community in the fMRI Toolbox of the Brainvisa software at `http://brainvisa.info`.

 ISSN: 2102-6238

each and every parcel $\mathscr{P}_\gamma$ by a single neurovascular impulse response $\boldsymbol{h}_\gamma$ and a NRL for each voxel and stimulus type. As illustrated in Fig. 1, this means that the HRF shape $\boldsymbol{h}_\gamma$ is assumed constant within $\mathscr{P}_\gamma$, while its magnitude a_j^m can vary in space ($j = 1 : J_\gamma$) and across experimental conditions ($m = 1 : M$), where M is the total number of stimulus types. Then, the generative BOLD model reads:

$$\forall j = 1 : J_\gamma, \quad \boldsymbol{y}_j = \sum_{m=1}^{M} \overbrace{a_j^m \boldsymbol{X}^m \boldsymbol{h}_\gamma}^{=\boldsymbol{s}_j^m} + \boldsymbol{P}\ell_j + \boldsymbol{b}_j, \tag{1}$$

where:

- $\boldsymbol{y}_j = (y_{j,n})_{n=1:N}$ denotes the fMRI signal measured in voxel V_j at times $n = 1{:}N$ (N is the number of scans).
- $\boldsymbol{X}^m = (x^m_{n-d\Delta t})_{n=1:N,d=0:D}$ is a $N \times (D+1)$ binary matrix coding for the occurrences of the m stimulus type, with Δt is the sampling period of the unknown HRF $\boldsymbol{h}_\gamma = (h_{d\Delta t,\gamma})_{d=0:D}$ in $\mathscr{P}_\gamma$.
- a_j^m stands for the NRL in voxel V_j for condition m. Hence, the activation time course associated to the mth stimulus type in voxel V_j is given by $\boldsymbol{h}_\gamma \times a_j^m$. Let also $\boldsymbol{A} = \left[\boldsymbol{a}^1 \,|\, \ldots \,|\, \boldsymbol{a}^M\right]$ be the whole NRL matrix in $\mathscr{P}_\gamma$ where $\boldsymbol{a}^m = (a_j^m)_{j=1:J_\gamma}$.
- $\boldsymbol{P}$ is a low frequency orthogonal matrix of size $N \times Q$. To each voxel is attached an unknown weighting vector ℓ_j to estimate the trend in V_j. We denote $\boldsymbol{L} = \left[\ell_1 \,|\, \ldots \,|\, \ell_{J_\gamma}\right]$ the set of low frequency drifts involved in $\mathscr{P}_\gamma$.
- $\boldsymbol{b}_j \in \mathbb{R}^N$ is the noise in V_j and follows a first-order autoregressive process: $\boldsymbol{b}_j \sim \mathscr{N}(\boldsymbol{0}, \sigma_j^2 \Lambda_j^{-1})$ where Λ_j is tridiagonal and depends on the AR parameter ρ_j [60].

Model (1) separates the dependency of the response on the hemodynamic convolution kernel $\boldsymbol{h}$ and the neuronal inputs $\boldsymbol{A}$, which kernel here is assumed to be a stick function. Conventional formulations conflate these two by setting $\boldsymbol{h}_\gamma$ to the canonical HRF shape $\boldsymbol{h}_{\mathrm{c}}$ [39] and try to estimate a single lumped parameter set (*e.g.*, $\boldsymbol{A}$) that reflects both neuronal and hemodynamic contributions. As illustrated in Fig. 1, our bilinear form enables us to separately model the neuronal hemodynamic contributions to the signal, while placing spatial constraints on the hemodynamic component through our parcellation scheme.

Although the noise structure is correlated in space [108, 107, 82], we neglect such spatial dependency and consider the fMRI time series $\boldsymbol{Y} = \left[\boldsymbol{y}_1 \,|\, \ldots \,|\, \boldsymbol{y}_{J_\gamma}\right]$ independent in space but *not* identically distributed. The reason is twofold: first, neglecting the spatial dependencies of noise is tenable when the BOLD signal model itself is flexible enough to account for HRF shape fluctuations. Indeed, part of the usually observed spatial correlation of the noise is due to a misspecification of the BOLD signal model. Second, the noise correlation is much lower than that of the evoked BOLD response. While its modelling introduces additional computational complexity, there is no evidence in the literature that ignoring this correlation induces strong bias on the sought parameters (see [82] for such comparison). Hence, the likelihood reads:

$$p(\boldsymbol{Y} \,|\, \boldsymbol{h}_\gamma, \boldsymbol{A}, \boldsymbol{L}, \theta_0) \propto \prod_{j=1}^{J_\gamma} \left|\Lambda_j\right|^{1/2} \sigma_j^{-N} \exp\left(-\frac{\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \Lambda_j \widetilde{\boldsymbol{y}}_j}{2\sigma_j^2}\right) \tag{2}$$

where $\theta_{0,j} = (\rho_j, \sigma_j^2)$, $\theta_0 = (\theta_{0,j})_{j=1:J_\gamma}$ and $\widetilde{\boldsymbol{y}}_j = \boldsymbol{y}_j - \sum_m \boldsymbol{s}_j^m - \boldsymbol{P}\ell_j$

 ISSN: 2102-6238

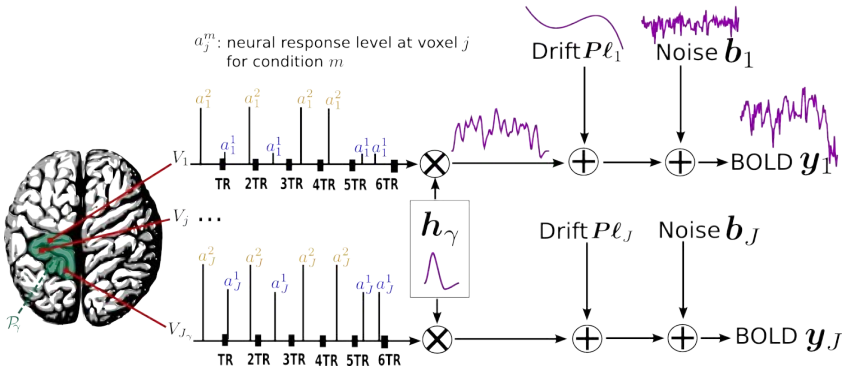

FIGURE 1. ***Parcel-based BOLD model.*** *The size of each parcel $\mathscr{P}_\gamma$ varies typically between a few tens and a few hundreds of voxels:* $80 \leqslant J_\gamma \leqslant 350$. *The number M of experimental conditions involved in the model usually varies from 1 to 5. Here,* $M = 2$ *and the NRLs* $(\boldsymbol{a}_j^1, \boldsymbol{a}_j^2)$ *corresponding to the first and the second conditions are surrounded by circles and squares, respectively. As illustrated, the NRLs take non-zero values (asynchronous paradigms) on time points that do not necessarily match the acquisition ones and may vary from one voxel to another. The HRF* $\boldsymbol{h}_\gamma$ *can be sampled at a period of 1s and estimated on a range of 20 to 25s (e.g.,* $D = 25$*). Most often, the LFD coefficients* ℓ_j *are estimated on a few components (*$Q = 4$*).*

Note that model (1) is bilinear in the sense that Eq. (1) linearly depends on $\boldsymbol{h}_\gamma$ when $\boldsymbol{A}$ is fixed and *vice-versa*. This means that the ML solution $(\boldsymbol{h}_\gamma^*, \boldsymbol{A}^*)$ cannot be distinguished from any other pair $(\boldsymbol{h}_\gamma^*/s, \boldsymbol{A}^* \times s)$ whatever the scale parameter $s > 0$. The Bayesian formalism is helpful to get rid of such identifiability problems and define a reference scale. In Section 4, we will introduce priors on $\boldsymbol{h}_\gamma$ and $\boldsymbol{A}$, which are helpful to fix this scale to an arbitrary value c. However, this value is not necessarily optimal for exploring the posterior distribution. Hence, instead of normalizing $\boldsymbol{h}_\gamma$ deterministically ($|\boldsymbol{h}_\gamma| = c$), it has been shown in [102, Appendix C] how this reference scale s can be selected to speed up the exploration of the posterior distribution. Note that for the priors involved in our inference scheme (see Section 4), this scalar parameter is quite easy to sample from since it follows a *Generalized Inverse Gaussian* density [92].

3.3. A parametric approach to habituation modelling

The stationary model (1) assumes that each trial k of a given experimental condition m evokes a BOLD response constant in shape and in magnitude. Recently, it was suggested that this might not always be the case [21]. In the GLM framework, the trial by trial variability that exhibits some repetition suppression effect of the BOLD response is usually captured by simply modelling the first and subsequent trials as separate experimental conditions. Here, we propose an extension of Eq. (1) that is able to account for this variability source in a sparser manner while assuming a constant HRF shape. In contrast to [20], our modelling of this variability source relies on a single normalised mean habituation parameter $r_{jm} \in [0, 1]$ for voxel V_j and condition m that introduces a *parametric* relationship between the trial-dependent NRLs a_{jk}^m. Our sparse model actually mimics the repetition suppression effectin the sense that decreasing NRLs over trials can be obtained in case of short ISIs. However, more flexible patterns can be observed in other paradigms (see Fig. 2 for details). A key feature is to progressively forget the past events. Hence, our parametric

 ISSN: 2102-6238

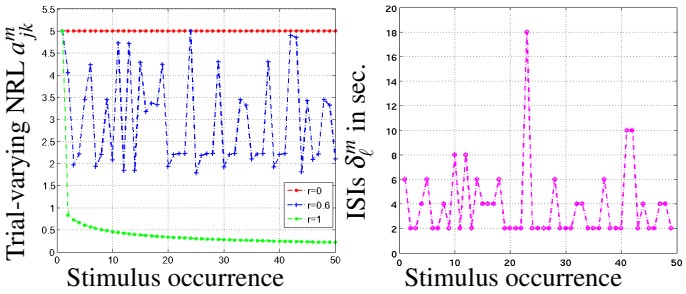

FIGURE 2. *(g) Illustration of the habituation effect on the trial-specific BOLD magnitudes a_{jk}^m in three different voxels corresponding to a zero (red curve), a medium (blue curve) and a large (green curve) normalised habituation speed r_j^m.*

habituation model depends on the paradigm as follows:

$$\forall k \geq 2,\ a_{jk}^m = \eta_{jk}^m a_{j1}^m \quad \text{with} \quad \eta_{jk}^m = \left(1 + \sum_{l=1}^{k-1} a_{j\ell}^m \boldsymbol{r}_{jm}^{\delta_\ell^m}\right)^{-1} \tag{3}$$

where $\delta_\ell^m = \tau_{\ell+1}^m - \tau_\ell^m$ and $(\tau_\ell^m)_\ell$ define the successive ISIs and onsets of the mth stimulus. In this way, we introduce a nonlinear dependence between the successive NRLs that satisfy a closed form updating rule across trials:

$$\forall k \geqslant 2, \quad \eta_{jk}^m = \left[(\eta_{j,k-1}^m)^{-1} + \eta_{j,k-1}^m a_{j1}^m \boldsymbol{r}_{jm}^{\delta_{k-1}^m}\right]^{-1}. \tag{4}$$

As shown in Eq. (3), there are particular constraints on the difference between the first and subsequent occurrences of particular trials. This allows one to parameterise the explicit dependency of repetition suppression on ISIs while still maintaining a large number of degrees of freedom in comparison to the standard GLM framework. The activation signal $\boldsymbol{s}_j^m$ in Eq. (1) thus becomes:

$$\boldsymbol{s}_j^m = \sum_{k=1}^{K_m} a_{jk}^m \boldsymbol{X}_k^m \boldsymbol{h}_\gamma = a_{j1}^m \widetilde{\boldsymbol{X}}_j^m \boldsymbol{h}_\gamma \quad \text{with} \quad \widetilde{\boldsymbol{X}}_j^m = \sum_{k=1}^{K_m} \eta_{jk}^m \boldsymbol{X}_k^m, \tag{5}$$

where $\eta_{j1}^m = 1$ and $\boldsymbol{X}_k^m$ is the kth trial-specific submatrix of $\boldsymbol{X}^m$. The ensuing region-based model of fMRI time series is depicted in Fig. 3. It clearly indicates that the habituation speed may vary in space and across stimulus types. As expected, when $r_{jm} \to 0$, the proposed extension becomes stationary since $a_{jk}^m \to a_{j1}^m$, $\forall k \geqslant 2$ (cf red line in Fig. 2(a)). When $r_{jm} \in (0,1)$, the sequence $(a_{jk}^m)_k$ is non-monotonous and the between-trial variability is strongly influenced by the ISI values (see blue line in Fig. 2(a) and Fig. 2(b)) . Finally, when $r_{jm} \to 1$, the sequence $(a_{jk}^m)_k \to 0$ whatever the ISIs, as shown by the green curve in Fig. 2.

In Fig. 3, our parcel-based model of the BOLD signal accounting for such habituation effect is illustrated. In comparison to Fig. 1, it clearly appears that the BOLD signal may decrease over successive trials. Also, the habituation effect may fluctuate over conditions and voxels meaning that our extension defined by Eqs.1 5 is able to simultaneously account for stationary and non-stationary BOLD responses.

