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, Thomas Vincent, Florence Forbes, Jérôme Idier & Philippe Ciuciu

Journal of Signal Processing Systems
for Signal, Image, and Video Technology
(formerly the Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology)

ISSN 1939-8018
Volume 65
Number 3

J Sign Process Syst (2011) 65:325-338
DOI 10.1007/s11265-010-0505-6

Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Springer

Author's personal copy

J Sign Process Syst (2011) 65:325–338
DOI 10.1007/s11265-010-0505-6

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 · Thomas Vincent · Florence Forbes · Jérôme Idier · Philippe Ciuciu

Received: 16 January 2010 / Revised: 29 June 2010 / Accepted: 29 June 2010 / Published online: 20 July 2010
© Springer Science+Business Media, LLC 2010

Abstract In this paper, we propose 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.

Keywords Markov random field · Potts fields · Partition function · fMRI · Bayesian inference · MCMC · Detection-estimation

The authors thank the Région Ile de France for funding.

L. Risser · T. Vincent · P. Ciuciu (✉)
NeuroSpin/CEA, 91191 Gif-sur-Yvette, France
e-mail: philippe.ciuciu@cea.fr

T. Vincent
e-mail: Thomas.Vincent@cea.fr

L. Risser · T. Vincent · P. Ciuciu
IFR 49, Institut d'Imagerie Neurofonctionnelle,
Paris, France

L. Risser · J. Idier
IRCCyN/CNRS, 1 rue de la Noë,
44300 Nantes, France

J. Idier
e-mail: Jerome.Idier@irccyn.ec-nantes.fr

L. Risser
Imperial College, Institute for Mathematical Sciences,
SW7 London, UK
e-mail: l.risser@imperial.ac.uk

F. Forbes
INRIA Rhônes-Alpes, MISTIS project,
665 av. de l'Europe, Montbonnot,
38334 Saint Ismier, France
e-mail: Florence.Forbes@inrialpes.fr

1 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 spatially smoothing the datasets at the expense of a loss of spatial resolution. A more challenging

Author's personal copy

approach works on the unsmoothed data by the introduction of 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, in symmetric Potts models, the temperature level controls the amount of spatial correlation. In the present fMRI application [1, 2], which aims to jointly detect the cerebral activations and estimate their dynamics from the 4D signal, 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 are 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 are automatically estimated, since such parameters may be different when considering different brain regions and their hand-tuning is reasonably unachievable. In [2], 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 aim to segregate activating, non-activating and deactivating voxels.

Section 2 is dedicated to the formulation of the partition function (PF) estimation problem for Potts fields. In Section 3 several techniques of PF estimation are presented. The main contribution of this paper lies in Section 4 where a fast extrapolation technique for Potts fields PF estimation 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 5. Conclusions are drawn in Section 6.

2 Problem Statement

Let us consider a grid characterized by a set of sites $\boldsymbol{s} = \{i = 1, \ldots, n\}$. 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:

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

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 [4]) and thus satisfies the two properties:

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

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 $\sum_{\boldsymbol{q} \in \{1, \ldots, L\}^n} \exp(\beta U(\boldsymbol{q}))$ and depends on the geometry of $\mathcal{G}$. Its exact evaluation in a reasonable amount of time is impossible except on tiny grids [11]. 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.

3 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

Author's personal copy

estimation of multiple PFs is necessary. To this end, after presenting some PF estimation techniques in this section, we propose in Section 4, 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.

3.1 Path-sampling Method

The path-sampling method is based on an importance sampling strategy [6]. It 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:

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

where $Z(\beta_0)$ is supposed already known and the $\{\boldsymbol{q}_m, m = 1...M\}$ are distributed according to $\mathbb{P}(\boldsymbol{q} \mid \beta_0)$ 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].

3.2 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:

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

where $\Pr_i^{\mathrm{MF}}(w_i \mid \beta) = \Pr(w_i \mid \beta, \overline{\omega}_{N_i})$ and $\Pr_i^{\mathrm{MF}}(w_i \mid \beta)$ 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): $\Pr^{\mathrm{MF}}(\boldsymbol{w} \mid \beta) = \exp\left(\beta U^{\mathrm{MF}}(\boldsymbol{w})\right) / Z^{\mathrm{MF}}(\beta)(\beta)$, where U^{MF} and $Z^{\mathrm{MF}}(\beta)$ 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 $\overline{\omega} = \{\overline{\omega}_i, i \in s\}$ 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:

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

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 Eqs. 17 and 20 of Appendix C, the mean field allows us to estimate the PF:

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

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

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

Author's personal copy

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

3.3 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 referred 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 regularization level β_k. 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 compute accurate estimates 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 larger the number of sites, the larger the log-PF whatever the regularity of the reference graphs. The dependence of $\log Z_{\mathcal{G}_p}$ on $s_{\mathcal{G}_p}$ at constant $c_{\mathcal{G}_p}$ becomes much more important at small β values. Hence, the bilinear extension of [11] 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.

