1

Min-max extrapolation scheme for fast estimation of 3D Potts field partition functions. Application to the joint detection-estimation of brain activity in fMRI

Laurent Risser[1,2,3,4], Thomas Vincent[1,2], Florence Forbes[5], Jérôme Idier[3] and Philippe Ciuciu[1,2]

Abstract

In this paper, we present a fast numerical scheme to estimate Partition Functions (PF) of symmetric Potts fields. Our strategy is first validated on 2D two-color Potts fields and then on 3D two- and three-color Potts fields. It is then applied to the joint detection-estimation of brain activity from functional Magnetic Resonance Imaging (fMRI) data, where the goal is to automatically recover activated, deactivated and inactivated brain regions and to estimate region-dependent hemodynamic filters. For any brain region, a specific 3D Potts field indeed embodies the spatial correlation over the hidden states of the voxels by modeling whether they are activated, deactivated or inactive. To make spatial regularization adaptive, the PFs of the Potts fields over all brain regions are computed prior to the brain activity estimation. Our approach is first based upon a classical path-sampling method to approximate a small subset of *reference* PFs corresponding to prespecified regions. Then, we propose an extrapolation method that allows us to approximate the PFs associated to the Potts fields defined over the remaining brain regions. In comparison with preexisting methods either based on a path-sampling strategy or mean-field approximations, our contribution strongly alleviates the computational cost and makes spatially adaptive regularization of whole brain fMRI datasets feasible. It is also robust against grid inhomogeneities and efficient irrespective of the topological configurations of the brain regions.

Index Terms

Markov Random Field; Potts fields; partition function; fMRI; Bayesian inference; MCMC; detection-estimation.

I. Introduction

In medical image analysis, one is often interested in recovering spatial structures. A simple but suboptimal approach to enhance signal-to-noise ratios (SNR) consists in filtering the datasets at the expense of a loss of spatial resolution. A more challenging approach works on the unsmoothed data by introducing some prior knowledge on the sought spatial structures. Spatial information is usually embedded in local interaction models such as Markov Random Fields (MRFs), which depend on a set of hyper-parameters. For instance, the temperature level controls the amount of spatial correlation in symmetric Potts models. In the considered fMRI application [1, 2], which aims to analyze 4D signals to jointly perform dynamics estimation and activation detection, the MRF definition is region-specific. Indeed, the Hemodynamic Filter (HF) modeling the impulse response of a brain activation in the fMRI signal, is assumed to be invariant only locally. Neuroimaging data is accordingly segmented into Γ functionally homogeneous irregular parcels [3], the order of Γ being several hundreds. This leads to a region-based analysis where Γ independent HFs have to be identified. Each model yields a HF shape estimate and spatial mixture models (SMM) are jointly expressed on the amplitude of the HF for every stimulus type to perform activation detection. SMMs in turn imply the involvement of discrete Potts fields to model spatial correlation. Therefore, several hundreds of temperature levels have to be estimated. Since optimal setting of such parameters may be different when considering different regions of the brain, all temperature levels cannot be fixed to the same value. Moreover, the hand-tuning of several hundreds of values is unachievable. In [4], an unsupervised and adaptive regularization scheme dedicated to two-class Potts fields and SMMs has been developed. Here, we extend this framework to L-class Potts fields with a direct application in functional brain imaging that aims at segregating activating, non-activating and deactivating voxels.

Section II is dedicated to the formulation of the partition function (PF) estimation problem for Potts fields. In Section III several techniques of PF estimation are presented. The main contribution of this paper lies in Section IV where a fast extrapolation technique to PF estimation of 3D Potts fields is proposed and validated both in the 2D and 3D context. The 2D situation offers the opportunity to provide a ground truth to the PF computation for the two-class Potts fields. The 3D extension is of particular importance since it matches our application setting. The application to the Joint Detection Estimation (JDE) of brain activity in fMRI is presented in Section V. Conclusions are drawn in Section VI.

[1] NeuroSpin/CEA, 91191 Gif-sur-Yvette, France
[2] IFR 49, Institut d'Imagerie Neurofonctionnelle, Paris, France
[3] IRCCyN/CNRS, 1 rue de la Noë, 44300 Nantes, France.
[4] Institute for Mathematical Sciences, Imperial College, SW7 London, UK.
[5] INRIA Rhônes-Alpes, MISTIS project, 665 av. de l'Europe, Montbonnot, 38334 Saint Ismier, France.
[1] firstname.lastname@cea.fr [3] firstname.lastname@irccyn.ec-nantes.fr [5] firstname.lastname@inrialpes.fr
The autors thank the Région Ile de France for funding.

II. Problem statement

Let us consider a grid characterized by a set of sites $\boldsymbol{s} = \{i = 1, \cdots, n\}$. A label $q_i \in \{1, , \ldots, L\}$ is associated to each site $i \in \boldsymbol{s}$ where L denotes the number of classes. A pair of adjacent sites i and j ($i \neq j$) is denoted by $i \sim j$ and is called a clique c. Equivalently, we denote the neighborhood of i as the set $N_i = \{j \in \boldsymbol{s}/i \sim j \text{ and } i \neq j\}$. The set of all cliques allows us to define an undirected graph denoted by $\mathcal{G}$. Let $\boldsymbol{q} = (q_1, q_2, \cdots, q_n) \in \{1, \ldots, L\}^n$ be the set of labels associated to $\boldsymbol{s}$. In what follows, we assume $\boldsymbol{q}$ to be distributed according to a symmetric Potts model:

$$\Pr(\boldsymbol{q}|\beta) = Z(\beta)^{-1} \exp\left(\beta U(\boldsymbol{q})\right), \tag{1}$$

where $U(\boldsymbol{q}) = \sum_{i\sim j} I(q_i = q_j)$ is the global "negative energy" and $I(A) = 1$ whenever A is true and 0 otherwise. The Gibbs distribution (1) defines a Markov random field (Hammersley-Clifford theorem) and thus satisfies the two properties:

$$\forall \boldsymbol{q}, \quad \Pr(q_i|q_{\boldsymbol{s}\setminus\{i\}}) = \Pr(q_i|q_j, j \in N_i) \quad \text{and} \quad \forall \boldsymbol{q}, \Pr(\boldsymbol{q}|\beta) > 0.$$

The inverse temperature $\beta \geq 0$ controls the amount of spatial correlation between the components of $\boldsymbol{q}$ according to $\mathcal{G}$. The partition function (PF) $Z(\beta)$ reads $\sum_{\boldsymbol{q}\in\{1,\cdots,L\}^n} \exp\left(\beta\, U(\boldsymbol{q})\right)$ and depends on the geometry of $\mathcal{G}$. Its exact evaluation in a reasonable amount of time is impossible except on tiny grids. Robust and fast estimation of $Z(\beta)$ is thus a key issue for numerous 3D medical imaging problems involving Potts models and more generally discrete MRFs. Due to our application, we applied the methods we present using 2 and 3-class Potts fields on 3D grids having a 6-connectivity system. The formula we present can however be used on general L-class Potts fields.

III. Partition function estimation

Several approaches have been designed to estimate a single PF [5–7]. Path-sampling [6] is an extension of importance sampling for estimating ratios of normalizing constants, by considering series of easy-to-sample unormalized intermediate densities. Such a strategy was proven efficient to tabulate the PF for 2-class Potts fields; see [8] for details. Algorithms with polynomial time complexity [7, 9] provide efficient alternatives to a single PF estimation. However, none of them is able to perform numerous PFs estimation for Potts fields of variable size and shape in a reasonable amount of time. Since several hundreds of such grids are manipulated in our fMRI application, fast estimation of multiple PFs is necessary. To this end, after presenting some PF estimation techniques in this section, we propose in section IV, a hybrid scheme which consists in resorting to path-sampling to get log-scale estimates $(\log \widehat{Z}_{\mathcal{G}_p}(\beta))_{p=1:P}$ in a small subset of *reference* graphs $(\mathcal{G}_p)_{p=1:P}$ out of the brain regions to be analyzed and then in using extrapolation formulas to obtain $\log \widetilde{Z}_{\mathcal{T}}(\beta)$ for the large remaining set of regions, each of them referenced by a *test* graph $\mathcal{T}$ for the sake of notational simplicity.