 ISSN: 2102-6238

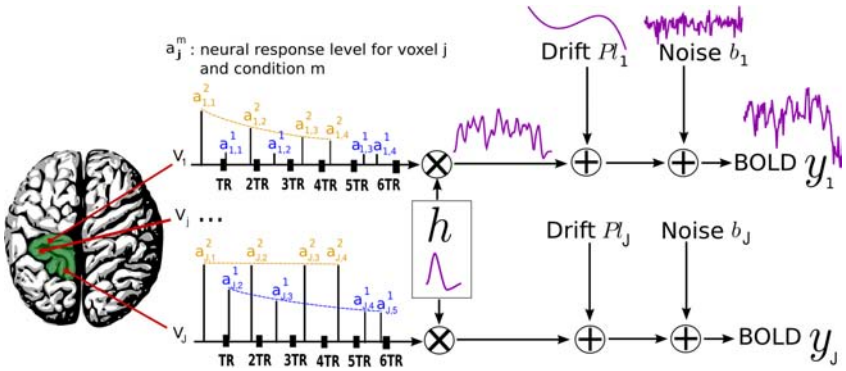

FIGURE 3. *Non-stationary ROI-based model accounting for voxel-based and stimulus-specific habituation effects.*

For the habituation model proposed in Eqs. 3- 5 and the same assumptions on the data $\boldsymbol{Y}$ in parcel $\mathscr{P}_\gamma$, the likelihood function $p(\boldsymbol{Y}\,|\,\boldsymbol{h}_\gamma, \boldsymbol{A}_1, \boldsymbol{R}, \boldsymbol{L}, \theta_0)$ reads as in Eq. 2, but $\widetilde{\boldsymbol{y}}_j$ depends on $\boldsymbol{s}_j^m$ which takes the form (5). Here, $\boldsymbol{A}_1$ and $\boldsymbol{R}$ make reference to the NRLs associated to the first trial and to the normalised habituation speed, respectively.

4. Bayesian priors

4.1. Hemodynamic filter

The Bayesian approach developed in [60] introduces proper priors on the unknown parameters $(\boldsymbol{h}_\gamma, \boldsymbol{A})$ or $(\boldsymbol{h}_\gamma, \boldsymbol{A}_1)$ in order to recover a robust estimate of brain activity (localisation and activation profile). Akin to [60, 14], we rest on a Gaussian prior process for the HRF i.e., $\boldsymbol{h}_\gamma \sim \mathcal{N}(\boldsymbol{0}, v_{\boldsymbol{h}}\boldsymbol{R})$ with $\boldsymbol{R} = (\boldsymbol{D}_2^{\mathrm{t}}\boldsymbol{D}_2)^{-1}$, which allows us to estimate a smooth HRF time course since $\boldsymbol{D}_2$ is the second-order finite difference matrix penalising therefore abrupt changes. Moreover, the extreme time points of the HRF can be constrained to zero if necessary [14].

4.2. Neural Response Levels

Regarding NRLs, we first deal with the stationary model. According to the maximum entropy principle we assume that different types of stimulus induce statistically independent NRLs *i.e.,* $p(\boldsymbol{A}\,|\,\theta_{\boldsymbol{a}}) = \prod_m p(\boldsymbol{a}^m\,|\,\theta^m)$ with $\theta_{\boldsymbol{a}} = (\theta^m)_{m=1:M}$. Vector θ^m denotes the set of unknown hyperparameters related to the mth stimulus type.

Mixture models are introduced to segregate activating voxels from non-activating and de-activating ones. To this end, let q_j^m be the allocation variable that states whether voxel V_j is activating ($q_j^m = 1$), de-activating ($q_j^m = -1$) or non-activating ($q_j^m = 0$) in response to stimulus m. The NRLs still remain independent conditionally upon $\boldsymbol{q}^m$. This means that $p(\boldsymbol{a}^m\,|\,\boldsymbol{q}^m, \theta^m) = \prod_j p(a_j^m\,|\,q_j^m, \theta^m)$ for every condition m. In the case of an IMM [59, 60], the marginal density of the NRLs factorizes over voxels and reads:

$$p(\boldsymbol{a}^m\,|\,\theta^m) = \prod_{j=1}^{J_\gamma} \sum_{i=-1}^{1} p(a_j^m\,|\,q_j^m, \theta^m)\,\Pr(q_j^m = i\,|\,\theta^m). \tag{6}$$

 ISSN: 2102-6238

An important feature of IMM lies in the definition of the mixing probability (or weight) $\Pr(q_j^m = i)$ which is constant over voxels and so independent of j. This means that IMM explicitely estimates the proportion of voxels in the activating and non-activating classes.

Instead, *spatial* mixture models are introduced here to favor clustered activations and deactivations and the mixing probabilities become space-varying. Hereafter, the marginal density no longer factorizes over voxels and reads:

$$p(\boldsymbol{a}^m \,|\, \theta^m) = \sum_{\boldsymbol{q}^m \in \{0,\pm 1\}^{J_\gamma}} \Big[\prod_{j=1}^{J_\gamma} p(a_j^m \,|\, q_j^m, \theta^m)\Big] \Pr(\boldsymbol{q}^m \,|\, \theta^m). \tag{7}$$

Spatial correlation is directly incorporated in the probabilities of activation through a *hidden* Potts field on the allocation variables $\boldsymbol{q}^m$, as already done in image analysis [49, 43] or in neuroimaging [91, 90]. Here, as opposed to IMM, the proportions of voxels for the different classes is not explicit. The prior density on the allocation variables reads:

$$\begin{aligned} \Pr(\boldsymbol{q}^m \,|\, \beta_m) &= Z(\beta_m)^{-1} \exp\big(\beta_m U(\boldsymbol{q}^m)\big) \\ \text{with} \qquad U(\boldsymbol{q}^m) &= \sum_{j \sim k} I(q_j^m = q_k^m) \end{aligned} \tag{8}$$

and $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. The notation $j \sim k$ means that the sum extends over all pairs (j,k) of neighbouring sites. The neighbouring system can be defined either in 3D in the brain volume intersecting region $\mathscr{P}_\gamma$ or in 2D along the cortical surface. In this paper, we only consider the 3D case using 6-connexity. Extensions to 18 or 26 neighborhood system are straightforward. The Potts field in (8) has no *external field*. Previous works have shown that anatomical prior information can be embedded in an external field to increase activation probability in the grey matter[91]. The parameter $\beta_m > 0$ in (8) controls the amount of spatial regularisation: large values of β_m associate higher probabilities to configurations containing clusters of like-valued neighboring binary variables. Since activation patterns within parcel $\mathscr{P}_\gamma$ should be different from one stimulus type to another, different parameters β_m are considered across m stimulus types. The normalization constant of the MRF, also called the *partition function* $Z(\cdot)$ reads:

$$Z(\beta_m) = \sum_{\boldsymbol{q}^m \in \{0,\pm 1\}^{J_\gamma}} \exp\big(\beta_m U(\boldsymbol{q}^m)\big) \tag{9}$$

and guarantees that the MRF defines a pdf.

In what follows, we assume that $(a_j^m \,|\, q_j^m = i) \sim \mathscr{N}(\mu_{i,m}, v_{i,m})$, with $i = 0,1$. We impose $\mu_{0,m} = 0$ for the mean of the NRLs in non-activating voxels, leading to $\theta^m = [v_{0,m}, \mu_{1,m}, v_{1,m}, \beta_m]$. Note that a Bernoulli-Gaussian formulation has also been tested in fMRI in [91]. This modelling corresponds to a degenerated mixture ($v_{0,m} = 0$).

4.3. Normalised Habituation Speed (NHS)

If we consider the forward model (1) based on Eqs. (3)-(5), the prior on NRLs involves $\boldsymbol{A}_1$ and the habituation parameters $\boldsymbol{R}$. In that case, we still assume that different stimulus type

 ISSN: 2102-6238

induce statistically independent NRLs and NHS: $p(\boldsymbol{A}_1, \boldsymbol{R} \,|\, \theta_{\boldsymbol{a}}) = \prod_m p(\boldsymbol{a}_1^m, \boldsymbol{r}^m \,|\, \theta^m)$. In the SMM context, this joint prior density admits the following expression:

$$p(\boldsymbol{a}_1^m, \boldsymbol{r}^m \,|\, \theta^m) = \sum_{\boldsymbol{q}^m \in \{0,\pm 1\}^{J_\gamma}} \left[\prod_{j=1}^{J_\gamma} p(\boldsymbol{a}_1^m, \boldsymbol{r}^m \,|\, q_j^m, \theta^m) \right] \Pr(\boldsymbol{q}^m \,|\, \theta^m). \tag{10}$$

In addition, the NRLs and NHSs still remain independent in space as well as independent of each other conditionally upon $\boldsymbol{q}^m$:

$$\forall m, \quad p(\boldsymbol{a}_1^m, \boldsymbol{r}^m \,|\, \boldsymbol{q}^m, \theta^m) = \prod_{j=1}^{J_\gamma} p(a_{j1}^m, r_j^m \,|\, q_j^m, \theta^m)$$

$$\text{with} \quad p(a_{j1}^m, r_j^m \,|\, q_j^m, \theta^m) = p(a_{j1}^m \,|\, q_j^m, \theta^m)\, p(r_j^m \,|\, q_j^m)$$

Since $p(a_{j1}^m \,|\, q_j^m, \theta^m)$ have been defined in Subsection 4.2, there only remains to precise the distribution $p(r_j^m \,|\, q_j^m)$. First, to get rid of identifiability problems, the NHSs $\boldsymbol{r}^m$ are constrained to zero in non-activating voxels: $\forall m, r_j^m = 0 \,|\, q_j^m = 0$. It seems a priori meaningless to fit habituation effect on noise-only time series. In contrast, for activating and deactivating voxels, the MHSs $\boldsymbol{r}^m$ are assumed identically and uniformly distributed: $(r_j^m \,|\, q_j^m = \pm 1) \sim \mathscr{U}([0,1])$. Note that this prior has not been tested yet in the deactivation context involving three-class SMMs. However, this could be easily handled considering that the BOLD signal reduction becomes less important under repeated stimulations. Our compound prior mixture therefore reads:

$$p(\boldsymbol{a}_1^m, \boldsymbol{r}^m \,|\, \theta^m) = \sum_{\boldsymbol{q}^m \in \{0,\pm 1\}^{J_\gamma}} \left[\prod_{j=1}^{J_\gamma} f_{i,m}(a_{j1}^m)\, p_i(r_j^m) \right] \Pr(\boldsymbol{q}^m \,|\, \beta^m). \tag{11}$$

4.4. *Noise and drift parameters*

To complete the Bayesian model, priors are required for all the remaining parameters. The noise and drift parameters, θ_0 and $\boldsymbol{L}$ respectively, are assumed independent in space: $p(\theta_0, \boldsymbol{L} \,|\, v_\ell) = \prod_j p(\theta_{0,j}) p(\ell_j \,|\, v_\ell)$ and without informative prior knowledge, the following priors are chosen: $\ell_j \sim \mathscr{N}(\boldsymbol{0}, v_\ell \boldsymbol{I}_Q)$ and $p(\rho_j, \sigma_j^2) = \sigma_j^{-1} I(|\rho_j| < 1)$ to ensure stability of the AR(1) noise process. Non-informative Jeffrey priors are retained for hyper-parameters such as the drift and HRF shape variances: $p(v_{\boldsymbol{h}}, v_\ell) = (v_{\boldsymbol{h}} v_\ell)^{-1/2}$.

4.5. *Mixture parameters*

Similarly, the prior considered for $v_{0,m}$ is $p(v_{0,m}) = v_{0,m}^{-1/2}$ because we *do* expect non-activating voxels in any parcel. Hence, class 0 should never be empty a priori. If this assumption is not tenable, we could introduce a conjugate prior (an inverse Gamma law) denoted as $\mathscr{IG}(a_{v_0}, b_{v_0})$, as already done for the variance parameters $(v_{\pm 1,m})$ of activating and deactivating voxels. We would then avoid degeneracy problem that could prevent its sampling. In the same way, a proper prior $\mathscr{N}(a_{\mu_{\pm 1}}, b_{\mu_{\pm 1}})$ is chosen for $\mu_{\pm 1,m}$ ($a_{\mu_{\pm 1}} = x$ and $b_{\mu_{\pm 1}} = y$). Finally, the prior on $\beta = (\beta_m)_{m=1:M}$ is independent and identically distributed (iid) across conditions and follows a uniform pdf over fixed range: $p(\beta_m) = \mathscr{U}([0, \beta_{\max}])$ with $\beta_{\max} = 1.6$. Interestingly, such proper prior defined over $[0, \beta_{\max}]$ allows us to easily compute Potts field partition functions on this discrete β-grid; see Section 6.

 ISSN: 2102-6238

5. Within-parcel inference of unsupervised SMMs

5.1. Role of the parcellation scheme

The parcellation scheme does not depend upon the posterior densities of the within-parcel optimisation: it is assumed known a priori and derived from a specific procedure as explained in Section 6. To simplify posterior inference, all parcels are treated as independent *latent* variables and the fMRI data are also independent conditionally on these parcel variables.

