

A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses



Linlin Zhang^a, Michele Guindani^b, Francesco Versace^c, Marina Vannucci^{a,*}

^a Department of Statistics, Rice University, Houston, USA

^b Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, USA

^c Department of Behavioral Science, The University of Texas MD Anderson Cancer Center, Houston, USA

ARTICLE INFO

Article history:

Accepted 10 March 2014

Available online 18 March 2014

Keywords:

Bayesian nonparametric
Dirichlet process prior
Discrete wavelet transform
fMRI
Long memory errors
Markov random field prior

ABSTRACT

In this paper we present a novel wavelet-based Bayesian nonparametric regression model for the analysis of functional magnetic resonance imaging (fMRI) data. Our goal is to provide a joint analytical framework that allows to detect regions of the brain which exhibit neuronal activity in response to a stimulus and, simultaneously, infer the association, or clustering, of spatially remote voxels that exhibit fMRI time series with similar characteristics. We start by modeling the data with a hemodynamic response function (HRF) with a voxel-dependent shape parameter. We detect regions of the brain activated in response to a given stimulus by using mixture priors with a spike at zero on the coefficients of the regression model. We account for the complex spatial correlation structure of the brain by using a Markov random field (MRF) prior on the parameters guiding the selection of the activated voxels, therefore capturing correlation among nearby voxels. In order to infer association of the voxel time courses, we assume correlated errors, in particular long memory, and exploit the whitening properties of discrete wavelet transforms. Furthermore, we achieve clustering of the voxels by imposing a Dirichlet process (DP) prior on the parameters of the long memory process. For inference, we use Markov Chain Monte Carlo (MCMC) sampling techniques that combine Metropolis–Hastings schemes employed in Bayesian variable selection with sampling algorithms for nonparametric DP models. We explore the performance of the proposed model on simulated data, with both block- and event-related design, and on real fMRI data.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Statistical methods play a crucial role in the analysis of fMRI data (Lazar, 2008; Lindquist, 2008; Poldrack et al., 2011), due to their complex spatial and temporal correlation structure. Common modeling approaches rely on general linear models formulations, as first proposed by Friston et al. (1994). In this paper, we propose a novel wavelet-based Bayesian nonparametric approach for modeling brain connectivity. More specifically, our goal is to provide a joint analytical framework that allows us to detect regions of the brain which exhibit neuronal activity in response to a stimulus and, simultaneously, infer the association, or clustering, of spatially remote voxels that exhibit fMRI time series with similar characteristics.

In fMRI experiments, neuronal activation in response to an input stimulus occurs in milliseconds and it is not observed directly. Instead, the blood oxygenation level dependent (BOLD) signal contrast is measured on the entire brain, since the metabolic process increases blood flow and volume in the activated areas following neuronal activation.

In the formulation we adopt, the stimulus pattern is convolved with a Poisson hemodynamic response function (HRF), with a voxel-dependent shape parameter that characterizes the delay between the onset of the stimulus and the arrival of blood to the activated brain regions. In addition, we assume correlated errors, in particular long memory, to take into account the time course structure of voxel responses, and then exploit the whitening properties of discrete wavelet transforms. Zarahn et al. (1997) and Aguirre et al. (1997) first suggested modeling the noise in fMRI data using a $1/f$ process. Fadili and Bullmore (2002) employed linear models assuming fractional Brownian motion (fBm) as the error term and derived wavelet-based approximate maximum likelihood estimates (MLE) of the model parameters. Meyer (2003) applied generalized linear models with drifts and errors contaminated by long-range dependencies. Jeong et al. (2013) employed a more general fractal structure for the error term and proposed a wavelet-based Bayesian approach for the estimation of the model parameters. When applied to data from a long memory process, discrete wavelet transformations result in wavelet coefficients that are approximately uncorrelated, leading to a relatively simple model in the wavelet domain that aids statistical inference, see Tewfik and Kim (1992), Craigmile and Percival (2005) and Ko and Vannucci (2006), among others. Bullmore et al. (2004) provide a nice review of wavelet-based methods for fMRI data.

* Corresponding author at: Department of Statistics, Rice University, 6100 Main Street, Houston, TX 77005, USA.

E-mail addresses: linlin.zhang@rice.edu (L. Zhang), mguindani@mdanderson.org (M. Guindani), fversace@mdanderson.org (F. Versace), marina@rice.edu (M. Vannucci).

A novel feature of our model is that we allow clustering the time course responses of distant brain regions via a Dirichlet process (DP) prior. In the fMRI literature, Thirion et al. (2007) and Jbabdi et al. (2009) have modeled fMRI profiles via an infinite mixture of multivariate Gaussian distributions with a Dirichlet process prior to cluster the model parameters, inducing a connectivity-based parcellation of the brain. In our approach, the DP prior model induces a clustering of the voxels that exhibit time series signals with similar variance and long-memory behavior. The induced spatio-temporal clustering can be viewed as an aspect of “functional” connectivity, as it naturally captures statistical dependencies among remote neurophysiological events (Friston, 1994, 2011). In addition, we detect activation in response to a stimulus by using mixture priors with a spike at zero on the coefficients of the regression model characterizing the association between response and stimulus. The selection of activated voxels takes into account the complex spatial correlation structure of the brain through a Markov random field (MRF) prior, thus capturing correlation among nearby voxels. For inference, we use Markov Chain Monte Carlo (MCMC) sampling techniques that combine Metropolis–Hastings schemes employed in Bayesian variable selection with sampling algorithms for nonparametric DP models (Neal, 2000; Savitsky et al., 2011).