A. path-sampling method

The path-sampling method is based on an importance sampling strategy [6]. This kind of strategies consists in studying distributions of interest using samples generated from other well-chosen distributions, called importance distributions. In the context of L-class Potts fields, we can use path-sampling to estimate $Z(\beta)$ for β around β_0 using:

$$Z(\beta) \simeq Z(\beta_0) \frac{1}{M} \sum_{m=1}^{M} \frac{\exp\left(\beta U(\boldsymbol{q}_m)\right)}{\exp\left(\beta_0 U(\boldsymbol{q}_m)\right)}. \tag{2}$$

where $Z(\beta_0)$ is considered already known and the $\{\boldsymbol{q}_m, m = 1...M\}$ are distributed according to $\mathbb{P}(\boldsymbol{q} \,|\, \beta_0)$ and sampled using the Swendsen-Wang algorithm [10]. Note that, to avoid numerical overflow, eq. 2 is usually used in logarithmic scale. In what follow, log-PF refers to the logarithm of a partition function. The PF of a Potts field can then be estimated robustly on a dense grid of β values by estimating iteratively the values of $Z(\beta_k)$ using $Z(\beta_{k-1})$, where $\beta_k = k\Delta\beta, k \in \{0, \cdots, K\}$, the first value $Z(0)$ being explicitly given by $Z(0) = L^n$, where n is the number of sites in the field. More details are given in [8].

B. Partition function estimation using the Mean-field theory

For comparison purpose, we recall here the background theory of mean-field approximation, developed for the general case of non-symmetric Potts fields, into a set of equations for the estimation of partition functions of symmetric Potts models. For a thorough analysis of the general framework the reader may refer to [9]. In this context, we use the following notations. The *negative energy* of a symmetric Potts model is equivalently defined by: $U(\boldsymbol{q}) = U(\boldsymbol{w}) = \sum_{i\sim j} w_i^{\mathrm{t}} w_j$ where each w_i is a binary vector of length L having only one non-zero component corresponding to the value of q_i (*e.g.,* $q_i = l$ implies $w_{il} = 1$ and $w_{il'} = 0$ for $l' \neq l$). The *negative energy* can then be written as:

$$U(\boldsymbol{q}) = U(\boldsymbol{w}) = \frac{1}{2} \sum_{i=1}^{n} w_i^{\mathrm{t}} \sum_{j\in N_i} w_j. \tag{3}$$

Let $\overline{\omega} = \{\overline{\omega}_i, i \in \boldsymbol{s}\}$, where $\sum_l \overline{\omega}_{il} = 1$, be the mean field associated to variables $\boldsymbol{W} = \{W_i, i \in \boldsymbol{s}\}$ at a given inverse temperature β. The mean field approximation $\Pr{}^{\mathrm{mf}}(\boldsymbol{w})$ of the Gibbs distribution (1) is defined as:

$$\Pr{}^{\mathrm{mf}}(\boldsymbol{w} \,|\, \beta) = \prod_{i=1}^{n} \Pr_i{}^{\mathrm{mf}}(w_i \,|\, \beta) \tag{4}$$

where $\Pr_i{}^{\mathrm{mf}}(w_i \,|\, \beta) = \Pr(w_i \,|\, \beta, \overline{\omega}_{N_i})$ and $\Pr_i{}^{\mathrm{mf}}(w_i \,|\, \beta)$ denotes the conditional probability of w_i given $\omega_j = \overline{\omega}_j$ for all $j \in N_i$. It follows straightforwardly an expression of $\Pr{}^{\mathrm{mf}}$ as a Gibbs distribution of the form (1): $\Pr{}^{\mathrm{mf}}(\boldsymbol{w}|\beta) = \exp\left(\beta U^{\mathrm{mf}}(\boldsymbol{w})\right) / Z^{\mathrm{mf}}(\beta)$, where U^{mf} and Z^{mf} denote the energy and the partition function, respectively, and are easy to compute given the factorization property (4). Their expressions are given in Appendix C. Using symmetries for all $i = 1, \ldots, n$ and considering the $\overline{\omega} = \{\overline{\omega}_i, i \in s\}$ as the means of the variables $\boldsymbol{W} = \{W_i, i \in s\}$ in a L-class Potts model and at a given inverse temperature β , we can write:

$$\overline{\omega}_{il} = \frac{\exp(\beta \sum_{j \in N_i} \overline{\omega}_{jl})}{\sum_{l'} \exp(\beta \sum_{j \in N_i} \overline{\omega}_{jl'})}, \qquad \forall i \in 1, \cdots, n \text{ and } \forall l \in 1, \cdots, L \tag{5}$$

The mean field approximation consists in solving this fixed point equation iteratively. Note that this equation can be solved sequentially over the sites. After convergence we use the solution $\overline{\omega}$ to define (4). Interestingly, as developed in [9] and shown in the equations (17) and (20) of Appendix C, the mean field allows us to estimate the PF:

$$Z^{\mathrm{mf}} = \prod_{i=1}^{n} \sum_{l=1}^{L} \exp(\beta \sum_{j \in N_i} \overline{\omega}_{jl}), \tag{6}$$

It can be shown (see Appendix C) that a better approximation is given by:

$$Z^{\mathrm{GBF}} = Z^{\mathrm{mf}} \exp(-\frac{\beta}{2} \sum_{i=1}^{n} \sum_{l=1}^{L} \overline{\omega}_{il} \sum_{j \in N_i} \overline{\omega}_{jl}), \tag{7}$$

The corresponding log-PF estimates $\log Z^{\mathrm{mf}}$ and $\log Z^{\mathrm{GBF}}$ are defined by directly taking the log in the above equations.

C. Linear/bilinear regression schemes for Potts fields

In [8], the authors have proposed a linear regression procedure to estimate quickly the log-PFs of 2-class Potts models for *test* grids, each of them being refered by $\mathcal{T}$. This approach consists in estimating, first, the log-PFs of P *reference* Potts fields defined on the grids $\mathcal{G}_p$ using a robust approach such as the path-sampling. The log-PFs are computed for values of β: $\beta_k = k\Delta\beta$. A linear regressor $\widehat{a_k}$ of the log-PF as a function of the number of cliques in $\mathcal{T}$ is then computed, for every regularization level β_k. Estimates of $\log \widetilde{Z}_{\mathcal{T}}(\beta)$ are then linearly computed using the estimated regression coefficients and the number of cliques in $\mathcal{T}$: $\log \widetilde{Z}_{\mathcal{T}}(\beta_k) = \widehat{a}_k c_{\mathcal{T}}$ at each β_k regularization level. A bilinear extension of this technique, which takes both the number of cliques $c_{\mathcal{G}_p}$ and sites $s_{\mathcal{G}_p}$ in the grid into account, has been developed in [11]. The regression coefficients ($\widehat{a}_k$, $\widehat{b}_k$ and $\widehat{c}_k$) are obtained by minimizing the least square criterion $\sum_{p=1}^{P} \| \log \widehat{Z}_{\mathcal{G}_p}(\beta_k) - a_k c_{\mathcal{G}_p} - b_k s_{\mathcal{G}_p} - c_k \|^2$. Then, β_k-dependent *bilinear* extrapolation formula applied to any 3D Potts field defined on a test grid $\mathcal{T}$: $\log \widetilde{Z}_{\mathcal{T}}(\beta_k) = \widehat{a}_k c_{\mathcal{T}} + \widehat{b}_k s_{\mathcal{T}} + \widehat{c}_k$. Note that the log-PF estimates for β values outside the β-grid are obtained using linear interpolation between its two closest values on the β-grid. This extension was shown efficient to provide accurate extrapolation results for *small and irregular graphs*[1] such as those appearing in our fMRI application. Indeed, as illustrated in Fig. 1, at a fixed number of cliques, the higher the number of sites, the higher the log-PF whatever the regularity of the reference graphs. This dependence of $\log Z_{\mathcal{G}_p}$ on $s_{\mathcal{G}_p}$ at constant values of $c_{\mathcal{G}_p}$ becomes much more important at small β values. Hence, our bilinear extension of [8] significantly improved estimation performance for small and irregular grids. However, the accuracy of linear/bilinear log-PF extrapolations strongly depends on the homogeneity and the number of *reference* grids: the less homogeneous the reference set, the larger the approximation error. These reasons motivate the development of a more reliable and versatile approach.

IV. Fast and robust extrapolation technique

A. Method

Our extrapolation technique proceeds in two steps: 1) Akin to [8], reference log-PFs $\log \widehat{Z}_{\mathcal{G}_p}(\beta_k)$ are estimated using path-sampling. The topological configurations of the reference grids $(\mathcal{G}_p)_{p=1:P}$ can be inhomogeneous to cover a maximum of situations that may occur when dealing with a brain parcellation. 2) For any test grid $\mathcal{T}$, the quantity $\log Z_{\mathcal{T}}$ is approximated

[1]Here, irregular refers to equally spaced Cartesian graphs of non-parallepipedic shape.

