Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 41 (2008) 941 – 969

A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI

Salima Makni,[a] Jérôme Idier,[b] Thomas Vincent,[c,d] Bertrand Thirion,[e]
Ghislaine Dehaene-Lambertz,[f,d] and Philippe Ciuciu[c,d,*]

[a]*Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), University of Oxford, John Radcliffe Hospital, Oxford, UK*
[b]*IRCCyN (CNRS), Nantes, France*
[c]*CEA, NeuroSpin, Gif-sur-Yvette, France*
[d]*IFR 49, Institut d'Imagerie Neurofonctionnelle, Paris, France*
[e]*INRIA Futurs, Orsay, France*
[f]*INSERM U562, NeuroSpin, Gif-sur-Yvette, France*

Received 21 September 2007; revised 12 December 2007; accepted 8 February 2008
Available online 26 February 2008

Within-subject analysis in fMRI essentially addresses two problems, i.e., the *detection* of activated brain regions in response to an experimental task and the *estimation* of the underlying dynamics, also known as the characterisation of Hemodynamic response function (HRF). So far, both issues have been treated sequentially while it is known that the HRF model has a dramatic impact on the localisation of activations and that the HRF shape may vary from one region to another. In this paper, we conciliate both issues in a region-based joint detection-estimation framework that we develop in the Bayesian formalism. Instead of considering function basis to account for spatial variability, spatially adaptive General Linear Models are built upon region-based non-parametric estimation of brain dynamics. Regions are first identified as functionally homogeneous *parcels* in the mask of the grey matter using a specific procedure [Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B., August 2006. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27 (8), 678–693.]. Then, in each parcel, prior information is embedded to constrain this estimation. Detection is achieved by modelling activating, deactivating and non-activating voxels through mixture models within each parcel. From the posterior distribution, we infer upon the model parameters using Markov Chain Monte Carlo (MCMC) techniques. Bayesian model comparison allows us to emphasize on artificial datasets first that inhomogeneous gamma-Gaussian mixture models outperform Gaussian mixtures in terms of sensitivity/specificity trade-off and second that it is worthwhile modelling serial correlation through an AR(1) noise process at low signal-to-noise (SNR) ratio. Our approach is then validated on an fMRI experiment that studies habituation to auditory sentence repetition. This phenomenon is clearly recovered as well as the hierarchical temporal organisation of the superior temporal sulcus, which is directly derived from the parcel-based HRF estimates.
© 2008 Elsevier Inc. All rights reserved.

Keywords: Bayesian modelling; fMRI; Gamma-Gaussian mixture model; Detection-estimation; Markov Chain Monte Carlo methods; Bayes factor; Model comparison

Introduction

Since the first report of the BOLD effect in human (Ogawa et al., 1990), functional magnetic resonance imaging (fMRI) has represented a powerful tool to non-invasively study the relation between cognitive stimulus and the hemodynamic (BOLD) response. Within-subject analysis in fMRI is usually addressed using a hypothesis-driven approach that actually postulates a model for the HRF and enable voxelwise inference in the General Linear Model (GLM) framework. In this formulation, the modelling of the BOLD response i.e., the definition of the *design matrix* is crucial. In its simplest form, this matrix relies on a spatially invariant temporal model of the BOLD signal across the brain meaning that the expected response to each stimulus is modelled by a single regressor. Assuming the neurovascular system as *linear* and *time-invariant* (LTI), this regressor is built as the convolution of a sparse spike train representing the stimulation signal and the canonical HRF, i.e., a composition of two gamma functions which reflects the BOLD signal best in the visual and motor cortices (Glover, 1999).

Intra-individual differences in the characteristics of the HRF have been exhibited between cortical areas in (Aguirre et al., 1998; Miezin et al., 2000; Neumann and Lohmann, 2003; Handwerker

[*] Corresponding author. Tel.: +33 1 6908 7785; fax: +33 1 6908 7855.
E-mail address: philippe.ciuciu@cea.fr (P. Ciuciu).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2008.02.017

Author's personal copy

et al., 2004). Although smaller than inter-individual fluctuations, this regional variability is large enough to be regarded with care. To account for these spatial fluctuations at the voxel level, one usually resorts to hemodynamic function basis. For instance, the canonical HRF can be supplemented with its first and second derivatives to model differences in time (Friston, 1998; Henson et al., 2002). To make the basis spatially adaptive, Woolrich et al. (2004a) have proposed a half-cosine parameterisation in combination to the selection of the best basis set. Although powerful and elegant, the price to be paid for such a flexible modelling lies in a loss of sensitivity of detection: the larger the number of regressors in the basis, the smaller the number of effective degrees of freedom in any subsequent statistical test. Crucially, in a GLM involving several regressors per condition, the Student-t statistic can no longer be used to infer on differences between experimental conditions. Rather, an *unsigned* Fisher statistic has to be computed, making direct interpretation of activation maps more difficult. Indeed, the null hypothesis is actually rejected whenever any of the contrast components deviates from zero and not specifically when the difference of the response magnitudes is far from zero.

In this paper, to facilitate cognitive interpretations, we argue in favour of a spatially adaptive GLM in which a local estimation of the HRF is performed. This allows us to factorise the expected BOLD response with a *single* regressor attached to each experimental condition and to enforce direct statistical comparisons based on response magnitudes. However, to conduct the analysis in an efficient and reliable manner, local estimation is performed at the scale of several voxels.

As mentioned earlier, the localisation of brain activation strongly depends on the modelling of the brain response and thus of its estimation. Of course, the converse also holds: HRF estimation is only relevant in voxels that elicit signal fluctuations correlated with the paradigm. Hence, detection and estimation are intrinsically linked to each other. The key point is therefore to tackle the two problems in a common setting, *i.e.,* to set up a formulation in which *detection* and *estimation* enter naturally and simultaneously. This setting cannot be the classical hypothesis testing framework. Indeed, the sequential procedure which first consists in estimating the HRF on a given dataset and then building a specific GLM upon this estimate for detecting activations in the same dataset, entails statistical problems in terms of sensitivity and specificity: the control of the false positive rate actually becomes hazardous due to the use of an erroneous number of degrees of freedom. We rather propose a Bayesian approach that provides an appropriate framework to address both detection and estimation issues in the same formalism.

The literature on Bayesian fMRI methods offers several approaches to adequately choose priors for detection. As introduced in (Everitt and Bullmore, 1999; Vaever Hartvig and Jensen, 2000; Penny and Friston, 2003) and further developed in (Smith et al., 2003; Woolrich et al., 2005; Ou and Golland, 2005; Woolrich and Behrens, 2006; Flandin and Penny, 2007), prior mixture models define an appropriate way to perform the classification or the segmentation of statistical parametric maps into activating, non-activating or deactivating brain regions. The pioneering contributions related to mixture modelling in fMRI (Everitt and Bullmore, 1999; Vaever Hartvig and Jensen, 2000) have proposed a voxel by voxel classification to decide whether the estimated effect is analogous to signal or noise in each voxel. Yet, the use of mixture modelling in a joint detection-estimation problem introduces specific concerns in comparison to the usual “hypothesis testing framework”. Indeed, our data are *not* the voxelwise z-statistics but rather the raw fMRI time courses, which are required for the estimation step.

As regards HRF estimation, various priors may be thought of depending on the underlying HRF model. Basically, three classes of models coexist. Parametric models appeared first in the literature (Friston, 1994; Lange, 1997; Cohen, 1997; Rajapakse et al., 1998; Kruggel and Von Crammon, 1999). In this setting, the estimation problem consists in minimising some criterion with respect to (w.r.t.) some parameters of a precise function (*e.g.,* Gaussian, gamma ,…). However, parametric models tend to introduce some bias in the HRF estimate, since it is unlikely that they capture the true shape variations of the brain dynamics. Moreover, the objective function to be minimized is often non-convex making the parameter estimates unreliable and dependent of the initialisation. Hence, more flexible *semi-parametric* approaches have been proposed later to capture these variations (Genovese, 2000; Gossl et al., 2001; Woolrich et al., 2004a). In a semi-parametric framework, the HRF time course is decomposed into different periods (initial dip, attack, rise, decay, fall, …), each of them being described by specific parameters. At the same time, *non-parametric* approaches or Finite Impulse Response (FIR) models have emerged in the fMRI literature as a powerful tool to infer on the HRF shape (Nielsen et al., 1997; Goutte et al., 2000; Marrelec et al., 2003, 2004; Ciuciu et al., 2003). Most of these works take place in the Bayesian setting and constrain the HRF to be temporally smooth, which warrants a stable estimation in case of ill-posed identification.

Whatever the model in use, most methods are massively univariate and therefore neglect the spatial structure of the BOLD signal. Early investigations have shown that estimating the HRF using regularised FIR models over a functionally homogeneous region of interest provides more reliable results (Gössl et al., 2001; Ciuciu et al., 2004). In the following, a region-based non-parametric model of the HRF is therefore adopted. Then, the critical issue arising consists in exhibiting a functionally homogeneous clustering of the fMRI datasets over the whole brain. To that end, the grey matter's mask is segregated into a few hundreds of connected Regions of Interest (ROIs), called *parcels*. The parcels are derived using the parcellation procedure proposed by Thirion et al. (2006), according to a compound criterion balancing spatial and functional homogeneity. The second step of our analysis solves for the detection-estimation problem over each parcel.

The rest of this paper is organised as follows. Section 2 details how anatomical information is handled and how parcels are built up. Then, the forward parcel-based model of the BOLD signal is derived and priors over the unknown parameters are specified. In Section 3, we explain the key steps of our inferential procedure based on MCMC methods, *posterior mean* (PM) HRF estimation and marginal Maximum *A Posteriori* (MAP) classification for detection. On artificial datasets, Section 4 reports the performance of our approach in terms of sensitivity-specificity trade-off depending on the mixture prior and the noise modelling. In Section 5, our joint detection-estimation approach is tested on real fMRI data acquired during an habituation study to auditory sentence repetition. On the same datasets, we also performed a classical GLM analysis employing the widely used Statistical parametric mapping (SPM) software[1]. The two approaches are then compared and the main differences are exhibited. The pros and cons of the proposed method are discussed in Section 6 and some future extensions are envisaged.

[1] http://www.fil.ion.ucl.ac.uk/spm/.

Author's personal copy

Methodology

Definition of functionnally homogeneous brain regions

Anatomical representation

The segmentation of the grey-white matter interface is performed on an anatomical T_1-weighted MRI image using the BrainVisa software[2] (Mangin et al., 1995). It provides us with the anatomical representation of the cortex. To accommodate the coarser spatial resolution of fMRI data (typically, 3.5 mm along each direction), the grey matter mask $\mathcal{M}_a$ is dilated using a sphere as structural element, with a radius equal to the resolution of functional images.

Concurrently, a functional mask $\mathcal{M}_f$ was computed from the motion-corrected[3] BOLD EPI volumes. Also, an average EPI volume was created. Then, we carried out a histogram analysis of this volume: a Gaussian density $\mathcal{N}(\mu, \sigma^2)$ was fitted on the main mode of the EPI signal of interest. A threshold defined as $\mu - 3\sigma$ was used to obtain the functional mask. Finally, the mask of interest where activation most likely occurs was built as $\mathcal{M}_s = \mathcal{M}_a \cap \mathcal{M}_f$.

Parcellation of the grey-matter

The volume in mask $\mathcal{M}_s$ was then divided in K functionally homogeneous parcels or ROIs using the parcellation technique proposed in (Thirion et al., 2006). The goal of this procedure is to segregate the brain into connected and functionally homogeneous components. For doing so, the parcellation algorithm relies on the minimisation of a compound criterion reflecting both the spatial and functional structures and hence the topology of the dataset. The spatial similarity measure favours the closeness in the Talairach coordinates system. The functional part of this criterion is computed on parameters that characterise the functional properties of the voxels. These parameters can be chosen either as the fMRI time series themselves or as the β-parameters estimated during a first-level SPM analysis. The latter choice is nothing but a projection onto a subspace of reduced dimension, *i.e.,* the feature space. Typically, the feature space is defined from a F-contrast in a SPM analysis.

The number of parcels K needs to be set by hand. The larger the number of parcels, the higher the degree of within-parcel homogeneity. Of course, there exists a trade-off between the within-parcel homogeneity and the signal-to-noise ratio (SNR). If the number of voxels is too small in a given parcel, the HRF estimation may become inaccurate, specifically in regions where no voxel elicits a specific response to any experimental condition. To objectively choose an adequate number of parcels, Thyreau et al. (2006) have used the Bayesian information criterion (BIC) and cross validation techniques on an fMRI study of ten subjects. They have shown converging evidence for $K \approx 500$ for a whole brain analysis and recommend $K = 200$ as a fair setting for a restricted analysis to the grey matter's mask leading to typical parcel sizes around a few hundreds voxels.

Parcel-based modelling of the BOLD signal

Vectors and matrices are displayed in lower and upper cases, respectively, both in bold font (*e.g.,* $\boldsymbol{y}$ and $\boldsymbol{P}$). Unless stated otherwise, subscripts i, j, m and n are respectively indexes over mixture components, voxels, stimulus types and time points. We refer the reader to Appendix A for the definitions of the non-standard probability density functions (pdf). Also, the pdf families are denoted using calligraphic letters (*e.g.,* $\mathcal{N}$ and $\mathcal{G}$ for the Gaussian and gamma densities).

The regional forward model of the BOLD signal introduced in (Makni et al., 2005) is used to account for voxel-dependent and task-related fluctuations of the magnitudes of the BOLD response. Hereafter, the latter magnitudes are called *Neural Response Levels* (NRLs). In short, this time-invariant model characterises each and every parcel by a single HRF shape and a NRL for each voxel and stimulus type. As shown in Fig. 1, this means that although the HRF shape is assumed constant within a parcel, the magnitude of the activation can vary in space and across experimental conditions. Let $\mathcal{P} = (V_j)_{j=1:J}$ be the current parcel and V_j a voxel in $\mathcal{P}$. Then, the generative BOLD model reads:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j, \qquad (1)$$

where

- $\boldsymbol{y}_j = (\boldsymbol{y}_{j,t_n})_{n=1:N}$ denotes the BOLD fMRI time course measured in voxel V_j at times $(t_n)_{n=1:N}$ (N is the number of scans) with $t_n = n$TR and TR is the time of repetition,
- $\boldsymbol{X}^m = \left(x^m_{t_n - d\Delta t}\right)_{n=1:N, d=0:D}$ is a $N \times (D+1)$ binary matrix corresponding to the arrival times for the mth condition. Δt is the sampling period of the HRF, usually smaller than TR. The onsets of the stimuli are put on the Δt-sampled grid by moving them to the nearest time points on this grid. Note that $\boldsymbol{X}^m$ can be adapted to paradigms having trial varying stimulus magnitudes or durations.
- Vector $\boldsymbol{h} = (h_{\mathrm{d}\Delta\mathrm{t}})_{d=0:D}$ represents the unknown HRF shape in parcel $\mathcal{P}$ ($D+1$ is the number of HRF coefficients). It actually seems reasonable to assume a single HRF shape in homogeneous parcels.
- a_{j}^m stands for the NRL in voxel V_j for condition m (M is the number of experimental conditions in the paradigm). Hence, the activation profile associated to the mth stimulus type in voxel V_j is computed as the product $\boldsymbol{h} \times a_j^m$.
- $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_Q]$ is the low frequency orthogonal matrix of size $N \times Q$. It consists of an orthonormal basis of functions $\boldsymbol{p}_q = (p_q(t_n))_{n=1:N}$. To each voxel is attached an unknown weighting vector $\boldsymbol{\ell}_j$ that has to be regressed in order to estimate the trend in V_j. We denote $\mathbb{l} = (\boldsymbol{\ell}_j)_{\mathrm{j}=1:J}$ the set of low frequency drifts involved in $\mathcal{P}$.
- $\mathbf{b}_j \in \mathbb{R}^N$ is the noise vector in voxel V_j . In (Woolrich et al., 2001; Worsley et al., 2002) an autoregressive (AR) noise model has been introduced to account for the serial correlation of the fMRI time series in the detection analysis. Importantly, when this temporal correlation is correctly estimated, the number of false positives decreases, yielding more conservative detection results. Similarly, in a joint detection-estimation framework, Makni et al. (2006b) have shown that the introduction of a spatially-varying first-order AR noise in model (1) provides a lower false positive rate. Hence, $\boldsymbol{b}_j$ is defined by $b_{j,t_n} = \rho_j b_{j,t_{n-1}} + \varepsilon_{j,t_n}, \forall j,\ t,$ with $\varepsilon_j \sim \mathcal{N}(0_N, \sigma^2_{\varepsilon_j}\boldsymbol{I}_N)$,, where $\mathbf{0}_N$ is a null vector of length N, and $\boldsymbol{I}_N$ is the identity matrix of size N.

Although the noise structure is correlated in space (Woolrich et al., 2004b) and non-stationary across tissues (Worsley et al., 2002), we do not essentially account for this correlation for two

[2] http://www.brainvisa.info.

[3] We applied the SPM2 motion-correction algorithm.

Author's personal copy

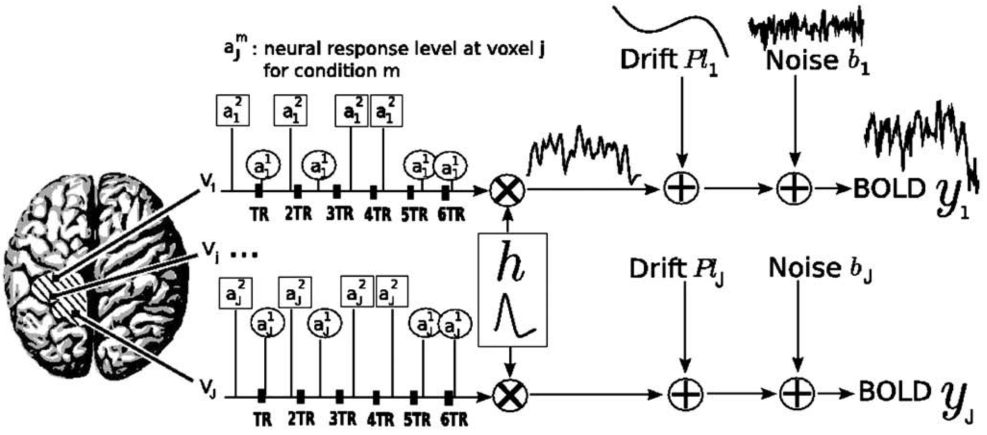


Fig. 1. Summary of the proposed regional BOLD model. The size of each parcel $\mathcal{P}$ varies typically between by a few tens and a few hundreds: $80 \leqslant J \leqslant 350$. The number M of experimental conditions involved in the model usually varies from 1 to 5. In our example, $M=2$, the NRLs $(\boldsymbol{a}_j^1, \boldsymbol{a}_j^2)$ 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 $\boldsymbol{h}$ 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 $\ell_j$are estimated on a few components ($Q=4$).

reasons. First, it is likely that a large part of the noise may be due to misspecification of the HRF. Second, we actually assume that the spatial correlation of the signal of interest is more important. Hence, the fMRI time series $\mathbb{y} = (\boldsymbol{y}_j)_{j=1:J}$ are supposed to be statistically conditionally independent. The likelihood then factors over voxels:

$$p\left(\mathbb{y}|\boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta}_0\right) = \prod_{j=1}^{J} p\left(\boldsymbol{y}_j|\boldsymbol{h}, \boldsymbol{a}_j, l_j, \theta_{0,j}\right) \propto \prod_{j=1}^{J} |\Lambda_j|^{1/2} \sigma_{\varepsilon_j}^{-N} \exp\left(-\sum_{j=1}^{J} \frac{\tilde{\boldsymbol{y}}_j^t \boldsymbol{\Lambda}_j \tilde{\boldsymbol{y}}_j}{2\sigma_{\varepsilon_j}^2}\right) \quad (2)$$

where $\theta_{0,j}=(\rho_j\ ,\ \sigma_{\varepsilon j}^2)$, $\boldsymbol{\theta}_0=(\theta_{0,j}\ )_{j=1:J}$ and $\tilde{\boldsymbol{y}}_j=\boldsymbol{y}_j - \Sigma_m\ a_j^m\ \boldsymbol{X}^m\boldsymbol{h}-\boldsymbol{P}\boldsymbol{\ell}_{\mathrm{j}}$. Note that $\sigma_{\varepsilon j}^{-2}\boldsymbol{\Lambda}_j$ defines the inverse of the autocorrelation matrix of $\boldsymbol{b}_j$. According to Kay (1988, Chap VI, p. 177), $\boldsymbol{\Lambda}_j$ is tridiagonal, with $(\boldsymbol{\Lambda}_j)_{1,1}=(\boldsymbol{\Lambda}_j)_{N,N}=1$, $(\boldsymbol{\Lambda}_j)_{n,n}=1+\rho_{\mathrm{j}}^2$ and $(\boldsymbol{\Lambda}_j\ )_{n+1,n}=(\boldsymbol{\Lambda}_j)_{n,n+1}=-\rho_j\ \forall n=2 : N-1$. Its determinant is given by $|\boldsymbol{\Lambda}_j|=1-\rho_j^2$. In what follows, we do not approximate Eq. (2) by dropping the term $|\boldsymbol{\Lambda}\mathrm{j}|^{1/2}$, as done in previous works (Roberts and Penny, 2002; Penny et al., 2003; Woolrich et al., 2004b). Indeed, when the AR parameter ρ significantly departs from zero (*e.g.,* $\rho \geq 0.4$), this approximation is biased and potentially far from the exact likelihood.

On the sole basis of the likelihood function (2), it seems impractical to identify the pair ($\boldsymbol{h}$, $\mathbb{a}$). Indeed, Maximum likelihood (ML) estimation of ($\boldsymbol{h}$, $\mathbb{a}$) is a bilinear inverse problem since (1) is linear w.r.t. $\boldsymbol{h}$ when $\mathbb{a}$ is fixed and *vice-versa*. Therefore, the ML solution ($\boldsymbol{h}^*$, $\mathbb{a}^*$) is not unique. For instance, every couple ($\boldsymbol{h}^*/s$, $\mathbb{a}^* \times s$) defines another pair of solutions in the ML sense whatever the scale parameter $s>0$. To get rid of identifiability problems and reach a more reliable estimation, in the Bayesian formalism we introduce suitable prior distributions attached to the unknown quantities ($\boldsymbol{h}$, $\mathbb{a}$).

Priors

The Hemodynamic response function

Akin to (Buxton and Frank, 1997; Goutte et al., 2000; Marrelec et al., 2003), the HRF is characterised as follows: *(i)* its variations are smooth; *(ii)* it is causal and returns to a baseline after a given time interval T ($h_t=0$, $\forall\ t<0$ and $t>T$); T is fixed by the user according to the experimental paradigm (usually 25 seconds).

Condition *(i)* may be fulfilled using an approximation of the second-order derivative $\|\partial^2\boldsymbol{h}\|^2$:

$$\left(\partial^2\boldsymbol{h}\right)_{d\tau} \approx \left(h_{(d+1)\tau} - 2h_{d\tau} + h_{(d-1)\tau}\right)/\tau^2,\ \forall d = 1 :\ D - 1.$$

In matrix form, we get $\partial^2\boldsymbol{h}=\boldsymbol{D}_2\boldsymbol{h}$.