Bayesian spatiotemporal model approaches that incorporate spatial correlation among brain responses have recently found successful applications in the analysis of fMRI data (Bowman et al., 2008; Flandin and Penny, 2007; Gössl et al., 2001; Harrison and Green, 2010; Kalus et al., 2013; Penny et al., 2005; Quirós et al., 2010; Smith and Fahrmeir, 2007; Woolrich et al., 2004). Gaussian Markov random field priors were imposed by Penny et al. (2005) on the regression coefficients of a general linear model, while Flandin and Penny (2007) used sparse spatial basis functions and Harrison and Green (2010) a conditional autoregressive (CAR) prior. Smith and Fahrmeir (2007) investigated spatial Bayesian variable selection linear regression models with an Ising prior for latent activation indicators, while Kalus et al. (2013) used a spatial probit prior of the CAR type. Gössl et al. (2001) and Woolrich et al. (2004) investigated spatio-temporal hierarchical Bayesian approaches incorporating the estimation of the HRF. Quirós et al. (2010) also parameterized the HRF with voxel-varying parameters and used Gaussian Markov random field priors on the activation characteristic parameters of the voxels. All these authors assumed independent structures for the error terms in their models, with the exception of Woolrich et al. (2004) and Penny et al. (2005), who imposed an autoregressive structure.

With respect to existing Bayesian approaches for the analysis of fMRI data, our modeling strategy combines several of the features of such approaches into a single modeling framework. In the time dimension, we allow for correlated errors and then employ wavelet transforms. Furthermore, we describe patterns of similar neuronal dynamics via the clustering of the voxel time courses induced by the DP. In addition, we account for the estimation of a voxel dependent delay parameter of the hemodynamic response function. In the spatial dimension, we capture dependence among nearby voxels via the MRF prior. Our inferential strategy results in the detection of activated regions of the brain and clustering of similar time courses. We show performances of our proposed model for both block and event-related design with simulated data, and for block design with synthetic data. We also consider a case study on a real fMRI experiment.

The rest of the paper is organized as follows. Model and prior distributions are presented in the *Statistical methods* section. Results obtained by applying the proposed model to simulated data and real fMRI data are examined in the *Results* section. The *Discussion* section concludes the paper.

Statistical methods

Regression model with correlated errors

Functional magnetic resonance imaging (fMRI) is a common tool for detecting changes in neuronal activity. It measures blood oxygenation

level-dependent (BOLD) contrast that depends on changes in regional cerebral blood-flow. In an fMRI experiment, the whole brain is scanned at multiple time points and a time series of BOLD response is acquired for each voxel of the brain while the subject performs a set of tasks. Let $Y_\nu = (Y_{\nu 1}, \dots, Y_{\nu T})^T$ be the $T \times 1$ vector of the response data, with $Y_{\nu i}$ the BOLD image intensity at time $i = 1, \dots, T$, for voxel ν , with $\nu = 1, \dots, V$. We model the BOLD response for a single subject with a linear regression model of the type

$$Y_\nu = X_\nu \beta_\nu + \varepsilon_\nu, \quad \varepsilon_\nu \sim N_T(0, \Sigma_\nu), \quad (1)$$

where X_ν is a known $T \times p$ covariate matrix and $\beta_\nu = (\beta_{\nu 1}, \dots, \beta_{\nu p})^T$ a $p \times 1$ vector of regression coefficients. Without loss of generality, we center the data at 0, and hence no intercept is needed in the model. In the applications, we consider the case $p = 1$, that is, only one regressor for one experimental condition is included in the model. Single-subject scanning is receiving renewed interest in the fMRI field, due to its use for presurgical purposes. In addition, single-subject posterior maps can be used in meta analyses for inter-subject investigations, similarly in spirit to other Bayesian approaches to fMRI modeling (Bowman et al., 2008; Jbabdi et al., 2009).

When measuring the change in the metabolism of BOLD contrast due to outside stimuli, the fMRI signal gets delayed hemodynamically (Buxton and Frank, 1997). The complete relationship between the neuronal activity and the vascular response is not fully known yet. A widely used model, to account for the lapse of time between the stimulus onset and vascular response, looks at an fMRI signal as the convolution of regional cerebral blood-flow response to stimulus with a hemodynamic response function. Here we model the covariate X_ν as the convolution of the stimulus pattern with a Poisson HRF, that is,

$$X_\nu = \int_0^t x(s) h_{\lambda_\nu}(t-s) ds, \quad (2)$$

where $x(s)$ represents the time dependent stimulus and $h_{\lambda_\nu}(t) = \exp(-\lambda_\nu t) \lambda_\nu^t / t!$ (Friston et al., 1994). The parameter λ_ν can be interpreted as the delay of the response with respect to the stimulus onset, and it is modeled as an unknown voxel-dependent parameter (Quirós et al., 2010).