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

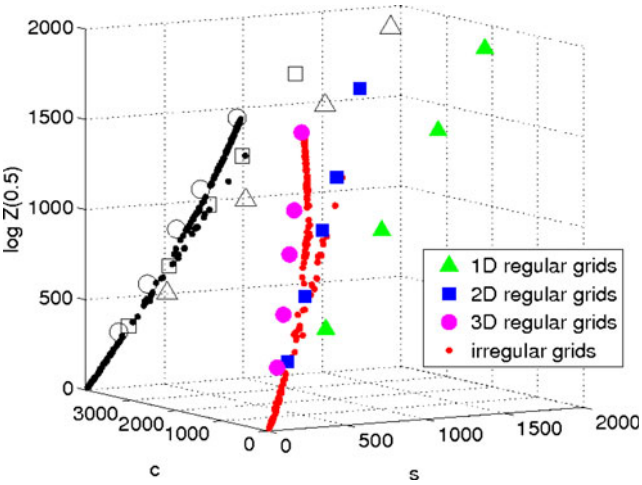

Figure 1 Examples of log-PF values for $\beta = 0.5$ as functions of the number of cliques c and sites s. The projection onto a plane $c = constant$ is depicted to show the dependence on s.

4 Fast and Robust Extrapolation Technique

4.1 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 diverse 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 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 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 to choose the closest reference grid $\mathcal{G}_{\mathrm{ref}}$ to $\mathcal{T}$ in combination with the approximation error criterion $\mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ defined by:

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

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

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 $\mathcal{G}_{\mathrm{ref}}$ is exhibited using a min-max principle, which consists in minimizing

Author's personal copy

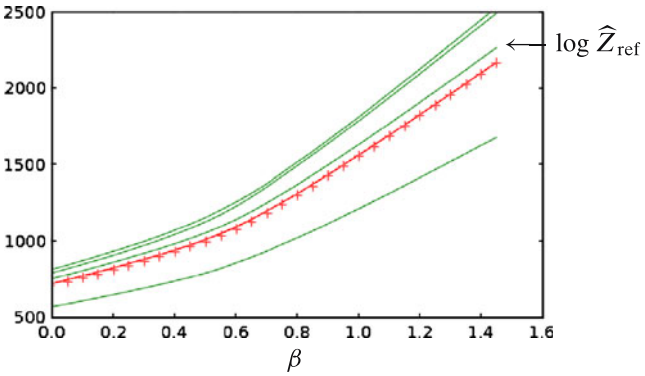

Figure 2 Path-sampled estimates of the *reference* log-PFs $\log \widehat{Z}_{\mathcal{G}_p}(\beta_k)$, $p = 1:4$ in green. Log-PF estimate found by path-sampling for the *test* field $\log \widehat{Z}_{\mathcal{T}}$ in red (ground truth). Our extrapolation method based on Eq. 8 provides a log-PF estimate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

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 homogeneities in $\mathcal{G}_p$ and $\mathcal{T}$ are similar. Hence, we get:

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

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

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: 1) an unbiased asymptotic approximation error[2] and 2) 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.

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

4.2 Assessment of the Method

4.2.1 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 Onsager's contribution [12], who derived an explicit expression for the log-PF of the 2-class Potts model on any 2D square grid under cyclic boundary constraints:[3]

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

where $u(\beta) = 2\sinh\beta/\cosh^2\beta$ and ψ 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 that another possible way to compute the log-PF is to determine for a small cyclic grid, say the 3×3 grid, the log-PF $\log Z_9(\beta)$ and then set, for all $N > 9$, $\log Z(\beta)$ to $\log Z(\beta) = N/9 \log Z_9(\beta)$. The log-PF $\log Z_9(\beta)$ can be computed exactly as shown in [9]. It follows that the log-PF of a cyclic square 2-class Potts fields can be written as:

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

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 compared to Onsager formula. In Fig. 3, we compare the estimation computed using Eq. 11 with the path-sampling and extrapolation approaches on a 2-class Potts field defined over a 30×30 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

[3] in 2D, the right is connected to the left and the top to the bottom.

Author's personal copy

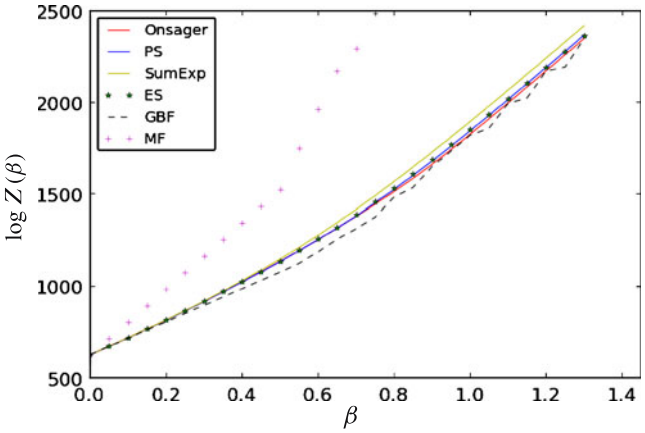

Figure 3 *Red*: True $\log Z(\beta)$ computed by Eqs. 11–12 for a 2D 2-class Potts field defined over a 30×30 cyclic regular grid. *Blue*: path-sampling estimate i.e., $\log \widehat{Z}(\beta)$. *Yellow*: log-PF obtained using the sum of exponentials (Eq. 13). *Green crosses* ($\times$): Extrapolation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

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.