Condition *(ii)* is ensured with a HRF $\boldsymbol{h}$ whose magnitude vanishes at first and last time points ($h_0=h_D=0$). Hence, $\boldsymbol{D}_2$ is the *truncated* second-order finite difference matrix of size $(D-1)\times(D-1)$ and $\|\partial^2\boldsymbol{h}\|^2=\boldsymbol{h}^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{h}$ with $\boldsymbol{R}=(\boldsymbol{D}_2^{\mathrm{t}}\ \boldsymbol{D}_2)^{-1}$ a symmetric positive definite matrix. The prior on $\boldsymbol{h}$ thus reads $\boldsymbol{h}\sim\mathcal{N}(\boldsymbol{0}, \sigma_{\boldsymbol{h}}^2\boldsymbol{R})$. To overcome the scale ambiguity problem mentioned earlier, we constrain the HRF to be of unitary norm ($\|\boldsymbol{h}\|=1$). Alternative constraints such as setting the value of the peak could be considered.

The "neural" response levels

Mixture models are often used as a second stage to segment the SPMs (*i.e.,* the statistical maps) resulting from a first-level temporal analysis of fMRI time series (Vaever Hartvig and Jensen, 2000; Everitt and Bullmore, 1999; Woolrich et al., 2005). This means that the data to be classified correspond to some normalised effect $\boldsymbol{c}^{\mathrm{t}}\hat{\boldsymbol{\beta}}$ / std ($\boldsymbol{c}^{\mathrm{t}}\hat{\boldsymbol{\beta}}$), where vector $\boldsymbol{c}$ defines a contrast of interest (typically a comparison between two experimental conditions) and $\hat{\boldsymbol{\beta}}$ is the vector of parameter estimates after fitting a GLM against the fMRI data.

In the present paper, as well as in (Makni et al., 2005), prior mixture models are used in a different way, closer to that proposed by Svensen et al. (2000). In the same spirit, a mixture model is introduced on the NRLs for every experimental condition m and not specifically on the linear combination $\boldsymbol{c}^{\mathrm{t}}\hat{\boldsymbol{\beta}}$. In (Makni et al., 2006a), it was stressed that a two-class Gaussian mixture model (GMM) may be inadequate for segregating noise from true activations. In particular, it can be shown that this kind of independent mixture may degenerate in the sense that the two probability density functions (pdf) overlap almost entirely if there are not enough activating voxels in the current parcel (see (Makni et al., 2005, §VII.)). For this reason, we have rather adopted an inhomogeneous prior mixture

Author's personal copy

Table 1
Model definition and notations

	GaGMM	GMM	GaGGaMM
AR(1) noise	$\mathcal{M}_1$	$\mathcal{M}_3$	
white noise	$\mathcal{M}_2$	$\mathcal{M}_4$	$\mathcal{M}_5$

Here, AR(1) stands for a first-order autoregressive noise in time whose parameters vary in space. In this respect, it is referenced as a spatially-varying AR(1) noise. The second noise model under study is a spatially-varying white noise. The columns describe the different NRLs priors: GMM stands for a *two-class Gaussian* mixture model (a zero-mean Gaussian density (G) for non-activating voxels and a Gaussian density (G) for activating voxels). GaGMM stands for a *two-class gamma-Gaussian* mixture model (a centred Gaussian density (G) for non-activating voxels and a gamma density (Ga) for activating voxels). GaGGaMM stands for a *three-class mixture* model composed of a zero-mean Gaussian density and two gamma densities (a gamma density (Ga) for activating voxels and a flipped gamma density (Ga) for deactivating voxels).

model. Among several possibilities (Gaussian-lognormal MM, Gaussian-truncated Gaussian MM, …), a gamma-Gaussian mixture model (GaGMM) has been retained for technical reasons that will become clearer in what follows. The non-activating voxels are still modelled using a zero-mean Gaussian pdf while a gamma distribution is used to enforce positivity of activating voxels. Akin to (Vaever Hartvig and Jensen, 2000; Woolrich et al., 2005), a three-class mixture prior model is actually considered to account for deactivating voxels. Since we assume that deactivation corresponds to a negative BOLD response, we use a flipped gamma density defined on the left part of the real line leading to define the GaGGaMM extension (see Table 1).

In our model, different stimulus types are supposed to induce statistically independent hemodynamic magnitudes or NRLs *i.e.,* $p(\mathbb{a}|\boldsymbol{\theta}_{\mathbb{a}}) = \Pi_m p(\boldsymbol{a}^m|\boldsymbol{\theta}^m)$ with $\mathbb{a} = (\boldsymbol{a}^m)_{m=1:M}, \boldsymbol{a}^m = \left(a_j^m\right)_{j=1:J}$ and $\boldsymbol{\theta}_{\mathbb{a}} = \left\{\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m\right\}$. Vector $\boldsymbol{\theta}^m$ denotes the set of unknown hyper-parameters related to the mth stimulus type. Because our mixture model is voxelwise, the prior pdf factors over voxels: $p(\boldsymbol{a}^m \mid \boldsymbol{\theta}^m) = \Pi_j\, p(a_j^m \mid \boldsymbol{\theta}^m)$. Importantly, the hyper-parameters are kept constant for all voxels in a given parcel because of the within-parcel homogeneity. These parameters may actually vary from one parcel to another. Let q_j^m be the *allocation variable* (the label) that indicates whether voxel V_j is activating ($q_j^m = 1$), deactivating ($q_j^m = -1$) or non-activating ($q_j^m = 0$) in condition m. The marginal density $p(a_j^m \mid \boldsymbol{\theta}^m)$ thus reads:

$$\begin{aligned} p\left(a_j^m|\boldsymbol{\theta}^m\right) &= \sum_{i=-1}^{1} \Pr\left(q_j^m = i|\boldsymbol{\lambda}_m\right) f\left(a_j^m|q_j^m = i, \boldsymbol{\theta}^m\right) \\ &= \sum_{i=-1}^{1} \lambda_{i,m} f_i\left(a_j^m|\theta^m\right), \end{aligned} \tag{3}$$

with $\boldsymbol{\lambda}_{\mathrm{m}} = (\lambda_{-1,m},\ \lambda_{0,m},\ \lambda_{1,m})$ and $f_0\left(a_j^m|\boldsymbol{\theta}^m\right) = \mathcal{N}(0, \upsilon_{0,m})$, $f_{\pm 1}\left(a_j^m|\boldsymbol{\theta}^m\right) = \mathcal{G}\left(\alpha_{\pm 1,\ m}, \beta_{\pm 1,m}\right)$. The $\lambda_{i,m}$ parameters define the prior probabilities of the three-class mixture on the NRLs ($\Sigma_i\, \lambda_{i,m} = 1$). For instance, $\lambda_{1,m}$ gives us the prior probability of being activated in response to condition m. Since the mixture is independent in space, we have $\lambda_{1,m} = Pr(q_j^m = 1 \mid \boldsymbol{\lambda}_m)$, $\forall j$. Note that $q_j^m \mid \boldsymbol{\lambda}_m$ follows a multinomial distribution over the 3-dimensional probability simplex, *i.e.,* $q_j^m \sim \mathcal{MN}_3(1; \boldsymbol{\lambda}_m)$ (see Appendix A). Hence, seven hyper-parameters are necessary to describe the prior mixture for each experimental condition m:

$$\boldsymbol{\theta}^m = \left\{\lambda_{\pm 1,m}, \alpha_{\pm 1,m}, \beta_{\pm 1,m}, \upsilon_{0,m}\right\}.$$

Compared to (Woolrich et al., 2005), we set the mean of the non-activating class to zero ($\mu_{0,m} = 0$, $\forall\ m$), while we do not need to place restrictions on the mode of the activation and deactivation gamma classes.

The nuisance variables

We assume that $\mathbb{l}$ is a random process independent of $\boldsymbol{h}$ such that $p\left(\mathbb{l}; \sigma_\ell^2\right) = \Pi_j p\left(\ell_j; \sigma_\ell^2\right)$ and $\ell_j \sim \mathcal{N}\left(\mathbf{0},\ \sigma_\ell^2 \boldsymbol{I}_Q\right)$.

The hyper-parameters

All the hyper-parameters are concatenated into the overall parameter vector $\boldsymbol{\Theta} = \left\{\boldsymbol{\theta}_0, \sigma_{\boldsymbol{h}}^2, \sigma_\ell^2, \boldsymbol{\theta}_{\mathbb{a}}\right\}$. Without informative prior knowledge, the following priors are retained for ($\sigma_{\boldsymbol{h}}{}^2$, σ_ℓ^2, $\boldsymbol{\theta}_0$):

$$p\left(\sigma_h^2, \sigma_\ell^2\right) = \left(\sigma_h \sigma_\ell\right)^{-1}, p(\boldsymbol{\theta}_0) = \prod_{j=1}^{J} p\left(\rho_j \sigma_{\varepsilon_j}^2\right) = \prod_{j=1}^{J} \sigma_{\varepsilon_j}^{-1} u([-1, 1])\left(\rho_j\right), \tag{4}$$

to ensure stability of the AR(1) noise process (Kay, 1988).

Mixture parameters. As regards variances $\upsilon_{0,m}$, an improper Jeffreys' prior $p(\upsilon_{0,m}) = \upsilon_{0,m}^{-1/2}$ is considered because we do expect non-activating voxels in a given parcel. Hence, class 0 should never be empty a priori. However, to avoid emptiness and subsequent degeneracy problems making the sampling of the posterior distribution of $\upsilon_{0,m}$ unfeasible, a conjugate prior could also be chosen, that is, an inverse gamma density $\mathcal{IG}\left(\upsilon_{0,m}, a_{\upsilon_0}, b_{\upsilon_0}\right)$, where ($b_{\upsilon_0}$, c_{υ_0}) are fixed values chosen in an appropriate way to make the prior flat enough.

The non-negativity of parameters $\alpha_{i,m}$ is guaranteed through the use of an exponential density $\mathcal{E}\left(\alpha_{i,m}; s_i\right) \equiv \mathcal{G}\left(\alpha_{i,m}; 1, s_i\right)$ as prior distribution (see Appendix A). For parameters $\beta_{i,m}$ we resort to the conjugate prior, given by a gamma density $\mathcal{G}\left(\beta_{i,m}; b_i, c_i\right)$ for $i = \pm 1$.

Mixture probabilities. As regards mixture probabilities $\boldsymbol{\lambda}_m \in [0, 1]^3$, a Dirichlet prior distribution is used as it is conjugate to the multinomial distribution used for labels, *i.e.,* $\mathcal{MN}_3\left(q_j^m|\boldsymbol{\lambda}_m\right)$. More exactly, a symmetric Dirichlet density $\mathcal{D}_3(\boldsymbol{\lambda}_m|\boldsymbol{\delta})$ is selected with $\boldsymbol{\delta} = \delta \mathbf{1}_3$ and $\delta > 0$ (see Appendix A).

The full prior density $p(\boldsymbol{\theta}^m)$ thus reads:

$$p(\boldsymbol{\theta}^m) = \upsilon_{0,m}^{-1/2} \frac{\Gamma(3\delta)}{3\Gamma(\delta)} \prod_{i=\pm 1} \lambda_{i,m}^{\delta-1} s_i \frac{c_i^{b_i}}{\Gamma(b_i)} \beta_{i,m}^{b_i-1} \exp\left(-s_i \alpha_{i,m} - c_i \beta_{i,m}\right). \tag{5}$$

Values of ($a_{\pm 1}$, $b_{\pm 1}$, $c_{\pm 1}$, $s_{\pm 1}$, δ) are fixed empirically but do not really influence the results in most cases[4]. These parameters make the sampling steps of ($\alpha_{\pm 1,m}$, $\beta_{\pm 1,m}$) always possible even when one of the two classes ±1 is empty, because the hyper-prior densities have been chosen proper.

The full posterior distribution

Combining data-driven information in each parcel with prior knowledge using Bayes' rule, we get the full posterior distribution, which is the keystone both for localising activations and

[4] Except potentially when the corresponding class is empty: $J_{i,m} = 0$ for $i = \pm 1$.

Author's personal copy

deactivations as well as for estimating the corresponding parcel-based HRF:

$$
\begin{aligned}
p\left(\boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\Theta} | \mathbb{y}\right) \propto\, & p\left(\mathbb{y} | \boldsymbol{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta}_0\right) p(\mathbb{a} | \boldsymbol{\theta}_{\mathbb{a}}) p\left(\boldsymbol{h} | \sigma_{\boldsymbol{h}}^2\right) p\left(\mathbb{l} | \sigma_{\ell}^2\right) p(\Theta) \\
& \propto \sigma_{\boldsymbol{h}}^{-D} \sigma_{\ell}^{-JQ} \prod_{j=1}^{J} \left(\frac{\left(1-\rho_j^2\right)^{1/2}}{\sigma_{\varepsilon_j}^{N+1}} \mathbb{1}_{(-1,1)}\left(\rho_j\right) \right) \times \ldots \\
& \times \exp\left(-\frac{\boldsymbol{h}^t \boldsymbol{R}^{-1} \boldsymbol{h}}{2\sigma_{\boldsymbol{h}}^2} - \sum_{j=1}^{J} \left(\frac{1}{2\sigma_{\varepsilon_j}^2} \widetilde{\boldsymbol{y}}_j^t \Lambda_j \widetilde{\boldsymbol{y}}_j + \frac{1}{2\sigma_{\ell}^2} ||\ell_j||^2 \right) \right) \\
& \times \prod_{m=1}^{M} \left(p(\boldsymbol{\theta}^m) \prod_{j=1}^{J} p\left(\mathrm{a}_m^j | \boldsymbol{\theta}^m\right) \right)
\end{aligned}
\tag{6}
$$

where $p(a_j^m \mid \boldsymbol{\theta}^m)$ and $p(\boldsymbol{\theta}^m)$ are defined by (3) and (5), respectively.

Note that the parcel-based HRF $\boldsymbol{h}$ can be identified if at least one voxel elicits activation in response to one or several experimental conditions involved in model(1). In addition, other identifiability problems may occur on hyper-parameters such as the mean and variance parameters. It is necessary that at least two voxels belong to each class in order to properly estimate the variances attached to the mixture components. In practice, there is no numerical problem because of the choice of proper priors for the hyper-parameters; see Subsection 3.1.2 for practical details.

Inference scheme

Our objective is to obtain an estimate of the joint posterior distribution of all unknown parameters, given the observed data. Exact and analytical approaches are not feasible with non-Gaussian models such as (6). Several competing inferential schemes are possible. For instance, approximations to the full posterior distribution can be derived in the Variational Bayes (VB) framework or using Taylor series expansion. In our context, given the bilinear structure of the generative BOLD model (6), the VB formulation would be feasible only at the expense of separability assumptions between $\mathbb{a}$ and $\boldsymbol{h}$ in the approximation of the posterior distribution. Further work is required to decide whether or not this hypothesis is tenable. Instead, we resort to a more computationally demanding but exact approach to simulate realisations of the full posterior distribution.

Gibbs sampling algorithm

To draw realisations of the full posterior distribution, a Gibbs sampler is implemented. This consists in building a Markov chain, whose stationary distribution is the joint posterior pdf (6), by sequentially generating random samples from the full conditional pdfs of all the unknown parameters and hyper-parameters; see (Liu, 2001; Robert, 2001) for a general introduction to MCMC.

As shown in Appendix B, direct sampling according to the full conditional distributions is only feasible for the HRF $\boldsymbol{h}$, the labels $\mathbb{q}$, the nuisance variables $\mathbb{l}$, the noise variances $\boldsymbol{\sigma_\varepsilon}$, the mixture probabilities $\boldsymbol{\lambda}_m$, and part of the hyper-parameters (scales σ_h and σ_ℓ, class 0 variances v_0 and shape parameters β_i for $i=\pm1$). In contrast, direct simulation is not tractable for the other parameters, *i.e.,* the NRLs $\mathbb{a}$ corresponding to classes ±1, the AR parameters $\boldsymbol{\rho}$ and the scale parameters $\boldsymbol{\alpha}_i$ of the gamma densities for $i=\pm1$. Therefore, single-component Metropolis-Hastings jumps (Hastings, 1970) are specifically designed. More precisely, separate jumps are proposed for each of the parameters in turn. To this end, suitable instrumental distributions regarding the parameters of interest are designed (see Appendix B for details).

Initialisation

Parameters are uniformly initialised. This means that we set up all voxels with the same noise statistical parameters ($\theta_{0,j}=\theta_0$, $\forall j$) and that we use the same starting values of mixture hyper-parameters ($\boldsymbol{\theta}^m=\boldsymbol{\theta}^*$, $\forall m$). In the first parcel, the HRF is initialised to the canonical shape (Glover, 1999). In the next ones, the HRF is set up using the mean of the estimates computed over the already processed neighbouring parcels. We resort to the same strategies for the labels and the corresponding NRLs when the parcel sizes match approximately. We have checked that this strategy provides shorter burn-in periods[5] and thus reduces the computation load.

Identifiability issues

To cope with these identifiability problems, we have carried out the following three steps procedure over the first iterations of our MCMC algorithm:

- initialise each parcel-specific HRF with a fixed shape in order to obtain a first estimate of labels $\hat{\mathbb{q}}$;
- check that the class of activating voxels is effectively not empty for at least one experimental condition in the current parcel $\mathcal{P}$:

–– If $\exists\ m \in \mathbb{N}_M^* = \{1, \ldots, M\}$ such that $\exists\, j{\in}\mathcal{P} | q_j^m = 1$ then release the HRF constraint to estimate the complete model *i.e.,* including the HRF shape;

–– otherwise, discard the current parcel: the HRF estimate is not reliable in $\mathcal{P}$. Since the corresponding NRLs are close to zero in that case there is no evoked activation due to the experimental paradigm.

Convergence diagnosis

We use a burn-in period of 500 iterations, followed by 1000 subsequent jumps and compute PM and MAP estimates every two jumps. Observations of the chain with different initial conditions confirmed that a burn-in of 500 jumps was sufficient. In addition, convergence has been checked by monitoring on-line the behaviour of the estimated values of some *scalar* parameters (*e.g.,* noise variances, AR parameters, …) from one iteration to another. These observations confirmed also that 1000 iterations were sufficient.

Computational load and parallel implementation

Our current implementation (PyHRF package) is in Python, while the most intensive computations (*e.g.,* computation of the inverse covariance matrix of $\boldsymbol{h}$) have been coded in C-language and interfaced with the Gnu Scientific Library (GSL)[6]. This allows us to take advantage of a parallel computing system available through the Seppo library (Simple Embarassingly Parallel Python[7]) and the Pyro (Python Remote Object) server. Using such a system, all the parameter estimates are obtained in about 2 mn for a parcel of mean size (250 voxels) for two experimental conditions ($M=2$).

[5] The burn-in period is the starting part of the Markov chain built by any MCMC algorithm which is used to ensure that the subsequent samples follow the equilibrium target distribution, *i.e.,* the posterior law.

[6] http://www.gnu.org/software/gsl.

[7] see http://www.its.caltech.edu/~astraw/seppo.html.

Author's personal copy

Since about 200 parcels are necessary to cover the grey matter's mask, a complete within-subject analysis takes about 2 hours when running four processes on a dual core bi-processors Pentium IV (2.7 GHz). PyHRF will be available in the next release of BrainVisa[8] in April 2008.

Derivation of parcel-based summaries

After convergence of the MCMC algorithm in each parcel $\mathcal{P}$, the samples of the quantities of interest are averaged over iterations to compute approximations of marginal posterior expectations:

$$\hat{x}_{\mathcal{P}}^{\mathrm{PM}} = \sum_{k=L_0}^{L_1} x^{(k)}/L, \quad L = L_1 - L_0 + 1, \quad \forall x \in \{h, \mathbb{a}, \mathbb{l}, \Theta\}, \tag{7}$$

where L_0 stands for the length of the burn-in period and L the effective number of iterations. For classification purpose, we proceed in two steps:

1. Compute the PM estimates $(\bar{p}_j^m)_i$ of $Pr(q_j^m = i \mid \boldsymbol{y}_j)$ for $i = -1, 0$ using the following expression:

   $$\left(\bar{p}_j^m\right)_i = \sum_{k=L_0}^{L_1} I\left[\left(q_j^m\right)^{(k)} = i\right]/L, \tag{8}$$

   where I stands for the identity function. Then, deduce $(\bar{p}_j^m)_1$ from the constraint of unitary probability mass: $(\bar{p}_j^m)_1 = 1 - (\bar{p}_j^m)_{-1} - (\bar{p}_j^m)_0$.
2. Sort the probabilities $(\bar{p}_j^m)_i$ and select the MAP estimate:

   $$\left(\hat{q}_j^m\right)^{\mathrm{MAP}} = \arg\max_i \Pr\left(q_j^m = i|\boldsymbol{y}_j\right) \approx \arg\max_i \left(\bar{p}_j^m\right)_i. \tag{9}$$

   whatever the number of components in the mixture. The MAP estimator is easily obtained in the two-class mixture case: V_j is non-activating ($(\hat{q}_j^m)^{\mathrm{MAP}} = 0$) for the mth condition if $(\bar{p}_j^m)_1 < 0.5$.

In combination with these PM estimates, one can attach uncertainty measures to the NRLs. More precisely, the error bars are derived as follows:

$$e_j^m = \sum_{k=L_1}^{L_2} \left(\sigma_{i,j}^m\right)^{(k)} /L \; with \; i = \left(\hat{q}_j^m\right)^{MAP}. \tag{10}$$

Interestingly, $\sigma_{0,j}^m$ is directly given by $\sqrt{\upsilon_{0,j}^m}$ since the full conditional posterior distribution of the zero class is Gaussian, *i.e.*, $\mathcal{N}\left(\mu_{0,j}^m, \upsilon_{0,j}^m\right)$. In contrast, the standard deviations (SD) $\sigma_{\pm1,j}^m$ require further computation since these full conditional densities are gamma-Gaussian (see Section A.5). As derived in Eq. (A.13), the variance of a gamma-Gaussian density admits a closed form expression, which gives $\sigma_{\pm1,j}^m$ after taking the square root. These SD estizmates are then plugged into (10) to get corresponding error bars e_j^m.

The stochastic algorithm is summarised in Table 2.