The temporal dependence between observations for a given voxel is taken into account by exploring the structure of the Σ_ν 's, i.e. the error terms in model (1). These are typically assumed autocorrelated and capture instrumental noise, such as head movement in the scanner. More specifically, two types of autocorrelation structures for fMRI signals have been typically assumed in the literature: first-order autoregressive and $1/f$ long memory structures (Smith et al., 1999). As estimation procedures for time series with autocorrelated residuals are computationally expensive, discrete wavelet transforms have often been used, as a tool to decorrelate fMRI data, see Fadili and Bullmore (2002), Meyer (2003), Turkheimer et al. (2003), and, more recently, Sanyal and Ferreira (2012) and Jeong et al. (2013). Let us write Σ_ν as $\Sigma_\nu(i, j) = [\gamma(|i - j|)]$ with $\gamma(h)$ the auto-covariance function of the process generating the data. Here, we model ε_ν as a long memory process, characterized by an autocovariance function with a slow decay, i.e.

$$\gamma(h) \sim Ch^{-\alpha}, \quad (3)$$

with $C > 0$, $0 < \alpha < 1$ and h large. We then employ discrete wavelet transformations (DWT) as a tool to “whiten” the data and simplify the dense covariance structure of the long memory process.

Wavelets are families of orthonormal basis functions that can be used to represent a generic function $f(x)$ as

$$f(x) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} d_{m,n} \psi_{m,n}(x), \quad d_{m,n} = \int_{-\infty}^{\infty} f(x) \psi_{m,n}(x) dx, \quad (4)$$

with $d_{m,n}$ coefficients describing features of the function $f(x)$ at locations indexed by n and scales indexed by m . The wavelet basis

functions $\psi_{m,n}(x)$ are defined as dilations and translations of a mother wavelet $\psi(x)$ as

$$\psi_{m,n}(x) = 2^{m/2} \psi(2^m x - n). \quad (5)$$

Wavelets have been successfully applied to the analysis of time series data. Discrete wavelet transforms (Mallat, 1989) can be used to reduce the data to a set of wavelet coefficients. Let $y = (y_1, y_2, \dots, y_T)^T$ be an observed time series, $T = 2^J$ with J a positive integer. The DWT of y can be written as $d = Wy$, where W is an orthogonal $T \times T$ matrix representing the wavelet transform. Furthermore, the covariance matrix of the wavelet coefficients d can be calculated as $\Sigma_d = W \Sigma_y W$ with $\Sigma_y(i, j) = [\gamma(|i - j|)]$ (Vannucci and Corradi, 1999).

After applying a DWT on both sides of model (1) we obtain the following model in the wavelet domain

$$Y_\nu^* = X_\nu^* \beta_\nu + \varepsilon_\nu^*, \quad \varepsilon_\nu^* \sim N_T(0, \Sigma_\nu^*), \quad (6)$$

where $Y_\nu^* = WY_\nu$, $X_\nu^* = WX_\nu$, and $\varepsilon_\nu^* = W\varepsilon_\nu$. As emphasized in the literature, wavelet coefficients tend to be less correlated than the original data. For long memory processes, in particular, such decorrelation properties are well documented (Craigmile and Percival, 2005; Ko and Vannucci, 2006; Tewfik and Kim, 1992). In our model (6) we can therefore assume the covariance matrix Σ_ν^* as a $(T \times T)$ diagonal matrix. We also write its diagonal elements as $\psi_\nu \sigma_{mn}^2$, indicating the variance of the n th wavelet coefficient at the m th scale. We follow Jeong et al. (2013) and adopt the variance progression formula for the covariance structure of wavelet coefficients of Wornell and Oppenheim (1992), that is,

$$\psi_\nu \sigma_{mn}^2 = \psi_\nu (2^{\alpha_\nu})^{-m}, \quad (7)$$

with ψ_ν the innovation variance and $\alpha_\nu \in (0, 1)$ the long memory parameter. This structure encompasses the general fractal process of type (3), including long memory processes.

Priors

One of the goals of our modeling approach is to detect regions of the brain that show activation to a given stimulus. In addition, we want to cluster the individual voxel time series and thus describe the patterns of association, or correlation, of neuronal activity in separate, possibly remote, voxels in the brain (Friston, 1994). We achieve both tasks via the choice of appropriate prior models.