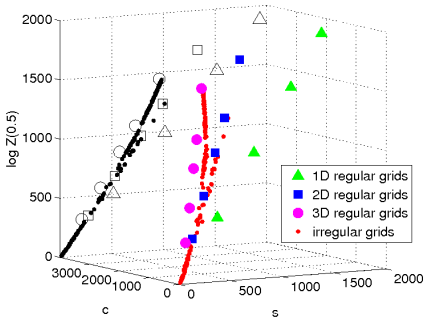

Fig. 1. Examples of log-PF values for $\beta = 0.5$ as functions of the number of cliques c and sites s. The projection onto a plane $c = \textit{constant}$ is depicted to show the dependence on s.

from a *single* reference log-PF estimated out of $(\log \widehat{Z}_{\mathcal{G}_p}(\beta))_{p=1:P}$ selected by an appropriate criterion. Let n_i be the number of neighbors for the site s_i of $\mathcal{T}$. We then define $r_{\mathcal{T}} = \sigma_{n,\mathcal{T}}/\mu_{n,\mathcal{T}}$ as a measure of grid homogeneity where $\mu_{n,\mathcal{T}}$ is the mean of the s_i over $\mathcal{T}$ and $\sigma_{n,\mathcal{T}}$ is the corresponding standard deviation: the smaller $r_{\mathcal{T}}$ the more regular $\mathcal{T}$. Our topological similarity measure, given by $\mathcal{L}_{\mathcal{T}}(\mathcal{G}_p) = \|r_{\mathcal{T}} - r_{\mathcal{G}_p}\|^2$, helps us choosing the closest reference grid $\mathcal{G}_{\text{ref}}$ to $\mathcal{T}$ in combination with the approximation error criterion $\mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ defined by:

$$\mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p) = \|\log Z_{\mathcal{T}}(\beta) - \log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p)\|^2 / \|\log Z_{\mathcal{T}}(\beta)\|^2$$

$$\text{with} \quad \log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p) = \Big(\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}(\log \widehat{Z}_{\mathcal{G}_p}(\beta) - \log L) + \log L\Big), \tag{8}$$

where $(c_{\mathcal{T}}, c_{\mathcal{G}_p})$ and $(n_{\mathcal{T}}, n_{\mathcal{G}_p})$ are the number of cliques and sites of the L-color Potts fields defined over $\mathcal{T}$ and $\mathcal{G}_p$, respectively. The reference grid $\mathcal{G}_{\text{ref}}$ is exhibited using a min-max principle, which consists in minimizing the maximal approximation error $\mathcal{A}(\beta, \mathcal{G}_p)$ with respect to all reference grids $(\mathcal{G}_p)_{p=1:P}$. In Appendix B, it is shown that $\mathcal{A}(0, \mathcal{G}_p) = \max_\beta \mathcal{A}(\beta, \mathcal{G}_p), \forall \mathcal{G}_p$, whenever the grid homogeneity in $\mathcal{G}_p$ and $\mathcal{T}$ is similar. Hence, we get:

$$\mathcal{G}_{\text{ref}} = \underset{(\mathcal{G}_p)_{p=1:P}}{\arg\min}\, \mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) \quad \text{subject to} \quad \mathcal{L}_{\mathcal{T}}(\mathcal{G}_p) \leq \epsilon \tag{9}$$

$$\text{and } \mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) \stackrel{\Delta}{=} \|(n_{\mathcal{T}} - 1) - c_{\mathcal{T}}(n_{\mathcal{G}_p} - 1)/c_{\mathcal{G}_p}\|^2 / n_{\mathcal{T}}^2 \tag{10}$$

where $\epsilon > 0$ is a positive threshold fixed by hand. Once $\mathcal{G}_{\text{ref}}$ has been identified, the log-PF estimate in $\mathcal{T}$ is thus given by $\log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_{\text{ref}})$ according to Eq. (8). Our extrapolation formula (8) is interestingly built up according to two principles: *i.)* an unbiased asymptotic approximation error[2] and *ii.)* an exact approximation of $(\log Z_{\mathcal{T}}(\beta))'$ for $\beta \to 0^+$. These principles are summarized in Appendix A. The choice of $\mathcal{G}_{\text{ref}}$ can then be seen as the choice of the grid, out of $\mathcal{G}_p$, for which the approximation error at $\beta = 0$ and β close to the phase change will be minimized using Eq. (8). The method is illustrated in Fig. 2 with $P = 4$ reference log-PFs of 3-class Potts fields and one test field. It appears that $\log \widehat{Z}_{\text{ref}}$ is the closest curve above the ground truth $\log \widehat{Z}_{\mathcal{T}}$ (in red) and that our log-PF estimate $\log \widetilde{Z}_{\mathcal{T}}$ represented by crosses (+) is superimposed on the path-sampled curve.

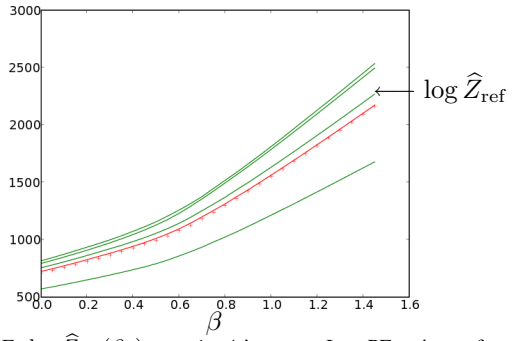

Fig. 2. Path-sampled estimates of the *reference* log-PFs $\log \widehat{Z}_{\mathcal{G}_p}(\beta_k), p = 1:4$ in green. Log-PF estimate found by path-sampling for the *test* field $\log \widehat{Z}_{\mathcal{T}}$ in red (ground truth). Our extrapolation method based on Eq. (8) provides a log-PF estimate $\log \widetilde{Z}_{\mathcal{T}}$ represented by crosses-line (+) and superimposed with the ground truth.

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

B. Assessment of the method

1) Binary 2D fields: We first validate our approach in a situation where the log-PF admits a closed form expression. This situation exists thanks to the Onsager's contribution [12], who derived a closed form expression for the log-PF of the 2-class Potts model on any 2D square grid under cyclic boundary constraints:

$$\log Z(\beta) = n(\beta + \log\left[2\cosh\beta\right] + \psi\left[u(\beta)\right]) \tag{11}$$

where $u(\beta) = 2\sinh\beta/\cosh^2\beta$ and the ψ function is a one dimensional log-elliptic integral:

$$\psi(u) = \frac{1}{2\pi}\int_0^{\pi} \log\left[\left(1+\sqrt{1-u^2\sin^2 x}\right)/2\right]\, dx \tag{12}$$

Therefore, the huge summation in Eq. (1) is equivalent to this far simpler one dimensional equation. Note that Onsager formulation depends on the graph through the number of sites n. This suggests than another possible way to compute the log-PF is to determine for a small grid, say the 3 x 3 grid, the log-PF $\log Z_9(\beta)$ and then set, for all $N > 9$, $\log Z(\beta)$ to $\log Z(\beta) = N/9 \log Z_9(\beta)$. This latter log-PF can be computed exactly by enumerating all possible configurations and is provided in [9]. It follows that the log-PF of a square 2-class Potts fields having cyclic boundary constraints can be written as:

$$\log Z(\beta) = \frac{n}{9}\log\left(102\exp\left(6\beta\right) + 144\exp\left(8\beta\right) + 198\exp\left(10\beta\right) + 48\exp\left(12\beta\right) + 18\exp\left(14\beta\right) + 2\exp\left(18\beta\right)\right) \tag{13}$$

This expression presents the advantage to be much easier to estimate than the formulation of Onsager and provides a good approximation to the log-PF when comparing to Onsager formula

In Fig. 3, we compare the exact calculation provided by Eq. (11) with the path-sampling and extrapolation approaches on a 2-class Potts field defined over a 30x30 regular grid having cyclic boundary constraints. We also compare the log-PF estimates obtained using the mean field and GBF approximations. It is shown that our extrapolation technique (red) is as accurate as possible since our estimate is superimposed on the estimate obtained using the path-sampling, considered as the ground truth (in blue). Moreover, it appears that these numerical approaches are very close to the accurate log-PFs given by Eqs (11) and (13). One can note that the log-PFs estimation obtained Eq. 13 is very close to the one obtained using Onsager's approach but is not equal. This could be due to approximation errors in Onsager's formula in very small fields. As explained in [9], the mean field approximation of the log-PF is not accurate, but its GBF generalization is below and close to the ground truth. Small oscillations of this latter approximation for large values of β are due to mean field approximation having not completely converged.