Table 2
Gibbs sampling algorithm in a given parcel $\mathcal{P}$

```
• Setting up: choose $\boldsymbol{h}^0$, $\mathbb{a}^0$, $\mathbb{l}^0$, $\boldsymbol{\theta}^0$, $\boldsymbol{\theta}_{\boldsymbol{a}}{}^0$.
• Iteration $k$: draw samples $\boldsymbol{h}^k, \mathbf{a}^k, \boldsymbol{\lambda}^k, (\varepsilon^2)^k, \boldsymbol{\theta}_a^k$ from the conditional posterior pdfs:
  -- HRF: $\boldsymbol{h}^k \sim \mathcal{N}(\boldsymbol{\mu_h}, \boldsymbol{\Sigma_h})$,
  -- HRF variance: $(\sigma_h^2)^k \sim \mathcal{IG}(D/2, \boldsymbol{h}^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{h}/2)$,
  -- NRLs: for every condition $m$ and every voxel $j$,
     -- $(u_j^m)^k \sim \mathcal{U}[0, 1]$; if $(u_j^m)^k \leqslant \lambda_{-1,j}^m$, then $(q_j^m)^k = -1$ else if $(u_j^m)^k \leqslant$
        $\lambda_{-1,j}^m + \lambda_{0,j}^m$ then $(q_j^m)^k = 0$, otherwise $(q_j^m)^k = 1$.
     -- $(a_j^m)^k|(q_j^m)^k = 0 \sim \mathcal{N}(\mu_{0,j}^m, \upsilon_{0,j}^m)$.
        $(a_j^m)^k|(q_j^m)^k = \pm1 \sim \mathcal{GN}(a_j^m|\alpha_{\pm1,m}, \mu_{\pm1,j}^m), \upsilon_{\pm1,j}^m)$.
  -- drift coefficients: $\forall_j, (\ell_j)^k \sim \mathcal{N}\left(\boldsymbol{\mu}_{\ell_j}, \sum_{\ell_j}\right)$
  -- Noise variances: $\forall_j, \left(\sigma_{\varepsilon_j}^2\right)^k \sim \mathcal{IG}\left((N+1)/2, ||\tilde{\boldsymbol{y}}_j||_{\boldsymbol{\Lambda}_j}^2/2\right)$.
  -- AR parameters: $\forall_j, (\rho_j)^k \sim \sqrt{1-\rho_j^2}\exp\left(-\frac{A_j}{2\sigma_{\varepsilon_j}^2}\left(\rho_j - \frac{B_j}{A_j}\right)^2\right)\mathbb{1}_{(-1,1)}(\rho_j)$.
  -- Mixture parameters: for every condition $m$,
     -- Weighting probabilities $\lambda_m$:
        $(\lambda_m)^k \sim \mathcal{D}(\delta'), \text{ with } \delta'_i = \delta + \underbrace{\mathrm{Card}\left[C_{i,m} = \left\{j \in 1:J|q_j^m = i\right\}\right]}_{= J_{i,m}},$
        $\forall i = -1:1$.
     -- Variance of NRLs for non-activating voxels: $(\upsilon_{0,m})^k \sim \mathcal{IG}(\eta_{0,m}^k, \nu_{0,m}^k)$.
     -- Shape parameters: $(\alpha_{\pm1,m})^k \sim \exp(J_{i,m}\tau_{i,m}\alpha_{i,m})/\Gamma(\alpha_{i,m})^{J_{i,m}}\mathbb{I}_{\mathbb{R}_+}(\alpha_{i,m})$.
     -- Scale parameters: $(\beta_{\pm i,m})^k \sim \mathcal{G}\left(J_{i,m}\alpha_{i,m} + b_i + 1, \sum_{j \in C_{i,m}} a_j^m + c_i\right)$.
• Iterate until convergence is achieved. PMEs of $\{\mathbf{h}, \mathbb{a}, \mathbb{l}, \boldsymbol{\theta}_\alpha\}$ are computed using (7).
• Classification is performed according to the MAP criterion using (8)-(9).
```

The parameters of the sampled distributions are derived in Appendix B.

Statistical comparisons for cognitive interpretation

Akin to the contrast definition in any GLM-based approach, statistical comparison between our task-related NRL estimates can be addressed in the proposed formalism. One might be interested in assessing *unsigned* or *signed* differences like using Fisher or Student-*t* tests, respectively in the classical hypothesis testing framework.

Let m and m' be the indexes of the conditions we plan to contrast across the brain. This contrast can be assessed by measuring how close the voxelwise marginal distributions $(p_j^m, p_j^{m'})$ of the NRLs $(a_j^m, a_j^{m'})$ are in every voxel V_j. Since these densities write as posterior mixtures, say $p_j^m = \Sigma_i \pi_i f_{i,j}^m$, we start with identifying the MAP estimates $(\hat{q}_j^m, \hat{q}_j^{m'})$ and then we compare the full conditional posterior densities $(f_{\hat{q}_j^m,j}^m, f_{\hat{q}_j^{m'},j}^{m'})$ instead of computing a distance between p_j^m and $p_j^{m'}$. Hence, three different (respectively, six) situations may arise depending on the mixture prior in use (two or three-class mixture, respectively). The different cases correspond to all possible combinations of the pair $(\hat{q}_j^m, \hat{q}_j^{m'})$:

a. if $\hat{q}_j^m = \hat{q}_j^{m'} = -1$, voxel V_j generates deactivations for both conditions. Comparing the NRLs $(a_j^m, a_j^{m'})$ is achievable by measuring how close $(f_{-1,j}^m, f_{-1,j}^{m'})$ are. This comparison therefore answers the question of deciding whether or not the deactivation is stronger for one condition w.r.t. the other (signed comparison) or if there is any difference between the two conditions (unsigned comparison).
b. if $\hat{q}_j^m = \hat{q}_j^{m'} = 0$, voxel V_j is non-activating for both conditions. Comparing the NRLs $(a_j^m, a_j^{m'})$ amounts to computing a criterion between $(f_{0,j}^m, f_{0,j}^{m'})$. The interesting comparison consists in

[8] http://brainvisa.info.

Author's personal copy

deciding whether or not there is some difference in the non-activating profile.

c. if $\hat{q}_j^m = \hat{q}_j^{m'} = 1$, both conditions elicit activations in V_j. By measuring how close $(f_{1,j}^m, f_{1,j}^{m'})$ are, we hope to know if activation occurring for condition m or m' is stronger or if there is any difference irrespective of its sign.
d. if $\hat{q}_j^m = -1$ and $\hat{q}_j^{m'} = 0$, V_j is deactivating in response to the mth stimulus type but is non-activating in response to the m'th condition. To quantify this decision, one can measure a signed or unsigned criterion between $(f_{-1,j}^m, f_{0,j}^{m'})$. By symmetry this case is equivalent to $\hat{q}_j^m = 0$ and $\hat{q}_j^{m'} = 1$.
e. if $\hat{q}_j^m = -1$ and $\hat{q}_j^{m'} = 1$ or vice-versa, V_j is activating in condition m' and deactivating in condition m. To quantify this decision, one can measure a signed or unsigned distance between $(f_{-1,j}^m, f_{1,j}^{m'})$.
f. if $\hat{q}_j^m = 0$ and $\hat{q}_j^{m'} = 1$ or vice-versa, V_j is activating in condition m' and non-activating in condition m. To quantify this decision, one can measure a signed or unsigned criterion between $(f_{0,j}^m, f_{1,j}^{m'})$.

Due to the use of mixture models, these comparisons can allow us to assess the null hypothesis $(H_0 : a_j^m = a_j^{m'})$ or the alternative one (*e.g.*, $H_1 : a_j^m \neq a_j^{m'}$ *or* $H_1 : a_j^m < a_j^{m'}$) depending on the computed criterion. The question is now to define what kind of signed or unsigned criteria we can implement to quantitatively discriminate the two underlying distributions $(f_{i,j}^m, f_{i',j}^{m'})$.

Unsigned task comparison

Unsigned comparison between $f_{i,j}^m$ and $f_{i',j}^{m'}$ can be computed using the Kullback-Leibler (KL) divergence *i.e.,*

$$D\left(f_{i,j}^m \| f_{i,j}^{m'}\right) = \int_{\mathbb{R}} f_{i,j}^m(a) \log \frac{f_{ij}^m(a)}{f_{ij}^{m'}(a)} da.$$

In the present case, its exact computation is only feasible when the two distributions are Gaussian *i.e.,* when $i = i' = 0$ (case b); see (A.3) in Appendix A for details. Otherwise, an approximation of $D(\cdot\|\cdot)$ has to be derived. For doing so, we proceed as follows. In cases (a, c, e), the sampling step of the NRLs $(a_j^m, a_j^{m'})$ relies on two Metropolis jumps, one for each NRL. The corresponding instrumental laws are truncated normal distributions (see (A.4) in Appendix B). Therefore, we approximate $f_{i,j}^m$ and $f_{i',j}^{m'}$ by these positive Gaussian distributions which mean and variance parameters are given in (A.5)-(A.6). We end up by applying the KL divergence formula (Eq. (A.3)) to these truncated Gaussian approximations. In cases *(d, f)*, we proceed similarly for the single activating or deactivating component.

Signed task comparison

To go one step further and recover a sign information regarding the difference $d_j^{m-m'} = a_j^m - a_j^{m'}$, we need to estimate its posterior probability distribution $f_j^{m-m'}$ from a histogram H_j^B (.) with B time bins $(\beta_b)_{b=1:B}$ constructed over the last 500 iterations (*i.e.,* the generated values $(d_j^{m-m'})^{(k)} = (a_j^m)^{(k)} - (a_j^{m'})^{(k)}$ in any voxel of the mask $\mathcal{M}_f$. The posterior cumulative distribution function (cdf) $F(\cdot)$ can then be easily estimated from H_j^B (.). Contrast-based posterior probability maps (PPMs) are thus given by looking at differences $d_j^{m-m'}$ above a given threshold α:

$$P\left(d_j^{m-m'} > \alpha\right) = 1 - F\left(d_j^{m-m'} \leqslant \alpha\right) = 1 - \int_{-\infty}^{\alpha} f_j^{m-m'}(t)\, dt \qquad (11)$$

$$\approx 1 - \sum_{n=1}^{d} H_j^B\left(\frac{\beta_n + \beta_{n+1}}{2}\right) \Delta\beta \text{ with } d < \alpha \leqslant d + 1, \qquad (12)$$

where $\Delta\beta = \beta_{n+1} - \beta_n$. Setting $\alpha = 0$, we actually find the voxels where $(a_j^m) > (a_j^{m'})$*. Finally, we can threshold* $P\,(d_j^{m-m'} > \alpha)$ at level η to retain the voxels which make the comparison significant at this level (*e.g.,* $\eta = 0.95$). Formally, the thresholded PPMs are given by $P\,(d_j^{m-m'} > \alpha) > \eta$. Note that this only provides uncorrected PPMs for multiple comparisons. The control of the familywise error is an open issue in the Bayesian formalism and is beyond the scope of this paper.

Results on synthetic data

Goal of the study

A comparison between two different prior mixture models has been done in (Makni et al., 2006a). In short, it has been shown that the gamma-Gaussian mixture model (*GaGMM*) introduced on the NRLs is more efficient than a two-class Gaussian mixture model (*GMM*) in terms of specificity: it provides a better control of the false positive rate. Similar conclusions have been drawn in (Makni et al., 2006b) when considering an AR(1) noise model instead of a white Gaussian one in combination with a *GMM* prior. As the two changes induce higher computation time, it is worth assessing which modelling effort is preferable i.e. leads to the more significant improvement: the introduction of an inhomogeneous prior mixture or the consideration of serial correlation. For doing so, the models described in Table 1 are tested on the same artificial fMRI dataset.

Artificial fMRI dataset

These data were obtained by first generating two sets of trials, each of them corresponding to a specific stimulus ($M = 2$). These binary time series were then multiplied by a stimulus-dependent scale factor. Here, the functionally homogeneous region $\mathcal{P}$ consisted of $J = 60$ voxels. The number of activating voxels $J_{1,m}$ was varied with the stimulus type m according to $(J_{1,1}, J_{1,2}) = (22, 30)$. Positive NRLs corresponding to activating voxels were simulated according to gamma pdfs:

activating voxels : $\boldsymbol{a}_j^1 \sim \mathcal{G}(\alpha_1 = 3, \beta_1 = 1)$, $\boldsymbol{a}_j^2 \sim \mathcal{G}(\alpha_2 = 10, \beta_2 = 2)$,
non-activating voxels : $\boldsymbol{a}_j^{1,2} \sim \mathcal{N}(0, \upsilon_{0,m} = 0.1)$.

Remark that the chosen gamma parameter values yields a lower SNR for condition 1 ($(\mu_1, \upsilon_1) = (3, 3)$ vs. $(\mu_2, \upsilon_2) = (5, 2.5)$). For all voxels, the binary stimulus sequence was convolved with the canonical HRF $\boldsymbol{h}_c$, whose exact shape appears in Fig. 2(a) in ■-line. An AR(1) noise $\boldsymbol{b}_j$ was then added to the stimulus-induced signal $\Sigma_m a_j^m \boldsymbol{X}^m \boldsymbol{h}$ in every voxel V_j. All AR parameters were set to the same value: $(\rho_j)_{j=1:J} = 0.4$, which is compatible with the serial correlation observed on actual fMRI time series. Also, a low SNR (SNR = 0.3) was considered in our simulations, in conformity with the real situation. Space-varying low-frequency drifts $\boldsymbol{P}\ell_j$ (generated from a cosine transform basis with coefficients ℓ_j drawn from a normal distribution) were also added to the fMRI time courses according to (1).

General comments

As shown in Figs. 2–5(a), all HRF estimates obtained using the four different models match the canonical time course $\boldsymbol{h}_c$ pretty

Author's personal copy

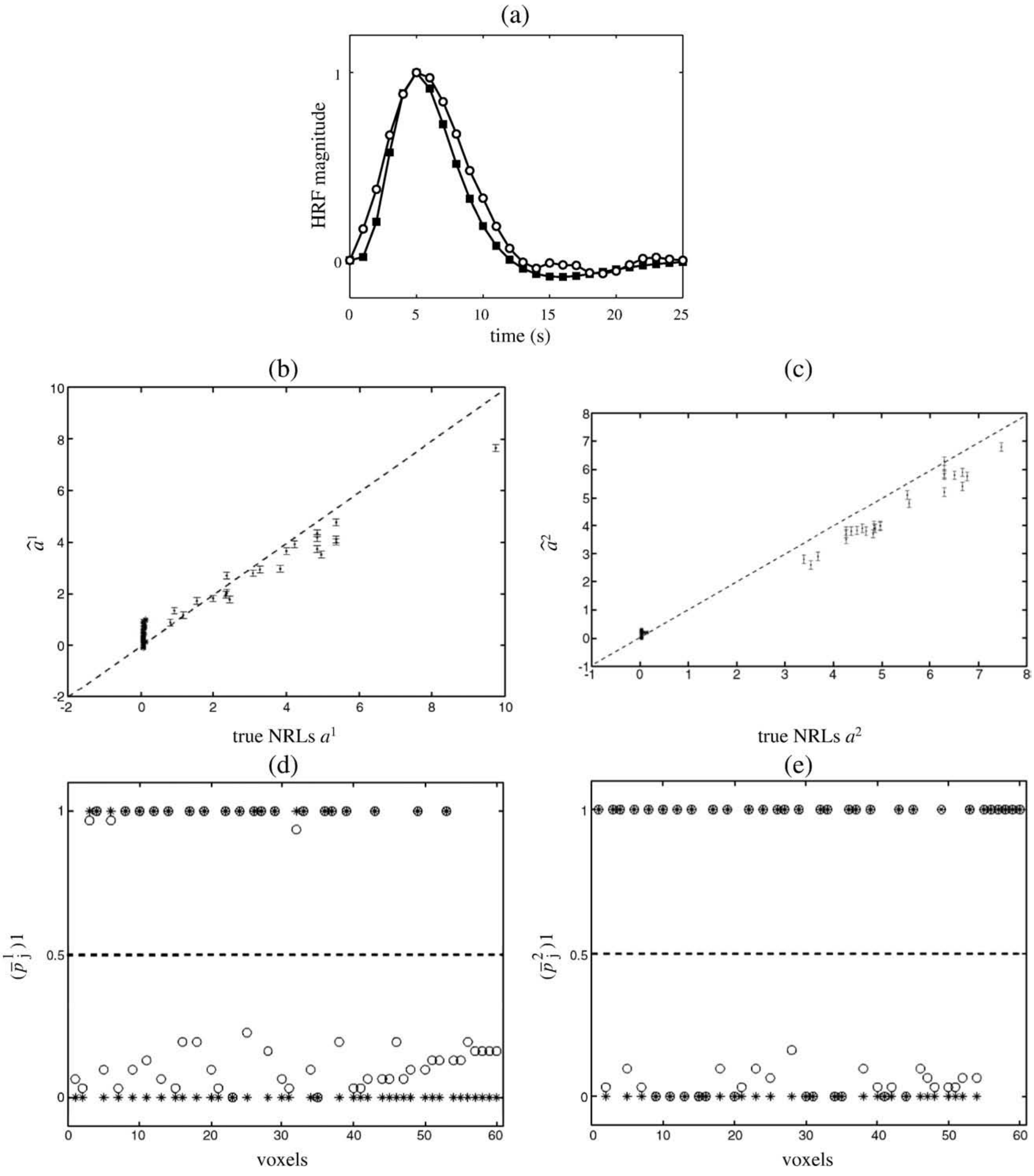


Fig. 2. Estimation results on the simulated data using model $\mathcal{M}_1$. (a) HRF results: Symbols ■ and ○ represent the true $\boldsymbol{h}_c$ and its corresponding HRF estimate, respectively. (b)-(c): NRL estimates for conditions 1 and 2, respectively. True values appear on the *x*-axis and estimated values on the *y*-axis. The error bars follow Eq. (10). (d)-(e): PM estimates of activation probabilities $\bar{p}_j^m$ (○ symbols) for the conditions 1 and 2, respectively. Symbols * depict the true class attached to each voxel.

well. Figs. 2–5(b) show the corresponding NRL estimates that we obtained from models $\mathcal{M}_1 - \mathcal{M}_4$, respectively in response to condition 1 while Figs. 2–5(c) summarize the same results for condition 2.

Since the artificial fMRI time courses were synthetised using a *GaGMM* prior and some correlated noise, it is not surprising that the estimation performed under model $\mathcal{M}_1$ provides the most accurate NRL estimates. Let us remark that the NRL estimates have a small but not negligible amount of bias, which is due to the bias/variance trade-off arising in the Bayesian approach in the non-asymptotic case. Nonetheless, we have checked that the bias tends to zero when the SNR increases.

Author's personal copy

Influence of the noise model

Figs. 2–3(b)-(c) illustrate the impact of the noise model: a more accurate estimation of the NRLs, with smaller error bars and lower mean square error, is observed in Figs. 2(b)-(c) compared to Figs. 3(b)-(c), that is for model $\mathcal{M}_1$ compared to model $\mathcal{M}_2$. This is a direct consequence of accounting for serial correlation in $\mathcal{M}_1$. The same conclusion holds when looking at Figs. 4 and 5(b)-(c), so irrespective of the prior mixture type. As regards the HRF estimate (compare Figs. 2 and 3(a)), the noise model has only little influence on the recovered shape, as already advocated in (Marrelec et al., 2003). As regards AR parameters, the estimated first order coefficients $(\rho_i)_i$ are close to the true values in every voxel for both models $\mathcal{M}_1 - \mathcal{M}_3$ (results not shown).

We also assessed the sensitivity and the specificity of the four models. Figs. 2–5(d)-(e) show the posterior mean estimates $(\overline{p}_j^m)_1$ of deciding that voxel V_j lies in class 1, *i.e.,* is activating for models $\mathcal{M}_1 - \mathcal{M}_4$ and conditions 1 and 2, respectively. These results confirm

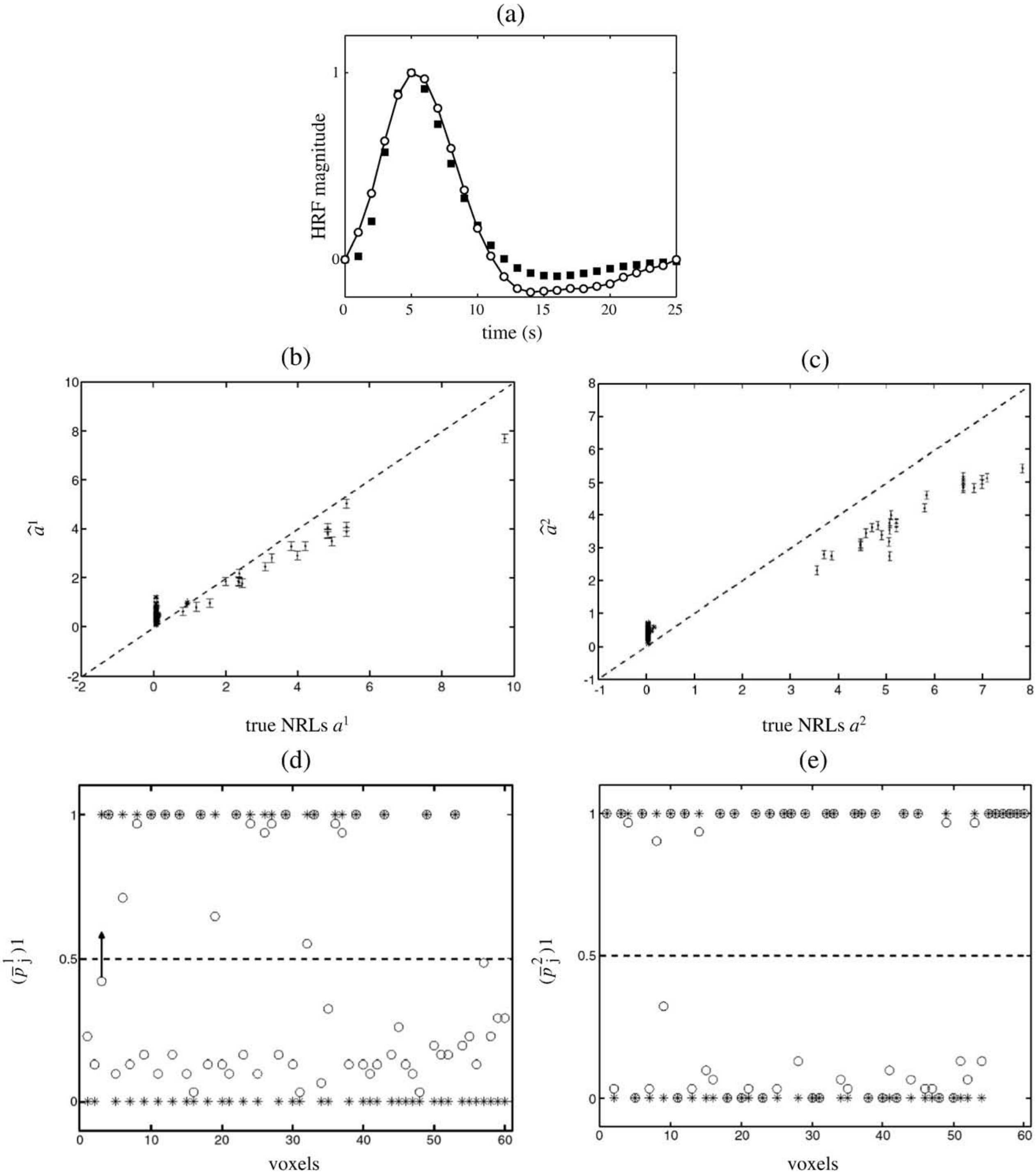


Fig. 3. Simulation results using model $\mathcal{M}_2$. The same legend as in Fig. 2 holds. Only one FN voxel is present, indicated with an upward arrow.

Author's personal copy

our expectations: the modelling of the temporal correlation significantly improves both the sensitivity and the specificity. A higher/lower value of $(\bar{p}_j^m)_1$ is obtained with $\mathcal{M}_1 - \mathcal{M}_3$ when V_j is truly activating/non-activating (compare Figs. 2 and 3(d)-(e) for *GaGMM* priors or Figs. 4 and 5(d)-(e) for *GMM* priors). This means that models $\mathcal{M}_1 - \mathcal{M}_3$ provide lower false positive (FP) and false negative (FN) rates than models $\mathcal{M}_2 - \mathcal{M}_4$, respectively. This effect is stronger in condition 1. This is in agreement with the idea that the precision of the noise model plays a more important role at a lower SNR.