The task of selecting activated voxels is equivalent to a problem of variable selection, that is the identification of the nonzero β_ν in model (6). Bayesian variable selection in linear regression models can be achieved by employing mixture priors with a spike at zero on the regression coefficients (Brown et al., 1998; George and McCulloch, 1993, 1997; Sha et al., 2004). A binary random variable γ_ν is introduced to identify whether a given voxel is activated or not, that is, $\gamma_\nu = 0$ if $\beta_\nu = 0$ and $\gamma_\nu = 1$ otherwise, and a mixture prior (commonly called *spike-and-slab prior*) is imposed on the coefficients β_ν as

$$\beta_\nu \sim \gamma_\nu N(0, \tau) + (1 - \gamma_\nu) \delta_0, \quad \nu = 1, \dots, V, \quad (8)$$

with δ_0 a point mass at zero and τ a relatively large variance term for the non-null component. In the brain network, neighboring voxels tend to be active or inactive together. Several authors have looked into Bayesian spatiotemporal models that incorporate spatial correlation structure into the priors (Flandin and Penny, 2007; Harrison and Green, 2010; Kalus et al., 2013; Penny et al., 2005; Quirós et al., 2010; Smith and Fahrmeir, 2007). In our model, we account for spatial correlation by placing a Markov random field (MRF) prior on the selection parameter

γ_ν of the mixture prior model. We follow Li and Zhang (2010) and Stingo et al. (2011, 2012) and parameterize the conditional probability of γ_ν as

$$P(\gamma_\nu | d, e, \gamma_k, k \in N_\nu) \propto \exp\left(\gamma_\nu \left(d + e \sum_{k \in N_\nu} \gamma_k\right)\right) \quad (9)$$

with N_ν the set of neighboring voxels of voxel ν . According to the prior (9), a voxel has a greater probability of being activated if more of the neighboring voxels are activated. The parameter $d \in (-\infty, \infty)$ represents the expected prior number of activated voxels and controls the sparsity of the model, whereas $e > 0$ affects the probability of identifying a voxel as active according to the activation of its neighbors and therefore acts as a smoothing parameter. Posterior inference may be very sensitive to the choice of d and e ; thus, some care is needed. Indeed, it is a general experience that allowing e to vary widely can lead to a phase transition problem, i.e. the expected number of variables equal to 1 can increase massively for small increments of e . This problem arises since Eq. (9) can only increase as a function of the number of γ_ν 's equal to 1. Thus, phase transition may lead to a drastic change in the proportion of detected activations, especially since the number of voxels is large. An empirical estimate of the phase transition value can be obtained using the algorithm proposed by Propp and Wilson (1996) and the values of d and e can then be chosen accordingly. In this paper, we follow Li and Zhang (2010) and Stingo et al. (2012), and treat d and e as fixed hyperparameters. More specifically, we first note that if a voxel does not have any neighbor, then its prior distribution reduces to an independent Bernoulli, with parameter $\exp(d)/[1 + \exp(d)]$, a prior often adopted in the Bayesian variable selection literature. This constitutes a lower bound on the prior probability of activation, and therefore, we set d to a small value that reflects our belief in a sparse model. As for e , any value of e below the phase transition point can be considered a good choice, with values closer to the phase transition point leading to higher prior probabilities of activations in a neighborhood of activated voxels.

We capture the association of neural activity over time by clustering voxels that exhibit fMRI time series signals with similar characteristics. More specifically, we impose a Dirichlet process (DP) prior on the parameters of the residual term of the model. The DP is a stochastic process commonly used in Bayesian nonparametric models. Here we use it as a tool to obtain simultaneous inference on the clusters of the voxel time courses as well as on the values of the parameters that define the specific long term process in each such cluster. The formal definition of a DP has been given in Ferguson (1973). Hjort et al. (2010) and Rodriguez and Müller (2013) provide exhaustive discussion of its properties and applications. For our purposes, it suffices to look at the Dirichlet process as a prior on a class of probability distributions. Let G indicate such random probability measure on the space of distributions. Then, we write $G \sim \text{DP}(\eta, G_0)$ to indicate that the model depends on the specification of two parameters, the base measure G_0 and the total mass parameter η . The base measure G_0 is the prior mean, $E(G) = G_0$. Typically, in Bayesian nonparametric inference, the unknown G is centered around a known parametric model. The total mass parameter determines, among other important properties, the variation of the random measure around the prior mean. Small values of η imply high uncertainty. In the following, we exploit a key property of the DP. Any realization G from a DP almost surely defines a discrete distribution. Let $\theta_i | G \sim G, i = 1, \dots, n$ be an i.i.d. sample from a distribution G , where $G \sim \text{DP}(\eta, G_0)$. Then, G can be almost surely written as a mixture of point masses $G = \sum_{h=1}^{\infty} w_h \delta_{\theta_h^*}$, with $w_h = V_h \prod_{j=1}^{h-1} (1 - V_j)$, $V_j \sim \text{Beta}(1, \eta)$, $j = 1, \dots, h$, and atoms $\theta_h^* \sim G_0, h = 1, \dots$. The discreteness of the DP is best appreciated by looking at the predictive distribution of θ_i conditional on all the other values $\theta_{-i} = \{\theta_j : j \neq i\}$

$$\theta_i | \theta_{-i} \sim \frac{\eta G_0 + \sum_{j \neq i} \delta_{\theta_j}}{V - 1 + \eta}. \quad (10)$$