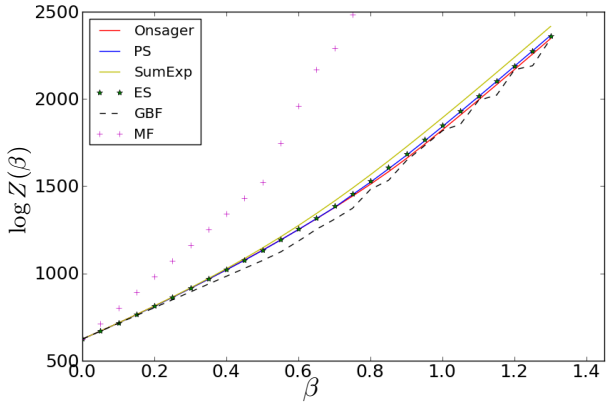

Fig. 3. Red: True $\log Z(\beta)$ computed by Eqs. (11)-(12) for a 2D 2-class Potts field defined over a 30x30 cyclic regular grid. Blue: path-sampling estimate *i.e.,* $\log \widehat{Z}(\beta)$. Yellow: log-PF obtained using the sum of exponentials (Eq. 13). Crosses ($\times$): Extrapolation estimates $\log \widetilde{Z}(\beta)$ from a reference set made up by 250 grids of various size and shape. Dashed line: log-PF obtained using the GBF approximation of Eq. 7. Crosses ($+$): log-PF obtained using the mean field approximation of Eq. 6.

2) 2- and 3-class 3D Potts fields: For validation purpose, we compared in [1] the log-PF estimates of 2-class Potts fields computed using our extrapolation techniques with those obtained using path-sampling, considered as the ground truth. This clearly shows that the bilinear and extrapolation methods outperform the linear one and that the extrapolation technique allows to mix successfully a much larger range of *reference* grids than linear and bilinear techniques. The extrapolation techniques provided indeed accurate log-PF approximations for a large range of grids by always using the same *reference* grids. Here, we show in table I, that using the same experimental protocol, the extrapolation technique also provides accurate log-PF approximations for 3-class Potts fields such as those used in our application.

Reference and *test* graphs are either regular or irregular. A total of 15 regular and large (more than 10^3 sites) *reference* graphs are considered with cubic, planar and curvilinear shapes. Irregular grids were extracted from regular bounding boxes in

which Potts field configurations were drawn using the β dependent Swendsen-Wang algorithm [13]. In each bounding box, we considered the largest connected component of sites having the same label as an irregular graph so that, the larger β the more regular and larger the created graphs. *Irregular reference* graphs were then computed using 170 bounding boxes of increasing size (from 10^3 to 15^3 sites) and regularization levels β within the range $[0.2, 0.7]$. *Regular test* graphs are divided into three subsets: 30 of them are *small* (less than 10^3 sites), 30 are *medium* size (between 10^3 and 15^3 sites) and 30 are *large* (more than 15^3 sites). *Irregular test* graphs are also subdivided into three subsets. Each of them contains 30 graphs obtained from bounding boxes of 16^3 sites, for $\beta = 0.2$, 0.4 and 0.5, respectively. For each graph $\mathcal{T}$, the difference between our approximation and the PS estimate $|\log \widetilde{Z}_{\mathcal{T}}(\beta) - \log Z_{PS}(\beta)|/\log Z_{PS}(\beta)$ is computed for each value of β. The maximum of these differences is considered and the average of such maxima is given in Table I using percentages.

TABLE I

MEAN MAXIMAL RELATIVE APPROXIMATION ERROR OVER REGULAR AND IRREGULAR TEST GRAPHS USING THE EXTRAPOLATION TECHNIQUE ON 2- AND 3-CLASS POTTS FIELDS. ERRORS ARE GIVEN IN PERCENTS.

Test grid		2-class	3-class
regular	small	0.639	2.76
	medium	2.77	2.80
	large	3.68	3.70
irregular	$\beta = 0.2$	0.375	0.633
	$\beta = 0.3$	0.281	0.959
	$\beta = 0.4$	0.621	0.747
	$\beta = 0.5$	0.693	1.80

Both for 2- and 3-class Potts fields the log-PF approximations are accurate. They are however slightly better for 2-class Potts fields than 3-class Potts fields in general. Note that in both cases, these results can be still improved by adding *reference* grids very close to the *test* grids in the database of *reference* grids. Similarly if no *reference* grid of the database is adapted to a *test* grid, this can be detected immediately by a large approximation error at $\beta = 0$ or too different grid homogeneities. Note that we also computed the log-PF estimates on the test grids using GBF approximations. In comparison with the ground truth, we obtained generaly 6% and 10% of maximum relative approximation error for 2- and 3-class Potts fields respectively. These maximum approximation errors were mostly observed for β values around the phase change, as presented in Fig.3.

C. A Monte Carlo study to hyper-parameter estimation

The last validation we examined addresses the estimation of the inverse temperature level (*i.e.,* β-estimation) in the Maximum Likelihood (ML) sense either from our log-PF estimate or from its path-sampled counterpart. This study has been conducted directly on *observed* 2D 2-class Potts fields. At each temperature level $\beta_k = k\Delta\beta$ with $\Delta\beta = 0.1$, we generated independently 100 2D Potts fields defined over the same grid. We tested different grid sizes (from 10^2 to 50^2) and showed that the number of voxels only influences the error bars on the β estimate. For a Potts field defined by Eq. (1), the ML estimate $\widehat{\beta}^{\text{ML}}$ is given by $\widehat{\beta}^{\text{ML}} = \arg\max_{\beta} [\beta U(\boldsymbol{q}) - \log Z(\beta)]$. In Fig. 4, we compared two ML estimators corresponding to the path-sampling and extrapolation method for estimating the log-PFs. As illustrated in Fig. 4, our extrapolation technique (red curve) retrieves the true regularization parameter for $\beta < 0.7$. For $0.7 < \beta < 1$, a very small bias is observed while for larger values, a more significant error occurs in comparison to a more precise path-sampling scheme (blue curve).

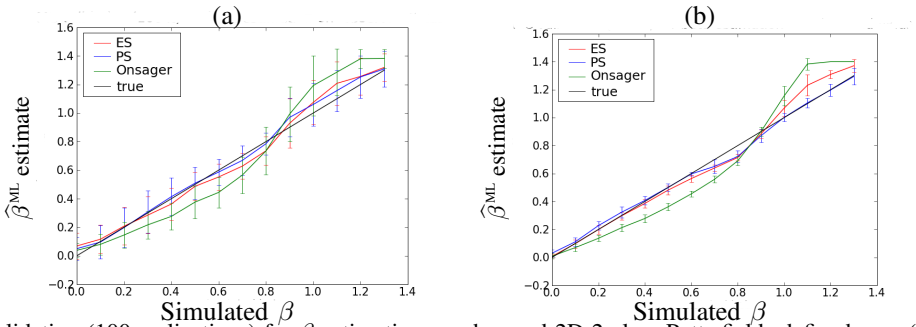

Fig. 4. Monte Carlo validation (100 realizations) for β-estimation on observed 2D 2-class Potts fields defined over (a) 10x10 and (b) 50x50 grids. Black: the ground truth is given by the first bisector. Red: ML estimate $\widehat{\beta}^{\text{ML}}$ relying on our log-PF extrapolation method. Blue: ML path-sampled estimate $\widehat{\beta}^{\text{ML}}$.

V. Application to fMRI data analysis

A. Problem statement

1) Parcel-based BOLD signal modeling: Our extrapolation algorithm was applied to the spatially adaptative regularization of the region-based Joint Detection-Estimation (JDE) of brain activity introduced in [14, 15]. The JDE approach relies on a prior parcellation of the brain into $\boldsymbol{\mathcal{P}} = (\mathcal{P}_\gamma)_{\gamma=1:\Gamma}$ functionally homogeneous and connected parcels [3] illustrated in Fig. 5. Every parcel $\mathcal{P}_\gamma$ comprising voxels $(V_j)_{j=1:J}$ is characterized by a single hemodynamic filter $\boldsymbol{h}_\gamma$. Within a given parcel $\mathcal{P}_\gamma$, voxel-dependent and stimulus-related fluctuations of the BOLD signal magnitude are encoded by $\mathbb{a} = (a_j^m)_{j=1:J,m=1:M}$, the response levels ($m$ stands for the stimulus type index). The fMRI time course measured in voxel V_j then reads:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{x}^m \star \boldsymbol{h}_\gamma + \boldsymbol{b}_j,$$

where $\boldsymbol{x}^m$ stands for the m^{th} binary stimuli vector and $\boldsymbol{b}_j$ stands for the noise component [14]. Within the Bayesian framework, prior probability density functions (pdfs) are introduced on $(\mathbb{a}, \boldsymbol{h})$ [14].