Influence of the mixture prior

Not surprisingly, the estimated NRLs are recovered more accurately using the true prior mixture ($\mathcal{M}_1 - \mathcal{M}_2$): compare Figs. 2–4(b)-(c) one to another for an AR(1) noise model and observe the difference in Figs. 4 and 5(b)-(c) for a white noise model. This effect is much more important at low SNR, *i.e.,* for condition 1. However, we have checked that when the true NRLs of the activating voxels follow a Gaussian distribution, the estimated shape and scale parameters of the gamma density in the *GaGMM* mixture provide close estimates

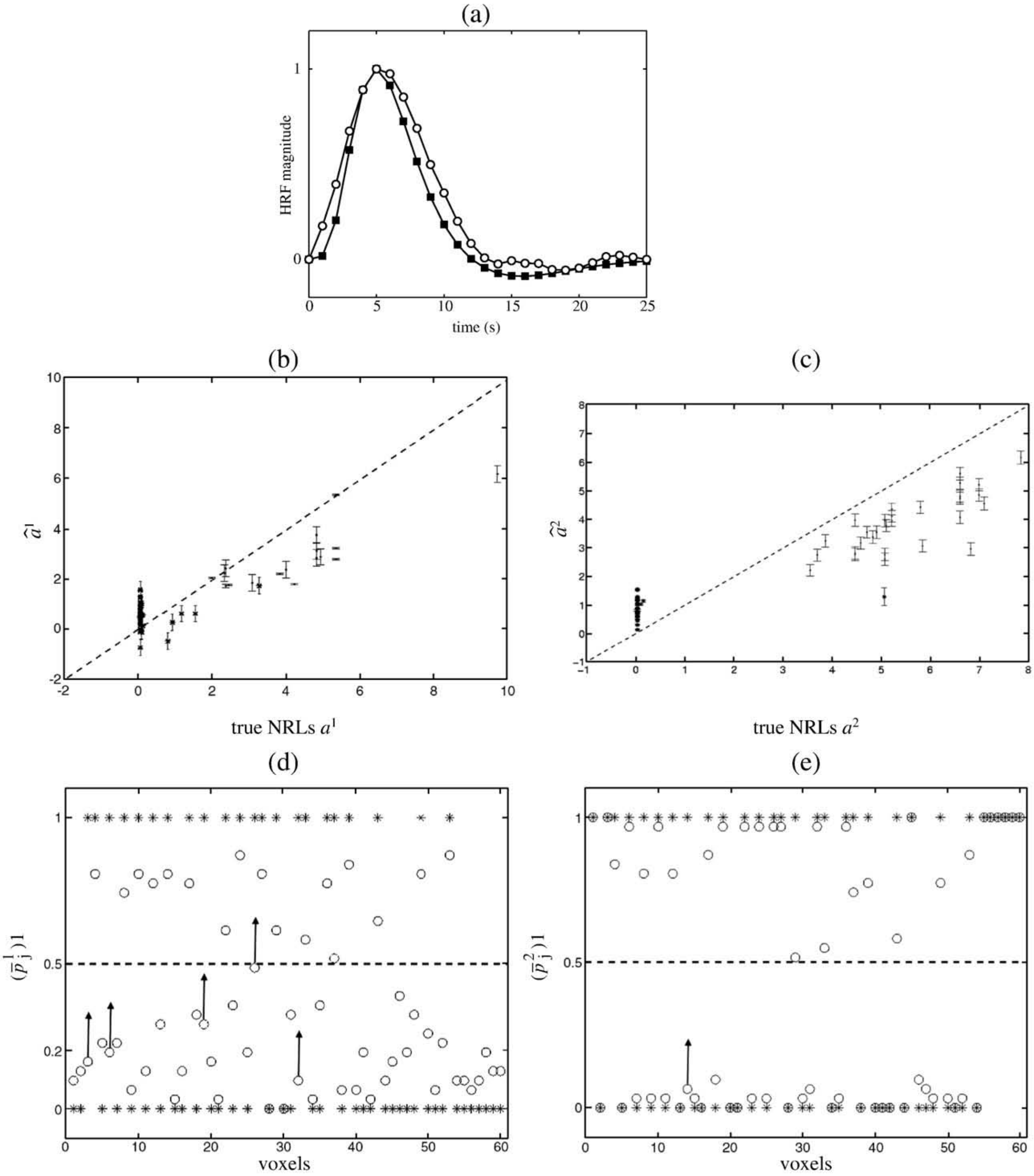


Fig. 4. Simulation results using model $\mathcal{M}_3$. The same legend as in Fig. 2 holds. FN voxels are indicated by upward arrows.

Author's personal copy

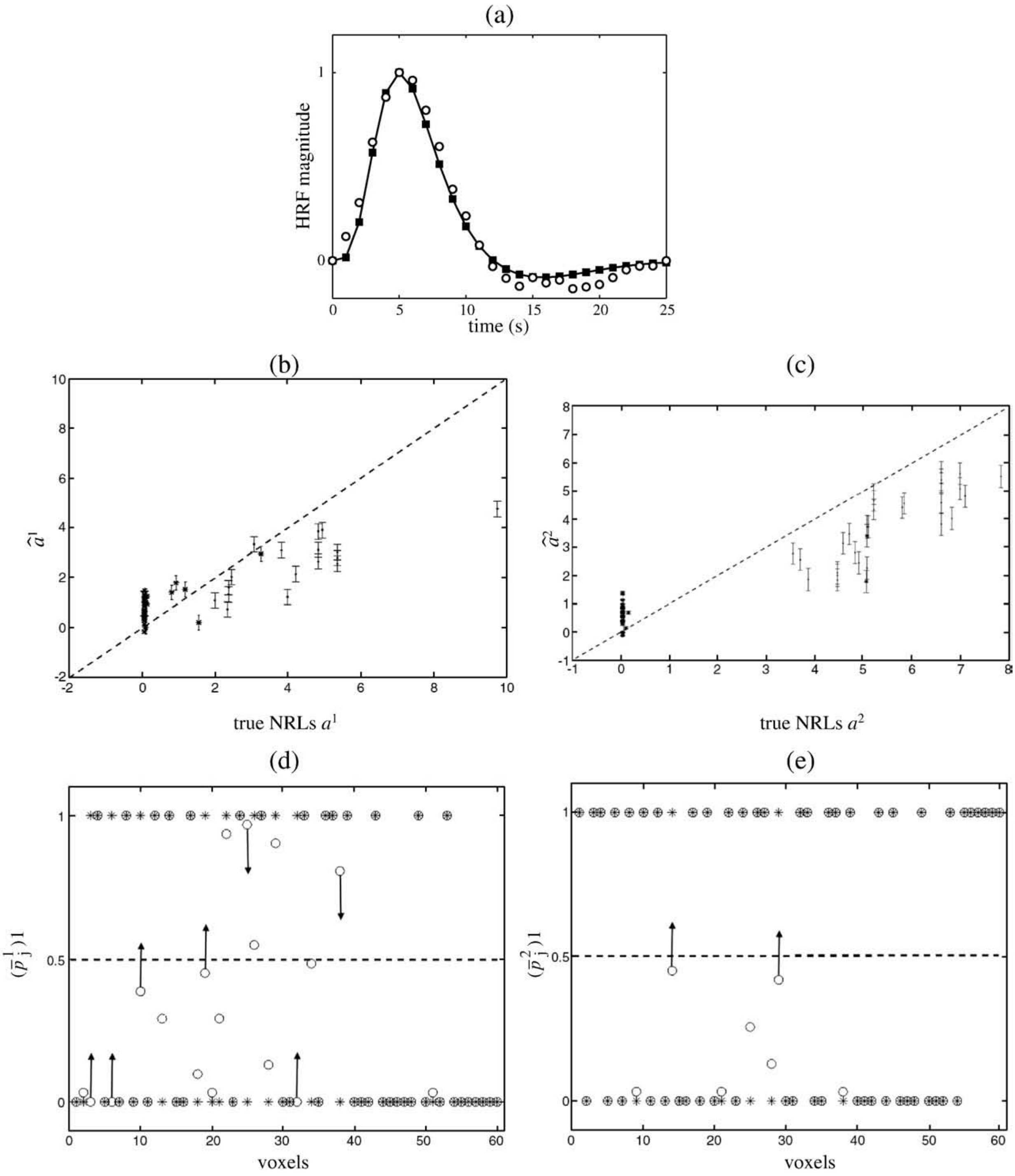


Fig. 5. Simulation results using model $\mathcal{M}_4$. The same legend as in Fig. 2 holds. FP and FN voxels are indicated by downward and upward arrows, respectively.

of the mean and variance parameters of the uncentered Gaussian distribution (results not shown).

We are now interested in assessing the differences between $\mathcal{M}_2$ and $\mathcal{M}_3$. The purpose of such a comparison is to decide whether or not a good mixture type provides more accurate and sensitive results than a precise noise modelling. Contrasting Figs. 3 and 4(b) allows us to note that $\mathcal{M}_2$ outperforms $\mathcal{M}_3$ in terms of accuracy of estimation for the first experimental condition. The NRLs attached to the non-activating voxels are over-estimated, leading to a much larger bias. In case of high SNR arising for the second condition, the comparison of Figs. 3 and 4(b) is less clear. The small NRLs are still over-estimated but the large ones are better estimated using $\mathcal{M}_2$ in some cases (*e.g.,* voxels 27, 58, 60). In terms of detection, Fig. 3(d) shows that a single false negative (voxel 3) is retrieved by model $\mathcal{M}_2$ for condition 1, while five FNs are found by model $\mathcal{M}_2$, as shown in Fig. 4(d) (voxels 3, 6, 19, 26, 32). Hence, model $\mathcal{M}_2$ achieves better results in terms of sensitivity and specificity. Therefore, we conclude that introducing an inhomogeneous prior

Author's personal copy

mixture is more powerful than modelling the serial correlation as regards both estimation and detection.

Receiver-operator-characteristic (ROC) curves have been also computed to quantitatively evaluate the differences between models $\mathcal{M}_1 - \mathcal{M}_4$. Fig. 6 illustrates and confirms that model $\mathcal{M}_1$ provides the most sensitive detection when specificity is fixed and a better specificity at a given sensitivity. These ROC curves also validate that model $\mathcal{M}_4$ is the less sensitive and the less specific out of the four models. Finally, model $\mathcal{M}_2$ outperforms $\mathcal{M}_3$ and provides better results in terms of sensitivity and specificity, irrespective of the stimulus type. Figs. 6(a)-(b) allows us to claim again that the noise model has a stronger impact in detection at low SNR since the distance between continuous and dotted lines is larger in Fig. 6(a) than in Fig. 6(b), except at very low specificity (0.1). This holds whatever the mixture type.

Deactivation modelling

Our purpose was to compare an inhomogeneous two-class mixture model with its three class extension. In the latter case, a third class is used to account for putative deactivation phenomenon arising for instance during sustained bursts of interictal epileptiform activity (Bagshaw et al., 2005; Bénar et al., 2006).

Suitable artificial fMRI datasets were simulated accordingly. We considered a ROI consisting of $J=60$ voxels. Let $J_{-1,m}$, $J_{0,m}$ $J_{1,m}$ be respectively the number of deactivating, non-activating and activating voxels in response to condition m. We set $(J_{1,1}, J_{-1,1})=(28, 19)$ and $(J_{1,2}, J_{-1,2})=(32, 12)$, so that $(J_{0,1}, J_{0,2})=(13, 16)$. We simulated the NRLs as follows:

$$\begin{aligned}&\text{activating voxels}: \boldsymbol{a}_j^1 \sim \mathcal{G}(3, 1),\ \boldsymbol{a}_j^2 \sim \mathcal{G}(5, 2)\\ &\text{non-activating voxels}: \boldsymbol{a}_j^{1,2} \sim \mathcal{N}(0, 0.1)\\ &\text{deactivating voxels}: -\boldsymbol{a}_j^1 \sim \mathcal{G}(5, 4),\ -\boldsymbol{a}_j^2 \sim \mathcal{G}(5, 4)\end{aligned}$$

The same procedure as before (see §4.1.1) was applied to simulate artificial fMRI time series. The only difference concerns the noise type, which is white, Gaussian and homogeneous in space to save computation time ($\forall j$, $\sigma_j^2=0.3$). Hence, model $\mathcal{M}_2$ and $\mathcal{M}_5$ (see Table 1) were tested and compared in terms of estimation, detection performance and evidence.

The HRF estimates corresponding to models $\mathcal{M}_2$-$\mathcal{M}_5$ are shown in Figs. 7 and 8(a), respectively. These estimated time courses appear very close to the true HRF shape. Figs. 7 and 8(b)-(c) show the NRL estimates related to conditions 1 and 2, computed for model $\mathcal{M}_2$ and $\mathcal{M}_5$, respectively. First, we observe that $\mathcal{M}_2$ provides under-estimated NRLs for activating voxels but over-estimated parameters for deactivating ones, irrespective of the stimulus type. The estimated error bars also appear significantly larger when deriving from $\mathcal{M}_2$. In contrast, model $\mathcal{M}_5$ provides more reliable NRL estimates with smaller error bars, as illustrated in Figs. 8-(b)(c). Also, the mean square error is decreased for the NRLs corresponding to deactivating and non-activating voxels.

Figs. 7(d)-(e) demonstrates that model $\mathcal{M}_2$ reports a few FN voxels (see upward arrows). All these voxels have small NRL coefficients, inducing their assignment to class 0. More importantly, we observe that the truly non-activating and deactivating voxels are mixed in class 0, irrespective of the condition. Figs. 8(d)-(e) reports the posterior mean estimates $(\bar{p}_j^m)_i$ (see (8)), which are then combined to get the final classification according to the MAP criterion $(\hat{q}_j^m)^{\text{MAP}}$ (see (9)). As indicated on these graphs, model $\mathcal{M}_5$ produces an accurate classification. Figs. 8(d)-(e) respectively show the presence of three FN voxels for condition 1 and only two FNs for condition 2. These classification errors could be explained by the low values taken by the true NRL coefficients in these voxels, making likely the assignment to class 0.

Finally, note that modelling the third class induces a higher computation time. In our simulations, inferring the parameters of models $\mathcal{M}_2$ and $\mathcal{M}_5$ takes about 6 and 11 minutes, respectively. If the ROI is large or if the experimental paradigm involves numerous conditions, it seems reasonable to start with a careful analysis of the paradigm to anticipate potential deactivations before inferring upon parameters of $\mathcal{M}_5$ instead of $\mathcal{M}_2$.

Bayesian model comparison

More formally, from a statistical point of view we compare models $\mathcal{M}_1 - \mathcal{M}_5$ by computing sample-based approximations to the model evidence $p(\mathrm{y}|\mathcal{M}_m)$. That allows us to derive Bayes factors BF_{mn} as ratios of model evidence (see Appendix C for computational details). Bayes factor provides us with good statistical summary for model comparison. As reported in Table 3, there is a strong evidence in favour of Model $\mathcal{M}_1$. More interestingly, our conclusion drawn from the parameter estimates are also confirmed when comparing $\mathcal{M}_2$ with $\mathcal{M}_3$ using Bayes factor (line 2, Table 3). This also holds for the comparison between the two-class and the three-class mixtures, $\mathcal{M}_2$ and $\mathcal{M}_5$ respectively (line 5, Table 3).

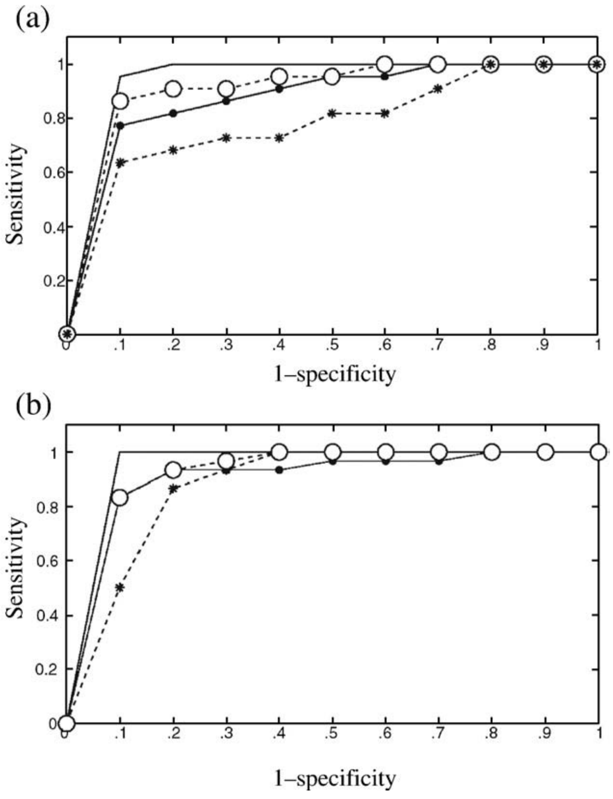


Fig. 6. (a)-(b): ROC curves associated to the four different models for condition 1 (a) and condition 2 (b), respectively. Continuous line, interrupted line with ○, continuous line with (a)-(b): ROC curves associated to the four different models for condition 1 (a) and condition 2 (b), respectively. Continuous line, interrupted line with ○, continuous line with ● and interrupted line with * represent the ROC curves for models $\mathcal{M}_1 - \mathcal{M}_4$, respectively.

Author's personal copy

Results on real fMRI data

fMRI experiment

MRI settings

The experiment was performed on a 3T whole-body system (Bruker, Germany) equipped with a quadrature birdcage radio frequency (RF) coil and a head-gradient coil insert designed for echo planar imaging (EPI). Functional images were obtained with a T2*-weighted GE-EPI sequence with an acquisition matrix at the 64×64 in-plane spatial resolution and 32 slices. A high-resolution ($1\times1\times1.2$ mm^3) anatomical image was also acquired for each subject using a 3-D gradient-echo inversion-recovery sequence.

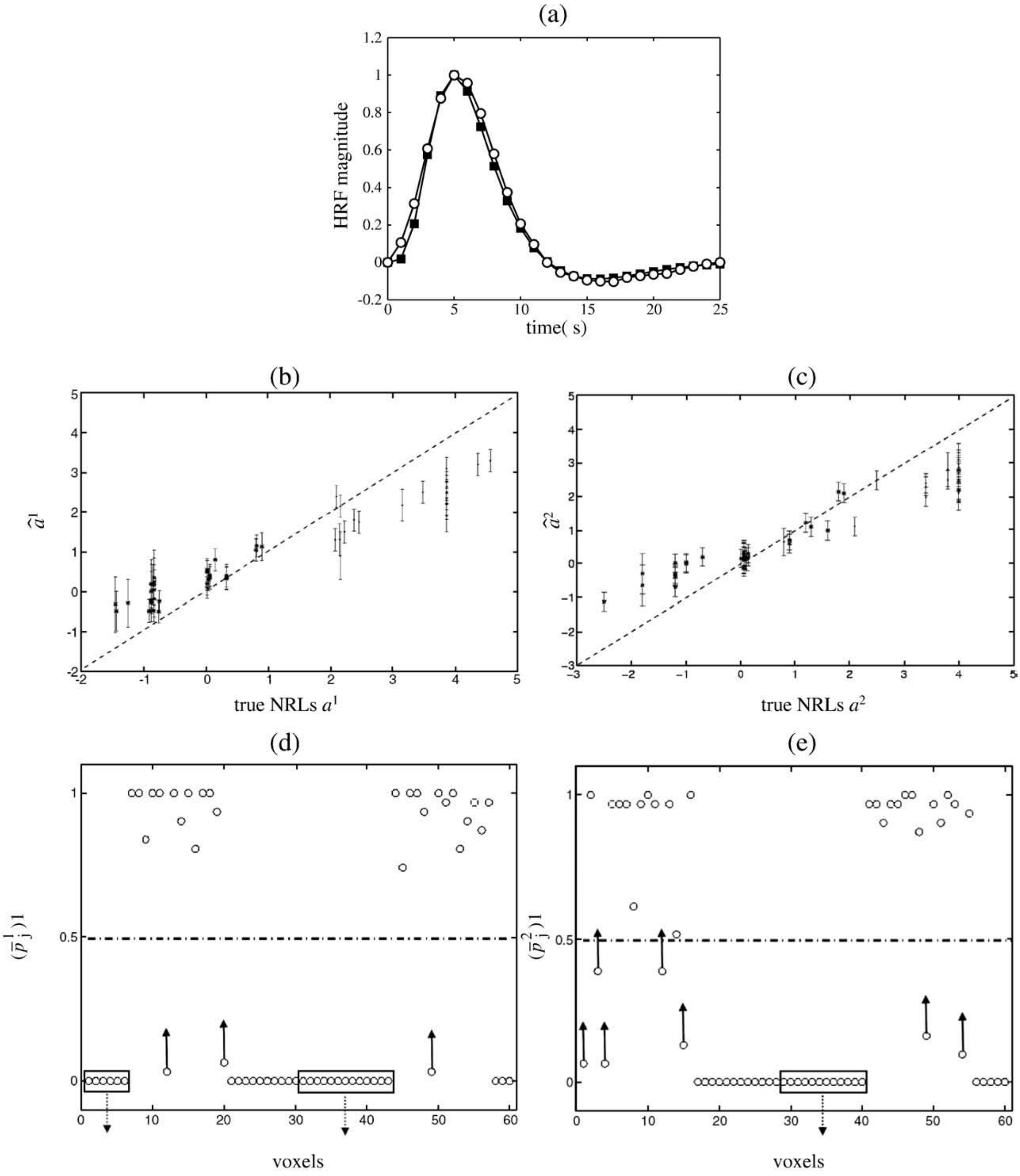


Fig. 7. Simulation results using model $\mathcal{M}_2$. FN voxels are indicated by upward arrows. The truly deactivating voxels that have been mixed in class 0 are surrounded by rectangle.

Author's personal copy

Experimental paradigm and contrast selection

The reader may refer to (Dehaene-Lambertz et al., 2006) for details about this fMRI experiment. In short, the motivation of this study was to measure the reduction in the neural activity subserving a cognitive representation when this representation is accessed twice (the so-called "repetition suppression" effect), resulting in a detectable adaptation of the measurable signal in fMRI (Grill-Spector and Malach, 2001; Naccache and Dehaene, 2001). The experiment consisted of a single session of $N=216$ scans lasting TR = 2.4 seconds each. Sixty sentences presented in a slow event-related design (SOA = 14.4 s) were recorded. Each sentence (S_1) could be repeated two (S_2), three (S_3) or four (S_4) times in a row. The main goal of our subsequent analysis was twofold. First, our primary interest was to exhibit regions which activation to a given sentence either decrease with repetition or keep a constant magnitude across the repetitions. Second, we were interested in inferring the hierarchical temporal organisation from the parcel-based HRF estimates along the superior temporal sulcus (STS).