Eq. (10) show that there is a positive probability of ties, that is the θ_i 's can assume a common value. Thus, the discreteness of the DP plays a crucial role in its use for clustering purposes, since the θ_i 's can form clusters. The larger the frequency of data assigned to a cluster, the higher the probability that the cluster will grow. In this representation, the parameter η acts as a weight; the larger is η , the higher is the probability that θ_i is sampled from the base measure G_0 and does not coincide with any of the other values, thus creating more clusters. In our model, we impose a DP prior on the parameters of the long memory process (7), (ψ_ν, α_ν) , as

$$\begin{aligned} (\psi_\nu, \alpha_\nu) | G &\sim G \\ G | \eta, G_0 &\sim DP(\eta, G_0) \\ G_0 &= IG(a_0, b_0) \times \text{Beta}(a_1, b_1) \end{aligned} \quad (11)$$

where the base measure G_0 is specified as the product of inverse gamma (IG) and beta distributions. Escobar and West (1995) discuss prior distributions on η . In practice, the impact of such prior typically decreases with V . Hence, when V is large, η is often fixed so to favor a parsimonious representation of the data or some other prior beliefs on their clustering behavior.

Finally, we complete our prior model by considering a uniform distribution on the interval (u_1, u_2) as a prior distribution for the delay parameter λ_ν in the hemodynamic function, as suggested by Quirós et al. (2010),

$$\lambda_\nu \sim U(u_1, u_2), \quad \nu = 1, \dots, V. \quad (12)$$

A graphical representation of our full model is shown in Fig. 1.

Posterior inference

The posterior distribution of all parameters of interest can be obtained by combining the prior information and the likelihood function via the Bayes theorem. For inference, we use Markov Chain Monte Carlo (MCMC) sampling techniques that sample the individual parameters conditional upon the others. Our algorithm combines Metropolis–Hastings schemes that use *add–delete–swap* moves, as employed in Bayesian variable selection (Savitsky et al., 2011), with sampling algorithms for nonparametric DP models (Neal, 2000). We give full details of the full conditional distributions and our implementation of the MCMC in the Appendix.

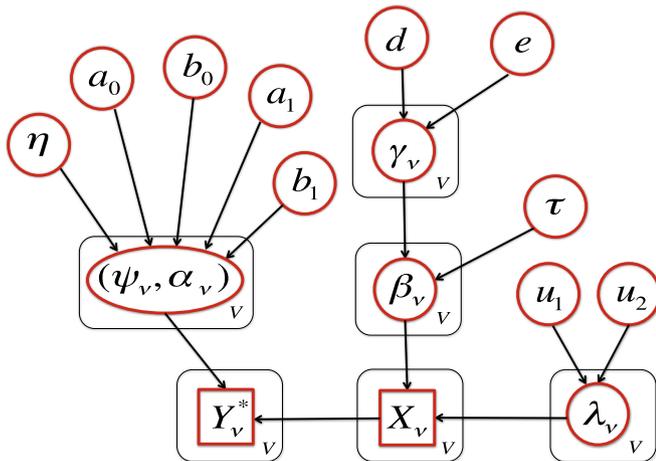


Fig. 1. Graphical representation of the model. Each node in a circle refers to a parameter of the model, nodes in squares are observables. The link between two nodes represents a direct probabilistic dependence.

Results

Simulation from the model

In this section we use simulated data generated from our model to investigate the performance of our strategy for posterior inference on the model parameters. We look at both block and event-related experimental designs and generate the data from model (6), that is directly in the wavelet domain. In all simulations presented below we considered $T = 256$ images of 30×30 voxels.

In the block design two different conditions, activity and rest, are alternating over time. For this, we generated a square wave signal as

$$x(t) = \begin{cases} 1, & kP < t < (k+1)P, k = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

with P the period of the signal, which we fixed at $P = 16$. In the event-related design the stimulus consisted of short discrete events whose timing and order were randomized. We set the number of stimuli to 20 and the duration of each event to 10. We obtained the covariate X_ν in model (1) by convolving the stimulus pattern with a Poisson hemodynamic function with delay parameter λ_ν . We sampled the λ_ν parameters from a uniform distribution between 0 and 8. We applied discrete wavelet transforms with Daubechies minimum phase wavelets, (Daubechies, 1992), with 4 vanishing moments, to obtain the covariate X_ν^* in model (6). We then set the selection parameters γ_ν 's in (8) as a 30×30 lattice with five rectangular active regions, see Fig. 3(a). Voxels inside the regions were assigned the value of 1, whereas voxels outside were assigned the value 0. This resulted in 296 active voxels and 604 inactive voxels. We simulated the β_ν parameters for the active voxels by sampling from a $N(0, \tau)$ with $\tau = 1$, and set those for the inactive voxels to 0. Finally, we set the values of (ψ_ν, α_ν) at $(0.1, 0.2)$, $(0.5, 0.5)$, and $(1, 0.8)$, leading to three different clusters of voxels. We consider the signal-to-noise ratio in logarithm decibel scale, defined as a function of the variance of the regression parameters versus the innovation variance of the error term, i.e. $\text{SNR}_{\text{dB}} = 10 \log_{10}(\tau/\psi)$. Thus, SNR_{dB} increases when the signal is allowed to vary more than the noise; $\text{SNR}_{\text{dB}} \approx 0$ corresponds instead to cases where the signal and the noise are