5.2. Posterior distribution under stationary BOLD model

Considering the constructed model and assuming no further prior dependence between parameters, Bayes' rule gives us for the stationary BOLD model (1):

$$\begin{aligned} p(\boldsymbol{h}_\gamma, \boldsymbol{A}, \boldsymbol{L}, \Theta \,|\, \boldsymbol{Y}) &\propto p(\boldsymbol{Y} \,|\, \boldsymbol{h}_\gamma, \boldsymbol{A}, \boldsymbol{L}, \theta_0)\, p(\boldsymbol{A} \,|\, \theta_{\boldsymbol{A}})\, p(\boldsymbol{h}_\gamma \,|\, v_{\boldsymbol{h}})\, p(\boldsymbol{L} \,|\, v_\ell)\, p(\theta_0) \\ &\quad p(\theta_{\boldsymbol{A}})\, p(v_{\boldsymbol{h}}, v_\ell) \\ &\propto v_{\boldsymbol{h}}^{-\frac{D}{2}} v_\ell^{-\frac{J_\gamma Q}{2}} \prod_{j=1}^{J_\gamma} \frac{(1-\rho_j^2)^{1/2}}{\sigma_j^{N+1}} I(|\rho_j| < 1) \\ &\quad \exp\Big(-\sum_{j=1}^{J_\gamma} \big[\frac{1}{2\sigma_j^2} \widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \Lambda_j \widetilde{\boldsymbol{y}}_j + \frac{1}{2v_\ell} \|\ell_j\|^2 \big] \Big) \\ &\quad \exp\Big(-\frac{\boldsymbol{h}_\gamma^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{h}_\gamma}{2v_{\boldsymbol{h}}} \Big) \prod_{m=1}^{M} p(\theta^m) p(\boldsymbol{a}^m \,|\, \theta^m). \end{aligned} \tag{12}$$

Our Bayesian model is too complex to be amenable to analytical calculations. Hence, we resort to a hybrid Gibbs sampling to sample the posterior distribution (12). The complete sampling procedure is detailed in [102, Table I]. The reader may also refer to [60, Appendix B]. PM estimates are then computed from these samples according to the following rule: $\widehat{x}^{\mathrm{PM}} = (T_{\mathrm{c}} - T_0)^{-1} \sum_{t=T_0+1}^{T_{\mathrm{c}}} x^{(g)}$, $\forall x \in \{\boldsymbol{h}_\gamma, \boldsymbol{A}, \Theta\}$ where T_0 stands for the length of the burn-in period (see Subsection 7.1). Also, for classification or detection purpose, the marginal maximum a posteriori criterion is employed: $(\widehat{q}_j^m)^{\mathrm{MAP}} = \arg\max_i \Pr(q_j^m = i \,|\, \boldsymbol{y}_j)$.

 ISSN: 2102-6238

5.3. *Posterior distribution under non-stationary BOLD model*

When resorting to the non-stationary BOLD signal $\boldsymbol{s}_j^m$ (see Eqs. (3)-(5)), the posterior distribution reads:

$$
\begin{aligned}
p(\boldsymbol{h}_\gamma, \boldsymbol{A}_1, \boldsymbol{R}, \boldsymbol{L}, \Theta \,|\, \boldsymbol{Y}) \propto\; & p(\boldsymbol{Y} \,|\, \boldsymbol{h}_\gamma, \boldsymbol{A}_1, \boldsymbol{R}, \boldsymbol{L}, \theta_0)\, p(\boldsymbol{A}_1, \boldsymbol{R} \,|\, \theta_{\boldsymbol{A}})\, p(\boldsymbol{h}_\gamma \,|\, v_{\boldsymbol{h}}) \\
& p(\boldsymbol{L} \,|\, v_\ell)\, p(\theta_0)\, p(\theta_{\boldsymbol{A}})\, p(v_{\boldsymbol{h}}, v_\ell) \\
\propto\; & v_{\boldsymbol{h}}^{-\frac{D}{2}} v_\ell^{-\frac{J_\gamma Q}{2}} \prod_{j=1}^{J_\gamma} \Big(\frac{(1-\rho_j^2)^{1/2}}{\sigma_j^{N+1}} I(|\rho_j| < 1) \Big) \\
& \exp\Big(- \sum_{j=1}^{J_\gamma} \big[\frac{1}{2\sigma_j^2} \widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \Lambda_j \widetilde{\boldsymbol{y}}_j + \frac{1}{2v_\ell} \|\ell_j\|^2 \big] \Big) \\
& \exp\big(- \frac{\boldsymbol{h}_\gamma^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{h}_\gamma}{2v_{\boldsymbol{h}}} \big) \prod_{m=1}^{M} p(\boldsymbol{a}_1^m, \boldsymbol{r}^m \,|\, \theta^m)\, p(\theta^m).
\end{aligned}
\tag{13}
$$

To simulate this posterior distribution, a hybrid Gibbs sampler including one-at-a-time Metropolis-Hastings (MH) moves has been implemented; see Subsection 5.5 and Appendix A for details.

5.4. *β sampling step*

Unsupervised spatial regularisation consists in automatically tuning parameter vector β from the dataset $\boldsymbol{Y}$ in a *given* parcel $\mathscr{P}_\gamma$. In the proposed hybrid Gibbs sampler (see [102, Table I]), this is implemented by adding a sampling block involving $p(\beta \,|\, \boldsymbol{Q})$ within the sampling loop. As shown in [102], a MH algorithm is designed to draw candidates but importantly this rests on the knowledge of the partition function $Z(\cdot)$, which can be estimated using importance sampling.

5.5. *One-at-a-time Metropolis-Hastings steps*

We have designed specific Single-component MH jumps for all parameters which the full conditional posterior distribution cannot be sampled directly. More precisely, separate jumps are proposed for each of the parameters in turn. To this end, suitable instrumental distributions regarding the parameters of interest are tuned. For AR parameters, akin to [60] we have paid attention to derive a *close-to-target* instrumental distribution by matching its mode to that of the full conditional posterior distribution $p(\rho_j \,|\, \boldsymbol{y}_j, \cdots)$. The acceptance ratio was tuned to about 70% limiting withdrawn realisations while enabling large jumps. Regarding the β-parameters, we resorted to a random walk MH step and consider a truncated Gaussian distribution as proposal i.e., $g(\cdot \,|\, x) \sim \mathscr{N}_{[0,\beta_{\max}]}(x, \xi_j^2)$ with the mean fixed on the current value x and a scale parameter ξ_j for each parameter that is updated every 10 jumps. At the tth update, ξ_j^t is updated using $\xi_j^{t+1} = \xi_j^t S1 + A + r/A + R$ where A and R are the numbers of accepted and rejected jumps since the last ξ_j^t update, respectively, S is the desired rejection rate, which was fixed at 0.7 [38, 51]. The same procedure has been successfully applied to the NHS parameters except that the proposal differs; see Appendix A.2 for details.

 ISSN: 2102-6238

6. Whole brain analysis: spatially adaptive USMM

We have described the Bayesian inference of within-parcel model parameters in the previous section. In this section, we describe how the parcels are defined. Crucially, the parcellation scheme does not depend upon the posterior estimates of the previous section. In other words, we first define our parcellation scheme and then optimise the model parameters for each parcel separately. In the next paragraph, we explain how the parcels are derived using a clustering algorithm that relies on a compound criterion balancing spatial and functional homogeneity.

6.1. Derivation of brain parcellation from fMRI data

As outlined in Section 3, our BOLD signal modelling is parcel-dependent and spatial regularisation differs over each parcel $\mathscr{P}_\gamma$. The methodology proposed in Section 5 to make this regularisation unsupervised is therefore applicable to all parcels separately making our whole brain analysis of fMRI data fully spatially adaptive. This makes sense given that the stimulus-specific SNRs also vary in space. Of course, the critical issue is to exhibit such functionally homogeneous parcellation of brain. To this end, several algorithms have been proposed [24, 99, 100]: they 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 can be computed either from the raw fMRI time series or from voxel-based hemodynamic features (time-to-peak and time-to-undershoot, peak and undershoot magnitudes, ...) [27], which can be extracted from nonparametric HRF estimates [14, 65]. As depicted in Figs. 4–5, the parcellations may be defined either in the volume or along the cortical surface.

Any parcel has a specific size and shape. Hence, the same Potts field defined over different parcels admits a distinct partition function since the latter depends on the number of voxels and cliques (here pairs of neighboring voxels) and the length of the parcel boundary. As unsupervised spatial regularisation in a given parcel needs partition function estimation, this estimation has to be repeated for each parcel to achieve spatially adaptive regularisation over the whole brain. For an averaged-size parcel of 250 voxels, partition function estimation requires around 10 seconds which results in a increase of about 30 minutes for a whole brain analysis, since several hundreds of parcels are typically necessary to cover the entire brain. In comparison with 1.5 hour for the complete analysis, it yields a large increase of 33%. Hence, we introduce fast numerical alternatives based on efficient approximations of 3D Potts field partition functions, so that the overall cost would be quite negligeable.

6.2. Multiple partition function estimation

To avoid any confusion, we reiterate that the estimation of the partition function is not part of the parcellation scheme but is required to implement the adaptive spatial constraints afforded by the parcellation, when inverting our model of hemodynamic responses in Section 5.

Several approaches have been designed to estimate a single partition function [66, 33]. However, none of them is able to perform multiple partition function estimation for Potts fields of variable size and shape in a reasonable amount of time. Since several hundreds of such grids are manipulated in the JDE context, fast estimation of multiple partition functions is necessary. To this

 ISSN: 2102-6238

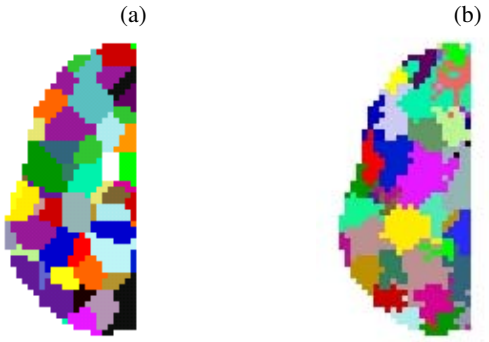

FIGURE 4. **3D Parcellations of the brain:** *(a) using the Voronoi based method - (b) as obtained with an optimal anatomo-functional parcellation [99].*

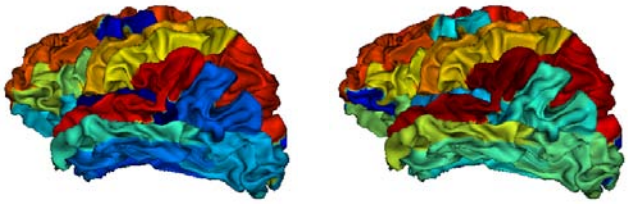

FIGURE 5. **Surface-based gyrii parcellation** *of the brain (left) and its decimated version (right) to guarantee parcels of homogeneous size [17, 100].*

end, we have proposed in [88] a hybrid scheme which consists first in resorting to path-sampling to get log-scale estimates $(\log \widehat{Z}_{\mathscr{G}_p}(\beta))_{p=1:P}$ in a small subset of *reference* grids $(\mathscr{G}_p)_{p=1:P}$ and then in using extrapolation formulas to obtain $\log \widetilde{Z}_{\mathscr{T}}(\beta)$ for the large remaining set of brain regions to be analysed, each of them referenced by a *test* graph $\mathscr{T}$ for the sake of notational simplicity.

6.2.1. *Fast and robust extrapolation technique*

The reliable extrapolation technique proposed in [87, 88] proceeds in two steps: 1) Akin to [85], reference log-PFs [5] $\log \widehat{Z}_{\mathscr{G}_p}(\beta_k)$ are estimated using path-sampling. The topological configurations of the reference grids $(\mathscr{G}_p)_{p=1:P}$ can be inhomogeneous to cover a maximum of situations that may occur when dealing with intra-subject parcellation. 2) For any test grid $\mathscr{T}$, the quantity $\log Z_{\mathscr{T}}$ is approximated from a *single* reference log-PF estimated by minimizing the maximal approximation error $\mathscr{A}(\beta, \mathscr{G}_p)$ with respect to all reference grids $(\mathscr{G}_p)_{p=1:P}$:

$$\mathscr{A}_{\mathscr{T}}(\beta, \mathscr{G}_p) = \|\log Z_{\mathscr{T}}(\beta) - \log \widetilde{Z}_{\mathscr{T}}(\beta, \mathscr{G}_p)\|^2 / \|\log Z_{\mathscr{T}}(\beta)\|^2$$
$$\text{with} \quad \log \widetilde{Z}_{\mathscr{T}}(\beta, \mathscr{G}_p) = \frac{c_{\mathscr{T}}}{c_{\mathscr{G}_p}} \left(\log \widehat{Z}_{\mathscr{G}_p}(\beta) - \log L\right) + \log L, \tag{14}$$

[5] log Partition-Function.