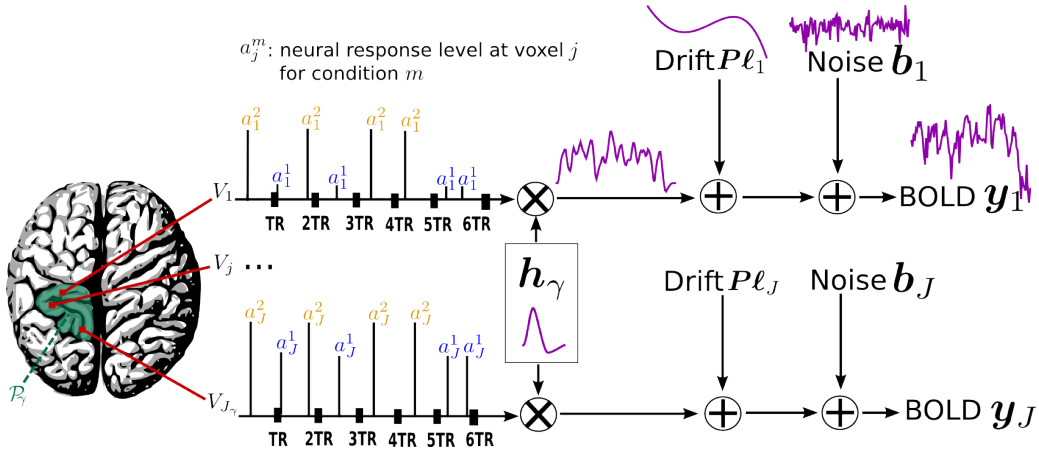

Fig. 5. **Parcel-based regional BOLD model**. The size of each parcel $\mathcal{P}_\gamma$ varies typically between by a few tens and a few hundreds of voxels: $80 \leqslant J_\gamma \leqslant 350$. The number M of experimental conditions involved in the model usually varies from 1 to 5. In our example, $M = 2$, 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. Note that our model accounts for asynchronous paradigms in which the onsets do not necessarily match acquisition time points. As illustrated, the NRLs take different values 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 ($Q = 4$).

2) Bayesian inference: The Bayesian approach developed in [15] introduces proper priors on the unknown parameters $(\boldsymbol{h}_\gamma, \mathbb{a})$ in order to recover a robust estimate of brain activity (localization and activation profile). Akin to [15, 16], the prior density for the HRF remains Gaussian, $\boldsymbol{h}_\gamma \sim \mathcal{N}(\boldsymbol{0}, v_{\boldsymbol{h}}\boldsymbol{R})$ with $\boldsymbol{R} = (\boldsymbol{D}_2^{\text{t}}\boldsymbol{D}_2)^{-1}$, which allows us to estimate a smooth HRF shape since $\boldsymbol{D}_2$ is the second-order finite difference matrix penalizing therefore abrupt changes. Moreover, the extreme time points of the HRF can be constrained to zero if necessary [16].

Regarding the NRLs $\mathbb{a}$, according to the maximum entropy principle we assume that different types of stimulus induce statistically independent NRLs *i.e.,* $p(\boldsymbol{A} \,|\, \boldsymbol{\theta}_{\boldsymbol{a}}) = \prod_m p(\boldsymbol{a}^m \,|\, \boldsymbol{\theta}^m)$ with $\boldsymbol{\theta}_{\boldsymbol{a}} = (\boldsymbol{\theta}^m)_{m=1:M}$. Vector $\boldsymbol{\theta}^m$ denotes the set of unknown hyperparameters related to the mth stimulus type. However, for certain classes of paradigms (*e.g.,*, attention and motor dual tasks; priming effects, ...) this assumption may potentially be unlikely. In such cases, a between-condition prior covariance matrix could be introduced, the difficulty lying in the choice of a relevant and know correlation model to limit the computational complexity.

Let q_j^m be the allocation variable that states whether voxel V_j is activating ($q_j^m = 1$), deactivating ($q_j^m = -1$) or non-activating ($q_j^m = 0$) in response to stimulus m. The NRLs still remains independent conditionally upon $\boldsymbol{q}^m$. This means that $p(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m, \boldsymbol{\theta}^m) = \prod_j p(a_j^m \,|\, q_j^m, \boldsymbol{\theta}^m)$ for every condition m. Spatial mixture models (SMM) are introduced to favor the recovery of activating and deactivating clusters. In this case, the marginal density $p(\boldsymbol{a}_m)$ does not factorize over voxels and reads:

$$p(\boldsymbol{a}^m \,|\, \boldsymbol{\theta}^m) = \sum_{\boldsymbol{q}^m} \Big[\prod_{j=1}^{J_\gamma} f(a_j^m \,|\, q_j^m, \boldsymbol{\theta}^m)\Big] \Pr(\boldsymbol{q}^m \,|\, \boldsymbol{\theta}^m). \tag{14}$$

Spatial correlation is directly incorporated in the probabilities of activation through a *hidden* three-class Potts MRF $\Pr(\boldsymbol{q}^m \,|\, \boldsymbol{\theta}^m)$, as already done in image analysis [17, 18] or in neuroimaging [19, 20]. For mathematical convenience, Gaussian densities are considered for modeling the conditional distribution of the NRLs: $f(a_j^m \,|\, q_j^m = i) \sim \mathcal{N}(\mu_{i,m}, v_{i,m})$. Parameters $\mu_{i,m}$ and $v_{i,m}$

define the prior mean and variance of class $i = 0, \pm 1$, respectively for the stimulus type m. The set $\boldsymbol{\theta}_m$ comprises four prior mixture parameters $\boldsymbol{\theta}_m = \{\mu_{0,m}, \mu_{\pm 1,m}, v_{0,m}, v_{\pm 1,m}, \beta_m\}$.

Samples of the full posterior pdf $p(\boldsymbol{h}, \mathbb{a}, \mathbb{q}, \boldsymbol{\Theta} \,|\, \mathbb{y})$ are simulated using a Gibbs sampler algorithm and posterior mean estimates are then computed from these samples. Here, we introduce the sampling of parameter β_m, which is achieved using a *symmetric* random walk Metropolis-Hasting step: At iteration k, a candidate $\beta_m^{(k+1/2)} \sim \mathcal{N}(\beta_m^{(k)}, \sigma_\epsilon^2)$ is generated and is accepted (*i.e.,* $\beta_m^{(k+1)} = \beta_m^{(k+1/2)}$) with probability: $\alpha(\beta_m^{(k)} \rightarrow \beta_m^{(k+1/2)}) = \min(1, A_{k,k+1/2})$, where the acceptation ratio $A_{k,k+1/2}$ follows from Eq. (1):

$$A_{k,k+1/2} = \frac{p(\beta_m^{(k+1/2)}|\boldsymbol{q}_m^{(k)})}{p(\beta_m^{(k)}|\boldsymbol{q}_m^{(k)})} = \frac{p(\boldsymbol{q}_m^{(k)}|\beta_m^{(k+1/2)})p(\beta_m^{(k+1/2)})}{p(\boldsymbol{q}_m^{(k)}|\beta_m^{(k)})p(\beta_m^{(k)})}$$
$$= \frac{Z(\beta_m^{(k)})}{Z(\beta_m^{(k+1/2)})} \exp\left((\beta_m^{(k+1/2)} - \beta_m^{(k)})U(\boldsymbol{q}_m^{(k)})\right),$$

using Bayes' rule and considering a uniform prior for β_m. The β_m sampling step then requires to estimate ratios of $Z(.)$ or log-PF differences for all $\mathcal{P}_\gamma$ parcels prior to exploring the full posterior pdf.

B. Results on synthetic fMRI data

In the context of the joint detection-estimation of brain activity, here we compare the supervised SMM (SSMM) to its unsupervised extension (USMM). Our synthetic fMRI time series have been generated at low signal-to-noise ratio considering true activation maps that do not derive from the Potts model . The underlying paradigm consisted of two stimulus types ($M = 2$) whose activation patterns are shown in Fig. 6(a)-(d), respectively.

Fig. 6(b)-(e) illustrates that a wrong choice of β-value ($\beta = 0.2$) in the supervised case (SSMM) induces a misspecification between the activating and deactivating voxels for the first experimental condition ($m = 1$) and between deactivating and non-activating voxels as the background is almost classified into deactivating and the non-activating class is almost empty. For $m = 2$, the situation is better but still noisy under SSMM. The situation is properly regularized by resorting to USMM: Fig. 6(c-f) yield estimated labels matching exactly the true ones for $m = 1, 2$.

These results are enforced by the estimates of the prior mixture components shown in Fig. 7 for the two conditions. In the SSMM case, we observed a degeneracy regarding the activating class for $m = 1$ (Fig. 7(a)), since the three Gaussian densities are superimposed. For $m = 2$, the mixture parameters of the SSMM are also problematic while less degenerated. Since we obtained $\widehat{\mu}_{-1}^2 \approx 0 = \mu_0^2$, this directly impacts the posterior classification towards the presence of false negatives. On the other hand, Fig. 7(b)-(d) illustrate a better distinction between the three components in the prior mixtures whatever the condition at hand ($m = 1, 2$). For $m = 1, 2$, we noticed that the distributions do not overlap (see Fig. 7(b)) with the USMM setting. Hence, we found an exact posterior classification.