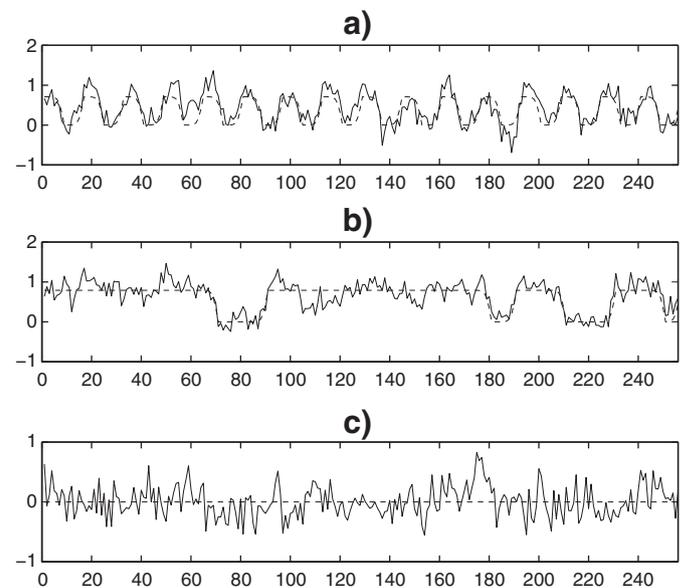


Fig. 2. Simulated data: Time series fitting for one active voxel under (a) block and (b) event-related designs and for (c) one inactive voxel, with $(\psi, \alpha) = (1, 0.8)$. The continuous black curves represent the simulated time series and the dashed black curves the fitted responses.

approximately of the same magnitude. Thus, the three clusters are characterized by decreasing values of the signal-to-noise ratio (respectively, 10, 3, and 0). The continuous black curves in Fig. 2 represent examples of simulated time series for active and nonactive voxels, obtained by applying the inverse wavelet transform to obtain $Y_{\nu} = W^T Y_{\nu}^*$, for the cluster $(\psi_{\nu}, \alpha_{\nu}) = (1, 0.8)$.

In the simulation described in this section, we specified $\tau = 5$ and $u_1 = 0, u_2 = 8$ for model fitting purposes. We also fixed the concentration parameter of the DP prior to $\eta = 1$ and set the base distribution as the product of a non-informative distribution on α , that is a $Beta(a_1 = 1, b_1 = 1)$, and a vague prior on ψ , that is an $IG(a_0 = 3, b_0 = 2)$. We set the MRF prior parameters to $d = -2.5$ and $e = 0.3$. We discuss sensitivity to these choices below. At every MCMC iteration we applied discrete wavelet transforms with Daubechies minimum phase wavelets with 4 vanishing moments. We ran MCMC chains with 10,000 iterations, discarding the

first 5,000 iterations as a burn-in. We assessed convergence by using the Raftery–Lewis diagnostic (Raftery and Lewis, 1992), as implemented in the R package “coda”.

We obtained posterior activation probability maps by setting the posterior probability threshold at 0.8, that is, an individual voxel was categorized as active if the posterior probability $p(\gamma_{\nu} = 1|y) > 0.8$, and categorized as inactive otherwise. This type of thresholds can be justified as the optimal rules in a Bayesian decision theory setting on the basis of loss functions that minimize a combination of false negative and false discovery counts or rates (see, among others, Bogdan et al., 2008; Guindani et al., 2009; Müller et al., 2004, 2007). Fig. 3(b) reports the posterior activation map for the block design obtained by assigning value 1 to those voxels with $p(\gamma_{\nu} = 1|y) > 0.8$ and value 0 otherwise. Our method does a good job at detecting the active voxels. In particular, a small number of active voxels are falsely identified as inactive, and all