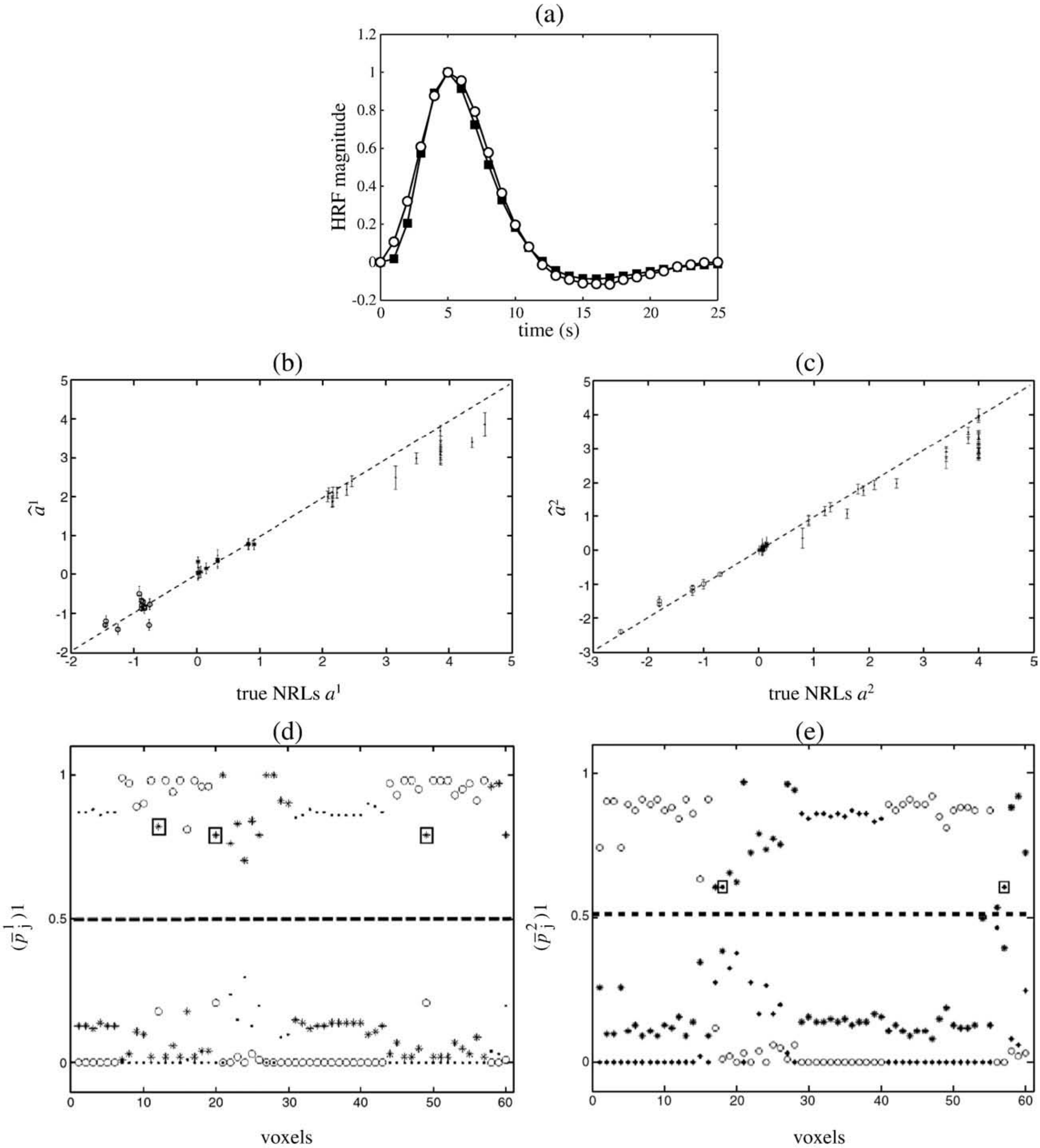


Fig. 8. Simulation results using model $\mathcal{M}_5$. The same legend as in Fig. 2 holds. FP and FN voxels are surrounded by rectangles (3 FNs in (d) and 2 FPs in (e)). ○, · and * symbols represent $(\bar{p}_j^m)_1$, $(\bar{p}_j^m)_{-1}$ and $(\bar{p}_j^m)_0$, respectively.

Author's personal copy

Table 3
Values of the integrated log-likelihood log $p(\mathbb{y} \mid \mathcal{M}_m)$ computed from a stabilized version of the harmonic mean identity (Raftery et al., 2007) for models $\mathcal{M}_m$ with $m \in \mathbb{N}_5^*$

Model m	log $p(\mathbb{y} \mid \mathcal{M}_m)$	Fig. #	log BF_{mn}, $n=1:4$			
$\mathcal{M}_1$	−199	Fig. 1	NR	18	316	400
$\mathcal{M}_2$	−217	Fig. 2	−18	NR	298	378
$\mathcal{M}_3$	−515	Fig. 3	−316	−298	NR	378
$\mathcal{M}_4$	−595	Fig. 4	−400	−378	−80	NR
$\mathcal{M}_2$	−700	Fig. 6	Log BF_{52}=356			
$\mathcal{M}_5$	−344	Fig. 7				

Model comparison based on the computation of Bayes factors log BF_{mn}= log $p(\mathbb{y} \mid \mathcal{M}_m)$ − log $p(\mathbb{y} \mid \mathcal{M}_n)$ for every pair (m, n). **NR** stands for **Not Relevant**.

Since the most significant habituation effect occurs between the first and second sentence repetitions, we modelled the four conditions S_1-S_4 but we only studied the contrast $S_1 - S_2$.

Pre-processings

As explained in Subsection 2.1.1, the grey matter's mask was first computed (see Fig. 9(a)) and then dilated using a 4 mm-radius sphere to account for the width of the cortical ribbon. Fig. 9(b) shows the result of this step. The resulting mask $\mathcal{M}_\mathrm{a}$ contains 19719 voxels at the fMRI resolution.

We checked that for nine out of ten subjects the raw fMRI data were motion-free approximately. All T_1-weighted MRI images were normalised onto the MNI template and functional images were transformed accordingly. fMRI volumes were also spatially smoothed using a Gaussian kernel with $FWHM=6$ mm along each direction. A first level analysis was performed for each subject using SPM2. The GLM modelled the four presentations of a given sentence with two regressors (the canonical HRF and its time derivative). Then, the parcellation was computed from the parameter estimates of this analysis. We chose a relevant F-contrast $\boldsymbol{c}=[1, 0, -1, 0, \cdots; 0, 1, 0, -1, \cdots, \ldots]$ to study the habituation effect etween the first and second presentations of a given sentence $(S_1 - S_2)$. Fig. 9(c) depicts an axial view of this parcellation for the same slice ($z=-4$ mm).

Our approach strongly relying on a functional homogeneity assumption, we started by comparing the results using increasing parcel numbers (from $K=100$ to $K=500$ parcels in $\mathcal{M}_s = \mathcal{M}_a \cup \mathcal{M}_f$. We checked that $K=200$ is large enough to guarantee a higher and sufficient degree of homogeneity. Here, the smallest and the largest parcels contained 44 and 190 voxels, respectively. Within each parcel, the degree of functional homogeneity was measured by computing a correlation matrix over the parameter estimates of the GLM. Note that this could also be done over the fMRI signals attached to each parcel.

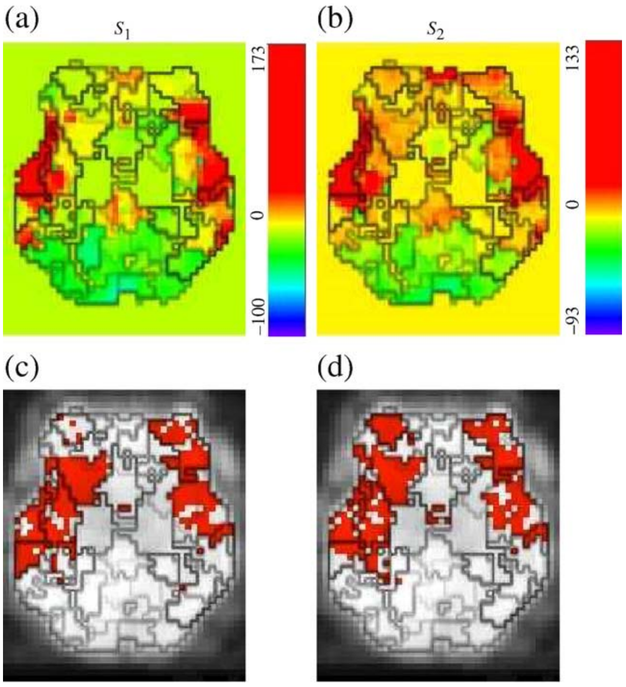


Fig. 10. (a)-(b) NRL estimates in one slice of subject 1's brain (at $z=-4$ mm) in response to S_1 (a) and S_2 (b). Values correspond to the NRL coefficients only for voxels belonging to $\mathcal{M}_s$. Otherwise they are equal to 0. (c)-(d): detection results in the same slice for S_1 and S_2, respectively. Voxels colour-coded in red are detected as activating. In black are the parcel borders that are superimposed to the different map results to show the parcellisation influence on the estimation of such parameters.

Results

Our method was tested on the nine datasets. Here, we only report results for Subject 1. Although the habituation effect and

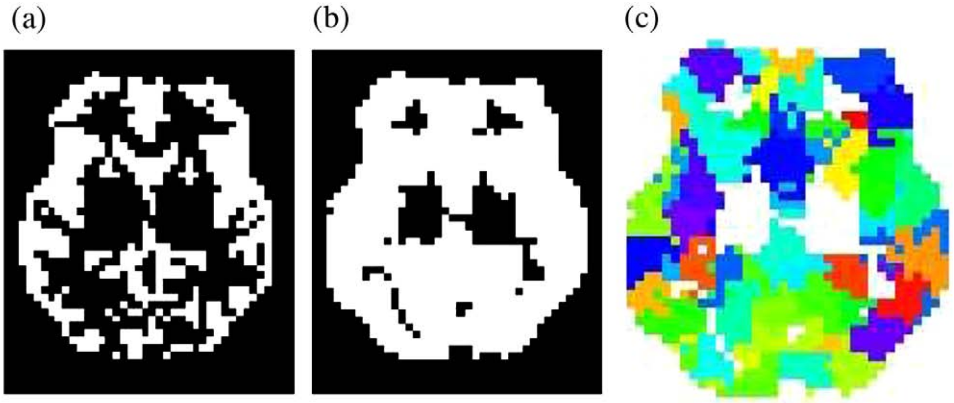


Fig. 9. (a): Slice of Subject 1's anatomical mask ($z=-4$ mm). (b): its dilated version $\mathcal{M}_a$ to match the functional resolution. (c): corresponding parcellation in the same slice. Each colour codes for a different parcel.

Author's personal copy

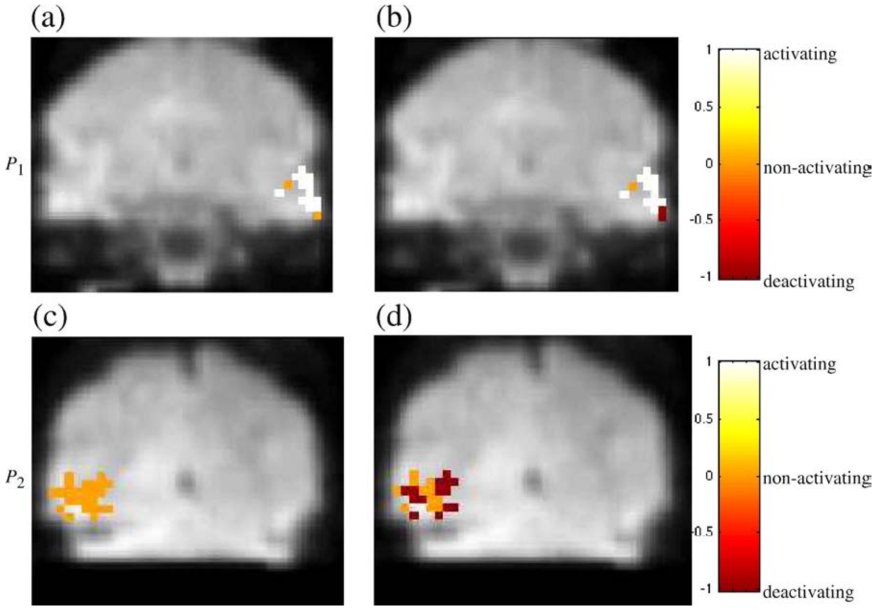


Fig. 11. Comparison of S_1-based classification maps between the *GaGMM* prior (left column, (a)-(c)) and its 3-class extension (right column, (b)-(d)) in two different parcels, $\mathcal{P}_1$ (top row, at $y=-16$ mm) and $\mathcal{P}_2$ (bottom row, at $y=-52$ mm). Square spots in white, orange and brown match with activating, non-activating and deactivating voxels, respectively.

brain dynamics (*i.e.,* the HRF shape) are subject to inter-individual variability in terms of spatial localisation and activation delay, the conclusions drawn for Subject 1 remain quite valid for the others.

In what follows, the proposed joint detection-estimation algorithm was applied to each parcel of Subject 1's brain. Figs. 10(a)-(b) shows the maps of the NRL estimates corresponding to conditions S_1 and S_2, respectively, in a given slice of the brain. In the same slice, Figs. 10(c)-(d) shows the activation probability maps attached to S_1 and S_2 (see (9) for details) our algorithm provides. Activating voxels appear in red colour.

Probing for putative deactivation

This first analysis was devoted to looking at putative deactivations, that is the presence of negative NRLs. We actually performed tests on all parcels to assess differences between the *GaGMM* prior and its 3-class extension. For illustrative purpose, Fig. 11 depicts the results of such a comparison on two parcels $\mathcal{P}_1$ and $\mathcal{P}_2$, composed of 129 and 135 voxels, respectively. Interestingly, the vast majority of voxels in $\mathcal{P}_1$ elicit a coherent activation in response to the first presentation of a sentence (S_1), while in $\mathcal{P}_2$, most voxels are non-activating. As shown in Fig. 11(a)-(b), the same S_1-based classification map is obtained in $\mathcal{P}_1$ irrespective of the mixture model. The same conclusion holds with respect to S_2 (results not shown). In $\mathcal{P}_2$, Figs. 11(c)-(d) illustrates that a few voxels move from the non-activating state to the deactivating one.

However, the corresponding NRL estimates are of small magnitudes indicating that this new classification may arise by chance. Bayesian model comparison statistically confirms our result since numerical evaluation of Bayes factors gives us log $BF_{52}=-1.2$ for

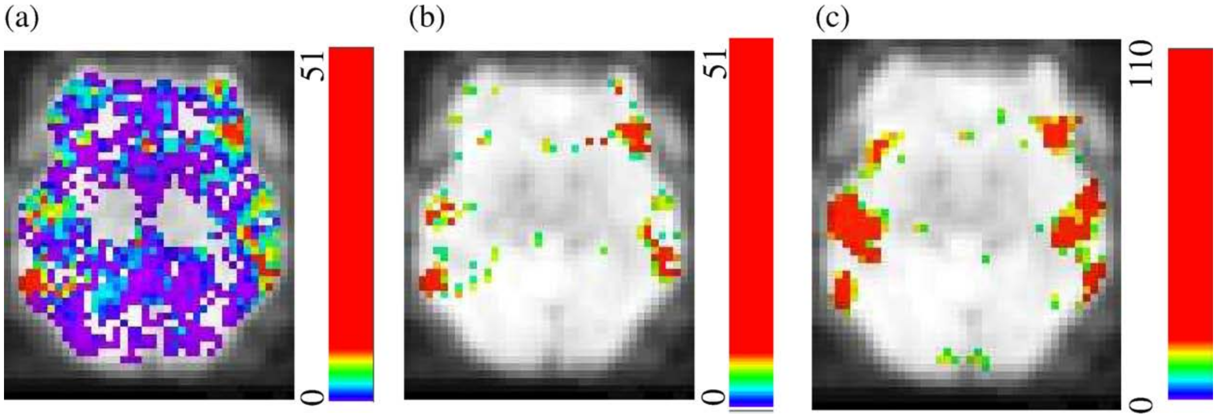


Fig. 12. Statistical maps are superimposed to a functional image and results are given in one slice of the brain (*z* Statistical maps are superimposed to a functional image and results are given in one slice of the brain ($z=-4$ mm). (a): KL-distance for voxels in $\mathcal{M}_s$ at $z=-4$ mm. (b): The MS most significant voxel KL-distance values. (c): The MS most significant voxel F-values.

Author's personal copy

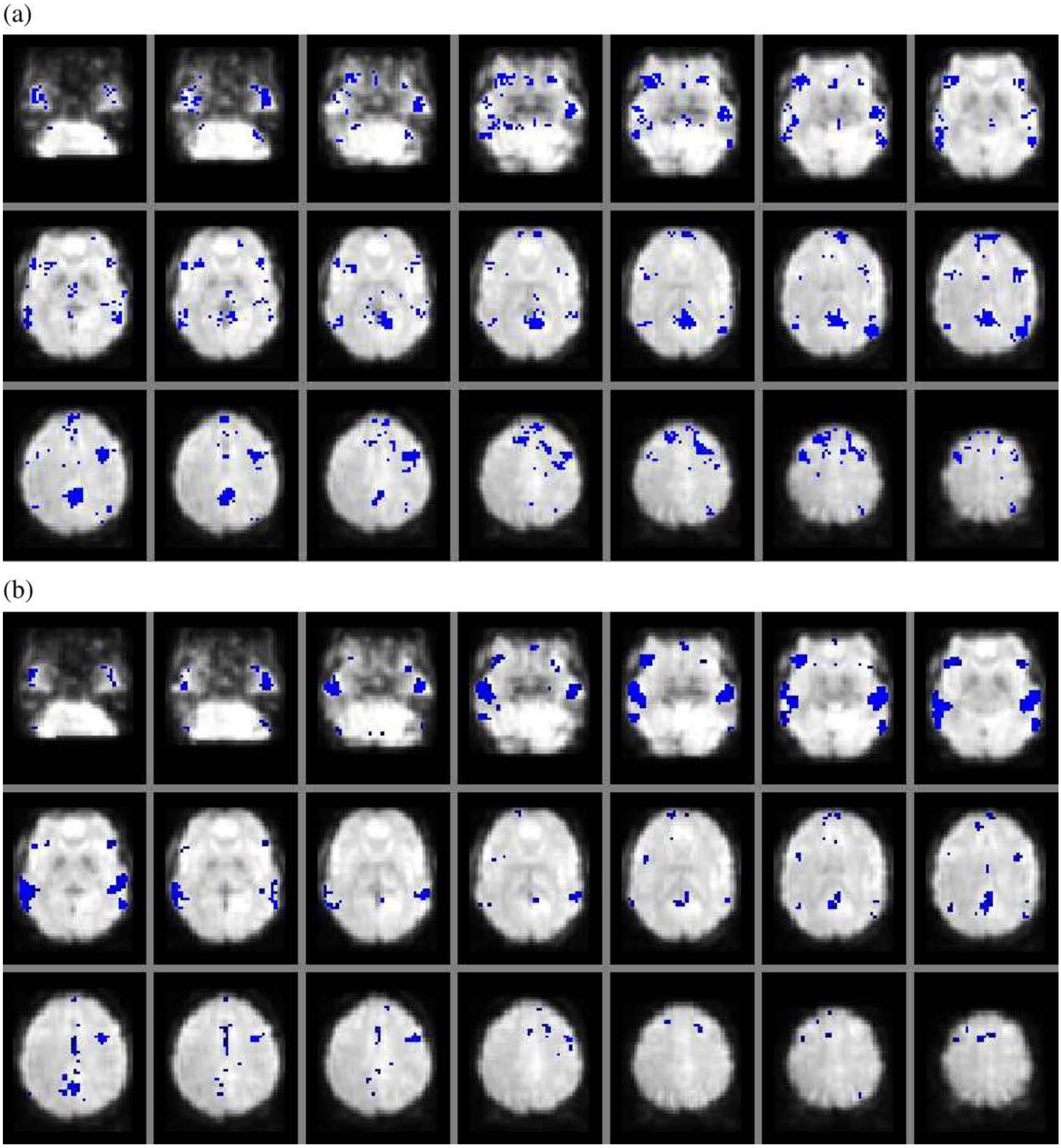
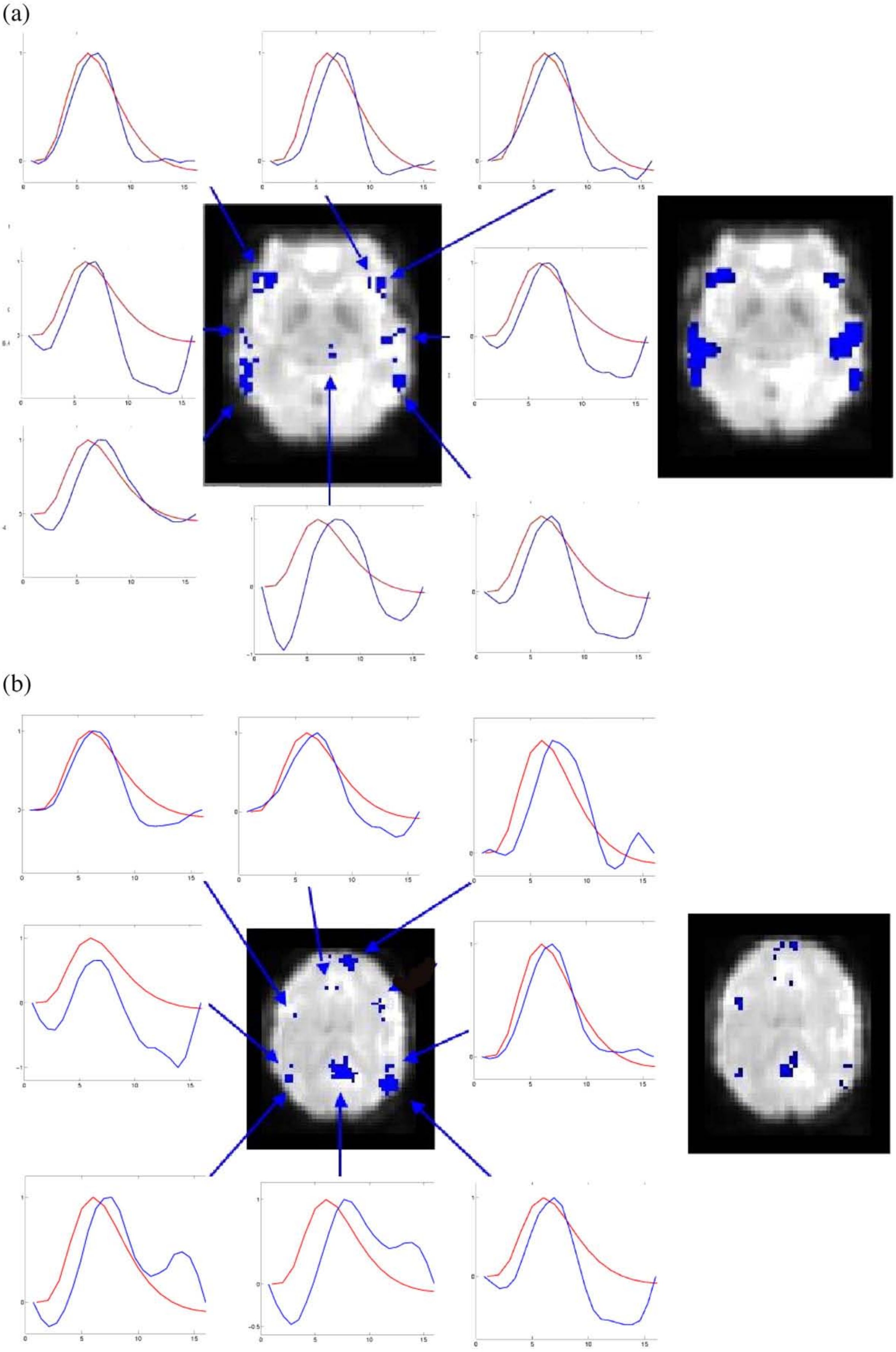


Fig. 13. From left to right and from top to bottom, brain slices organised along increasing axial axis (the bottom of the brain appear first in the superior left corner). Study of the contrast S_1. Study of the contrast $S_1 > S_2$. (a): thresholded PPM (see (11)) at $\eta = 0.999$; (b): thresholded SPM t-map at T=3.09 (corrected for multiple comparisons) obtained using SPM2.