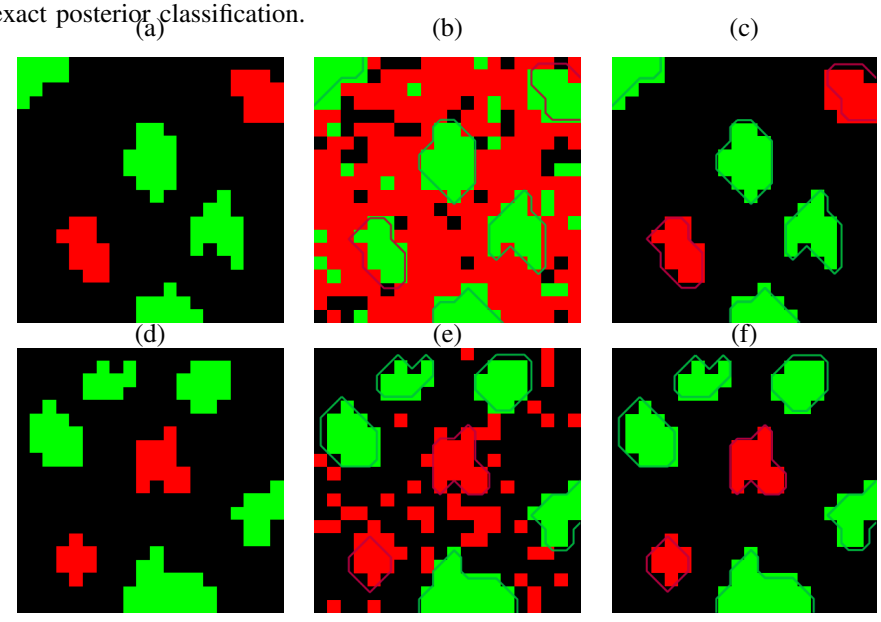

Fig. 6. Estimated labels in the Maximum A Posteriori sense. **Top row:** $(\widehat{\boldsymbol{q}}^1)^{\text{MAP}}$. **Bottom row:** $(\widehat{\boldsymbol{q}}^2)^{\text{MAP}}$. **(a)-(c)**: Label maps obtained using the supervised SMM (SSMM) for $m = 1$ and $m = 2$, respectively with $\beta^{(1)} = \beta^{(2)} = 0.2$. **(b)-(d)**: Label maps obtained using the unsupervised SMM (USMM) approach. Deactivating, non-activating and activating voxels are color coded in red, blue and green, respectively.

Fig. 8 illustrates the behaviour of the SSMM when varying the fixed β value. A wrong tuning yields a significant decrease in the right classification rate of non-activating (blue) and deactivating labels (red) for $\beta \leqslant 0.8$ et $\beta \geqslant 1.2$ whereas an optimal

tuning of β lies in the range $[0.8 - 1.5]$. The unsupervised version provides an estimate $\widehat{\beta} = 0.92$ (see $\diamond$) which is in this optimal range and yields optimal classification rates.

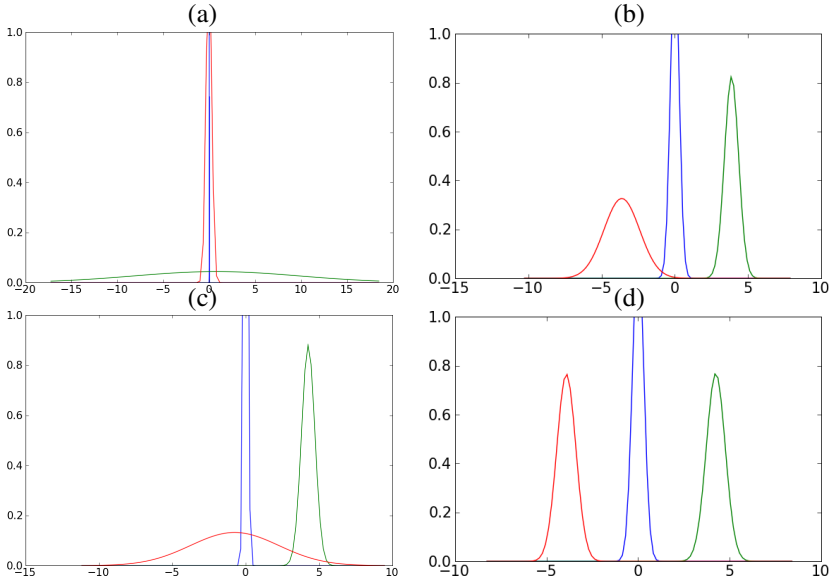

Fig. 7. Estimated components of the three-class Gaussian prior mixture. **Top row:** first experimental condition ($m = 1$). **Bottom row:** second experimental condition ($m = 2$). **(a)-(c)**: Mixture components yielded by the SSMM approach for $m = 1$ and $m = 2$, with $\beta^{(1)} = \beta^{(2)} = 0.2$; **(b)-(d)**: Mixture components given by the USMM approach. Deactivating, non-activating and activating voxels are color coded in red, blue and green, respectively.

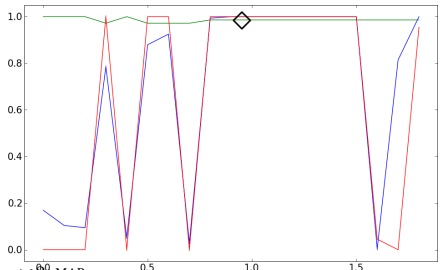

Fig. 8. Rate of right classification of $(\widehat{\boldsymbol{q}}^1)^{\text{MAP}}$ and $(\widehat{\boldsymbol{q}}^2)^{\text{MAP}}$ in the supervised case for different fixed β-values. Activating, non-activating and deacctivating classes are color coded in red, blue and green, respectively.

C. Results on real fMRI data

We applied the JDE procedure to real fMRI data recorded during an experiment designed to map auditory, visual and motor brain functions, which 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, declined in 10 experimental conditions (auditory phrase, visual phrase, left auditory or visual clic ...).

We compare three versions of the JDE procedure: Independent Mixture Models (IMM), Supervised SMM (SSMM, $\beta = 0.8$) and unsupervised SMM (USMM), in order to assess the impact of the adaptive spatial correlation model. We also assess the behaviour of a 3-color compared to a 2-color Potts prior on labels. Fig. 9 shows normalized contrast maps ($\widehat{\boldsymbol{a}}^{\text{LAC}} - \widehat{\boldsymbol{a}}^{\text{RAC}}$) of auditory induced left versus right clic (LAC vs. RAC). As expected, the activations lie in the contralateral right motor cortex. Here, only USMM is more sensitive illustrating thus the advantage of an *adaptive* spatial correlation model. The same constrat maps are observed whether we resort to a 2-color or 3-color prior Potts field. Indeed, we do not expect any deactivation arising from these data. The purpose is only illustrative so as to show that our spatially adaptive approach may generalize from a 2-color to a 3-color Potts prior.

For the 2-color prior, $\widehat{\beta}^{\text{PM}}$ estimates for the left auditory clic was 0.56 so that the supervised setting of SSMM with $\beta = 0.8, L = 2$ leads to too much correlation and less sensitive results. Interestingly, still for the 2-color prior, Fig. 9 also depicts the parcel-dependent maps of the PM $\widehat{\beta}$ estimates for the RAC and LAC experimental conditions. The gain in sensitivity in the USMM contrast map ($\widehat{\boldsymbol{a}}^{\text{LAC}} - \widehat{\boldsymbol{a}}^{\text{RAC}}$) may be explained by a difference in the amount of spatial regularization introduced between the two conditions involved in the contrast. A lower regularization level is estimated ($\widehat{\beta}^{\text{LAC}} \approx 0.5$ vs. $\widehat{\beta}^{\text{RAC}} \approx 0.75$) in parcels located in the right motor cortex since the BOLD signal is stronger for the LAC than for the RAC condition in