 ISSN: 2102-6238

where $(c_{\mathscr{T}}, c_{\mathscr{G}_p})$ and $(n_{\mathscr{T}}, n_{\mathscr{G}_p})$ are the number of cliques and sites of the L-color Potts fields defined over $\mathscr{T}$ and $\mathscr{G}_p$, respectively. In [103, Appendix A], it has been shown that $\mathscr{A}(0,\mathscr{G}_p) = \max_\beta \mathscr{A}(\beta,\mathscr{G}_p), \forall \mathscr{G}_p$, whenever the grid homogeneity in $\mathscr{G}_p$ and $\mathscr{T}$ is similar. Hence, we get:

$$\mathscr{G}_{\text{ref}} = \underset{(\mathscr{G}_p)_{p=1:P}}{\arg\min}\, \mathscr{A}_{\mathscr{T}}(0,\mathscr{G}_p) \quad \text{subject to} \quad \mathscr{L}_{\mathscr{T}}(\mathscr{G}_p) \leq \varepsilon$$

$$\text{and} \quad \mathscr{A}_{\mathscr{T}}(0,\mathscr{G}_p) \stackrel{\Delta}{=} \|(n_{\mathscr{T}}-1) - c_{\mathscr{T}}(n_{\mathscr{G}_p}-1)/c_{\mathscr{G}_p}\|^2 / n_{\mathscr{T}}^2$$

where $\varepsilon > 0$ is a positive threshold fixed by hand. Once $\mathscr{G}_{\text{ref}}$ has been identified, the log-PF estimate in $\mathscr{T}$ is thus given by $\log \widetilde{Z}_{\mathscr{T}}(\beta,\mathscr{G}_{\text{ref}})$ according to Eq. (14). Our extrapolation formula (14) is derived according to two principles: *i.)* an unbiased asymptotic approximation error [6] and *ii.)* an exact approximation of $(\log Z_{\mathscr{T}}(\beta))'$ for $\beta \to 0^+$; see [103, 88] for details.

In [88], we also compared the accuracy of our extrapolation approach to alternative mean-field like approximations [26] both in 2D and 3D. To get such results, we studied Potts fields defined on small grids and computed the ground truth partition function value both by path-sampling as well as using Onsager's formulae in the 2D context [71]. These results were also confirmed on larger fieds.

7. Results on Real fMRI datasets

We applied the JDE procedure to real *unsmoothed* fMRI data recorded during an experiment designed to map auditory, visual and motor brain functions as well as higher cognitive tasks such as number processing and language comprehension. It consisted of a single session of $N = 125$ scans lasting $TR = 2.4$ s each, yielding 3-D volumes composed of $64 \times 64 \times 32$ voxels. The paradigm was a fast event-related design comprising sixty auditory, visual and motor stimuli, defined in ten experimental conditions (auditory and visual sentences, auditory and visual calculations, left/right auditory and visual clicks, horizontal and vertical checkerboards).

7.1. MCMC convergence assessment

In practice, observations of the chain with different initial conditions confirmed that a burn-in of $T_0 = 5000$ iterations was sufficient followed by $T_c = 10^4$ subsequent jumps. In addition, convergence has been checked by monitoring on-line (see Appendix A.1) the behaviour of the estimated values of some *scalar* parameters (eg, noise variances, AR parameters, mixture parameters...) from one iteration to another. These observations also confirmed that 10^4 iterations were a fair compromise between convergence and numerical cost. Posterior mean estimates $(\widehat{\boldsymbol{A}}_\gamma, \widehat{\boldsymbol{h}}_\gamma)_{\gamma=1}^{\Gamma}$ have thus been computed over 10^4 realizations of the MCMC procedure.

7.2. 3D JDE analysis

We compare the three versions of the JDE procedure: IMM, SSMM ($\beta = 0.8$) and USMM, in order to assess the impact of the adaptive spatial correlation model. Fig. 6 shows normalised

[6] $\lim_{\beta \to +\infty} \mathscr{A}_{\mathscr{T}}(\beta,\mathscr{G}_p) = 0$.

 ISSN: 2102-6238

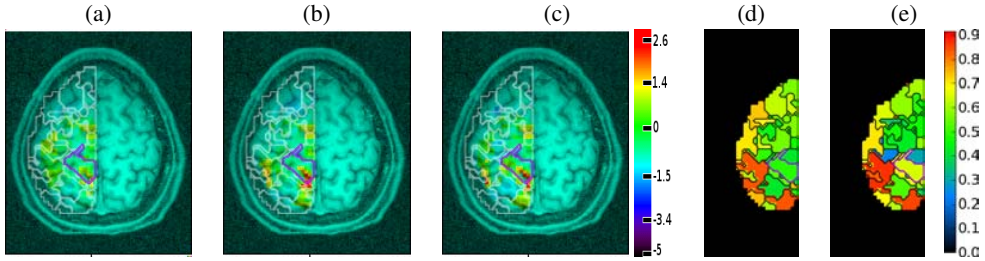

FIGURE 6. ***(a)-(c):*** *Comparison of the IMM (a), SSMM (b) and USMM (c) models wrt the Auditory Computation vs. Sentence (AC/S) normalised contrast maps:* $\left(\widehat{\boldsymbol{a}}^{\mathrm{AC}} - \widehat{\boldsymbol{a}}^{\mathrm{AS}}\right) / \mathrm{std}\left(\widehat{\boldsymbol{a}}^{\mathrm{AC}} - \widehat{\boldsymbol{a}}^{\mathrm{AS}}\right)$. *Stronger activations and deactivations appear in red and purple, respectively. The most activating parcel located in the parietal cortex is surrounded in magenta.* ***(d)-(e):*** *correspond to the parcelwise and condition-specific* β*-maps ((d): AC and (e): AS). Neurological orientation: left is left.*

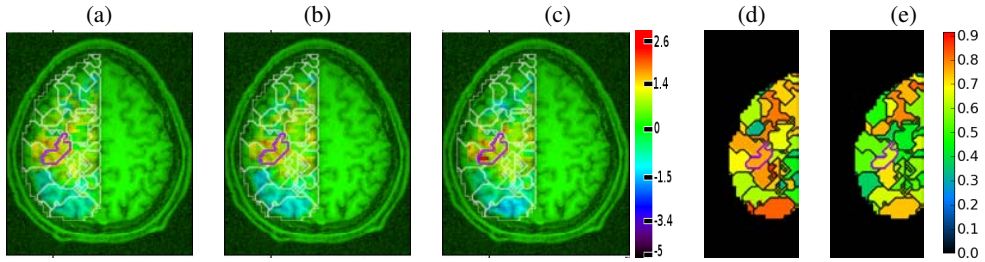

FIGURE 7. ***(a)-(c):*** *Comparison of the IMM (a), SSMM (b) and USMM (c) models wrt the Right vs. Left Auditory Click (R/LAC) normalised contrast maps:* $\left(\widehat{\boldsymbol{a}}^{\mathrm{RAC}} - \widehat{\boldsymbol{a}}^{\mathrm{LAC}}\right) / \mathrm{std}\left(\widehat{\boldsymbol{a}}^{\mathrm{RAC}} - \widehat{\boldsymbol{a}}^{\mathrm{LAC}}\right)$. *Stronger activations and deactivations appear in red and purple, respectively. The most activating parcel located in the left motor cortex is surrounded in magenta.* ***(d)-(e):*** *correspond to the parcelwise and condition-specific* β*-maps ((d): RAC and (e): LAC). Neurological orientation: left is left.*

contrasts maps of auditory computation (AC) versus auditory sentence (AS), where the modelling of spatial correlation seems to lead more sensitive results as activations in the parietal cortex are highlighted with SSMM and USMM whereas they are not with IMM. Moreover, these activations are coherent with the anatomy since they seem to follow the posterior part of the cingulate sulcus, which implication in numbers processing has been identified [44]. In another respect, Fig. 7 shows normalised contrast maps of auditory induced right click (RAC) versus auditory induced left click (LAC). As expected, the activations lie in the contralateral left motor cortex. Here, only USMM is more sensitive and we illustrate the advantage of an *adaptive* spatial correlation model. Indeed, estimated $\widehat{\beta}^{\mathrm{PM}}$ with USMM for the left auditory click was 0.56 so that the supervised setting of SSMM with $\beta = 0.8$ leads to too much correlation and less sensitive results.

Interestingly, Figs. 6-7 also depict the parcel-dependent maps of $\widehat{\beta}^{\mathrm{PM}}$ estimates for the RAC and LAC experimental conditions. The gain in sensitivity in the USMM contrast map $(\widehat{\boldsymbol{a}}^{\mathrm{RAC}} - \widehat{\boldsymbol{a}}^{\mathrm{LAC}})$ results from a difference in the amount of spatial regularisation introduced between the two conditions involved in the contrast. In parcels located in the left motor cortex, the BOLD signal is known to be stronger for the RAC than for the LAC condition. This is a possible interpretation of the lower regularisation level estimated for RAC compared to LAC ($\widehat{\beta}_{\mathrm{LAC}} \approx 0.5$ vs. $\widehat{\beta}_{\mathrm{RAC}} \approx 0.75$) in the activating region outlined in Fig. 7.

 ISSN: 2102-6238

Finally, following an illustrative purpose, Fig. 8 shows estimated HRF shapes in the most active regions for the two contrasts of interest. The time course $\widehat{\boldsymbol{h}}_{26}$ strongly departs from the canonical shape: its Time-To-Peak (TTP) is shifted by around 2 seconds. The bumped tail of $\widehat{\boldsymbol{h}}_{232}$ can be explained by several hypotheses. There may be a sort of periodic scheme in the stimulus involved by the concerned activating regions. This may also be a particular behaviour of the local vascular system which responds to the undershoot. In any case, we would have to resort to a specific paradigm for a precise investigation.

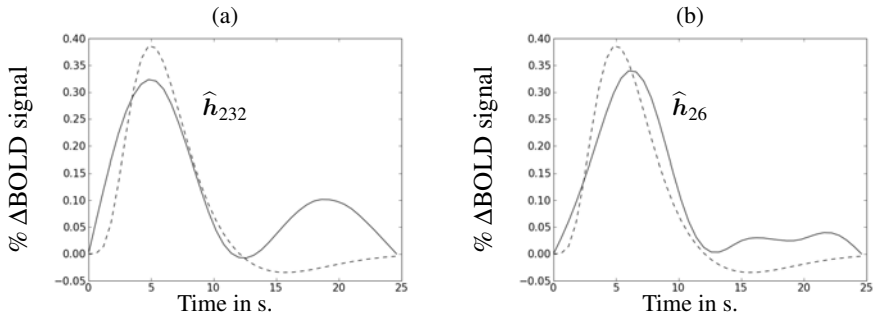

FIGURE 8. ***(a)**: HRF time course estimate $\widehat{\boldsymbol{h}}_{232}$ (solid line) for the mostly activated parcel in Fig. 6 superimposed to the canonical HRF (dashed line). **(b)**: HRF time course estimate $\widehat{\boldsymbol{h}}_{26}$ for the mostly activated parcel in Fig. -7 superimposed to the canonical HRF (dashed line).*

For the three-color Potts prior the parcel-dependent $\widehat{\beta}^{\mathrm{PM}}$ maps shown in Fig. 9 appear more homogeneous in space, especially for the LAC condition, and higher in absolute value. This observation can be explained as follows. When considering a three-class label configuration space for the Potts field, a Potts field realization involving only two states out of three has a lower probability than the same realization defined on a two-class Ising field. Hence, one way to enforce voxels in activating and non-activating states in the Potts case consists in increasing the β-value in comparison to the Ising case. This is exactly what happens in Fig. 9.

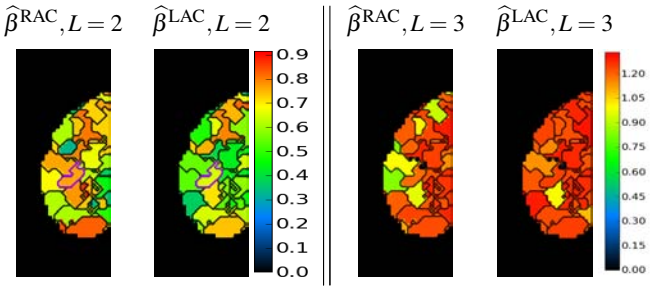

FIGURE 9. *Comparison of parcel-dependent $\widehat{\beta}$ maps computed for the RAC and LAC conditions and for the two-color ($L=2$, left panels also appearing in Fig. 7) and three-color ($L=3$, right panels) Potts fields.*

Moreover, it has been checked on observed Potts fields that changing the state space from two to three-classes generates an increase of the ML estimates of the β-parameters. This result is actually stronger on highly correlated configurations.