4.2.2 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 the 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. Table 1 shows that by 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 extracted 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 1.

Table 1 Mean maximal relative approximation error over regular and irregular test graphs using the extrapolation technique on 2- and 3-class Potts fields.

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

Errors are given in percents.

For both 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 generally 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.

Author's personal copy

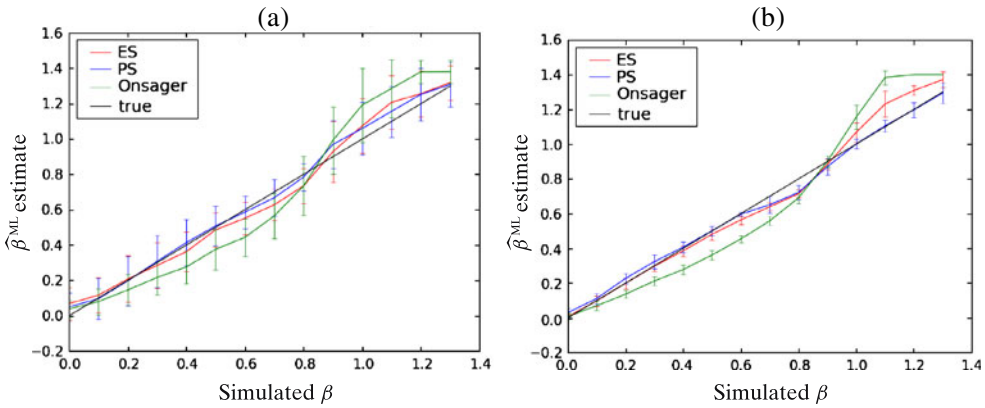

Figure 4 Monte Carlo validation (100 realizations) for β-estimation on observed 2D 2-class Potts fields defined over (**a**) 10×10 and (**b**) 50×50 grids. *Black*: the ground truth is given by the first bisector. *Red*: ML estimate $\widehat{\beta}^{\mathrm{ML}}$ relying on our log-PF extrapolation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

4.3 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 $\widehat{\beta}^{\mathrm{ML}}$ is given by $\widehat{\beta}^{\mathrm{ML}} = \arg\max_{\beta} \left[\beta U(\boldsymbol{q}) - \log Z(\beta)\right]$. 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).

5 Application to fMRI Data Analysis

5.1 Problem Statement

5.1.1 Parcel-based BOLD Signal Modeling

Our extrapolation algorithm was applied to the spatially adaptive 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 $\mathcal{P} = (\mathcal{P}_\gamma)_{\gamma=1:\Gamma}$ 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].

5.1.2 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 $\boldsymbol{R} = (\boldsymbol{D}_2^{\mathrm{t}}\boldsymbol{D}_2)^{-1}$, which allows us to estimate a smooth HRF shape since $\boldsymbol{D}_2$ is the second-order finite difference matrix that penalizes 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., $p(\boldsymbol{A} \mid \boldsymbol{\theta}_{\boldsymbol{a}}) = \prod_m p(\boldsymbol{a}^m \mid \boldsymbol{\theta}^m)$ 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 be questioned. In such cases, a between-condition prior covariance matrix could be introduced, the difficulty lying in the choice of a relevant correlation model to limit the computational complexity.