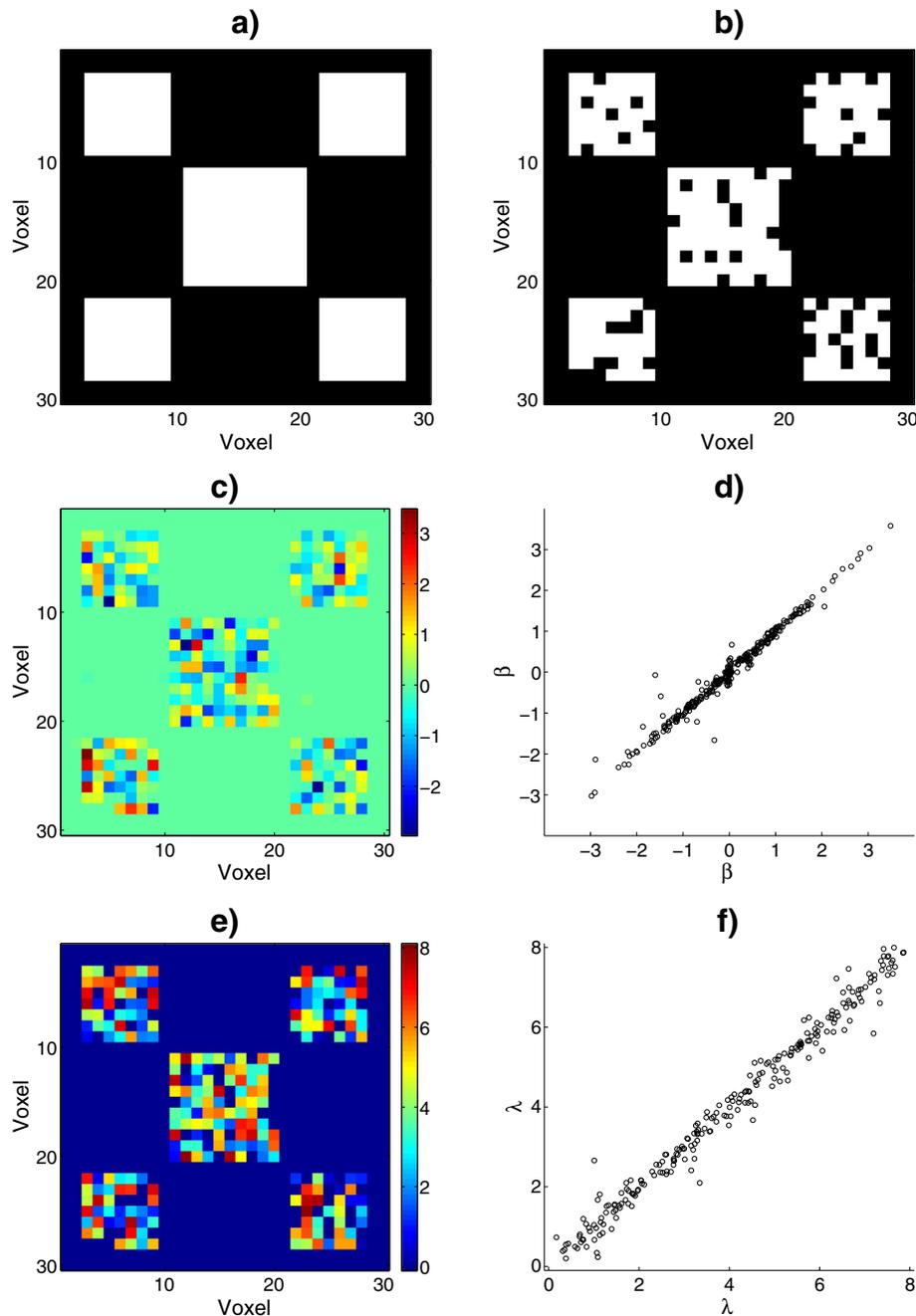


Fig. 3. Simulated data with block design: (a) true map of the activation indicators γ ; (b) posterior activation map obtained by assigning value 1 to those voxels with $p(\gamma_{\nu} = 1|y) > 0.8$ and value 0 otherwise; (c) posterior mean map of β ; (d) scatter plot of posterior mean estimates vs. true values for β ; (e) posterior mean map for λ ; (f) same as (d) for λ .

inactive voxels are correctly identified in both designs. Results are overall comparable across clusters, although, as expected, the accuracy slightly decreases when the noise level increases. Fig. 3(c)–(f) reports the posterior mean estimates for the parameters β and λ in the block design. Our method produces good estimates, close to the true values of all parameters. We refer the reader to the supplementary materials for the analogous maps in the event-related design. We also investigated the fit of the response time series obtained by our model. Fig. 2 shows the fit of the time series as dashed black curves under either the block and the event-related design. Once again, we notice the good fit of our estimates.

Our approach also allows us to infer a clustering of the time course responses of distant voxels via the DP prior on the parameters of the residual term of the model, ψ and α . For example, a single estimate of the sample clustering allocations can be obtained from those configurations visited during the MCMC as follows (Tadesse et al., 2005). First, an estimate of the number of clusters is computed as the value, say K , most frequently visited by the MCMC sampler, then all MCMC configurations with K clusters are relabeled, for example by using the algorithm of Stephens (2000). Finally the sample allocations are estimated by computing, for each sample, the posterior probability that the sample belongs to cluster k , for $k = 1, \dots, K$, and assigning the sample to the cluster with largest posterior probability. Fig. 4(a) shows the resulting clustering of the voxels for the block design. In order to formally compare our results with the ground truth, we consider the Normalized Mutual Information (NMI), an information theoretic based measure of clustering comparison proposed by Strehl and Ghosh (2002), which is defined as follows. Let C and C' denote two alternative clustering allocations of size K and L , respectively, and let $fr(i, j')$ denote the relative frequency of voxels commonly assigned to cluster C_i in C and to cluster $C'_{j'}$ in C' . Similarly, let $fr(i)$ denote the frequency of voxels assigned to cluster C_i in C , and $fr(j')$ the frequency of voxels assigned to cluster $C'_{j'}$ in C' . Then we can define the mutual information between the two clustering allocations as