 ISSN: 2102-6238

7.3. 2D surface-based JDE analysis

Surface-based analysis requires three preprocessing steps: *i.)* the segmentation of the grey-white matter interface from the T_1-weighted anatomical MRI in order to extract the mesh of the cortical surface. *ii.)* the projection of fMRI signals acquired in the volume onto this cortical surface and *iii.)* the derivation of a cortical parcellation. Step *i.)* is easily addressed using the anatomical pipeline (T_1 **MRI toolbox**) in the BrainVISA software [7]. Step *ii.)* corresponds to mapping voxels in the original acquisition space to nodes of the cortical surface. To achieve this goal, we resort to the method proposed in [72], which is also available in BrainVISA but in the **Cortical Surface** toolbox [8]. The third step may also be addressed using the latter toolbox or resorting to the Freesurfer package [9]. We chose the latter solution to derive cortical parcellations in Fig 5.

Regarding JDE analysis, we only focus on the USMM version. The goal here is to illustrate the between-region hemodynamics variability at the subject level. To this end, we carried out the JDE algorithm on the over-segmented gyrii parcellation depicted in Fig. 5. From the surfaced-based HRF estimates, we extracted two parameters of interest: first, the TTP corresponding to the time point at which the HRF maximum is reached and second, the Full Width at Half-Maximum (FWHM), which describes the duration of activation before baseline return. The cortical mapping of these parameters is shown in Fig. 10. Interestingly, an antero-posterior TTP gradient is observed in the sense of slower responses in the frontal lobe. Also, it is worth noticing that the TTP fluctuates more across regions than the FWHM. The latter appears quite stable even for distant regions (pre-frontal and parietal gyrii).

As shown in Fig. 11(a)-(b), the vertical checkerboard stimulus elicits activation in the primary visual cortices (occipital lobe) whose orientation is coherently organized with the stimulus orientation due to the retinotopy mapping. Of course, the horizontal checkerboard visual stimulus elicits coherent horizontal activation in the same areas (results not shown). Also, we grouped the experimental conditions involving visual and auditory stimulus whatever the underlying cognitive task (reading, computing, ...). Fig. 11(a)-(b) show the Auditory-Visual contrast in the right and left hemispheres of a given subject, respectively. As expected, the major activation lie in the superior temporal sulci, with a stronger and larger BOLD response in the left hemisphere because of the lateralization of the language organization in the brain.

8. Bayesian model comparison

The fitting or inversion of forward models of hemodynamic responses in a Bayesian setting produces the integrated likelihood (or evidence) for a model (which is generally used for model comparison or selection), and the posterior or conditional density on the unknown model parameters, given that model. The evidence is used for inference on model-space and the posterior density is used for inference about the parameters conditioned upon a particular model. In this paper, we focussed on the inversion of models and how to access the posterior density using sampling techniques. Here, we close this review with a brief discussion on approximations to the integrated likelihood using appropriate techniques.

[7] `http://brainvisa.info`
[8] `http://olivier.coulon.perso.esil.univmed.fr/brainvisa.html`
[9] `http://surfer.nmr.mgh.harvard.edu`

 ISSN: 2102-6238

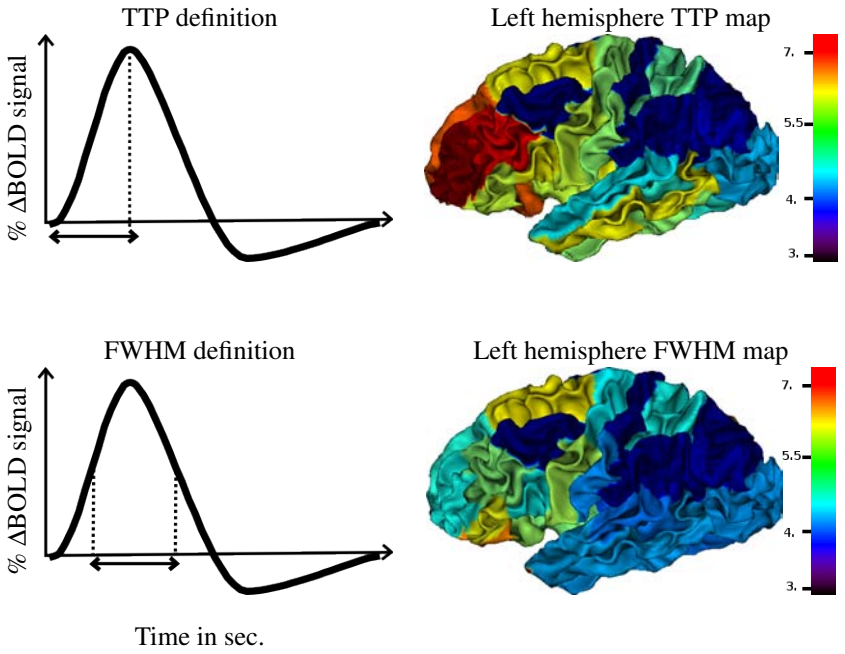

FIGURE 10. *Within-subject variability on the cortical surface (left hemisphere) of two HRF parameters.* **Top:** *Time-To-Peak (TTP) and* **Bottom:** *Full Width at Half-Maximum (FWHM).*

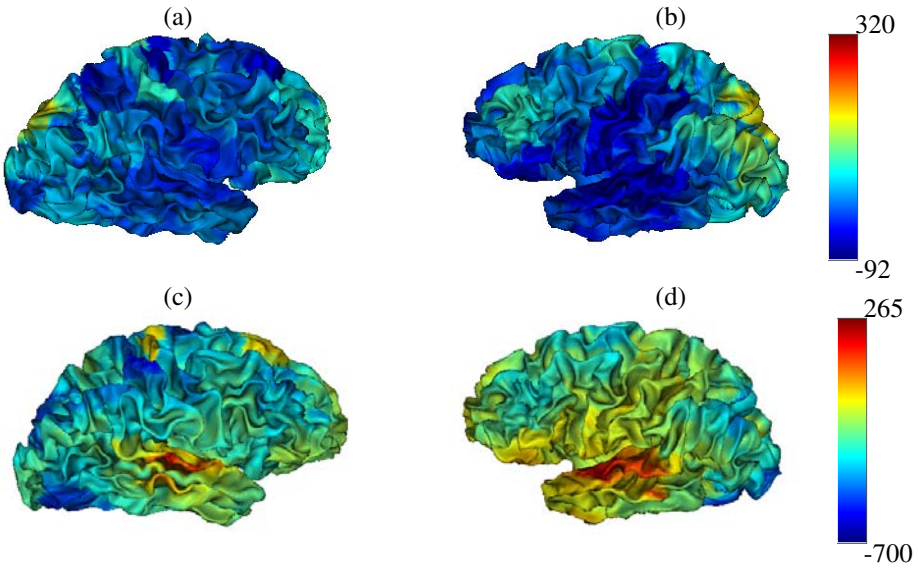

FIGURE 11. *Surface-based contrast maps derived from the USMM version of the JDE approach:* **(a)-(b):** *Contrast maps associated to the vertical checkerboard condition in the right and left hemispheres, respectively.* **(c)-(d):** *Contrast maps associated to the Auditory-Visual conditions in the right and left hemispheres, respectively.*

 ISSN: 2102-6238

8.1. *Why do we need model comparison or selection?*

We have chosen to frame our approach in terms of a joint detection-estimation framework, where both the detection and estimation rest upon the posterior density on the model parameters. The distinction between detection and estimation depends on how one uses the posterior density. First, we summarise what are the main motivations for model comparison and selection in the JDE formalism:

- **Sensitivity to the parcellation:** Since our inference depends on the input parcellation, it seems of primary interest to compare different configurations at the whole brain level in order to assess the impact of changing the parcellation on the HRF and NRL estimates and hence to decide which one provides the most functionally homogeneous brain regions. This may help us improving the experimental set up to acquire complementary fMRI data that aims at deriving the optimal parcellation.
- **sparse BOLD model:** In any parcel $\mathscr{P}_\gamma$, it is of primary interest to derive the sparsest BOLD signal model in the sense of the smallest set of experimental conditions that elicit activation in $\mathscr{P}_\gamma$. The problem therefore consists in selecting first the smallest integer $S \leqslant M$ and then in deriving the most relevant set of stimulus types $\mathscr{M}_S = \{m_1, \ldots, m_S\}$ involved in Eq. (1) to explain the measured fMRI time series at best; see [20] for details.
- **BOLD stationarity:** Even if the proposed non-stationary BOLD model in Subsection 3.3 may encompass the stationary one, its estimation is very time consuming given its voxel-dependent nature (see Eq. (5)). Hence, it only makes sense to consider the habituation modelling in parcels eliciting significant and strong repetition-suppression effect (eg, in the superior temporal sulcus if repeated sentences are delivered over headphones).
- **Negative BOLD response:** Select the best SMM prior regarding its number of components: two (activating and non-activating voxels) vs. three class-mixture in which deactivating voxels enter in the model; see [88] for further details about this comparison.
- **Noise modeling:** Compare the influence of different temporal noise models (white, AR(1), AR(p),...) on the sensitivity-specificity trade-off or consider spatially correlated noise models as done in [77]. Compare the detection power of conditional random fields in comparison with that of MRF [82], since CRFs enable the introduction of spatial dependencies between the measured fMRI signals, ie the modelling of spatially correlated noise process.
- **Prior MRF:** Compare the influence of a symmetric prior 3D MRF to its asymmetric counterpart in which the external field involves the sengmentation of anatomical MRI data in three tissue types (white and grey matters, cerebrospinal fluid).

To perform Bayesian model comparison and selection between different JDE results, we proceed first by carrying out parameter estimation for each model separately before comparing them using Bayes factors and thus approximations to model evidence or integrated likelihood, as explained below.

8.2. *Bayes factor computation*

Bayesian model comparison is usually addressed through the computation of Bayes factors:

$$BF_{ST} \triangleq \frac{p(\boldsymbol{Y} \,|\, \mathscr{M}_S)}{p(\boldsymbol{Y} \,|\, \mathscr{M}_T)} = \frac{\int p(\boldsymbol{Y} \,|\, \theta_S, \mathscr{M}_S)\, p(\theta_S \,|\, \mathscr{M}_S)\, d\theta_S}{\int p(\boldsymbol{Y} \,|\, \theta_T, \mathscr{M}_T)\, p(\theta_T \,|\, \mathscr{M}_T)\, d\theta_T},$$

 ISSN: 2102-6238

where $\mathscr{M}_S$ and $\mathscr{M}_T$ refer to tw different models, for instance when $(S,T) \in \mathbb{N}_M^*$ in order to compare the number of experimental conditions involved in Eq. (1). Bayes factor is computed as the ratio of *model evidences*, ie $p(\boldsymbol{Y} \,|\, \mathscr{M}_S)$ and $p(\boldsymbol{Y} \,|\, \mathscr{M}_T)$ of models $\mathscr{M}_S$ and $\mathscr{M}_T$. Several algorithms exist to estimate model evidences [68, 54, 10, 11, 42]. However, we focus on techniques that perform this estimation from the outputs of our MCMC algorithm since the JDE methodology relies to date on such simulation techniques.

The methodology proposed in [68, 54] and further developed in [79] based on the *harmonic mean identity*: $p(\boldsymbol{Y} \,|\, \mathscr{M}_S)^{-1} = \mathrm{E}\left[p(\boldsymbol{Y} \,|\, \theta_S, \mathscr{M}_S)^{-1} \,|\, \boldsymbol{Y}, \mathscr{M}_S\right]$, which provides the practitioner with the following model evidence estimate:

$$\widehat{p}_{\mathrm{HM}}(\mathscr{M}_S) = \left[\frac{1}{G}\sum_{g=1}^{G} \frac{1}{p(\boldsymbol{Y} \,|\, \theta_S^{(g)}, \mathscr{M}_S)}\right]^{-1}. \tag{15}$$

Although this estimator is a simulation-consistent estimate of $p(\boldsymbol{Y} \,|\, \mathscr{M}_S)$, it is not stable because of the inverse likelihood does not have finite variance. Hence, in [32] it has been proposed to resort to

$$\widehat{p}_{\mathrm{GD}}(\mathscr{M}_S) = \left[\frac{1}{G}\sum_{g=1}^{G} \frac{t(\theta_S^{(g)})}{p(\boldsymbol{Y} \,|\, \theta_S^{(g)}, \mathscr{M}_S)\pi(\theta_S^{(g)} \,|\, \mathscr{M}_S)}\right]^{-1}$$

where $t(\cdot)$ is a density with tails thinner than the product of the prior and the likelihood. It can be shown that $\widehat{p}_{\mathrm{GD}}(\mathscr{M}_S) \to p(\boldsymbol{Y} \,|\, \mathscr{M}_S)$ as G becomes large without the instability of $\widehat{p}_{\mathrm{HM}}(\mathscr{M}_S)$. Nonetheless, this approach requires a tuning function, which can be quite difficult to determine in high-dimensional problems, and subsequent monitoring to ensure that the numbers are stable. Its application therefore requires a stabilising procedure (see [79] for its theoretical foundations and [60, 12] for an application to the JDE context) that depends on the problem at hand and thus prevents its general use for all abovementioned model comparisons. Other attempts to modify the harmonic mean estimator, though requiring samples from both the prior and the posterior distributions, have been discussed in [68].