Author's personal copy

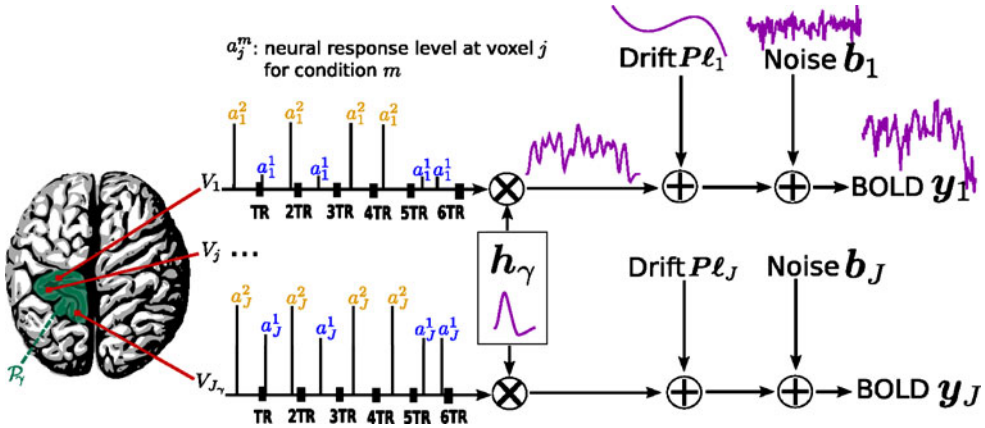

Figure 5 Parcel-based regional BOLD model. The size of each parcel $\mathcal{P}_\gamma$ varies typically between by a few tens and a few hundreds of voxels: $80 \leqslant J_\gamma \leqslant 350$. The number M of experimental conditions involved in the model usually varies from 1 to 5. In our example, $M = 2$, the NRLs (a_j^1, 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 h_γ can be sampled at a period of 1 s and estimated on a range of 20 to 25 s (e.g., $D = 25$). Most often, the LFD coefficients ℓ_j are estimated on a few components ($Q = 4$).

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 $p(\boldsymbol{a}^m \mid \boldsymbol{q}^m, \boldsymbol{\theta}^m) = \prod_j p(a_j^m \mid q_j^m, \boldsymbol{\theta}^m)$ 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:

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

Spatial correlation is directly incorporated in the probabilities of activation through a *hidden* three-class Potts MRF $\Pr(\boldsymbol{q}^m \mid \boldsymbol{\theta}^m)$, 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: $f(a_j^m \mid q_j^m = i) \sim \mathcal{N}(\mu_{i,m}, v_{i,m})$. Parameters $\mu_{i,m}$ and $v_{i,m}$ define the prior mean and variance of class $i = 0, \pm 1$, respectively for the mth stimulus type. 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 $p(\boldsymbol{h}, \mathbb{a}, \mathbb{q}, \boldsymbol{\Theta} \mid \mathbb{y})$ 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: $\alpha(\beta_m^{(k)} \to \beta_m^{(k+1/2)}) = \min(1, A_{k,k+1/2})$, 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.

5.2 Results on Synthetic fMRI Data

In the context of the joint detection-estimation of brain activity, 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. 6a–d, respectively.

Figure 6b–e illustrate 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 ($m = 1$)

Author's personal copy

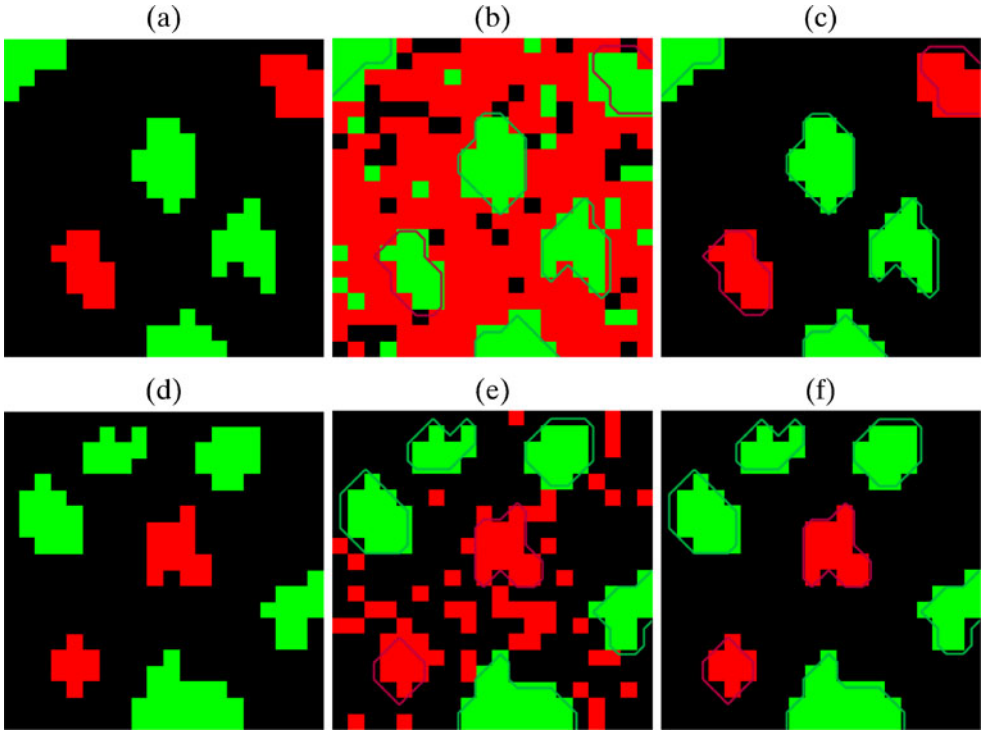

Figure 6 Estimated labels in the Maximum A Posteriori sense. *Top row*: $(\widehat{\boldsymbol{q}}^1)^{\mathrm{MAP}}$. *Bottom row*: $(\widehat{\boldsymbol{q}}^2)^{\mathrm{MAP}}$. (**a**)–(**c**): Label maps obtained using the supervised SMM (SSMM) for $m = 1$ and $m = 2$, respectively with $\beta^{(1)} = \beta^{(2)} = 0.2$. (**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.