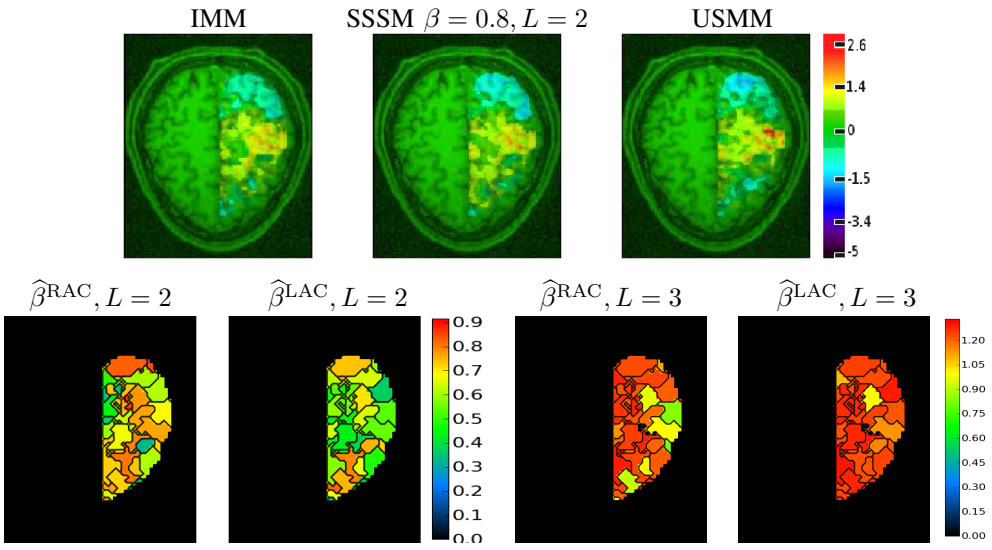

Fig. 9. Comparison of the IMM, SSMM and USMM models wrt the estimated normalized constrat maps: left auditory clic (LAC) versus right auditory clic (RAC): $\widehat{\boldsymbol{a}}^{\mathrm{LAC}} - \widehat{\boldsymbol{a}}^{\mathrm{RAC}}$. $\widehat{\beta}$ parcel-dependent maps computed for the LAC and RAC conditions and for 2-color ($L = 2$) and 3-color $L = 3$ Potts fields.

these regions. For the 3-color Potts prior the interpretation of $\widehat{\beta}^{\mathrm{PM}}$ estimates is more delicate as the regularization impacts 3 states. Still we can notice that USMM-$L = 3$ is able to spatially adapt the amount of spatial regularization.

On these real fMRI data, our extrapolation scheme provides log-PFs estimate for a brain parcellation $(\mathcal{P}_\gamma)_{\gamma=1:300}$ and $(\mathcal{G}_p)_{p=1:50}$ reference grids. In terms of computational complexity, these log-PF estimates were computed in about ten seconds, a very appealing approach in comparison to path-sampling, which requires about one hour for estimating all log-PF estimates for a negligable gain in accuracy (less than 3%). Finally, we did not observe any significant difference between the USSM effect maps derived using path-sampling and our extrapolation scheme (results not shown).

VI. Conclusion

In this work, we extended a joint detection-estimation of brain activity framework which enables the processing of *unsmoothed* fMRI data, described in [1, 2]. The latter approach considers two possible states at each voxel: activating or non-activating. Here, we generalize the model to consider a third state: de-activation, in order to take into account putativve negative BOLD effects. The spatial regularization that was performed by using 2-colors Potts fields then requires 3-colors Potts fields. In order to make spatially adaptive regularization feasible, the joint detection-estimation technique requires a reliable and fast estimation of 3D Potts field partition function. The problem of joint detection-estimation is indeed solved independently in a large number of pre-defined brain regions of different shape, each region requiring the estimation of a partition function, referred as *test* partition function. Therefore, we extended the extrapolation algorithm dedicated to the approximation of partition functions of 2-colors Potts fields to 3-colors Potts fields. This extension is the heart of the paper. The extrapolation algorithm exploits a set of *reference* log-PF estimates on *reference* grids, robustly pre-computed using path-sampling. Interestingly, the maximum approximation error of every *test* log-PF is controlled using the extrapolation technique so that a robust estimation can be performed whenever no suitable *reference* candidate is found in the set of *reference* log-PFs. Obviously, efficiency is conditionned by the number of *reference* grids, and more importantly by their similarity to the topologies encountered in the conducted analysis. The technique were shown accurate on 2D and 3D grids as well as regular or non-regular grids. It were also compared with a PF estimation technique based on the mean field theory [9]. In general, for the estimation of the log-PFs on grids of a similar type (2D, 3D, regular, irregular, small, large,...), about ten problem-specific reference grids are enough to provide good log-PF estimates. In the end, using our fast extrapolation technique, the computational burden remains acceptable since a whole brain data analysis at the subject level takes about 1h30. The application to real fMRI data showed a gain in statistical sensitivity for the unsupervised version whether we resort to 3-color or 2-color Potts priors. In order to test their reproducibility, the promising subject-level results have to be confirmed in group studies. The technique will also be performed on epileptic fMRI datasets, for which we expect de-activations. This will emphasize and assess the usefullness of the 3-color Potts prior model compared to the 2-color prior.

References

[1] 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 *12thProc. MICCAI'09*, G.-Z. Yang, Ed., London, UK, Sep. 2009, LNCS 5761, pp. 975–983, Springer Verlag Berlin Heidelberg.

[2] T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of within-subject fMRI time series," in revision to IEEE Trans. Med. Imag., LNAO, NeuroSpin/CEA, Gif-sur-Yvette, France, Oct. 2009.
[3] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline, "Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets," *Hum. Brain Mapp.*, vol. 27, no. 8, pp. 678–693, Aug. 2006.
[4] 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.
[5] X.L. Meng and W.H. Wong, "Simulating ratios of normalizing constants via a simple identity: a theoretical exploration," *Statistica Sinica*, vol. 6, pp. 831–860, 1996.
[6] A. Gelman and X.-L. Meng, "Simulating normalizing constants: from importance sampling to bridge sampling to path sampling," *Statistical Science*, vol. 13, pp. 163–185, 1998.
[7] M. Jerrum and A. Sinclair, "Polynomial-time approximation algorithms for the Ising model," *SIAM J. Comput.*, vol. 22, pp. 1087–1116, 1993.
[8] Adrien Trillon, Jérôme Idier, and Pierre Peureux, "Unsupervised Bayesian 3D reconstruction for non-destructive evaluation using gammagraphy," in *EUSIPCO*, Lausanne, Suisse, Aug. 2008.
[9] Florence Forbes and Nathalie Peyrard, "Hidden Markov Random Field model selection criteria based on mean field-like approximations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1089–1101, Sep. 2003.
[10] R. H. Swendsen and J. S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Physical Review Letters*, vol. 58, pp. 86–88, 1987.
[11] 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*, Cairo, Egypt, Nov. 2009.
[12] L. Onsager, "A two-dimensional model with an order-disorder transition," *Phys. Rev.*, vol. 65, no. 3& 4, pp. 117–149, Feb. 1944.
[13] David M. Higdon, James E. Bowsher, Valen E. Johnson, Timothy G. Turkington, David R. Gilland, and Ronald J. Jaszczak, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 516–526, Oct. 1997.
[14] T. Vincent, P. Ciuciu, and J. Idier, "Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI," in *32th Proc. IEEE ICASSP*, Honolulu, Hawaii, Apr. 2007, vol. I, pp. 325–328.
[15] 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*, vol. 41, no. 3, pp. 941–969, July 2008.
[16] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, Ch. Pallier, and H. Benali, "Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1235–1251, Oct. 2003.
[17] D. M. Higdon, "Auxiliary variable methods for Markov chain Monte Carlo with applications," *J. Amer. Statist. Assoc.*, vol. 93, no. 442, pp. 585–595, June 1998.
[18] Peter J. Green and Sylvia Richardson, "Hidden Markov models and desease mapping," *J. Amer. Statist. Assoc.*, vol. 97, no. 460, pp. 1–16, Dec. 2002.
[19] M. Smith, B. Pütz, D. Auer, and L. Fahrmeir, "Assessing brain activity through spatial Bayesian variable selection," *Neuroimage*, vol. 20, pp. 802–815, 2003.
[20] D. Smith and M. Smith, "Estimation of binary Markov random fields using Markov Chain Monte Carlo," *J. Comput. and Graph. Stats.*, vol. 15, no. 1, pp. 207–227, 2006.
[21] David Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, USA, 1 edition, September 1987.

APPENDIX

A. Properties of our log-PF estimate

The first property deals with the asymptotic behavior ($\beta \to \infty$) of the log-PF of a symmetric L-color Potts field:

$$\lim_{\beta\to\infty} \left(\log Z(\beta) - \beta c\right) = \log L \tag{15}$$

It is quite straightforward to demonstrate that when $\beta \to \infty$ only homogeneous configurations of $\boldsymbol{q}$ have a significant weight in the evaluation of $Z(\beta)$. In Potts MRFs, such configurations arise whenever all sites are equal to a given label leading first to $\sum_{k\sim j} I(q_{jk} = q_j) = c$ and finally to Eq. (15). Applying Eq. (15) to the extrapolation context allows one to derive the following proposition.