$\mathcal{P}_1$ and log $BF_{52} = -.5$ for $\mathcal{P}_2$. These results show that there is less evidence in this dataset for supporting model $\mathcal{M}_5$. Therefore, the introduction of the third class in the mixture is not necessary to analyse these data, particularly in the brain regions involved in the treatment of phonological stimuli (language comprehension). In the rest of the paper, we restrict ourselves to the *GaGMM* prior.

Fig. 14. (a): Single-slice comparison (z=0 mm, slice 7 in Fig. 13) for the contrast $S_1 > S_2$. **left panels**: thresholded PPM at $\eta = 0.999$; **right panels**: thresholded SPM t-map at T=3.09 (corrected for multiple comparisons) obtained from the SPM analysis. (b): same comparison in slice z=24 mm (slice 13 in Fig. 13). HRF estimates (in blue) in these regions compared to the canonical hemodynamic response function used in SPM (in red).

Author's personal copy



Author's personal copy

Unsigned comparison between conditions

The result of our KL-distance map is shown in Fig. 12(a) in one slice of the brain ($z = -4$ mm). In Figs. 12(b)-(c), the KL-distance map is then compared to the standard F-test map by extracting the most significant, say MS, voxels from both volumes, according to the corresponding criteria (the largest KL divergence and the highest F value). This thresholding procedure shows common features in the activation patterns but also discrepancies in the temporal lobes that elicit different responses. On this slice, the KL-based criterion provides less activations. This may be due either to our approximation of the KL divergence or to the criterion itself.

Signed comparison between conditions

The mapping of the habituation phenomenon calls for *signed* comparisons since we are looking for voxels where a significant decrease of NRLs between S_1 and S_2 can be observed. Fig. 13 shows such a comparison over the whole brain between the PPM derived using the proposed methodology and the thresholded T-map obtained using SPM2. The activated regions (in blue) for the contrast $S_1 > S_2$ elicit therefore a higher response when a sentence is presented only once.

Statistical differences appear between the proposed PPM and the corresponding SPM. On the one hand, in the majority of slices we observe more activations on the PPM shown in Fig. 13(a). To a certain extent, these differences can be explained by the shape variations of the HRF estimate and its deviation from the canonical shape prescribed in SPM. Moreover, the correction for multiple comparisons used for the t-map may dramatically reduce the number of activating voxels. In contrast, no correction has been applied over the PPM.

On the other hand, in the temporal lobes (top row, right slices of Fig. 13(b)), sensitivity of detection seems better on the Student-t map: the activated clusters appear larger. Note also the presence of isolated activating voxels in Fig. 13(a). It is likely that this reflects the presence of false positives. To circumvent this issue and decrease the FP rate in the PPM, a spatial correlation model between neighbouring voxels can be introduced; see (Vincent et al., 2007b,a) and the discussion.

In Fig. 14, we represent the estimated HRFs in parcels corresponding to areas where we notice sensitivity differences. It clearly appears that in some parcels our HRF estimates exhibit unexpected timing properties. For instance, in central regions ($x = 0$,

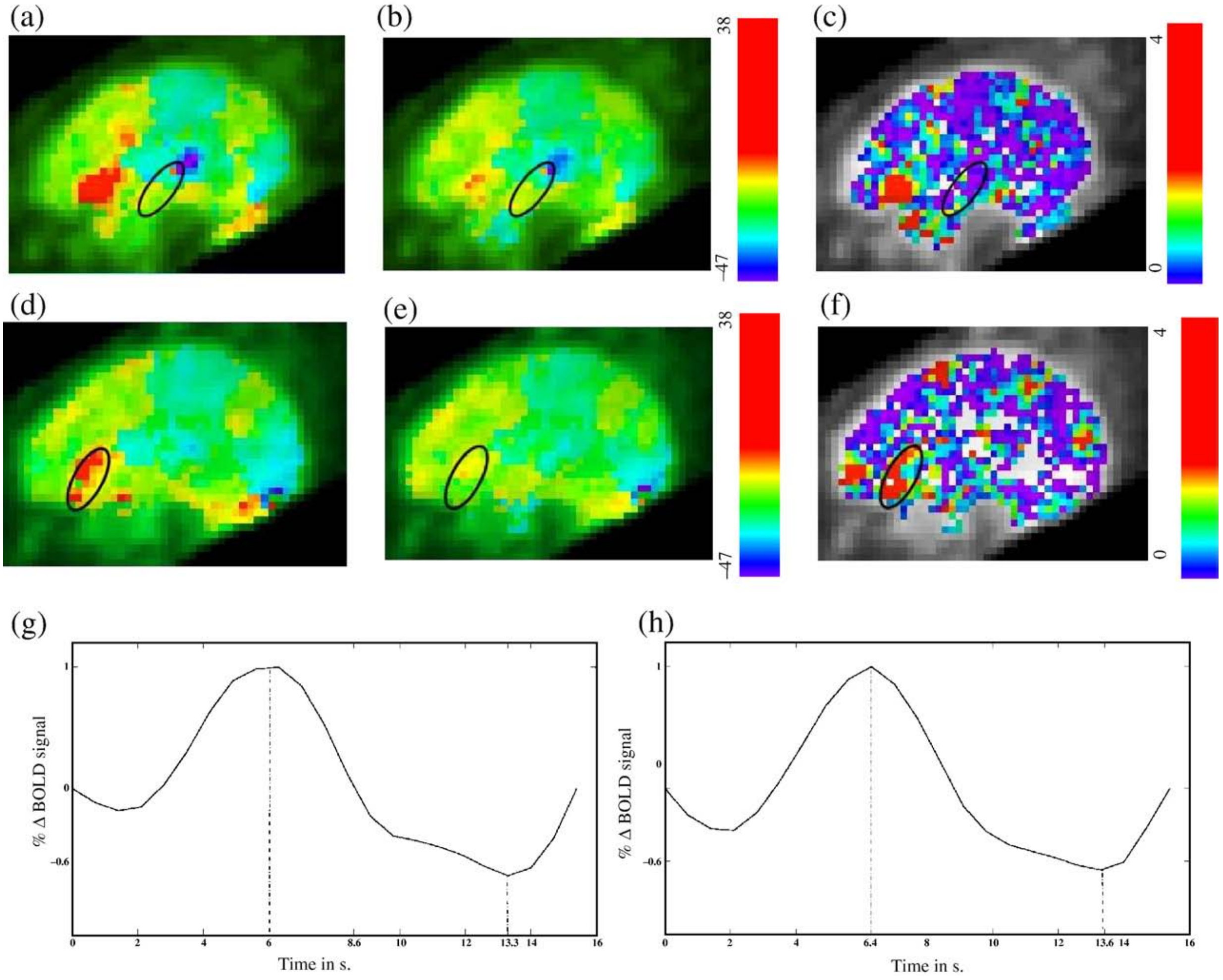


Fig. 15. (a)-(b) and (d)-(e): NRL estimates in the sagittal slices located (a)-(b) and (d)-(e): NRL estimates in the sagittal slices located at $x = -48$ mm and $x = -40$ mm, respectively. (a)-(b) provide the magnitudes in response to S_1 while (d)-(e) give us the NRLs in response to S_2. (c)-(f): KL-distance maps between the corresponding NRLs. Images are superimposed to functional data. The parcels containing the Heschl's gyrus (top row) and Broca's area (bottom row) are surrounded in black and referenced as $\mathcal{P}_{He}$ and $\mathcal{P}_{Br}$ in the text. (g)-(h): HRF estimates in $\mathcal{P}_{He}$ and $\mathcal{P}_{Br}$, respectively.

Author's personal copy

$y<0$), we obtained initial dips that were difficult to predict in advance. This requires further analysis (see Discussion). In regions where the HRF estimate is very close to the canonical shape, the PPM and SPM provide similar activation patterns. Finally, some regions were also detected as activating ($S_1>S_2$) by both methods while there is no evidence in the literature to suppose a priori that they elicit responses to auditory stimuli (see for instance Fig. 14(b) along the interhemispheric axis).

Habituation and temporal organisation

We focus on superior temporal regions ranging from the primary auditory cortex (Heschl's gyrus) to associative areas (middle and posterior STS). Although it has been shown in (Dehaene-Lambertz et al., 2006) that repetitions affect both amplitude and delay of responses, we only model habituation effect on the NRLs by considering the different sentence presentations as different conditions. This procedure is not optimal but remains quite simple.

The first interesting region is Heschl's gyrus located in the primary auditory cortex around voxel with coordinates (−48, −12, 0) mm in the standard Talairach space. This area shows the same response magnitude each time a sentence was presented. Figs. 15(a)-(b) shows the NRL estimates in parcel $\mathcal{P}_{He}$, which is circled in black for S_1 and S_2, respectively. In every voxel of $\mathcal{P}_{He}$, these magnitude parameters are very close to each other making the KL-distance between the marginal posterior distributions of S_1 and S_2 close to zero (see Fig. 15(c)). Hence, the measured difference between S_1 and S_2 is not statistically significant. Summary statistics computed over the NRL estimates in $\mathcal{P}_{He}$ are reported in Table 4 (left col.) and confirm our first analysis quantitatively. The same study was done in $\mathcal{P}_{Br}$ containing Broca's area and centered around voxel (−40, 24, 0) mm in the Talairach space. Figs. 15(d)-(e) clearly indicates a strong decrease in the NRLs between S_1 and S_2. The higher value of the KL divergence reported in Fig. 15(f) confirms the presence of a significant habituation effect in $\mathcal{P}_{Br}$. Our quantitative analysis (see Table 4, right col.) also shows the same trend as outlined by the strong discrepancies between the mean and standard deviations of S_1 and S_2.

In (Dehaene-Lambertz et al., 2006), a sine-wave GLM was designed and fitted to study the speed of habituation. It allowed one to exhibit a temporal organisation of the temporal lobe with fastest responses located in Heschl's gyrus and slowest ones in temporal poles. Here, given the proposed methodology, the temporal organisation is studied more directly by measuring and sorting the timing properties of the parcels $\mathcal{P}_{He}$ and $\mathcal{P}_{Br}$.

In these parcels, we also investigated the temporal aspects of the STS organisation by measuring different features (time-to-peak T_{peak}, time-to-undershoot $T_{\text{undershoot}}$) on the estimated HRFs (see Figs. 15 (g)-(h)). We computed these quantities over each parcel ($\mathcal{P}_{He}$ and $\mathcal{P}_{Br}$ and their four closest neighbours) separately before averaging them to get a mean estimate. We found that the responses in $\mathcal{P}_{He}$ ($T_{\text{peak}}=6$ s. and $T_{\text{undershoot}}=13.3$ s.) occur and return to the baseline earlier than the responses in $\mathcal{P}_{Br}$ ($T_{\text{peak}}=6.4$ s. and $T_{\text{undershoot}}=13.6$ s.). After averaging these quantities over the five parcels, we obtained congruent results ($T_{\text{peak}}=6.1$ s. and $T_{\text{undershoot}}=13.2$ s. around $\mathcal{P}_{He}$ vs. $T_{\text{peak}}=6.5$ s. and $T_{\text{undershoot}}=13.5$ s. around $\mathcal{P}_{Br}$) meaning that the region embedding Hesch's gyrus elicits brain activations faster than the region including Broca's area. This confirms more directly what has been already derived in (Dehaene-Lambertz et al., 2006) although no statistical test is provided to assess the significance of this result. The next question concerns of course the putative reasons of these earlier responses in Hesch's gyrus. There is a large evidence in these datasets for supporting faster neurodynamics as the main origin of these results instead of faster hemodynamics. This can be checked for instance using complementary analysis like a phase analysis conducted in (Dehaene-Lambertz et al., 2006).

Table 4
Summary statistics in parcels containing Heschl’s gyrus and Broca’s area

Statistics	$\mathcal{P}_{He}$, Heschl’s gyrus		$\mathcal{P}_{Br}$, Broca’s area	
	S_1 ($m=1$)	S_2 ($m=2$)	S_1 ($m=1$)	S_2 ($m=2$)
$\max_j\,(\hat{a}_j^m)^{PM}$	39.45	31.29	46.57	22.33
$\min_j\,(\hat{a}_j^m)^{PM}$	0.35	−0.88	−4.21	−0.32
Mean $(\hat{a}_j^m)^{PM}$	7.96	3.76	19.81	9.91
Median $(\hat{a}_j^m)^{PM}$	5.09	1.44	19.95	10.49
Std $(\hat{a}_j^m)^{PM}$	7.62	6.72	8.95	4.92

Discussion

In this paper, we have proposed an original method to perform a parcel-based joint detection-estimation of brain activity from fMRI data. It has been shown on simulated datasets that a gamma-Gaussian mixture as prior pdf on the NRLs outperforms a Gaussian mixture in terms of sensitivity/specificity trade-off. It has also been reported that the noise model has an influence over this compromise, particularly at lower SNR: a first-order AR model provides lower false positive and negative rates in comparison with a white noise model.

Our method extends previous works (Makni et al., 2005, 2006b,a) to deal with anatomically informed whole brain analysis. As already done in (Smith et al., 2003; Nieto-Castanon et al., 2003; Flandin et al., 2002), analysis was constrained to the mask of the grey matter obtained from a segmentation of the T1-weighted MRI. Our approach also relies on functional homogeneity assumptions at a regional scale that can be assessed either from the fMRI time series themselves or from the GLM parameter estimates. To meet these conditions, we resort to an automatic parcellation technique developed in (Thirion et al., 2006) but alternative clustering strategies may be thought of. Our approach therefore depends on this prior decomposition making the global within-subject analysis a two-steps procedure. The quality of the parcels will have an impact on the model fitting and a slight modification of the parcellation may generate different results especially in case of identifiability problems. By varying the number of parcels, we have checked that our results remain quite stable for different parcellations. The solution to this problem actually lies in the coupling of the parcellation procedure with our detection-estimation approach. This a very appealing direction of research but remains beyond the scope of the present work. At the expense of an increased computational complexity, the two steps could be merged in a combined approach through a hierarchical Bayesian model: one might be interested in improving the parcellation from the results of the detection-estimation stage using an iterative strategy: neighbouring parcels would be grouped if their underlying hemodynamics share similar features. The algorithm should take place in the context of reversible jumps MCMC to properly handle fusion/segragation moves between parcels Green (1995); Richardson and Green (1997). Besides, the parcellation identification issue could also be attacked using triplet Markov fields (Benboudjema and Pieczynski, 2007), which seem suitable for modelling nonstationarities in image segmentation.

A strong feature of our approach is the possibility to derive parcel-based HRF time courses throughout the brain. It allows us to assess the spatial variability of the HRF shape and to check that this shape greatly fluctuates across parcels. Since the parcellation

Author's personal copy

procedure is derived at the group level, one is able to compare subject specific HRFs in a given parcel. Doing so, we have noticed that the between-subject variability in the HRF shape seems to be larger than the within-subject spatial variability, as already suggested in (Handwerker et al., 2004).

Our results also suggest that the modelling of spatial HRF fluctuations is important to segregate brain regions involved in the experimental paradigm (Heschl's gyrus, Broca's area, …). The adaptation effect was particularly evident between the first and second sentences. The pattern of adaptation was different across regions with a set of regions demonstrating the same response each time a sentence was presented (i.e., Heschl's gyrus) and regions showing a more or less strong decrease between the first and second presentation (eg, Broca's area, superior STS).

Nonetheless, there exist fMRI experiments for which the proposed approach may fail because of the inhomogeneity of the HRF shape both in space and across conditions at a regional scale. This may occur when neurodynamic and hemodynamic fluctuations intermix. Indeed, recent studies of the fine structure of the fusiform face area (FFA) (Grill-Spector et al., 2006) have shown that the FFA is actually highly heterogeneous. It appears that the FFA is composed in reality of several small scale subregions that respond strongly, not only to faces, but also to cars and sculptures. The subregions discovered in (Grill-Spector et al., 2006) were associated with very distinct HRFs. For such studies, the best way to capture these HRF fluctuations is to perform a voxel-based non-parametric analysis on the basis of which statistical comparisons can be done across conditions, as demonstrated in (Ciuciu et al., 2003; Marrelec et al., 2004).

We have also distinguished some differences between our PPMs and the SPMs derived from a classical GLM-based analysis. Our results confirm the interest of a simultaneous procedure for detecting and estimating brain activity. The proposed procedure actually improves the sensitivity of detection in some regions where the temporal characteristics (time-to-peak, time-to-undershoot, …) of our HRF estimate deviate from those of the canonical shape. Unfortunately, this effect is not systematic on the datasets since in other regions a loss of sensitivity was observed. The reasons underlying this unexpected decrease have to be identified. To elucidate this issue, future work will be devoted to the comparison of a degraded version of our joint detection-estimation procedure with the actual one. The degraded version corresponds to an inference scheme where the HRF is maintained fixed and is not sampled at every iteration of the Gibbs sampler. We will quantify statistically this sensitivity difference on simulated data where we know exactly the “ground” truth. For doing so, we could generate fMRI data over the whole brain using the fMRI simulator developed by (Drobnjak et al., 2006).

In the present paper, we get rid of the scale ambiguity problem due to the bilinearity of (1) w.r.t. the pair ($\boldsymbol{h}$, $\mathbb{a}$) by imposing a unitary norm constraint over $\boldsymbol{h}$. This could have a dramatic impact on the convergence of our sampling scheme and then on the recovered HRF shape by distorting the target distribution of the MCMC scheme. Alternative strategies may first consist in cancelling this normalisation step. However, the resulting sampling scheme is too slow to converge in a reasonable amount of time (Veit and Idier, 2007). An efficient alternative has been proposed in (Veit and Idier, 2007). It has been applied to the joint detection-estimation of brain activity in (Ciuciu et al., 2007). It consists in adding to the MCMC procedure a sampling step of a positive scalar parameter s coding for the HRF scale. It can be shown that its sampling is fast, follows a *generalized inverse Gaussian* distribution in case of Gaussian mixtures and guarantees the theoretical convergence of the generated Markov chain to the posterior distribution. Deriving the target distribution of this scale parameter for inhomogeneous mixtures is beyond the scope of this paper.

To conclude about the real impact of our within-subject analysis, inference should take place at the group level. In other words, we should compare the results of two random effect analyses (RFX) based on the same group statistics (*e.g.,* mean effect) and statistical test (Student *t*-test). The first RFX analysis would correspond to the gold standard, in which the input data are given by the normalised effects of a standard individual SPM analysis. The second analysis would take the results of our algorithm for each subject as inputs. Of course, this is only feasible in case of multi-subjects parcellation, what is currently obtained using the procedure described in (Thirion et al., 2006, 2007).

From a methodological point of view, we have shown that our joint detection-estimation technique is able to identify deactivations in the brain. This is owing to the introduction of a third class in the prior mixture model associated to the NRLs. Nonetheless, we did not exhibit real deactivations on the analysed datasets. In the future, we should therefore validate the 3-class extension on specific datasets. A good candidate could be a dataset acquired during an event-related auditory paradigm in which silence events are presented randomly to compare activations to a baseline derived from such events. As already shown in (Ciuciu et al., 2003), silence events may generate deactivations in the temporal lobe if they are presented to the subject when the gradients of the scanner are switched off. This will be the subject of further work.

Smoothing the data spatially provides $\mathbb{a}$ reliable manner for recovering clusters of activation instead of isolated spots, at the expense of a loss of resolution. To avoid this preprocessing, the proposed method could be extended by introducing spatial correlation in the prior model. This could be done either on the NRLs ($\mathbb{a}$) or on the underlying states (labels $\mathbb{q}$). We argue in favour of the second solution for simplicity reasons. As already derived for Gaussian mixtures in (Vincent et al., 2007b,a), it is quite simple to sample from an Ising (2-class model) or Potts (3-class model) Markov random field (MRF) that enforce neighbouring voxels to be classified in the same state (*e.g.,* activating). This approach actually seems more reasonable in terms of computational load than considering edge-preserving MRF based on non-quadratic potentials (Green, 1990; Geman and McClure, 1987). Also, for computational reasons this extension has been developed in a supervised framework meaning that the hyper-parameter encoding spatial regularity of the hidden MRF is set by hand. Future work will be focused on an spatially adaptive extension in which this parameter is estimated as well, as already done in (Woolrich et al., 2005; Woolrich and Behrens, 2006).

Another extension that could be introduced at little expense concerns the analysis on the cortical surface, as proposed in (Andrade et al., 2001). This will probably improve the sensitivity of detection by constraining the analysis to the cortical surface. Such study needs first a segmentation of the anatomical MRI, then an extraction of the grey-white matter interface (*e.g.,* as a mesh), and finally requires an interpolation of the fMRI signal on the nodes of the mesh (see for instance (Grova et al., 2006) for a suitable approach).

Finally, the model presented here assumes that the NRLs are constant in time. Hence, to account for putative habituation effects, it requires to model repetitions of the same stimulus as different

Author's personal copy

experimental conditions, what may be not optimal in terms of sensitivity of detection. To account for trial-varying NRLs due to adaptation or learning effects arising either as a direct consequence of the paradigm or as a alteration of subject's arousal, the proposed model can be generalised in a way that makes the number of unknown parameters not too large. As proposed in (Ciuciu et al., 2006), habituation can be modelled at the voxel level by a pair of parameters: the NRL to the first trial of the stimulus and a mean habituation speed across the consecutive trials that follows a hyperbolic parametric model depending on the inter-stimulus intervals.

Hopefully, all these additional points will induce improvements in the detection-estimation results and will help to a better comprehension of brain functions.

Acknowledgments

The authors are grateful to Guillaume Flandin, Tim Behrens and Mark Woolrich for fruitful discussion and helpful comments.

Appendix A. Densities

We give the definitions of the densities used throughout this paper. We also provide numerical recipes for efficient simulations according to complex distributions when necessary.