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: Figure 6c–f yield estimated labels exactly matching the true ones for $m = 1, 2$.

These results are a consequence of the estimates of the prior mixture components shown in Fig. 7 for the two conditions. In the SSMM case, we observed

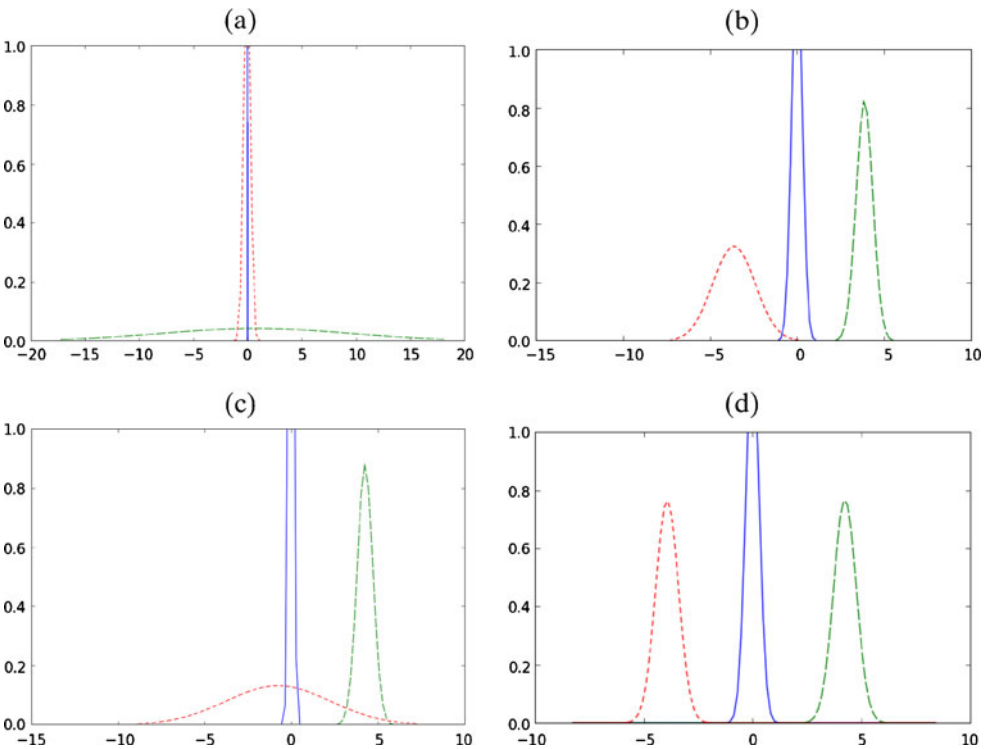

Figure 7 Estimated components of the three-class Gaussian prior mixture. *Top row*: first experimental condition ($m = 1$). *Bottom row*: second experimental condition ($m = 2$). (**a**)–(**c**): Mixture components yielded by the SSMM approach for $m = 1$ and $m = 2$, with $\beta^{(1)} = \beta^{(2)}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

Author's personal copy

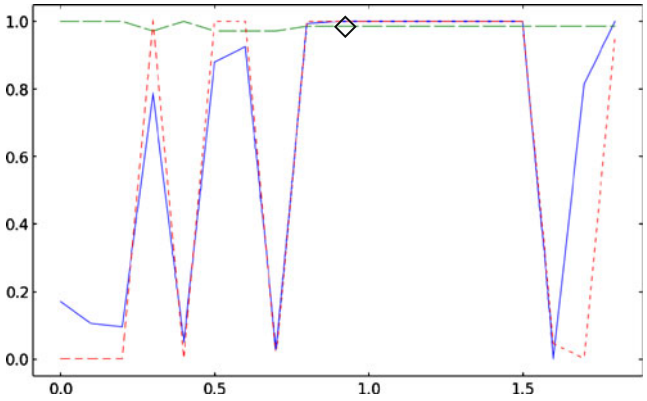

Figure 8 Rate of right classification of $(\widehat{\boldsymbol{q}}^1)^{\mathrm{MAP}}$ and $(\widehat{\boldsymbol{q}}^2)^{\mathrm{MAP}}$ in the supervised case for different fixed β-values. Activating, non-activating and deacctivating classes are color coded in red (*short dash*), blue (*continuous*)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

a degeneracy regarding the activating class for $m = 1$ (Fig. 7a), 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 $\widehat{\mu}^2_{-1} \approx 0 = \mu^2_0$, this directly impacts the posterior classification in regard to the presence of false negatives. On the other hand, Fig. 7b–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. 7b) with the USMM setting. Hence, we found an exact posterior classification.