To overcome such difficulty, Chib has proposed in [10] an efficient procedure to approximate $p(\boldsymbol{Y} \,|\, \mathscr{M}_S)$ as soon as latent variables $\boldsymbol{Z}_S$ may enter in the formulation. In other words, using the data augmentation principle, the joint posterior distribution $p(\boldsymbol{Z}_S, \theta_S \,|\, \boldsymbol{Y}, \mathscr{M}_S)$ is sampled using Gibbs sampling. At convergence, we get $(\theta_S^{(g)}, \boldsymbol{z}^{(g)})_{g=1\ldots G} \sim p(\theta_S, \boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \mathscr{M}_S)$ and the computation of $\log p(\boldsymbol{Y} \,|\, \mathscr{M}_S)$ can thus be done using:

$$\forall \theta^* \in \Theta_S,\ \log p(\boldsymbol{Y} \,|\, \mathscr{M}_S) = \log p(\boldsymbol{Y} \,|\, \theta^*, \mathscr{M}_S) + \log p(\theta^* \,|\, \mathscr{M}_S) - \log p(\theta^* \,|\, \boldsymbol{Y}, \mathscr{M}_S) \tag{16}$$

$$p(\theta^* \,|\, \boldsymbol{Y}, \mathscr{M}_S) = \int p(\theta^* \,|\, \boldsymbol{Z}_S, \boldsymbol{Y}, \mathscr{M}_S)\, p(\boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \mathscr{M}_S)\, d\boldsymbol{Z}_S$$

where $p(\theta^* \,|\, \boldsymbol{Y}, \mathscr{M}_S)$ is approximated by

$$\widehat{p}(\theta^* \,|\, \boldsymbol{Y}, \mathscr{M}_S) = \frac{1}{G}\sum_{g=1}^{G} p(\theta^* \,|\, \boldsymbol{z}_S^{(g)}, \boldsymbol{Y}, \mathscr{M}_S).$$

 ISSN: 2102-6238

and the samples $(\boldsymbol{z}_S^{(g)})_{g=1\ldots G}$ are considered as outputs of the Gibbs sampler because the condition $\boldsymbol{z}_S^{(g)} \sim p(\boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \mathscr{M}_S)$ is met asymptotically ($G \to +\infty$). Since Eq. (16) is satisfied $\forall \theta^* \in \Theta_S$, it holds in particular for the MAP estimate under model $\mathscr{M}_S$.

In the JDE framework, we first note that our MCMC scheme is a hybrid MH-within-Gibbs sampling algorithm. Fortunately, Chib's approach has been generalized to the MH case in [11]. Nonetheless, whatever the choice of $\boldsymbol{Z}_S$ and θ_S the *two-blocks* version of Chib's algorithm cannot be implemented as such for JDE analysis since we are not able to simulate according to all conditional posterior distributions. We need to decompose $\theta_S = (\theta_{S1}, \theta_{S2})$, which leads to the following procedure: *i.)* sample $p(\theta_{S1} \,|\, \theta_{S2}, \boldsymbol{Z}_S, \boldsymbol{Y}, \mathscr{M}_S)$, *ii.)* sample $p(\theta_{S2} \,|\, \theta_{S1}, \boldsymbol{Z}_S, \boldsymbol{Y}, \mathscr{M}_S)$, and *iii.)* sample $p(\boldsymbol{Z}_S \,|\, \theta_{S1}, \theta_{S2}, \boldsymbol{Y}, \mathscr{M}_S)$. The goal is still to estimate $p(\theta_S^* \,|\, \boldsymbol{Y}, \mathscr{M}_S)$ as follows:

$$
\begin{aligned}
p(\theta_S^* \,|\, \boldsymbol{Y}, \mathscr{M}_S) &= p(\theta_{S1}^* \,|\, \boldsymbol{Y}, \mathscr{M}_S)\, p(\theta_{S2}^* \,|\, \theta_{S1}^*, \boldsymbol{Y}, \mathscr{M}_S) \\
p(\theta_{S1}^* \,|\, \boldsymbol{Y}, \mathscr{M}_S) &= \int p(\theta_{S1}^* \,|\, \theta_{S2}, \boldsymbol{Z}_S, \boldsymbol{Y}, \mathscr{M}_S)\, p(\theta_{S2}, \boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \mathscr{M}_S)\, d\theta_{S2}\, d\boldsymbol{Z}_S \qquad (17) \\
p(\theta_{S2}^* \,|\, \theta_{S1}^*, \boldsymbol{Y}, \mathscr{M}_S) &= \int p(\theta_{S2}^* \,|\, \theta_{S1}^*, \boldsymbol{Z}_S, \boldsymbol{Y}, \mathscr{M}_S)\, p(\boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \theta_{S1}^*, \mathscr{M}_S)\, d\boldsymbol{Z}_S. \qquad (18)
\end{aligned}
$$

As already done in the simple "two-blocks" case, we approximate Eq. (17) using Monte-Carlo integration:

$$
\widehat{p}(\theta_{S1}^* \,|\, \boldsymbol{Y}, \mathscr{M}_S) = \frac{1}{G_1} \sum_{g=1}^{G_1} p(\theta_{S1}^* \,|\, \theta_{S2}^{(g)}, \boldsymbol{Z}_S^{(g)}, \boldsymbol{Y}, \mathscr{M}_S)
$$

assuming that $\{\theta_{S2}^{(g)}, \boldsymbol{z}_S^{(g)}\}$ are drawn according to $p(\theta_{S2}, \boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \mathscr{M}_S)$. The main difficulty now lies in Eq. (18) because our Gibbs sampler generates realizations of $\boldsymbol{Z}_S$ which are distributed according to $p(\boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \mathscr{M}_S)$ and not to $p(\boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \theta_{S1}^*, \mathscr{M}_S)$. Hence, we propose to set $\theta_{S1} = \theta_{S1}^*$ after G_1 iterations of Gibbs sampling and then to continue the alternating simulation of $p(\theta_{S2} \,|\, \theta_{S1}^*, \boldsymbol{Z}_S, \boldsymbol{Y}, \mathscr{M}_S)$ and $p(\boldsymbol{Z}_S \,|\, \theta_{S1}^*, \theta_{S2}, \boldsymbol{Y}, \mathscr{M}_S)$. Then, asymptotically, we have $\boldsymbol{z}_S^{(G_1+g)} \sim p(\boldsymbol{Z}_S \,|\, \boldsymbol{Y}, \theta_{S1}^*, \mathscr{M}_S)$ and

$$
\widehat{p}(\theta_2^* \,|\, \theta_1^*, \boldsymbol{Y}, \mathscr{M}_S) = \frac{1}{G_2} \sum_{g=1}^{G_2} p(\theta_2^* \,|\, \theta_1^*, \boldsymbol{Z}_S^{(G_1+g)}, \boldsymbol{Y}, \mathscr{M}_S).
$$

Importantly, the asymptotic convergence of this scheme to the model evidence has been proved in [10, 11]. In the JDE framework (IMM case considering white noise assumptions to enable the simulation by Gibbs sampler without any MH step [12]), the choice of $\boldsymbol{Z}_S = (\boldsymbol{h}_\gamma, \boldsymbol{Q}, \boldsymbol{L})$ and $\theta_S = (\boldsymbol{A}, \theta_{\boldsymbol{A}}, \sigma^2, \sigma_{\boldsymbol{h}}^2)$ for every parcel $\mathscr{P}_\gamma$ has led to a closed form expression of the log-likelihood $\log p(\boldsymbol{Y} \,|\, \theta_S \mathscr{M}_S)\ \forall S \in \mathbb{N}_M^*$, which corresponds to the first term in the right hand side of Eq. (16). Hence, the proposed approximation scheme to model evidence can be integrated within our Gibbs sampler as follows:

1. Simulation of the missing data $\boldsymbol{Z}_S$ at iteration g:
   - $\boldsymbol{h}_\gamma \sim p\left(\cdot \,|\, \boldsymbol{Y}, \boldsymbol{L}^{(g-1)}, \boldsymbol{A}^{(g-1)}, (\sigma^2)^{(g-1)}, (\sigma_{\boldsymbol{h}}^2)^{(g-1)}\right)$
   - $\forall j, \ell_j \sim p\left(\cdot \,|\, \boldsymbol{y}_j, \boldsymbol{h}_\gamma^{(g)}, \boldsymbol{a}_j^{(g-1)}, (\sigma_j^2)^{(g-1)}\right)$
   - $\forall m, \boldsymbol{q}^m \sim p\left(\cdot \,|\, (\lambda_1^m)^{(g-1)}\right)$ where $\lambda_1^m = \Pr(q_j^m = 1), \forall j \in \mathscr{P}_\gamma$.

 ISSN: 2102-6238

2. Simulation of the parameters θ_S at iteration g:
   - Block 1: $\theta_{S1} = (\sigma^2_{\boldsymbol{h}}, \boldsymbol{A})$: simulate $\sigma^2_{\boldsymbol{h}} \sim p(\cdot \,|\, \boldsymbol{h}_\gamma^{(g)})$ and $\forall m \in \mathscr{M}_S$ sample $(\boldsymbol{a}^m)^{(g)} \sim p\left(. \,|\, \boldsymbol{Y}, (\boldsymbol{q}^m)^{(g)}, \boldsymbol{h}_\gamma^{(g)}, \boldsymbol{L}^{(g)}, (\sigma^2)^{(g-1)}\right)$. Note that the parallel computing over voxels for a given m is only feasible in the IMM case.
   - Block 2: $\theta_{S2} = (\sigma^2, \theta_{\boldsymbol{A}})$: Draw $\forall j, \sigma^2_j \sim p\left(\cdot \,|\, \boldsymbol{y}_j, \boldsymbol{h}_\gamma^{(g)}, \boldsymbol{a}_j^{(g)}\right)$ and then proceed to the simulation of mixture parameters θ^m, $\forall m$:
     - Simulate $\lambda_1^m \sim p\left(\cdot \,|\, (\boldsymbol{q}^m)^{(g)}\right)$
     - Simulate $v_i^m \sim p\left(\cdot \,|\, (\boldsymbol{a}^m)^{(g)}, (\boldsymbol{q}_m)^{(g)}\right)$ for $i = 0, 1$
     - Simulate $\mu_1^m \sim p\left(\cdot \,|\, (\boldsymbol{a}^m)^{(g)}, (\boldsymbol{q}_m)^{(g)}, (v_1^m)^{(g)}\right)$

The technical details are available in [12]. The extension of this structure to correlated noise processes is currently under development but remains feasible. However, the generalisation to SMM mixtures is not straightforward because the log-likelihood $\log p(\boldsymbol{Y} \,|\, \theta_S, \mathscr{M}_S)$ do not factorise over voxels and thus no longer admits a closed-form expression. To this end, we will pay attention to alternative strategies based on mean-field variational approximations for computing model evidences.

9. Discussion and conclusion

In the present paper, we review the most advanced version of the JDE approach to analyse fMRI data at the subject level. We focussed on several aspects. First, unsupervised and spatially adaptive regularisation has been integrated in the Bayesian formalism to make the recovery of activation clusters feasible even from the unsmoothed fMRI time series. This extension has been achieved by estimating parcel-dependent regularisation parameters β, which requires a precise estimation of the underlying 3D Potts field partition function. To this end, an extrapolation algorithm based on a few path-sampled partition function estimates has been used over the vast majority of parcels, either in the 3D or the 2D context.

Second we have highlighted, using real fMRI data, that a misspecification of a fixed β value (SSMM) can lead to wrongly estimated activation label maps so that spatial correlation modelling does not bring any advantage compared to IMM. This limitation of SSMM was finessed by the USMM approach where a more relevant setting of β was found. The optimal setting of β actually varies when considering different regions of the brain. In this respect, we identified regions where SSMM-$\beta = 0.8$ provides the same effect maps as IMM whereas USMM was more sensitive ($\hat{\beta} = 0.56$). Hence, our claim is that our regularisation scheme is not only unsupervised but *spatially adaptive*. As a remark, the sensitivity gain compared to IMM was mainly observed for low contrasts between conditions. For contrast involving higher response levels (for example auditive vs. visual conditions), there were no noticeable gain in resorting to USMM. Indeed, in this case there is enough information within data and spatial regularisation may be useless. To summarise, our approach enables a finer recovering of subtle contrasts by adapting the spatial regularisation to varying contrast-to-noise ratios as well as various underlying activation patterns.

Third, our approach relies on the definition of an appropriate spatial scale which should be small enough so that the HRF shape invariance can be assumed and big enough so that we benefit from enough HRF reproductibility. This optimal trade-off, which impacts temporal modelling, is fulfilled by the parcellation scheme we rely on. If we resort to such partitioning for HRF

 ISSN: 2102-6238

modelling, one could also think of taking the whole brain as the spatial support for the detection part, so that we would consider only one SMM to model response levels. However, this would obvioulsy prevent us from making regularisation spatially adaptive. Even if parcellation is well justified for HRF modelling from a physiological viewpoint, it is still questionable for response levels modelling. Indeed, two regions may be well suited to explain different vascular system properties so that we consider two different HRFs but they may not be suited for underlying activations which may span these two regions for example. From a pratictal viewpoint, we did not observe any impact of parcel boundaries that would prevent activation clusters to span different regions.