A.1. Multivariate normal density

The multivariate Normal density for d-dimensional variable $\boldsymbol{x}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right). \quad \text{(A.1)}$$

The general formula of Kullback-Leibler (KL) divergence between a test density, say $q(\boldsymbol{x})$ and a reference density $p(\boldsymbol{x})$ is

$$KL(q||p) = \int q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} d\boldsymbol{x}. \quad \text{(A.2)}$$

For multivariate normal densities $q(\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu}_q,\boldsymbol{\Sigma}_q\right)$ and $p(\boldsymbol{x}) = N\left(\boldsymbol{x}|\boldsymbol{\mu}_q,\boldsymbol{\Sigma}_p\right)$, Eq. (A.2) takes the following form:

$$KL(q||p) = \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_p|}{|\boldsymbol{\Sigma}_q|} + \frac{1}{2}tr\left(\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\Sigma}_q\right) + \frac{1}{2}\left(\boldsymbol{\mu}_q-\boldsymbol{\mu}_p\right)^t\boldsymbol{\Sigma}_p^{-1}\left(\boldsymbol{\mu}_q-\boldsymbol{\mu}_p\right) - \frac{d}{2}.$$

For univariate normal densities $q(x) = \mathcal{N}\left(x|\mu_q,\sigma_q\right)$ and $p(x) = \mathcal{N}\left(x|\mu_p,\sigma_p^2\right)$, the KL (A.2) distance becomes:

$$KL(q||p) == \frac{\left(\sigma_p^2-\sigma_q^2\right)^2 + \left(\mu_p-\mu_q\right)^2\left(\sigma_p^2+\sigma_q^2\right)}{4\sigma_p^2\sigma_p^2}. \quad \text{(A.3)}$$

A.2. Positive normal density

The truncated normal distribution for a scalar variable $x \in \mathbb{R}_+$ with *parameters* (m, υ) expresses as follows:

$$\begin{cases} \mathcal{N}^+(x|m,\upsilon) = C^{-1}\exp\left(-(x-m)^2/2\upsilon\right)\mathbb{I}_{\mathbb{R}_+}(x), \\ C = \sqrt{\pi\upsilon}\left[1+\text{erf}\left(m/\sqrt{2\upsilon}\right)\right]/\sqrt{2}, \end{cases} \quad \text{(A.4)}$$

where erf is the error function (Abramowitz and Stegun, 1970, p. 297): $\text{erf}(z) = 2/\sqrt{\pi}\int_0^z e^{-t^2}dt$. Parameter m defines the mode of the density if $m>0$. Note that the knowledge of C is not required for simulating a realisation of the density. Indeed, its mean μ and variance σ^2 are given by

$$\mu = m + \sqrt{\frac{2v}{\pi}}\frac{\exp(-m^2/2v)}{1+\text{erf}\left(m/\sqrt{2v}\right),} \quad \text{(A.5)}$$

$$\sigma^2 = \upsilon + \frac{m^2}{4} - \left[\mu - \frac{m}{2}\right]^2. \quad \text{(A.6)}$$

Hence, the standard inversion technique of the cumulative distribution function (Devroye, 1986; Gelfand et al., 1992) may be used. First, it consists in simualting an uniform variate u~$\mathcal{U}([0,1])$ and then in calculating:

$$x = m + \sqrt{2\upsilon}\ \text{erf}^{-1}\left(u + \text{erf}\left(m/\sqrt{2\upsilon}\right)(u-1)\right),$$

The erf function is approximated numerically in practice. In cases where the approximation error becomes important (*i.e.,* when $|m|$ is too large), this simulation method is inefficient. Instead, we use efficient alternatives which are based on accept-reject algorithms (Robert, 1995), the most powerful relies on multiple instrumental distributions (Mazet et al., 2005)[9].

A.3. Gamma density

The gamma density for variable $x\in\mathbb{R}_+$ with shape parameter $\alpha>0$ and scale parameter $\beta>0$, is defined by

$$\mathcal{G}(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x)\mathbb{I}_{\mathbb{R}_+}(x) \quad \text{(A.7)}$$

where $\Gamma(x)$ is the gamma function defined as

$$\Gamma(x) = (x-1)\Gamma(x-1) = \int_0^\infty t^{x-1}e^{-t}dt. \quad \text{(A.8)}$$

The mean and variance are respectively given by E $[x]=\alpha/\beta$ and var $[x]=\alpha/\beta^2$, while the mode of the distribution is given by $M_0=(\alpha-1)/\beta$. Particular cases of the gamma distribution are the *Erlang distribution* $\mathcal{G}(x|\alpha,1)$, the *exponential* distribution $\mathcal{E}(x|\beta) = \mathcal{G}(x|1,\beta)$ and the chi-squared distribution $\mathcal{G}(x|v/2,1/2)$ denoted by χ_v^2. The Erlang distribution may be used in pratice as a preliminary step for simulating a gamma variate x~$\mathcal{G}(\alpha,\beta)$ since we get a right sample if $x=u/\beta$ and u~$\mathcal{G}(\alpha,1)$.

Note also that the inverse gamma distribution, denoted by $\mathcal{IG}(\alpha,\beta)$ throughout the paper is the distribution of x^{-1} when x~$\mathcal{G}(\alpha,\beta)$.

A.4. Beta density

The Beta density for variable $x \in [0, 1]$ with shape parameter $\alpha>0$ and scale parameter $\beta>0$, is defined by

$$\mathcal{B}e(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\mathbb{I}_{[0,1]}(x) \quad \text{(A.9)}$$

[9] This code available on-line at http://www.iris.cran.uhp-nancy.fr/francais/si/Personnes/Perso Mazet/rpnorm-fr.htm.

Author's personal copy

where B (α, β) is the Beta function:

$$B(x)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

The mean and variance are respectively given by

$$E[x]=\frac{\alpha}{\alpha+\beta},\quad \text{and}\quad \mathrm{var}[x]=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

The mode of the Beta distribution evolves w.r.t. the domains of (α, β):

$$M_0=\begin{cases}\dfrac{\alpha-1}{\alpha+\beta-2} & \text{if } \alpha>1 \text{ and } \beta>1\\ 0 \text{ and } 1 & \text{if } \alpha<1 \text{ and } \beta<1\\ 0 & \text{if } \begin{cases}\alpha<1 \text{ and } \beta\geq 1\\ \alpha=1 \text{ and } \beta>1\end{cases}\\ 1 & \text{if } \begin{cases}\alpha\geqslant 1 \text{ and } \beta<1\\ \alpha>1 \text{ and } \beta=1\end{cases}\\ \text{does not exist if } \alpha=\beta=1 & \end{cases}$$

Importantly, for simulating Beta-distributed random variate $x\sim\mathcal{B}e(\cdot|\alpha,\beta)$, we proceeds as follows. First, generate two exponential variables $e_1\sim\mathcal{E}(\cdot|\alpha)$ and $e_2\sim\mathcal{E}(\cdot|\beta)$. Second, compute x as $e_1/(e_1+e_2)$.

A.5. gamma-Gaussian density

The non-standard gamma-Gaussian density for variable $x\in\mathbb{R}_+$ with shape $\alpha>0$, mean $\mu>0$ and variance parameters $\upsilon>0$, is defined by

$$\mathcal{GN}(x|\alpha,\mu,\upsilon)=K_\alpha^{-1}x^{\alpha-1}\exp\left(-(x-\mu)^2/2\upsilon\right)\mathbb{I}_{\mathbb{R}_+}(x)\qquad \text{(A.10)}$$

where K_α is the normalising constant:

$$\begin{aligned}K_\alpha&=\int_0^{+\infty}x^{\alpha-1}\exp\left(-(x-\mu)^2/2\upsilon\right)dx\\&=\exp\left(-\mu^2/4\upsilon\right)\upsilon^{\frac{\alpha}{2}}\Gamma(\alpha)\mathcal{D}_\alpha\left(-\mu/\sqrt{\upsilon}\right)\end{aligned}\qquad \text{(A.11)}$$

The last equation follows from (Gradshteyn and Ryzhik, 1994, p. 337, Eq 3.462-1) that relies upon the gamma function (see (A.8)) and the Parabolic Cylinder functions D_v (Gradshteyn and Ryzhik, 1994, p. 885, Eq 7.711-1 and p. 1065)[10]. Importantly, the first two centered moments of a gamma-Gaussian random variable $x\sim\mathcal{GN}(x|\alpha,\mu,\upsilon)$ can be computed analytically:

$$\mathrm{E}[x|\alpha,\mu,\upsilon]=\frac{1}{K_\alpha}\int_0^{+\infty}x^\alpha\exp\left(-(x-\mu)^2/2\upsilon\right)dx=\frac{K_{\alpha+1}}{K_\alpha}\qquad \text{(A.12)}$$

$$\mathrm{var}[x|\alpha,\mu,\upsilon]=\mathrm{E}\left[x^2|\alpha,\mu,\upsilon\right]-\mathrm{E}[x|\alpha,\mu,\upsilon]^2=\frac{K_{\alpha+2}}{K_\alpha}-\frac{K_{\alpha+1}^2}{K_\alpha^2}.\qquad \text{(A.13)}$$

Simulating gamma-Gaussian random variables $x\sim\mathcal{GN}(\cdot|\alpha,\mu,\upsilon)$ is not straightforward and thus more effortful as compared to sampling from standard laws. Following (Moussaoui et al., 2006), to solve for this problem, we resort to a Metropolis-Hastings algorithm which needs the specification of an instrumental distribution q (Hastings, 1970; Robert, 2001). To avoid high rejection rate, this instrumental pdf has to be chosen to fit the target distribution $f=\mathcal{GN}(\cdot|\alpha,\mu,\upsilon)$ at best. In this regard, expression (A.10) is useful to characterise$f$ in terms of mode, mean or variance from which the instrumental distribution q may be adjusted. Calculating the first-order derivative of (A.10) w.r.t. x and equating to zero, the mode of f is obtained as the solution of the following second order equation:

$$x^2-\mu x-\upsilon(\alpha-1)=0,\quad \text{subject to}\quad x\geqslant 0.$$

Let us denote $\Delta=\mu^2+4\upsilon$ ($\alpha-1$). The mode of f, which is non-negative by definition, expresses as follows:

$$v=\begin{cases}0 & \text{if } \Delta<0\\ \max\left(\left(\mu+\sqrt{\Delta}\right)/2,0\right) & \text{otherwise.}\end{cases}\qquad \text{(A.14)}$$

Hence, the instrumental density q is taken as a truncated normal distribution $\mathcal{N}^+(x|v,\upsilon)$, which is easier to sample from as detailed in Appendix A.2.

A.6. Multinomial density

The density of a multinomial discrete distribution for variable $\boldsymbol{x}=\{x_1,\ \dots,\ x_N\}$ with parameters $\boldsymbol{\pi}=\{\pi_1,\ \dots,\ \pi_N\ \}$ is defined by

$$\mathcal{MN}_M(\boldsymbol{x}|\boldsymbol{\pi})=\frac{M!}{\prod_{i=1}^N x_i!}\prod_{i=1}^N\pi_i^{x_i}\mathbb{I}_{\sum_i x_i=M}$$

where $x_i\geq 0$, $\pi_i>0$ and $\Sigma_{i=1}^{\mathrm{N}}\ \pi_i=1$.

A.7. Dirichlet density

Let $\boldsymbol{\delta}=\{\delta_1,\ \dots,\ \delta_N\}$ be some positive parameters. The probabilistic density function of the N-state Dirichlet distribution for variable $\boldsymbol{\pi}=\{\pi_1,\dots,\pi_N\}$ satisfying $\pi_i\geqslant 0$ with parameters $\boldsymbol{\delta}$, is defined by

$$\mathcal{D}_N(\boldsymbol{\pi}|\boldsymbol{\delta})=\frac{\Gamma\left(\sum_{i=1}^N\delta_i\right)}{\prod_{i=1}^N\Gamma(\delta_i)}\prod_{i=1}^N\pi_i^{\delta_i-1}\mathbb{I}_{\sum_i\pi_i=1},$$

where $\Gamma(x)$ is, as before, the gamma function (A.8). Parameters δ_i are *prior observation counts* for events governed by π_i. The Dirichlet distribution is the conjugate prior of the parameters of a multinomial distribution. One special case is the *symmetric* Dirichlet distribution where $\delta_i=\delta_0\ \forall i$. In this case, the density becomes

$$\mathcal{D}_N(\boldsymbol{\pi}|\delta_0)=\frac{\Gamma(N\delta_0)}{\Gamma(\delta_0)^N}\prod_{i=1}^N\pi_i^{\delta_0-1}.$$

The real vector (X_1/S, … , X_N/S) follows a Dirichlet distribution denoted as $\mathcal{D}_N$ ($\cdot|\ \boldsymbol{\delta}$) if $X_i\sim\mathcal{G}(\delta_i,\ \beta)$ are independent, and $S=\Sigma_i X_i$. This holds true for any β, so in practice we choose $\beta=1$. This result is very useful in practice for simulating realisations of a Dirichlet process.

Appendix B. Computational details for the MCMC procedure

In this section, we derive the full conditional distributions of the quantities ($\boldsymbol{h}$, ($\mathbb{a}$, $\mathbb{q}$), ℓ and Θ) to be sampled. When the sampling

[10] See also http://mahieddine.ichir.free.fr for implementation of Parabolic Cylinder functions.

Author's personal copy

procedure w.r.t. a given parameter cannot be implemented as a Gibbs sampling step, we provide the reader with the derivations of some relevant instrumental distribution needed in the corresponding Metropolis Hastings move.

B.1. The HRF **h** *and its scale* σ_h^2

Let us denote $\boldsymbol{S}_j = \Sigma_m a_j^m \boldsymbol{X}^m$. $\boldsymbol{h}$ is $\mathcal{N}$ ($\boldsymbol{\mu}_h$, Σ_h)-distributed with:

$$\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1} = \sigma_{\boldsymbol{h}}^{-2}\boldsymbol{R}^{-1} + \sum_{j=1}^{J} \boldsymbol{S}_j^t \boldsymbol{\Lambda}_j \boldsymbol{S}_j \text{ and } \boldsymbol{\mu}_{\boldsymbol{h}} = \boldsymbol{\Sigma}_{\boldsymbol{h}} \sum_{j=1}^{J} \sigma_{\varepsilon_j}^{-2} \boldsymbol{S}_j^t \boldsymbol{\Lambda}_j \left(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j\right). \tag{B.1}$$

Variance $\sigma_{\boldsymbol{h}}^2$ is simulated according to $p(\sigma_{\boldsymbol{h}}^2|\boldsymbol{h}) = \mathcal{IG}(D/2, \boldsymbol{h}^t\boldsymbol{R}^{-1}\boldsymbol{h}/2)$.

B.2. The nuisance variables $\boldsymbol{\ell}$ *and their scale* $\sigma_{\boldsymbol{\ell}}^2$

Vectors $\boldsymbol{\ell}_j$ being independent ($j = 1 : J$), they can be sampled in parallel according to $N\left(\boldsymbol{\mu}_{\ell_j}, \boldsymbol{\Sigma}_{\ell_j}\right)$ where

$$\boldsymbol{\Sigma}_{\ell_j}^{-1} = \sigma_{\ell}^{-2} I_Q + \sigma_{\varepsilon_j}^{-2} \boldsymbol{P}^t \boldsymbol{\Lambda}_j \boldsymbol{P} \text{ and } \boldsymbol{\mu}_{\ell_j} = \sigma_{\varepsilon_j}^{-2} \boldsymbol{\Sigma}_{\ell_j} \boldsymbol{P}^t \boldsymbol{\Lambda}_j \left(\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h}\right). \tag{B.2}$$

Variance $\sigma_{\boldsymbol{\ell}}^2$ is simulated according to $\mathcal{IG}((QJ+1)/2, \Sigma_j \|\ell_j\|^2/2)$.

B.3. The voxelwise mixtures ($\mathbb{q}$, $\mathbb{a}$)

Although we do not introduce any spatial correlation between the NRLs, the latter are sampled one-at-a time since the distribution $p(a_j^m \mid rest)$[11] depends on $\boldsymbol{a}_j^{\backslash m} = [a_j^1, , \ldots , a_j^{m-1}, a_j^{m+1}, \ldots , a_j^M]$ due to the linearity of model (1) with respect to $\boldsymbol{a}_j$. The sampling of the NRLs is therefore implemented through two nested loops, the inner corresponding to the stimulus types (*e.g.,* index m) and the outer to voxels (*e.g.,* index j). Since the full conditional posterior $p(a_j^m \mid \text{rest})$ is a mixture, its sampling can be achieved in two steps. The first one consists in drawing a realisation of class q_j^m while the second one proceeds conditionally on class q_j^m. To carry out the first step, we need to identify the posterior mixture in voxel j and for condition m:

$$p\left(a_j^m|rest\right) \propto \exp\left(-\frac{1}{2\sigma_{\varepsilon_j}^2}||\boldsymbol{e}_{j,m} - a_j^m \boldsymbol{g}_m||_{\boldsymbol{\Lambda}_j}^2\right) \sum_{i=-1}^{1} \lambda_{i,m} f_i\left(a_j^m|\boldsymbol{\theta}_{i,m}\right) \tag{B.3}$$

where $\boldsymbol{g}_m = \boldsymbol{X}^m \boldsymbol{h}$, and $\boldsymbol{e}_{j,m} = y_j - \boldsymbol{P}\boldsymbol{\ell}_j - \Sigma_{n \neq m} a_j^n \mathbf{g}_n$.

After straightforward calculations including the normalisation of (B.3), we get the following expression:

$$p\left(a_j^m|rest\right) = \sum_{i=-1}^{1} \lambda_{i,j}^m f_i\left(a_j^m|\theta_{i,m}^m\right), \quad \text{with :} \tag{B.4}$$

$$f_0\left(a_j^m|\boldsymbol{\theta}_{0,j}^m\right) = \mathcal{N}\left(a_j^m|\mu_{0,j}^m, \upsilon_{0,j}^m\right) \tag{B.5}$$

$$f_1\left(a_j^m|\boldsymbol{\theta}_{1,j}^m\right) = \mathcal{GN}\left(a_j^m|\alpha_{1,m}, \mu_{1,j}^m, \upsilon_{1,j}^m\right) \tag{B.6}$$

$$f_{-1}\left(a_j^m|\boldsymbol{\theta}_{-1,j}^m\right) = -\mathcal{GN}\left(-a_j^m|\alpha_{-1,m}, -\mu_{-1,j}^m, \upsilon_{-1,j}^m\right) \tag{B.7}$$

The mixing probabilities are given by

$$\lambda_{i,j}^m = \left(1 + \sum_{i' \neq i} \widetilde{\lambda}_{i',j}^m / \widetilde{\lambda}_{i,j}^m\right)^{-1}, \quad \forall i = -1 : 1$$

$$\widetilde{\lambda}_{0,j}^m = \lambda_{0,m}\left(\upsilon_{0,j}^m/\upsilon_{0,m}\right)^{1/2} \exp\left(\left(\mu_{0,j}^m\right)^2/2\upsilon_{0,j}^m\right),$$

$$\widetilde{\lambda}_{i,j}^m = \lambda_{i,m}\frac{\beta_{i,m}^{\alpha_{i,m}}}{\Gamma(\alpha_{i,m})} K_{i,j}^m \exp\left(\left(\mu_{i,j}^m\right)^2/2\upsilon_{i,j}^m\right), \quad \text{for} \quad i \in \{-1, 1\},$$

and depend on the normalising constants $K_{i,j}^m$, for $i = \pm 1$, of the corresponding gamma-Gaussian densities; see (A.11) for its closed-form expression.

The parameters of the conditional posterior densities $f_i(a_j^m \mid \boldsymbol{\theta}_{i,j}^m)$ are given by:

$$\begin{cases} \upsilon_{1,j}^m = \upsilon_{-1,j}^m = \sigma_{\varepsilon_j}^2 \left(\boldsymbol{g}_m^t \boldsymbol{\Lambda}_j \boldsymbol{g}_m\right)^{-1}, \quad \upsilon_{0,m}^m = \left(\upsilon_{0,m}^{-1} + \left(\upsilon_{1,j}^m\right)^{-1}\right)^{-1} \\ \mu_{i,j}^m = \upsilon_{1,j}^m\left(\sigma_{\varepsilon_j}^{-2} \boldsymbol{g}_m^t \boldsymbol{\Lambda}_j \boldsymbol{e}_{m,j} - i\beta_{i,m}\right), \quad \forall i = -1 : 1. \end{cases} \tag{B.8}$$

Sampling the class q_j^m first amounts to generating $u_j^m \sim \mathcal{U}([0, 1])$ and then to applying the following rules:

$$q_j^m \begin{cases} -1 \text{ if } \quad u_j^m \leqslant \lambda_{-1,j}^m, \\ 0 \text{ if } \quad \lambda_{-1,j}^m < u_j^m \leqslant \lambda_{-1,j}^m + \lambda_{0,j}^m, \\ 1 \text{ otherwise.} \end{cases}$$

Once q_j^m is correctly set, it remains to sample from the conditional distribution $f_i(\cdot \mid \boldsymbol{\theta}_{i,j}^m)$ as suggested by (B.4). If $q_j^m = 0$, this operation is straightforward because $f_0(\cdot \mid \mu_{0,j}^m, \upsilon_{0,j}^m)$ is Gaussian (*cf.* (B.5)). However, if $q_j^m = \pm 1$, this operation is computationally more expensive since are gamma-Gaussian; see Appendix A.5 for details. Strictly speaking, the sampling of $f_{-1}(\cdot \mid \boldsymbol{\theta}_{1,j}^m)$ consists first in simulating a realisation of $(-a_j^m)$ using a well-suited positive normal density and then negating that realisation.

Interestingly, when $\alpha_{i,m} = 1$, which corresponds to taking an exponential prior for the NRL distribution, the conditional posterior density $f_i(\cdot \mid \boldsymbol{\theta}_{i,j}^m)$ is exactly a truncated normal distribution with parameters equal to those of p_i. In that case, the Metropolis-Hastings is not necessary since all proposals are accepted (the acceptation rate equals to 1). The sampling of the truncated normal density can be achieved efficiently as detailed in Appendix A.

B.4. Mixture probabilities

Since we have $\Pr(q_j^m = i \mid \boldsymbol{\lambda}_m) = \lambda_{i,m}$, for $i = -1 : 1$ and because the prior is conjugate, *i.e.,* a symmetric Dirichlet distribution $D(\boldsymbol{\lambda}_m \mid \delta)$ with $\delta > 0$, the full conditional posterior distribution is also Dirichlet and reads:

$$p\left(\boldsymbol{\lambda}_m|q_j^m = i, \delta\right) \propto Pr\left(q_j^m = i|\boldsymbol{\lambda}_m\right)p(\boldsymbol{\lambda}_m|\delta) \sim D(\boldsymbol{\delta}'), \text{ with } \delta_i' = \delta + 1, \text{ and } \delta_l' = \delta, \ \forall l \neq i.$$

[11] *rest* stands for the “remaining variables”.

Author's personal copy

The spatial correlation being not modelled in (6), we may write $\Pr(\boldsymbol{q}^m = \boldsymbol{\kappa} \mid \boldsymbol{\lambda}_m) = \Pi_j \Pr(q_j^m = \kappa_j \mid \boldsymbol{\lambda}_m)$, with $\kappa_j = -1:1$ and the joint posterior distribution $p(\boldsymbol{\lambda}_m \mid \boldsymbol{\kappa}, \boldsymbol{\delta})$ is given by

$$p(\boldsymbol{\lambda}_m | \boldsymbol{j}, \boldsymbol{\delta}) \propto \boldsymbol{\Pi}_j \Pr\left(q_j^m = \kappa_j | \boldsymbol{\lambda}_m\right) p(\boldsymbol{\lambda}_m | \boldsymbol{\delta}) \sim D(\boldsymbol{\delta}') \\ \text{with} \quad \delta_i' = \delta + \underbrace{\mathrm{Card}\left[C_{i,m} = \left\{j \in 1 : J | q_j^m = i\right\}\right]}_{= J_{i,m}}. \tag{B.9}$$

B.5. Mixture hyper-parameters

Variance $\upsilon_{0,m}$ is very easy to sample because $p(\upsilon_{0,m} \mid z^m) = \mathcal{IG}((J_{0,m} - 1)/2,\ \nu_{0,m}/2)$, where $\nu_{0,m} = \Sigma_{j \in C_{0,m}} (a_j^m)^2$.

For the two other classes, we proceed to the sampling of the scale and shape gamma distribution parameters $\alpha_{\pm 1,m}$ and $\beta_{\pm 1,m}$, respectively. Following (Moussaoui et al., 2006), we use a Metropolis-Hastings step for $\alpha_{\pm 1,m}$ parameters with a gamma instrumental density defined below. Simulating parameters $\beta_{\pm 1,m}$ is easier since they follow a gamma distribution.