Figure 8 illustrates the behavior of the SSMM when varying the fixed β value. A wrong tuning yields a significant decrease of the right classification rate of non-activating (blue) and deactivating labels (red) for $\beta \leqslant 0.8$ and $\beta \geqslant 1.2$ 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 $\diamond$) that belongs to this optimal range and yields optimal classification rates.

5.3 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 behavior of a 3-color compared to a 2-color Potts prior on labels. Figure 9 shows normalized contrast maps ($\widehat{\boldsymbol{a}}^{\mathrm{LAC}}$–$\widehat{\boldsymbol{a}}^{\mathrm{RAC}}$) 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, which illustrates the superiority of an *adaptive* spatial correlation model. The same contrast 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

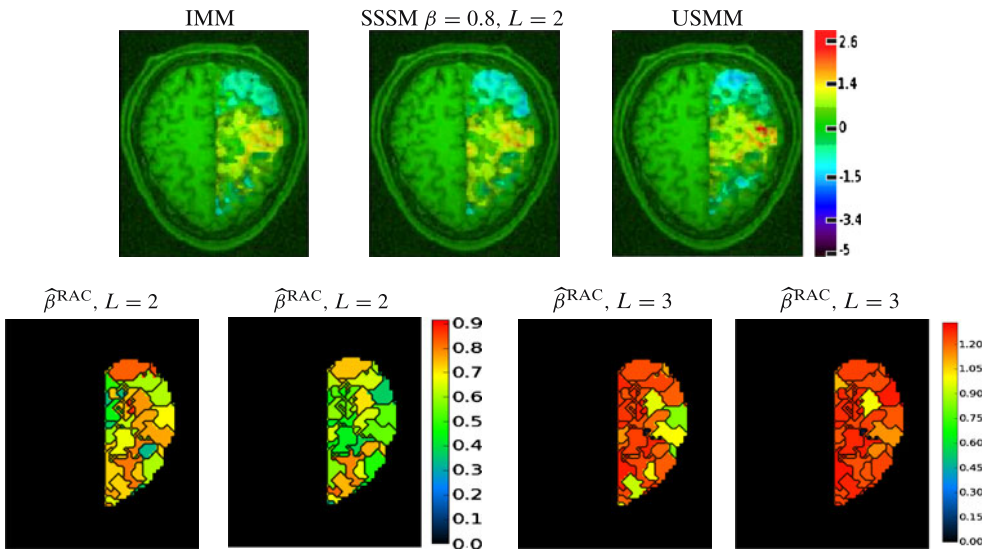

Figure 9 Comparison of the IMM, SSMM and USMM models wrt the estimated normalized constrat maps: left auditory clic (LAC) versus right auditory clic (RAC): $\widehat{\boldsymbol{a}}^{\mathrm{LAC}} - \widehat{\boldsymbol{a}}^{\mathrm{RAC}}$. $\widehat{\beta}$ parcel-dependent maps computed for the LAC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

Author's personal copy

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, $\widehat{\beta}^{\mathrm{PM}}$ estimates for the left auditory clic was 0.56, so 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 $(\widehat{a}^{\mathrm{LAC}} - \widehat{a}^{\mathrm{RAC}})$ 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 ($\widehat{\beta}^{\mathrm{LAC}} \approx 0.5$ vs. $\widehat{\beta}^{\mathrm{RAC}} \approx 0.75$) in parcels located in the right motor cortex since the BOLD signal is stronger for the LAC than for the RAC condition in 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 estimates 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 negligible 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).

6 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 putative negative BOLD effects. The spatial regularization that was performed by using 2-colors Potts fields requires therefore 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. 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 was shown accurate on 2D and 3D grids as well as regular or non-regular grids. It was 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 1 h 30. 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 usefulness of the 3-color Potts prior model compared to the 2-color prior.

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:

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

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.

Author's personal copy

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:

$$\lim_{\beta\to\infty}\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}\left[\log Z_{\mathcal{G}_p}(\beta)-\beta c_{\mathcal{G}_p}\right]=\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}\log L$$

$$\Leftrightarrow\lim_{\beta\to\infty}\left[\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}\left(\log Z_{\mathcal{G}_p}(\beta)-\log L\right)-\beta c_{\mathcal{T}}\right]=0$$

$$\Leftrightarrow\lim_{\beta\to\infty}\left[\log\widetilde{Z}_{\mathcal{T}}(\beta,\mathcal{G}_p)-\beta c_{\mathcal{T}}\right]=\log L.$$

□

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 $(\log Z(\beta))'=\mathrm{E}\left[U(\boldsymbol{q})\,|\,\beta\right]$.