Also, we have illustrated that the JDE framework enables the study of brain deactivations considering three-class prior hidden Potts model instead of Ising ones. On the proposed example, we did not show any difference in the contrast maps indicating that there is no deactivation in the data set under consideration. Future work will investigate pathological data (eg like in epilepsy) on which it is known a priori that deactivations occur.

Finally, we have demonstrated the interest of running the JDE analysis on the cortical surface to investigate the between-region hemodynamics fluctuations. At the subject level, it has been shown that an antero-posterior TTP gradient exists with earlier BOLD responses in the occipital cortex. Further analysis will investigate the origin of these findings with multimodal MRI data (Arterial Spin Labeling), the goal being to disentangle the role of the vascular circuitry from the neuronal sources of activation. This also needs to be validated at the group level.

Ongoing work will validate the current method on group analysis to appraise the impact of our spatial regularisation scheme to the sensitivity of group level effect maps. To this end, parcellation at the group level could also be performed to account for between-subject variability [16]. Indeed, linking together brain parcels rather than voxels allows us to overcome the drawbacks of normalisation due to spatial variability.

As future direction of research, one could also produce a meaningful functional parcellation related to hemodynamic parameters. After treating an over-segmented parcellation, spatially connected parcels which share the same labels could then be merged together and then be subsequently used in another JDE iteration to derive more reliable hemodynamic parameters. This might require more constraints like the use of a semi-parametric approach regarding the hemodynamic filter to specify its shape among a finite class of FIR models whose dimension should be small compared to number of parcels. Hence, several parcels could have the same HRF shape. Also, the use of VB approximations might maintain closed-form updating rules for the different unknows in the inference scheme as well as a computationally realistic numerical strategy. This would enable the optimisation of the full joint posterior probability of the parcellation scheme and the unknown parameters related to the JDE approach.

Appendix A: Details on our hybrid Gibbs sampler

A.1. Convergence diagnosis of our hybrid Gibbs-MH algorithm

Convergence monitoring has been performed component-wise using parallel sampling as detailed in [33]. For each estimand ϕ (eg, $\phi =$), we draw B parallel sequences of length C (we typically took $B = 10$ and $C = 50$), each sample being denoted $\phi^{[bc]}$, with $b = 1 : B$ and $c = 1 : C$. We

 ISSN: 2102-6238

then compute the between-sequence variance (BV), and the Within-sequence Variance (WV) as follows:

$$\mathrm{BV} = \frac{C}{B-1}\sum_{b=1}^{B}\left(\bar{\phi}^{[\mathrm{b}\cdot]} - \bar{\phi}^{[\cdot\cdot]}\right)^2 \text{ with } \bar{\phi}^{[\mathrm{b}\cdot]} = \frac{1}{C}\sum_{c=1}^{C}\phi^{[\mathrm{bc}]} \text{ and } \bar{\phi}^{[\cdot\cdot]} = \frac{1}{B}\sum_{b=1}^{B}\bar{\phi}^{[\mathrm{b}\cdot]}$$

$$\mathrm{WV} = \frac{1}{B}\sum_{b=1}^{B}(s^2)^{[\mathrm{b}]}, \quad \text{with} \quad (s^2)^{[\mathrm{b}]} = \frac{1}{C-1}\sum_{c=1}^{C}\left(\phi^{[\mathrm{bc}]} - \bar{\phi}^{[\mathrm{b}\cdot]}\right)^2$$

We then calculate $\sqrt{\widehat{R}} = \sqrt{1+\frac{1}{C}\left(\frac{\mathrm{BV}}{\mathrm{WV}}-1\right)}$ for each scalar estimand. These quantities are supposed to decline to 1 as the sampling converges. We stop the algorithm when all are close enough to 1, e.g., smaller than 1.1, and remove α percent of each chain to account for a burn-in period.

A.2. Metropolis-Hastings steps

The MH move specifically designed for Potts field parameters β has been intensively studied in [102], while the one dedicated to AR noise parameters ρ has been detailed in [60]: the calibration of the instrumental distribution was not too difficult in both cases.

Here, we study the MH step that concerns the normalised habituation speeds $\boldsymbol{R}$. The full conditional posterior $\pi_1^\star(r_j^m) = p(r_j^m \,|\, \boldsymbol{y}_j, q_j^m = 1, rest)$ reads:

$$\pi_1^\star(r_j^m) \propto \exp\left(-\|\boldsymbol{y}_j - C_{j\setminus m} - \boldsymbol{s}_j^m\|^2_{\Lambda_j}/2\sigma_j^2\right)\mathscr{U}_{(0,1)}(r_j^m),$$

where $\boldsymbol{s}_j^m$ depends on r_j^m and $C_{j\setminus m} = \boldsymbol{P}\ell_j - \sum_{n\neq m}\boldsymbol{s}_j^n$ is independent of r_j^m, it cannot be sampled directly. Hence, we resort to a MH step with an uncentered Laplacian density, truncated over [0,1], as proposal: $f(r\,|\,r_0) = Z_{\beta,r_0}^{-1}e^{-\beta|r-r_0|}\mathbb{1}_{[0,1]}(r)$, where $r_0 = r_j^{m,(t-1)}$. At iteration t, the MH acceptance ratio is given by $\alpha(r_0 \to r) = \min\left[1, \frac{\pi_1^\star(r)}{\pi_1^\star(r_0)}\frac{Z_{\beta,r_0}}{Z_{\beta,r}}\right]$, which requires to generate the trial-varying NRLs (Eq. (3)) at points r and r_0. This can be hopefully done efficiently using Eq. (4). Note that detection is performed according to the MAP criterion: $(\widehat{q}_j^m)^{\mathrm{MAP}} = \arg\max_i \Pr(q_j^m = i\,|\,\boldsymbol{y}_j)$. The NHS estimate is derived as follows: in every non-activating voxel ($(\widehat{q}_j^m)^{\mathrm{MAP}} = 0$), we impose $\widehat{r}_j^m = 0$. For activating voxels ($(\widehat{q}_j^m)^{\mathrm{MAP}} = 1$), the MHS estimate is computed as the average of samples $r_j^{m,(g)}$ over iterations g that satisfy $q_j^{(g),m} = 1$. The goal is to avoid mixing effects with zero-valued MHS samples in case of label switching.

Acknowledgments

We are very grateful to the reviewers for their constructive comments and recommandations. We also want to express our gratitude to Jérôme Idier and Florence Forbes for fruitful discussions. This work has been supported by Region Ile de France.

References

[1] G. K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human BOLD hemodynamic responses. *Neuroimage*, 7:574, 1998.

 ISSN: 2102-6238

[2] P. Baraldi, A. Manginelli, M. Maieron, D. Liberati, and A. Porro. An ARX model-based approach to trial by trial identification of fMRI-BOLD responses. *Neuroimage*, 37:189–201, 2007.

[3] Adrian Barbu and Song-Chun Zhu. Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1239–1253, Aug. 2005.

[4] Matthew Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University College of London, London, United Kingdom, May 2003.

[5] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, Nov. 2001.

[6] R. B. Buxton, E. C. Wong, and Frank. L. R. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.*, 39:855–864, June 1998.

[7] Richard B. Buxton, Kãmil Uludağ, David J. Dubowitz, and Thomas T. Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23, Supplement 1:S220–S233, 2004.

[8] Ramon Casanova, Srikanth Ryali, John Serences, Lucie Yang, Robert Kraft, Paul J Laurienti, and Joseph A Maldjian. The impact of temporal regularization on estimates of the bold hemodynamic response function: a comparative analysis. *Neuroimage*, 40(4):1606–1618, May 2008.

[9] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Processing*, 6(2):298–311, Feb. 1997.

[10] Siddharta Chib. Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.*, 90:1313–1321, 1995.

[11] Siddharta Chib and Ivan Jeliazkov. Marginal likelihood from the Metropolis-Hastings output. *J. Amer. Statist. Assoc.*, 96(453):270–281, 2001.

[12] P. Ciuciu and S. Donnet. Bayesian model comparison and selection in neuroimaging from JDE results. Research Rep., LNAO, NeuroSpin/CEA, Gif-sur-Yvette, France, May 2010.

[13] P. Ciuciu and J. Idier. A Half-Quadratic block-coordinate descent method for spectral estimation. *Signal Processing*, 82(7):941–959, July 2002.

[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.*, 22(10):1235–1251, Oct. 2003.

[15] P. Ciuciu, S. Sockeel, T. Vincent, and J. Idier. Modelling the neurovascular habituation effect on fMRI time series. In *34th Proc. IEEE ICASSP*, pages 433–436, Taipei, Taiwan, Apr. 2009.

[16] P. Ciuciu, T. Vincent, A.-L. Fouque, and A. Roche. Improved fMRI group studies based on spatially varying non-parametric BOLD signal modeling. In *5th Proc. IEEE ISBI*, pages 1263–1266, Paris, France, May 2008.

[17] C. Clouchoux, O. Coulon, J.-L Anton, , J.-F. Mangin, and J. Régis. A new cortical surface parcellation model and its automatic implementation. In *Proc. 9th MICCAI*, LNCS 4191, pages 193–200, Copenhaegen, Denmark, Oct. 2006. Springer Verlag.

[18] Mark S. Cohen. Parametric analysis of MRI data using linear systems methods. *Neuroimage*, 6:93–103, 1997.

[19] X. Descombes, F. Kruggel, and D. Y. von Cramon. Spatio-temporal fMRI analysis using Markov Random Fields. *IEEE Trans. Med. Imag.*, 17(6):1028–1039, Dec. 1998.

[20] S. Donnet, M. Lavielle, and J.-B. Poline. Are fMRI event-related response constant in time? A model selection answer. *Neuroimage*, pages 1169–1176, Apr. 2006.

[21] JR Duann, TP Jung, WJ Kuo, TC Yeh, S Makeig, Hsieh JC, and Sejnowski TJ. Single-trial variability in event-related BOLD signals. *Neuroimage*, 15(4):823–35, Apr. 2002.

[22] B. S. Everitt and E. T. Bullmore. Mixture model mapping of brain activation in functional magnetic resonance images. *Hum. Brain Mapp.*, 7:1–14, 1999.

[23] Sylvia Fernández and Peter J. Green. Modelling spatially correlated data via mixtures: a Bayesian approach. *J. R. Statist. Soc. B*, 64(4):805–826, 2002.

[24] G. Flandin, F. Kherif, X. Pennec, G. Malandain, N. Ayache, and J.-B. Poline. Improved detection sensitivity of functional MRI data using a brain parcellation technique. In *Proc. 5th MICCAI*, LNCS 2488 (Part I), pages 467–474, Tokyo, Japan, Sep. 2002. Springer Verlag.

[25] G. Flandin and W. D. Penny. Bayesian fMRI data analysis with sparse spatial basis function priors. *Neuroimage*, 34(3):1108–1125, Feb. 2007.

[26] F. Forbes and N. Peyrard. Hidden Markov Random Field model selection criteria based on mean field-like approximations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1089–1101, Sep. 2003.

 ISSN: 2102-6238

[27] 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*, pages 445–448, Taipei, Taiwan, Apr. 2009.

[28] K. J. Friston, A. Mechelli, R. Turner, and C. J. Price. Nonlinear responses in fMRI: the balloon model, Volterra kernels, and other hemodynamics. *Neuroimage*, 12:466–477, June 2000.

[29] K.J. Friston. Statistical parametric mapping. In R.W. Thatcher, M. Hallet, T. Zeffiro, E.R. John, and M. Huerta, editors, *Functional Neuroimaging : Technical Foundations*, pages 79–93, 1994.

[30] KJ Friston, L Harrison, and W Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–302, Aug. 2003.

[31] K.J. Friston, J. Mattout, N. Trullijo-Barreto, J. Ashburner, and W. Penny. Variational free energy and the Laplace approximation. *Neuroimage*, 33:220–234, 2007.

[32] A. E. Gelfand and D. Dey. Bayesian model choice: asymptotic and exact calculations. *J. R. Statist. Soc. B*, 56:501–514, 1994.

[33] A. Gelman and X.-L. Meng. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*, 13:163–185, 1998.

[34] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):367–383, Mar. 1992.

[35] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Processing*, 4(7):932–946, July 1995.

[36] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov. 1984.

[37] C.R. Genovese. A Bayesian time-course model for functional magnetic resonance imaging data (with discussion). *J. Amer. Statist. Assoc.*, 95:691–719, 2000.

[38] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo in practice*. Chapman and Hall, London, United Kingdom, 1996.

[39] G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9:416–429, 1999.

[40] C. Gössl, D. P. Auer, and L. Fahrmeir. Bayesian spatio-temporal modeling of the hemodynamic response function in BOLD fMRI. *Biometrics*, 57:554–562, June 2001.

[41] C. Goutte, F. A. Nielsen, and L. K. Hansen. Modeling the haemodynamic response in fMRI using smooth filters. *IEEE Trans. Med. Imag.*, 19(12):1188–1201, Dec. 2000.

[42] P. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732, 1995.

[43] P J. Green and S. Richardson. Hidden Markov models and desease mapping. *J. Amer. Statist. Assoc.*, 97(460):1–16, Dec. 2002.