The posterior density of each hyper-parameter $\alpha_{i,m}$ takes the form

$$p\left(\alpha_{i,m} | rest\right) \propto \prod_{j \in C_{i,m}} \frac{\beta_{i,m}^{\alpha_{i,m}}}{\Gamma\left(\alpha_{i,m}\right)} \left(a_j^m\right)^{\alpha_{i,m} - 1} p\left(\alpha_{i,m} | s_i\right) \propto g\left(\alpha_{i,m}\right)^{J_{i,m}} \mathbb{I}_{\mathbb{R}_+}\left(\alpha_{i,m}\right), \tag{B.10}$$

where

$$g\left(\alpha_{i,m}\right) = \exp\left(\tau_{i,m} \alpha_{i,m}\right) / \Gamma\left(\alpha_{i,m}\right), \\ \tau_{i,m} = \ln \beta_{i,m} + \sum_{j \in C_{i,m}} \left(a_j^m - s_i\right) / J_{i,m}.$$

This posterior distribution does not belong to a known family, so its simulation requires a MH jump. Akin to (Moussaoui et al., 2006), to obtain a good instrumental law $q(\alpha_{i,m})$, we propose to approximate function $g(\alpha_{i,m})$ using a gamma density $\mathcal{G}(t_{i,m},\ u_{i,m})$. More precisely, parameters $(u_{i,m},\ t_{i,m})$ of this density are determined in order for its mode and inflexion points match those of function $g(\alpha_{i,m})$. After some simple manipulations, we obtain:

$$t_{i,m} = 1 + \alpha_{\mathrm{mode}}^2 / \left(\alpha_{\mathrm{mode}} - \alpha_{\mathrm{infl}}\right)^2, \ u_{i,m} = \alpha_{\mathrm{mode}} / \left(\alpha_{\mathrm{mode}} - \alpha_{\mathrm{infl}}\right)^2, \tag{B.11}$$

where α_{mode} and α_{infl} are the mode and the superior inflexion point ($\alpha_{\mathrm{infl}} > \alpha_{\mathrm{mode}}$) of $g(\alpha_{i,m})$. Calculating the first and second derivatives of $g(\alpha_{i,m})$ yields these two non-linear equations that implicitly define α_{mode} and α_{infl}:

$$\psi\left(\alpha_{\mathrm{mode}}\right) = \tau_{i,m} \quad \text{and} \quad \psi^{(1)}\left(\alpha_{\mathrm{infl}}\right) = \left(\psi\left(\alpha_{\mathrm{infl}}\right) - \tau_{i,m}\right)^2, \tag{B.12}$$

where ψ is the digamma function defined by $\psi(x) = \frac{d}{dx} \log \Gamma(x)$ and $\psi^{(1)}$ is its first derivative (trigamma function). Details about these functions are provided in (Abramowitz and Stegun, 1970, p. 253). The resolution of the two Eq.s (B.12) is done using a root finding numerical method (*cf.* (Press et al., 1992, Ch. 9)). Finally, the posterior density (B.10) is simulated using a Metropolis-Hastings algorithm with a instrumental density $q(\alpha_{i,m})$ chosen as a gamma distribution $\mathcal{G}(t_{i,m}^{'}, u_{i,m}^{'})$ whose parameters are given by

$$t_{i,m}' = J_{i,m}\left(t_{i,m} - 1\right) + 1, \ u'_{i,m} = J_{i,m} u_{i,m}. \tag{B.13}$$

The sampling of shape parameters $\beta_{i,m}$ is done according to the full conditional posterior distribution

$$p\left(\beta_{i,m} | rest\right) \propto \beta_{i,m}^{\left(J_{i,m}\alpha_{i,m} + b_i\right)} \exp\left(-\beta_{\mathrm{i,m}}\left(\sum_{j \in C_{i,m}} a_j^m + c_i\right)\right) \\ \sim \mathcal{G}\left(J_{i,m}\alpha_{i,m} + b_i + 1, \quad \sum_{j \in C_{i,m}} a_j^m + c_i\right). \tag{B.14}$$

B.6. Noise variances

Sampling the noise variances $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2$ can be performed in parallel. Drawing a noise variance is straightforward because $p\left(\sigma_{\varepsilon_j}^2 | rest\right) = \mathcal{IG}\left((N+1)/2, || \widetilde{\mathbf{y}}_j ||_{\Lambda_j}^2 / 2\right)$.

B.7. AR parameters

For each voxel V_j, we have:

$$p_j\left(p_j\right) = p\left(p_j | rest\right) \propto \sqrt{1 - p_j^2} \ \exp\left(-a_j\left(p_j - m_j\right)^2\right) \mathbb{1}_{(-1,1)}\left(p_j\right), \tag{B.15}$$

where $a_j = A_j / 2\sigma_{\varepsilon_j}^2$ and $m_j = B_j \ / \ A_j$, with $A_j = \sum_{n=2}^{N-1} \widetilde{y}_{j,n}^2$ and $B_j = \sum_{n=1}^{N-1} \widetilde{y}_{j,n} \widetilde{y}_{j,n+1}$.

The density p_j is log-concave, unfortunately it does not seem to belong to a referenced family of pdf, from which an efficient sampling technique would be available. Here, we propose to resort to a Metropolis-Hastings independence algorithm that uses a beta pdf $g_j \sim Be(\zeta_j,\ \kappa_j)$ defined over $(-1,\ 1)$[12] as the instrumental distribution:

$$g_i(\rho) \propto (1 + \rho)^{\zeta_{j-1}} (1 - \rho)^{\kappa_{j-1}}, \quad \forall |\rho| < 1. \tag{B.16}$$

The parameters ζ_j and κ_j have to be tuned in an appropriate way, so that g_j approximates p_j as closely as possible. Here, ζ_j and κ_j are chosen in such a way that log g_j and log p_j have the same curvature around a common maximizer over $(-1,\ 1)$. Let us first remark that the maximizer r_j of log p_j is uniquely defined by

$$2a_j\left(r_j - m_j\right)\left(1 - r_j^2\right) + r_j = 0, \quad |r_j| < 1.$$

Moreover, r_j takes an explicit expression, as the root of a polynomial of degree three. Then, $(\zeta_j\ ,\ \kappa_j)$ can be found by solving

$$\begin{cases} \left(\log g_j\right)'\left(r_j\right) = 0 \\ \left(\log g_j\right)''\left(r_j\right) = \left(\log p_j\right)''\left(r_j\right) \end{cases}$$

which is a linear system. After some straightforward calculations, the solution can be expressed as follows:

$$\zeta_j = a_j\left(1 + r_j\right)^2\left(1 + m_j - 2r_j\right) + 3/2 \\ \kappa_j = a_j\left(1 - r_j\right)^2\left(1 - m_j + 2r_j\right) + 3/2$$

It can be practically checked that $g_j(\rho)$ and $p_j(\rho)$ take very similar values for all $\rho \in (-1,\ 1)$. Therefore, the proposal ρ'_j (sampled from g_j) has a high acceptance probability min $\{1, p_j\ (\rho'_j) g_j\ (\rho_j) / p_j(\rho_j) g_j\ (\rho'_j\ )$o. In practice, the worse acceptation ratio that we observed was about 0.92.

[12] If $x \in (0,\ 1)$ and $x \sim Be(\zeta,\ \kappa)$ then $\rho = 2x - 1$ is said $Be(\zeta,\ \kappa)$-distributed over $(-1,\ 1)$.

Author's personal copy

Appendix C. Bayesian model comparison

We have introduced Bayesian model comparison through the computation of Bayes factors:

$$\mathrm{BF}_{mn} \triangleq \frac{p\left(\mathrm{y}|\mathcal{M}_m\right)}{p\left(\mathrm{y}|\mathcal{M}_n\right)} = \frac{\int p\left(\mathrm{y}|\boldsymbol{\theta}_m, \mathcal{M}_m\right)p\left(\boldsymbol{\theta}_m|\mathcal{M}_m\right)\mathrm{d}\boldsymbol{\theta}_m}{\int p\left(\mathrm{y}|\boldsymbol{\theta}_n, \mathcal{M}_n\right)p\left(\boldsymbol{\theta}_n|\mathcal{M}_n\right)\mathrm{d}\boldsymbol{\theta}_n},$$

with $(m,n) \in \mathbb{N}_4^*$.

These quantities are thus computed as the ratios of integrated likelihoods or *model evidence* of the different models. Of course, they are approximated from the MCMC outputs using the metholodogy proposed in (Kass and Raftery, 1995; Chib, 1995; Chib and Jeliazkov, 2001) and further developed in (Raftery et al., 2007). The latter relies on the *harmonic mean identity*:

$$\frac{1}{p\left(\mathrm{y}|\mathcal{M}_*\right)} = \mathrm{E}\left[\frac{1}{p\left(\mathrm{y}|\boldsymbol{\theta}, \mathcal{M}_*\right)}|\mathrm{y}, \mathcal{M}_*\right]$$

This suggests that the model evidence can be approximated by the harmonic mean of the likelihoods $p(\mathrm{y} \mid \boldsymbol{\theta}^{(l)}, \mathcal{M}_*)$ based on L draws $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(L)}$ from the posterior distribution $p(\boldsymbol{\theta} \mid \mathrm{y}, \mathcal{M}_*)$:

$$p\widehat{(\mathrm{y}|\mathcal{M}_*)} = \left[\frac{1}{L}\sum_{l=L_0}^{L_1}\frac{1}{p\left(\mathrm{y}|\boldsymbol{\theta}^{(l)}, \mathcal{M}_*\right)}\right]^{-1} \quad (\mathrm{C.1})$$

with $L=L_1-L_0+1$. These sample might come out of a standard MCMC implementation. Although $p\widehat{(\mathrm{y} \mid \mathcal{M}_*)}$ is consistent as the sample size L increases, its precision is not guaranteed: it may have an infinite variance. Therefore, we have implemented a *stabilized version* of $p\widehat{(\mathrm{y} \mid \mathcal{M}_*)}$ which is presented in detail in (Raftery et al., 2007). In short, it consists in replacing $p(\mathrm{y} \mid \boldsymbol{\theta}^{(l)}, \mathcal{M}_*)$ by $p(\mathrm{y} \mid f(\boldsymbol{\theta}^{(l)}), \mathcal{M}_*)$ in (C.1) such that f is a measurable function of $\boldsymbol{\theta}$ and a *dimension reduction* transformation. Since the voxels are assumed independent in space, one may proceed separately for each voxel: $p(\mathrm{y} \mid f(\boldsymbol{\theta}^{(l)}), \mathcal{M}_*) = \Pi_j\, p(\boldsymbol{y}_j \mid f(\boldsymbol{\theta}_j^{(l)}), \mathcal{M}_*)$. More precisely, this means that f can be derived by integrating out analytically the NRLs $\mathbf{a}_j$ and the noise variance $\sigma_{\varepsilon_j}^2$ for each voxel V_j . This seems sufficient to ensure that var $[(\mathrm{p}(\mathrm{y} \mid f(\boldsymbol{\theta}), \mathcal{M}_*))^{-1} \mid \mathrm{y}] < \infty$. Hence, in practice we consider the following estimator:

$$p\left(\widehat{\mathrm{y}|\mathcal{M}_*}\right) = \left[\frac{1}{L}\sum_{l=L_o}^{L_1}\frac{1}{\Pi_j p\left(\boldsymbol{y}_j|f\left(\boldsymbol{\theta}_j^{(l)}\right), \mathcal{M}_*\right)}\right]^{-1}.$$

Doing so, we have computed the log-evidence log $p(\mathrm{y} \mid \mathcal{M}_m)$ of models $\mathcal{M}_1$-$\mathcal{M}_5$ once they have been fitted against the first set of artificial data. Then, the logarithms of Bayes factors have been derived:

$$\log BF_{mn} = \log p\left(\mathrm{y}|\mathcal{M}_m\right) - \log p\left(\mathrm{y}|\mathcal{M}_n\right), \quad \forall (m,n) \in \mathbb{N}_4^*.$$

The same procedure has been applied to models $\mathcal{M}_2 - \mathcal{M}_5$ with artificial data eliciting deactivations.

References

Abramowitz, M., Stegun, I.A., 1970. Handbook of mathematical functions. Dover publications, New York, ny.

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of humain bold hemodynamic responses. NeuroImage 7, 574.

Andrade, A., Kherif, F., Mangin, J.-F., Worsley, K., Paradis, A.-L., Simon, O., Dehaene, S., Poline, J.-B., 2001. Detection of fMRI activation using cortical surface mapping. Hum. Brain Mapp. 12, 79–93.

Bagshaw, A.P., Hawco, C., Bénar, C.-G., Kobayashi, E., Aghakhani, Y., Dubeau, F., Pike, G.B., Gotman, J., 2005. Analysis of the EEG-fMRI response to prolonged bursts of interictal epileptiform activity. NeuroImage 24, 1099–1112.

Bénar, C.-G., Grova, C., Kobayashi, E., Bagshaw, A.P., Aghakhani, Y., Dubeau, F., Gotman, J., 2006. EEG-fMRI of epileptic spikes: Concordance with EEG source localization and intracranial EEG. NeuroImage 30, 1161–1170.

Benboudjema, D., Pieczynski, W., August 2007. Unsupervised statistical segmentation of nonstationary images using triplet Markov fields. IEEE Trans. Pattern Anal. Mach. Intell. 29 (8), 1367–1378.

Buxton, R., Frank, L., 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J. Cereb. Blood Flow Metab. 17 (1), 64–72.

Chib, S., 1995. Marginal likelihood from the Gibbs output. J. Am. Stat. Assoc. 90, 1313–1321.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis-Hastings output. J. Am. Stat. Assoc. 96 (453), 270–281.

Ciuciu, P., Poline, J.-B., Marrelec, G., Idier, J., Pallier, C., Benali, H., October 2003. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. IEEE Trans. Med. Imag. 22 (10), 1235–1251.

Ciuciu, P., Idier, J., Roche, A., Pallier, C., April 2004. Outlier detection for robust region-based estimation of the hemodynamic response function in event-related fMRI. 2nd Proc. IEEE ISBI. IEEE, Arlington, VA, pp. 392–395.

Ciuciu, P., Idier, J., Makni, S., June 11-15 2006. Modeling non-linear and non-stationary effects of the BOLD response using mixture models in fMRI. Proc. 12th HBM CD-Rom. Elsevier, Florence, Italy.

Ciuciu, P., Idier, J., Veit, T., Vincent, T., September 2007. Application du rééchantillonnage stochastique de l'échelle en détection-estimation de l'activité cérébrale par IRMf. Actes du 21e colloque GRETSI. GRETSI, Troyes, France, pp. 373–376.

Cohen, M.S., 1997. Parametric analysis of MRI data using linear systems methods. NeuroImage 6, 93–103.

Dehaene-Lambertz, G., Dehaene, S., Anton, J.-L., Campagne, A., Ciuciu, P., Dehaene, G.P., Denghien, I., Jobert, A., Le Bihan, D., Sigman, M., Pallier, C., Poline, J.-B., 2006. Functional segregation of cortical language areas by sentence repetition. Hum. Brain Mapp. 27, 360–371.

Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer Verlag, New York, USA. Available from http://jeff.cs.mcgill.ca/~luc/rnbookindex.html.

Drobnjak, I., Gavaghan, D., Suli, E., Pitt-Francis, J., M., J., 2006. Development of a functional magnetic resonance imaging simulator for modeling realistic rigid-body motion artifacts. Magn. Reson. Med. 56, 364–380.

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

Flandin, G., Penny, W.D., February 2007. Bayesian fMRI data analysis with sparse spatial basis function priors. NeuroImage 34 (3), 1108–1125.

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

Friston, K., 1994. Statistical parametric mapping. In: Thatcher, R., Hallet, M., Zeffiro, T., John, E., Huerta, M. (Eds.), Functional Neuroimaging: Technical Foundations, pp. 79–93.

Friston, K., 1998. Imaging neuroscience: Principles or maps? Proc. Natl. Acad. Sci. U. S. A. 95, 796–802.

Gelfand, A., Smith, A., Lee, T.-M., 1992. Bayesian analysis of constrained parameter and truncated problems using Gibbs sampling. J. Am. Stat. Assoc. 87 (518), 523–532.

Geman, S., McClure, D., 1987. Statistical methods for tomographic image reconstruction. Proceedings of the 46th Session of the ICI. Bulletin of the ICI, vol. 52, pp. 5–21.

Author's personal copy

Genovese, C., 2000. A Bayesian time-course model for functional magnetic resonance imaging data (with discussion). J. Am. Stat. Assoc. 95, 691–719.

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.

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

Goutte, C., Nielsen, F.A., Hansen, L.K., December 2000. Modeling the haemodynamic response in fMRI using smooth filters. IEEE Trans. Med. Imag. 19 (12), 1188–1201.

Gradshteyn, I., Ryzhik, I., 1994. Table of Integrals, Series, and Products, 5th ed. Academic Press, 1250 Sixth Avenue, San Diego CA.

Green, P.J., March 1990. Bayesian reconstructions from emission tomography data using a modified EM algorithm. IEEE Trans. Med. Imag. 9 (1), 84–93.

Green, P.J., 1995. Reversible jump MCMC computation and Bayesian model determination. Biometrika 82, 711–732.

Grill-Spector, K., Malach, April 2001. fmri-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol. (Amst) 107 (1-3), 293–321.

Grill-Spector, K., Sayres, R., Ress, D., Sep 2006. High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nat. Neurosci. 9 (9), 1177–1185.

Grova, C., Makni, S., Flandin, G., Ciuciu, P., Gotman, J., Poline, J.-B., 2006. Anatomically informed interpolation of fMRI data on the cortical surface. NeuroImage 31, 1475–1486.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21, 1639–1651.

Hastings, W.K., January 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97.

Henson, R., Price, C., Rugg, M., Turner, R., Friston, K., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. NeuroImage 15 (1), 83–97.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90, 773–795.

Kay, S.M., 1988. Modern Spectral Estimation. Prentice-Hall, Englewood Cliffs, nj.

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

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

Liu, J., 2001. Monte Carlo strategies in scientific computing. Springer series in Statistics. Springer-Verlag, New-York.

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

Makni, S., Ciuciu, P., Idier, J., Poline, J.-B., May 2006a. Bayesian joint detection-estimation of brain activity using MCMC with a Gamma-Gaussian mixture prior model. Proc. 31th Proc. IEEE ICASSP, vol. V. IEEE, Toulouse, France, pp. 1093–1096.

Makni, S., Ciuciu, P., Idier, J., Poline, J.-B., April 2006b. Joint detection-estimation of brain activity in fMRI using an autoregressive noise model. 3th Proc. IEEE ISBI. IEEE, Arlington, VA, pp. 1048–1051.

Mangin, J.-F., Frouin, V., Bloch, I., Régis, J., López-Krahe, J., 1995. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. J. Math. Imaging Vis. 5, 297–318.

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

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

Mazet, V., Brie, D., Idier, J., 2005. Simulation of positive normal variables using several proposal distributions. IIEEE workshop on statistical signal processing. Bordeaux, France.

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

Moussaoui, S., Brie, D., Mohammad-Djafari, A., Carteret, C., November 2006. Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. IEEE Trans. Signal Process. 54 (11), 4133–4145.

Naccache, L., Dehaene, S., 2001. Unconscious semantic priming extends to novel unseen stimuli. Cognition 80, 215–229.

Neumann, J., Lohmann, G., 2003. Bayesian second-level analysis of functional magnetic resonance images. NeuroImage 20 (2), 1346–1355.

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

Nieto-Castanon, A., Ghosh, S., Tourville, J., Guenther, F., 2003. Region of interest based analysis of functional imaging data. NeuroImage 19 (4), 1303–1316.

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

Ou, W., Golland, P., July 2005. From spatial regularization to anatomical priors in fMRI analysis. IPMI. Glenwood Springs, Colorado.

Penny, W., Friston, K., Apr 2003. Mixtures of general linear models for functional neuroimaging. IEEE Trans. Med. Imag. 22 (4), 504–514.

Penny, W.D., Kiebel, S., Friston, K.J., 2003. Variational Bayesian inference for fMRI time series. Neuroimage 19 (3), 727–741.

Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. Numerical Recipes in C- The Art of Scientific Computing, vol. 15.7. Cambridge University Press, Ch, pp. 699–706.

Raftery, A.E., Newton, M.A., Satagopan, J.M., Krivitsky, P.N., 2007. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In: Bernardo, J., Bayarri, M., Berger, O., David, A., Heckermann, D., Smith, A., West, M. (Eds.), Bayesian statistics 8. Oxford University Press, pp. 1–45.

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

Richardson, S., Green, P.J., 1997. On Bayesian analysis of mixtures with an unknown number of components (with discussion). J. R. Stat. Soc., B 59 (4), 731–792.

Robert, C., 1995. Simulation of truncated normal variables. Stat. Comput. 5, 121–125.

Robert, C.P., 2001. The Bayesian Choice, Springer Texts in StatisticsSecond Edition. Springer Verlag, New York, ny.

Roberts, S.J., Penny, W.D., September 2002. Variational Bayes for generalized autoregressive models. IEEE Trans. Signal Process. 50 (9), 2245–2257.

Smith, M., Pütz, B., Auer, D., Fahrmeir, L., 2003. Assessing brain activity through spatial Bayesian variable selection. NeuroImage 20, 802–815.

Svensen, M., Kruggel, F., von Crammon, D., January 2000. Probabilistic modeling of single-trial fMRI data. IEEE Trans. Med. Imag. 19, 19–35.

Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B., August 2006. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27 (8), 678–693.

Thirion, B., Pinel, P., Tucholka, A., Roche, A., Ciuciu, P., Mangin, J.-F., Poline, J.-B., September 2007. Structural analysis of fMRI data revisited: Improving the sensitivity and reliability of fMRI group studies. IEEE Trans. Med. Imag. 26 (9), 1256–1269.

Thyreau, B., Thirion, B., Flandin, G., Poline, J.-B., May 2006. Anatomo-functional description of the brain: a probabilistic approach. Proc. 31th Proc. IEEE ICASSP, vol. V. IEEE, Toulouse, France, pp. 1109–1112.

Author's personal copy

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

Veit, T., Idier, J., September 2007. Rééchantillonnage de l'échelle dans les algorithmes MCMC pour les problémes inverses bilinéaires. Actes du 21^{e} colloque GRETSI. GRETSI, Troyes, France, pp. 1233–1236.

Vincent, T., Ciuciu, P., Idier, J., August 2007a. Application and validation of spatial mixture modelling for the joint detection-estimation of brain activity in fMRI. Proc. of the 29th IEEE EMBS Annual international conference. Lyon, France, pp. 5218–5222.

Vincent, T., Ciuciu, P., Idier, J., April 2007b. Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI. 32th Proc. IEEE ICASSP, vol. I. IEEE, Honolulu, Hawaii, pp. 325–328.

Woolrich, M., Behrens, T., October 2006. Variational Bayes inference of spatial mixture models for segmentation. IEEE Trans. Med. Imag. 25 (10), 1380–1391.

Woolrich, M., Ripley, B., Brady, M., Smith, S., December 2001. Temporal autocorrelation in univariate linear modelling of fMRI data. NeuroImage 14 (6), 1370–1386.

Woolrich, M., Jenkinson, M., Brady, J., Smith, S., February 2004a. Fully Bayesian spatio-temporal modelling of fMRI data. IEEE Trans. Med. Imag. 23 (2), 213–231.

Woolrich, M., Jenkinson, M., Brady, J.M., Smith, S., 2004b. Constrained linear basis set for HRF modelling using variational Bayes. NeuroImage 21 (4), 1748–1761.

Woolrich, M., Behrens, T., Beckmann, C., Smith, S., January 2005. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. IEEE Trans. Med. Imag. 24 (1), 1–11.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A., January 2002. A general statistical analysis for fMRI data. NeuroImage 15 (1), 1–15.