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 $\mathrm{E}\left[U(\boldsymbol{q})\,|\,\beta=0\right]=\sum_{k\sim j}1/L$. Finally, by equating the two expressions, we get:

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

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* $\mathrm{E}_{\mathcal{T}}\left[U(\boldsymbol{q})\,|\,\beta\right]/c_{\mathcal{T}}\neq\mathrm{E}_{\mathcal{G}_p}\left[U(\boldsymbol{q})\,|\,\beta\right]/c_{\mathcal{G}_p}$, $\forall\beta>0$ (**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 $\mathbb{R}_+$:

$$\frac{d\mathcal{E}_{\mathcal{T}}(\beta,\mathcal{G}_p)}{d\beta}$$

$$=L\underbrace{\left(\log Z_{\mathcal{T}}(\beta)-\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}(\log Z_{\mathcal{G}_p}(\beta)-\log L)-\log L\right)}_{f_1(\beta)}$$

$$\times\underbrace{\left(\mathrm{E}_{\mathcal{T}}\left[U\,|\,\beta\right]-\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_p}}\mathrm{E}_{\mathcal{G}_p}\left[U\,|\,\beta\right]\right)}_{f_2(\beta)},$$

□

where $\mathcal{E}_{\mathcal{T}}(\beta,\mathcal{G}_p)$ is strictly monotonous on $\mathbb{R}_+$ if $f_{1,2}(\beta)\neq0$, $\forall\beta>0$. According to **Hyp. 2**, we directly obtain $f_2(\beta)\neq0$. Moreover, it is easy to notice that $f_1(\beta)=\pm\sqrt{\mathcal{E}_{\mathcal{T}}(\beta,\mathcal{G}_p)}$. Hence, $f_1(0)\neq0$ 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: $f_1'(\beta)=f_2(\beta)\neq0$, $\forall\beta>0$. Function f_1 being continue, its sign is then contant over $\mathbb{R}_+$ and then $f_1(\beta)\neq0$, $\forall\beta>0$. 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** is 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 $U^{\mathrm{MF}}(\boldsymbol{w})=\sum_{i=1}^{n}w_i^{\mathrm{t}}\sum_{j\in N_i}\bar{w}_j$ where $\bar{w}_j=\mathrm{E}^{\mathrm{MF}}[W_j]$ so that the zeroth order PF approximation based on the the mean field approximation is given by:

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

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

Author's personal copy

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

$$Z(\beta) \geqslant Z^{\mathrm{MF}}(\beta) \exp\left(\mathrm{E}^{\mathrm{MF}}\left[U(W) - U^{\mathrm{MF}}(W)\right]\right). \quad (19)$$

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 $\exp(-\mathrm{E}^{\mathrm{MF}}[U(W) - U^{\mathrm{MF}}(W)])$, the right hand side of inequality (19) denoted by $Z^{\mathrm{GBF}}(\beta)$ in what follows:

$$Z^{\mathrm{GBF}}(\beta) = Z^{\mathrm{MF}}(\beta) \exp\left(\mathrm{E}^{\mathrm{MF}}\left[U(W) - U^{\mathrm{MF}}(W)\right]\right)$$

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

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

References

1. Risser, L., Vincent, T., Ciuciu, P., & Idier, J. (2009). Robust extrapolation scheme for fast estimation of 3D Ising field partition functions. Application to within subject fMRI data analysis. In *12th Proc. MICCAI, LNCS* (Vol. 5761, pp. 975-983). London, UK.
2. Vincent, T., Risser, L., & Ciuciu, P. (2010). Spatially adaptive mixture modeling for analysis of fmri time series. *IEEE Transactions on Medical Imaging, 29*(4), 1059–1074.
3. Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., & Poline, J.-B. (2006). Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Human Brain Mapping, 27*(8), 678–693.
4. Hammersley, J. M., & Clifford, P. (1990). Markov random fields in statistics. In G. R. Grimmett & D. J. A. Welsh (Eds.), *Disorder in physical systems: A volume in honour of John M. Hammersley* (pp. 19–32). Oxford University Press.
5. Meng, X. L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica, 6*, 831–860.
6. Gelman, A., & Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science, 13*, 163–185.
7. Jerrum, M., & Sinclair, A. (1993) Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing, 22*, 1087–1116.
8. Trillon, A., Idier, J., & Peureux, P. (2008). Unsupervised Bayesian 3D reconstruction for non-destructive evaluation using gammagraphy. In *EUSIPCO 16th European Signal Processing Conference*. Lausanne, Switzerland.
9. Forbes, F., & Peyrard, N. (2003). Hidden Markov random field model selection criteria based on mean field-like approximations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(9), 1089–1101.
10. Swendsen, R. H. & Wang, J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters, 58*, 86–88.
11. Risser, L., Idier, J., Ciuciu, P., & Vincent, T. (2009). Fast bilinear extrapolation of 3D ISING field partition function. Application to fMRI image analysis. In *Proc. IEEE ICIP*. Cairo, Egypt.
12. Onsager, L. (1944). A two-dimensional model with an order-disorder transition. *Physical Review, 65*(3&4), 117–149.
13. Higdon, D. M., Bowsher, J. E., Johnson, V. E., Turkington, T. G., Gilland, D. R., & Jaszczak, R. J. (1997). Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging, 16*(5), 516–526.
14. Vincent, T., Ciuciu, P., & Idier, J. (2007). Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI. In *32th Proc. IEEE ICASSP* (Vol. I, pp. 325–328). Honolulu, Hawaii.
15. Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., & Ciuciu, P. (2008). A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *NeuroImage, 41*(3), 941–969.
16. Ciuciu, P., Poline, J.-B., Marrelec, G., Idier, J., Pallier, Ch., & Benali, H. (2003). Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on Medical Imaging, 22*(10), 1235–1251.
17. Higdon, D. M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association, 93*(442), 585–595.
18. Green, P. J. & Richardson, S. (2002). Hidden Markov models and desease mapping. *Journal of the American Statistical Association, 97*(460), 1–16.
19. Smith, M., Pütz, B., Auer, D., & Fahrmeir, L. (2003). Assessing brain activity through spatial Bayesian variable selection. *NeuroImage, 20*, 802–815.
20. Smith, D., & Smith, M. (2006). Estimation of binary Markov random fields using Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics, 15*(1), 207–227.
21. Chandler, D. (1987). *Introduction to modern statistical mechanics* (1st Ed). USA: Oxford University Press.