[44] O. Gruber, P. Indefrey, H. Steinmetz, and A. Kleinschmidt. Dissociating Neural Correlates of Cognitive Components in Mental Calculation. *Cereb. Cortex*, 11(4):350–359, 2001.

[45] 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*, 21:1639–1651, 2004.

[46] L. M. Harrison, W. Penny, J. Daunizeau, and K. J. Friston. Diffusion-based spatial priors for functional magnetic resonance images. *Neuroimage*, 41(2):408–423, Jun 2008.

[47] 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*, 15(1):83–97, 2002.

[48] R. Henson, M. Rugg, and K. Friston. The choice of basis function in event-related fMRI. volume 13, page 149, 2001.

[49] D. M. Higdon. Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Amer. Statist. Assoc.*, 93(442):585–595, June 1998.

[50] A.P. Holmes and I. Ford. A Bayesian approach to significance testing for statistical images from PET. In *Quantification of Brain Function, Tracer Kinematics in Image Analysis in Brain PET*, Excerpta Medica International Congress Series (1030), pages 521–531, New York, USA, 1993. Elsevier Science.

[51] J. Idier. *Bayesian approach to Inverse problems*. ISTE Ltd and John Wiley & Sons Inc, Apr. 2008.

 ISSN: 2102-6238

[52] F. C. Jeng and J. W. Woods. Compound Gauss-Markov random fields for image estimation. *IEEE Trans. Signal Processing*, 39(3):683–697, Mar. 1991.

[53] Leigh A. Johnston, Eugene Duff, Iven Mareels, and Gary F. Egan. Nonlinear estimation of the BOLD signal. *Magn. Reson. Med.*, 40(2):504–514, Apr. 2008.

[54] R. E. Kass and Adrian E. Raftery. Bayes factors. *J. Amer. Statist. Assoc.*, 90:773–795, 1995.

[55] S.J Kiebel, J. Daunizeau C. Phillips, and K.J. Friston. Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG. *Neuroimage*, 39:728–741, 2008.

[56] F. Kruggel and D. Y. Von Crammon. Modeling the hemodynamic response in single-trial functional MRI experiments. *Magn. Reson. Med.*, 42:787–797, 1999.

[57] N. Lange. Empirical and substantive models, the Bayesian paradigm, and meta-analysis in functional brain imaging. *Hum. Brain Mapp.*, 5:259–263, 1997.

[58] J.S. Liu. *Monte Carlo strategies in scientific computing*. Springer series in Statistics. Springer-Verlag, New-York, 2001.

[59] S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline. Joint detection-estimation of brain activity in functional MRI: a multichannel deconvolution solution. *IEEE Trans. Signal Processing*, 53(9):3488–3502, Sep. 2005.

[60] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3):941–969, July 2008.

[61] Salima Makni, Christian Beckmann, Steve Smith, and Mark Woolrich. Bayesian deconvolution fMRI data using bilinear dynamical systems. *Neuroimage*, 42(4):1381–1396, Oct. 2008.

[62] André C Marreiros, Stefan J Kiebel, and Karl J Friston. A dynamic causal model study of neuronal population dynamics. *Neuroimage*, 51(1):91–101, May 2010.

[63] G. Marrelec, H. Benali, P. Ciuciu, M. Pélégrini-Issac, and J.-B. Poline. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD MRI using basic physiological information. *Hum. Brain Mapp.*, 19(1):1–17, May 2003.

[64] G. Marrelec, P. Ciuciu, M. Pélégrini-Issac, and H. Benali. Estimation of the hemodynamic response function in event-related functional MRI: Directed acyclic graphs for a general Bayesian inference framework. In Chris Taylor and J. Alison Noble, editors, *Proc. 18th IPMI*, LNCS-2732, pages 635–646, Ambleside, United Kingdom, 2003. Springer Verlag.

[65] G. Marrelec, P. Ciuciu, M. Pélégrini-Issac, and H. Benali. Estimation of the hemodynamic response function in event-related functional MRI: Bayesian networks as a framework for efficient Bayesian modeling and inference. *IEEE Trans. Med. Imag.*, 23(8):959–967, Aug. 2004.

[66] X.L. Meng and W.H. Wong. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6:831–860, 1996.

[67] F. M. Miezin, L. Maccotta, J. M. Ollinger, S. E. Petersen, and R. L. Buckner. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*, 11:735–759, 2000.

[68] M.A. Newton and E. Raftery. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J. R. Statist. Soc. B*, 56:3–48, 1994.

[69] F. A. Nielsen, L. K. Hansen, P. Toft, C. Goutte, N. Lange, S. C. Stroher, N. Morch, C. Svarer, R. Savoy, B. Rosen, E. Rostrup, and P. Born. Comparison of two convolution models for fMRI time series. *Neuroimage*, 5:S473, 1997.

[70] S. Ogawa, T. Lee, A. Kay, and D. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA*, 87(24):9868–9872, 1990.

[71] L. Onsager. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65(3& 4):117–149, Feb. 1944.

[72] G Operto, R. Bulot, J.-L. Anton, and O. Coulon. Anatomically informed convolution kernels for the projection of fMRI data on the cortical surface. In *Proc. 9th MICCAI*, LNCS 4191, pages 300–307, Copenhaegen, Denmark, Oct. 2006. Springer Verlag.

[73] Wanmei Ou and Polina Golland. From spatial regularization to anatomical priors in fMRI analysis. In *IPMI, Glenwood Springs, Colorado*, July 2005.

[74] Wanmei Ou and Polina Golland. Combining spatial priors and anatomical information for fMRI detection. *Medical Image Analysis*, 14(3):318–331, June 2010.

 ISSN: 2102-6238

[75] W. Penny, Z. Ghahramani, and K. Friston. Bilinear dynamical systems. *Philos Trans R Soc Lond B Biol Sci*, 360(1457):983–993, May 2005.

[76] W. D. Penny, N. Trujillo-Barreto, and K. J. Friston. Bayesian fMRI time series analysis with spatial priors. *Neuroimage*, 23(2):350–362, 2005.

[77] W.D. Penny, G. Flandin, and N. Trujillo-Bareto. Bayesian Comparison of Spatially Regularised General Linear Models. *Human Brain Mapping*, 28(4):275–293, 2007.

[78] Will D Penny, Klaas E Stephan, Jean Daunizeau, Maria J Rosa, Karl J Friston, Thomas M Schofield, and Alex P Leff. Comparing families of dynamic causal models. *PLoS Comput Biol*, 6(3):e1000709, 2010.

[79] Adrian E. Raftery, Michael A. Newton, Jaya M. Satagopan, and Pavel N. Krivitsky. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In J.M. Bernardo, M.J. Bayarri, O. Berger, A.P. David, D. Heckermann, A.F.M. Smith, and M. West, editors, *Bayesian statistics 8*, pages 1–45. Oxford University Press, 2007.

[80] J. C. Rajapakse, F. Kruggel, J. M. Maisog, and D.Y. Von Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. *Hum. Brain Mapp.*, 6:283–300, 1998.

[81] J. C. Rajapakse and J. Piyaratna. Bayesian approach to segmentation of statistical parametric maps. *IEEE Trans. Biomed. Eng.*, 48:1186–1194, 2001.

[82] Jagath C Rajapakse, Choong Leong Tan, Xuebin Zheng, Susanta Mukhopadhyay, and Kanyan Yang. Exploratory analysis of brain connectivity with ICA. *IEEE Eng Med Biol Mag*, 25(2):102–111, 2006.

[83] J.J. Riera, J. Bosch, O. Yamashita, R. Kawashima, N. Sadato, T. Okada, and T. Ozaki. fMRI activation maps based on the NN-ARx model. *Neuroimage*, 23:680–697, 2004.

[84] J.J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and R. Kawashima. A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signal. *Neuroimage*, 21:547–567, 2004.

[85] L. Risser, J. Idier, and P. Ciuciu. Bilinear extrapolation scheme for fast estimation of 3D ising field partition function. Application to fMRI time course analysis. In *16th Proc. IEEE ICIP*, pages 833–836, Cairo, Egypt, Nov. 2009.

[86] L. Risser, T. Vincent, and P. Ciuciu. Schéma d'extrapolation de fonctions de partition de champs de potts. application à l'analyse d'images en IRMf. In *Actes du 22^{e} colloque GRETSI*, Dijon, France, Sep. 2009.

[87] L. Risser, T. Vincent, P. Ciuciu, and J. Idier. Robust extrapolation scheme for fast estimation of 3D Ising field partition functions. application to within-subject fMRI data analysis. In G.-Z. Yang, editor, *12thProc. MICCAI'09*, LNCS 5761, pages 975–983, London, UK, Sep. 2009. Springer Verlag Berlin Heidelberg.

[88] L. Risser, T. Vincent, F. Forbes, J. Idier, and P. Ciuciu. Min-max extrapolation scheme for fast estimation of 3D Potts field partition functions. application to the joint detection-estimation of brain activity in fMRI. *Journal of Signal Processing Systems*, in press, June 2010.

[89] Amir Shmuel, Mark Augath, Axel Oeltermann, and Nikos K Logothetis. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci*, 9(4):569–577, Apr 2006.

[90] D. Smith and M. Smith. Estimation of binary Markov random fields using Markov Chain Monte Carlo. *J. Comput. and Graph. Stats.*, 15(1):207–227, 2006.

[91] M. Smith, B. Pütz, D. Auer, and L. Fahrmeir. Assessing brain activity through spatial Bayesian variable selection. *Neuroimage*, 20:802–815, 2003.

[92] Hichem Snoussi and J. Idier. Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures. *IEEE Trans. Signal Processing*, 54(9):3257–3269, Sep. 2006.

[93] R.C. Sotero and N.J. Trullijo-Barreto. Modelling the role of excitatory and inhibitory neuronal activity in the generation of the BOLD signal. *Neuroimage*, 35:149–165, 2007.

[94] R.C. Sotero, N.J. Trullijo-Barreto, J. Jiménez, F. Carbonell, and R. Rodrìguez-Rojas. Identification and comparison of stochastic metabolic/hemodynamic models (smhm) for the generation of the BOLD signal. *J. Comput. NeuroSci.*, 26:251–269, 2009.

[95] K. E. Stephan, W. D. Penny, R. J. Moran, H. E M den Ouden, J. Daunizeau, and K. J. Friston. Ten simple rules for dynamic causal modeling. *Neuroimage*, 49(4):3099–3109, Feb. 2010.

[96] Klaas E Stephan, Lee M Harrison, Stefan J Kiebel, Olivier David, Will D Penny, and Karl J Friston. Dynamic causal models of neural system dynamics:current state and future extensions. *J Biosci*, 32(1):129–144, Jan. 2007.

 ISSN: 2102-6238

[97] Klaas E Stephan, Lee M Harrison, Will D Penny, and Karl J Friston. Biophysical models of fMRI responses. *Curr Opin Neurobiol*, 14(5):629–635, Oct. 2004.

[98] Markus Svensen, Frithjof Kruggel, and D.Y. von Crammon. Probabilistic modeling of single-trial fMRI data. *IEEE Trans. Med. Imag.*, 19:19–35, Jan. 2000.

[99] 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.*, 27(8):678–693, Aug. 2006.

[100] A. Tucholka, B. Thirion, M. Perrot, P. Pinel, J.-F. Mangin, and J.-B. Poline. Probabilistic anatomo-functional parcellation of the cortex: how many regions? In *11thProc. MICCAI, LNCS Springer Verlag*, New-York, USA, 2008.

[101] N. Vaever Hartvig and J. Jensen. Spatial mixture modeling of fMRI data. *Hum. Brain Mapp.*, 11(4):233–248, 2000.

[102] T. Vincent, L. Risser, and P. Ciuciu. Spatially adaptive mixture modeling for analysis of within-subject fMRI time series. *IEEE Trans. Med. Imag.*, 29(4):1059–1074, Apr. 2010.

[103] T. Vincent, L. Risser, P. Ciuciu, and J. Idier. Spatially unsupervised analysis of within-subject fMRI data using multiple extrapolations of 3D Ising field partition functions. In *2009 IEEE international workshop on Machine Learning for Signal Processing*, Grenoble, France, Sep. 2009.

[104] M. Woolrich and T. Behrens. Variational Bayes inference of spatial mixture models for segmentation. *IEEE Trans. Med. Imag.*, 25(10):1380–1391, Oct. 2006.

[105] M. Woolrich, T. Behrens, Ch. Beckmann, and S. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Trans. Med. Imag.*, 24(1):1–11, Jan. 2005.

[106] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith. Fully Bayesian spatio-temporal modelling of fMRI data. *IEEE Trans. Med. Imag.*, 23(2):213–231, Feb. 2004.

[107] M. Woolrich, M. Jenkinson, J. M. Brady, and S. Smith. Constrained linear basis set for HRF modelling using variational Bayes. *Neuroimage*, 21(4):1748–1761, 2004.

[108] K.J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A general statistical analysis for fMRI data. *Neuroimage*, 15(1):1–15, Jan. 2002.

 ISSN: 2102-6238