Proposition 1 $\lim_{\beta\to\infty}\mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p) = 0$*, so* $\log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ *defined in Eq.* (8) *provides an asymptotically unbiased estimate of* $\log Z_{\mathcal{T}}(\beta)$*,* $\forall \mathcal{G}_p$*.*

Proof: First, applying Eq. (15) to $\mathcal{G}_p$ and using Eq. (8), we get:

$$\begin{aligned}
&\lim_{\beta\to\infty} \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}} \left[\log Z_{\mathcal{G}_p}(\beta) - \beta c_{\mathcal{G}_p}\right] = \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}} \log L \\
\Leftrightarrow &\lim_{\beta\to\infty} \Big[\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}} \left(\log Z_{\mathcal{G}_p}(\beta) - \log L\right) - \beta c_{\mathcal{T}}\Big] = 0 \\
\Leftrightarrow &\lim_{\beta\to\infty} \left[\log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p) - \beta c_{\mathcal{T}}\right] = \log L.
\end{aligned}$$

Applying Eq. (15) to $\log Z_{\mathcal{T}}(\beta)$, we obtain $\lim_{\beta\to\infty} \left[\log Z_{\mathcal{T}}(\beta) - \log \widetilde{Z}_{\mathcal{G}_p}(\beta, \mathcal{G}_p)\right] = 0$ $\forall \mathcal{G}_p$.

The second property gives us the expression of the first-order derivative of the log-PF at $\beta = 0$. On the one hand, following [13], it can be shown that $(\log Z(\beta))' = \mathbb{E}\left[U(\boldsymbol{q})|\beta\right]$. On the other hand, for $\beta = 0$, all sites are independent and follow a uniform Bernoulli distribution. Hence, for each clique $j \sim k$ the L-homogeneous configurations $(q_j, q_k) = (l, l)$ for $l \in \{1, \ldots, L\}$ contribute to U with the same weight of $1/(2L)$. We therefore obtain $\mathbb{E}(U(\boldsymbol{q}) \,|\, \beta = 0) = \sum_{k\sim j} 1/L$. Finally, by equating the two expressions, we get:

$$(\log Z(0))' \triangleq d \log Z(\beta)/d\beta|_{\beta=0} = c/L \,. \tag{16}$$

From Eq. (8), we get $\left(\log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p)\right)' = \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}} \left(\log Z_{\mathcal{G}_p}(\beta)\right)'$, hence Eq. (16) allows us to derive that $\forall \mathcal{G}_p$, $(\log \widetilde{Z}_{\mathcal{T}}(0, \mathcal{G}_p))'$ provides an unbiased estimate of $(\log Z_{\mathcal{T}}(0))'$.

B. Maximal Approximation Error

We give here a sufficient condition involving that the approximation errors $\mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ of 2-class Potts fields defined over $\mathcal{T}$ and $\mathcal{G}_p$ achieve their largest value at $\beta = 0$.

Proposition 2 *$\forall \mathcal{G}_p$, if $(s_{\mathcal{T}} - 1)/c_{\mathcal{T}} \neq (s_{\mathcal{G}_p} - 1)/c_{\mathcal{G}_p}$ (**Hyp. 1**) and $\mathbb{E}_{\mathcal{T}}(U(\boldsymbol{q}) \,|\, \beta)/c_{\mathcal{T}} \neq \mathbb{E}_{\mathcal{G}_p}(U(\boldsymbol{q}) \,|\, \beta)/c_{\mathcal{G}_p}$, $\forall \beta > 0$ (**Hyp. 2**) then $\mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) = \max_{\beta \in \mathbb{R}_+} \mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p)$, which expression is given by Eq. (10).*

Proof:Let $err_{\mathcal{T}}(\beta, \mathcal{G}_p)$ be the unnormalized approximation error: $\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p) = (\log Z_{\mathcal{T}}(\beta) - \log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p))^2$. We prove that $\mathcal{E}_{\mathcal{T}}(0, \mathcal{G}_p) = \max_{\beta \in R_+} err_{\mathcal{T}}(\beta, \mathcal{G}_p)$ by showing that $\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ is a strictly decreasing function on $\mathbb{R}_+$:

$$\frac{d\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)}{d\beta} = L \underbrace{\left(\log Z_{\mathcal{T}}(\beta) - \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}(\log Z_{\mathcal{G}_p}(\beta) - \log L) - \log L\right)}_{f_1(\beta)} \times \underbrace{\left(\mathbb{E}_{\mathcal{T}}(U|\beta) - \frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}\mathbb{E}_{\mathcal{G}_p}(U|\beta)\right)}_{f_2(\beta)}$$

$\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ is strictly monotonous on $\mathbb{R}_+$ if $f_{1,2}(\beta) \neq 0 \; \forall \beta > 0$. According to **Hyp. 2**, we directly obtain $f_2(\beta) \neq 0$. Moreover, it is easy to notice that $f_1(\beta) = \pm\sqrt{\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)}$. Hence, $f_1(0) \neq 0$ according to **Hyp. 1** and $\lim_{\beta \to \infty} f_1(\beta) = 0$ by definition of $\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)$. Furthermore, according to the value of $(\log Z(\beta))'$ and **Hyp. 2**, we get: $f_1'(\beta) = f_2(\beta) \neq 0$, $\forall \beta > 0$. Function f_1 being continue, its sign is then contant over $\mathbb{R}_+$ and then $f_1(\beta) \neq 0$, $\forall \beta > 0$. As a consequence, $\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ is then stricly monotonous for $\beta > 0$. Again, according to **Hyp. 1**, we obtain $\mathcal{E}_{\mathcal{T}}(0, \mathcal{G}_p) > 0$.

Since by definition $\lim_{\beta \to \infty} \mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p) = 0$, function $\mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ is therefore strictly decreasing on $\mathbb{R}_+$ and finally $\mathcal{E}_{\mathcal{T}}(0, \mathcal{G}_p) = \max_{\beta \in \mathbb{R}_+} \mathcal{E}_{\mathcal{T}}(\beta, \mathcal{G}_p)$. Since $\log Z(\beta)$ is a strictly increasing function of β, its inverse is strictly decreasing on $\mathbb{R}_+$, so $\mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) = \max_{\beta \in \mathbb{R}_+} \mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p)$. Note finally that **Hyp. 2** were empirically observed on numerous graphs of different size and shape. The only exceptions were found when the topological similarity measure $\mathcal{L}_{\mathcal{T}}(\mathcal{G}_p)$ was large.

C. Mean field-based partition function estimates

Following [9], from the mean-field theory we recall here how to derive the zeroth and first order approximations for the partition function with a special concern to the L-color symmetric Potts model.

The negative energy $U^{\mathrm{mf}}(\boldsymbol{w})$ actually reads $U^{\mathrm{mf}}(\boldsymbol{w}) = \sum_{i=1}^{n} w_i^{\mathrm{t}} \sum_{j \in N_i} \bar{w}_j$ with $\bar{w}_j = \mathrm{E}^{\mathrm{mf}}\big[W_j\big]$. Hence,

$$\mathrm{E}^{\mathrm{mf}}\big[U^{\mathrm{mf}}(\boldsymbol{W})\big] = \frac{1}{2}\sum_{i=1}^{n} w_i^{\mathrm{t}} \sum_{j \in N_i} w_j = \frac{1}{2}\mathrm{E}^{\mathrm{mf}}\big[U^{\mathrm{mf}}(\boldsymbol{W})\big]$$

so that the zeroth order PF approximation based on the the mean field approximation is given by:

$$Z^{\mathrm{mf}} = \sum_{\boldsymbol{w}} \exp(\beta U^{\mathrm{mf}}(\boldsymbol{w})) \tag{17}$$

$$= \prod_{i=1}^{n} \sum_{w_i} \exp(\beta w_i^{\mathrm{t}} \sum_{j \in N_i} \bar{w}_j) \tag{18}$$

The first order approximation is based upon the Gibbs-Bogoliubov-Feynman [21] inequality that states:

$$Z \geqslant Z^{\mathrm{mf}} \exp(\mathrm{E}^{\mathrm{mf}}\big[U(W) - U^{\mathrm{mf}}(W)\big]). \tag{19}$$

The mean field model (4) is optimal among models with factorization property in the sense that it maximizes the lower bound in inequality (19) for such models. When considering the Taylor expansion around zero of $\exp(-\mathrm{E}^{\mathrm{mf}}\big[U(W) - U^{\mathrm{mf}}(W)\big])$, the right hand side of inequality (19) denoted by Z^{GBF} in what follows

$$Z^{\mathrm{GBF}} = Z^{\mathrm{mf}} \exp(-\mathrm{E}^{\mathrm{mf}}\big[U(W) - U^{\mathrm{mf}}(W)\big])$$

can be seen as a first order approximation of Z and so much closer than the zeroth order approximation Z^{mf}. From Eq. (17), the first order approximation of Z can be deduced:

$$Z^{\mathrm{GBF}} = Z^{\mathrm{mf}} \exp(-\frac{\beta}{2}\sum_{i=1}^{n} \bar{w}_i \sum_{j \in N_i} \bar{w}_j). \tag{20}$$