Laurent Risser graduated in Mathematical Engineering, then in Computer Science for Image Analysis and obtained in 2007 his PhD in Biomedical Image Analysis at University Paul Sabatier (Toulouse, France). He then held a first post-doctoral experience with Neurospin/CEA Saclay (France), where he worked on the analysis of the brain activity in functional MRI. He currently holds a post-doctoral research assistant position at Imperial College London with the Institute for Mathematical Sciences and the Biomedical Image Analysis group and develops solutions of Computational Anatomy applied to cardiac and cerebral 3D volumetric images. He has published fifteen journal papers and conference proceedings, mainly in biomedical image analysis but also in signal processing, biology and neuroscience.

Author's personal copy

Thomas Vincent received his Msc degree in bio-informatics and applied mathematics in 2006 from university of Cergy-Pontoise. He will defend his PhD thesis on September 2010 with at the University of Paris-Sud (France). He has pursued his research at Neurospin/LNAO (CEA, Life Science Division) under Ph. Ciuciu's supervision. His main subject focuses on spatial regularization models applied to the joint detection-estimation problem of brain activity from fMRI data by developing efficient adaptive approaches. He is also involved in characterizing cerebral hemodynamics over the cortical surface in the perspective of multi-modal data fusion with EEG and NIRS modalities. He is coauthoring four journal papers and twelve conference proceedings, mainly in biomedical image analysis and signal processing.

Florence Forbes was born in Monaco in 1970. She received both the B.Sc. and M.Sc. degrees in computer science and applied mathematics from the Ecole Nationale Sup'erieure d'Informatique et Mathématiques Appliquées de Grenoble, Grenoble, France, and the Ph.D. degree in applied probabilities from University Joseph Fourier, Grenoble, France in 1996. Since 1998 she has been a research scientist with the Institut National de Recherche en Informatique et Automatique (INRIA), Grenoble Rhône-Alpes, Montbonnot, France, where she is the head of the MISTIS team since 2003. Her research activities include Bayesian image analysis, Markov models and hidden structure models.

Jérôme Idier was born in France in 1966. He received the diploma degree in electrical engineering from École Supérieure d'Électricité, Gif-sur-Yvette, France, in 1988 and the Ph.D. degree in physics from University of Paris-Sud, Orsay, France, in 1991.

Since 1991, he joined the Centre National de la Recherche Scientifique. He is currently a Senior Researcher at the Institut de Recherche en Communications et Cybernétique in Nantes. His major scientific interests are in probabilistic approaches to inverse problems for signal and image processing. He is serving as an Associate Editor for the IEEE Transactions on Signal Processing and for the Journal of Electronic Imaging, copublished by SPIE and IS&T.

Philippe Ciuciu received his PhD from University Paris-Sud in 2000 for a work done at *Groupe Problèmes Inverses* (L2S, CNRS) in spectral analysis and radar Doppler imaging. He then became a post-doctoral fellow (2000-2001) in the *fMRI signal processing group* at Service Hospitalier Frédéric Joliot (CEA, Life Science Division) before being appointed on a permanent position by the same institute in November 2001. Since 2007 he have been with Neurospin centre. In 2008, he has been nominated *Principal Investigator* of the neurodynamics research program. His research interests include hemodynamic characterization from functional MRI data, parallel MRI SENSE-like reconstruction using wavelet-based regularization schemes and analysis of scale invariance in brain dynamics from multimodal imaging modalities (MEG/EEG/fMRI) in the context of his young researcher ANR-SCHUBERT grant.