$$I(C, C') = \sum_{i=1}^K \sum_{j'=1}^L fr(i, j') \log_2 \frac{fr(i, j')}{fr(i)fr(j')},$$

and, correspondingly, the NMI as

$$NMI(C, C') = \frac{I(C, C')}{\sqrt{H(C)H(C')}},$$

with $H(C) = I(C, C)$. It follows that $0 \leq NMI(C, C') \leq 1$ with $NMI(C, C') = 1$ for $C \equiv C'$. In our simulation, the NMI for the block design and the event-related design are 0.9068 and 0.9552, respectively. Both values are close to 1, indicating a strong similarity between the estimated clustering allocation and ground truth. We notice that the DP prior model allows us to

Table 1

Simulated data: Accuracy, precision, false positive rates (FPR) and false negative rates (FNR) for different choices of the variance structure of the error term and different priors, for block design (B) and event-related design (E). The choice ψ_{I_T} corresponds to a model with an uncorrelated error. Results are percentages averaged over 30 replicated datasets.

	Σ_v	$\psi_{I_T} (2^{\alpha v})^{-m}$		ψ_{I_T}
		$IG(3, 2) \times Be(1, 1)$	$IG(3, 2) \times Be(2, 3)$	
B	Accuracy	93.24	93.33	87.09
	Precision	99.82	99.83	75.64
	FPR	0.07	0.07	14.19
	FNR	20.42	20.15	10.30
E	Accuracy	91.54	91.50	86.77
	Precision	99.90	99.94	73.89
	FPR	0.04	0.02	16.07
	FNR	25.64	25.80	7.43

mostly recover the true clustering of the voxels, with the map showing three large clusters and a few misclassified voxels. As for the inference on the parameters ψ and α , voxel-wise estimates can be calculated by averaging over the MCMC samples after burn in, at each voxel. The resulting posterior mean maps for the block design are shown in Fig. 4. Again, we refer to the supplementary material for the posterior mean maps of α and ψ in the event-related design. We notice that our estimates well recover the true values of the parameters.

In order to further assess the performance of our estimation method, we repeated the simulation 30 times. Table 1 reports results in terms of accuracy (i.e. the percentage of voxels that are correctly identified), precision (i.e. the proportion of active voxels correctly identified against the number of all those identified as active), false positive rate (FPR) (i.e. the proportion of active voxels falsely identified against all the inactive voxels) and false negative rate (FNR) (i.e. the proportion of nonactive voxels falsely identified against all the active voxels), for both block and event-related design, averaged over the 30 replicates. Results are reported for two choices of the joint base prior distribution on (ψ, α) in (11). In the same table we compare results with those obtained by fitting our model with a simpler choice of the variance structure of the error term, i.e., $\Sigma_v = \psi_{I_T}$ that, unlike (7), ignores the temporal correlation of the data. As it should be expected, this model shows worse performances, with lower accuracy, precision and false negative rate, and a higher false positive rate.

Data simulated with different noise levels, that is different values of α and ψ , gave similar results to those reported here, though, as expected, the accuracy lowers when the noise level increases.

We also evaluate the performance of our model assuming that the true HRFs are different than the Poisson HRF used in model fitting. Table 2 reports accuracy, precision, FPR and FNR for one simulated data set where we considered the canonical HRF (Worsley et al., 2002) as well as the inverse logit HRF proposed by Lindquist and

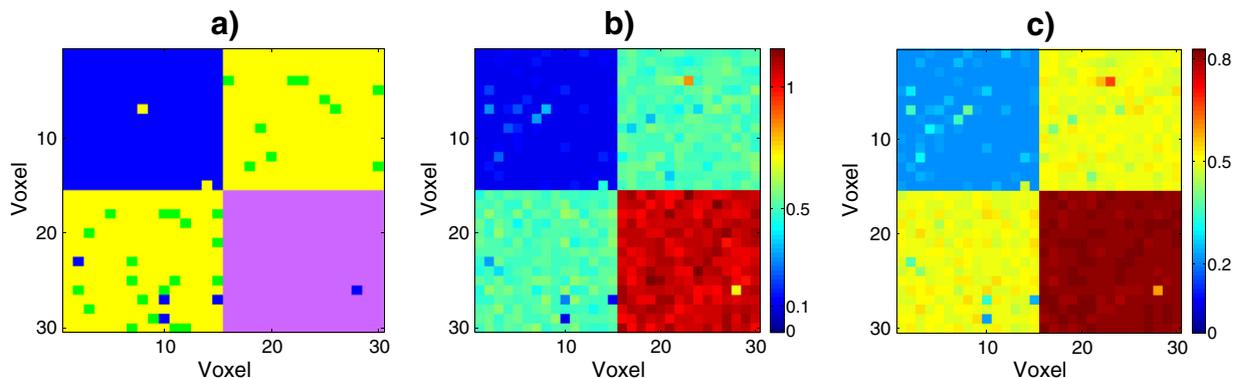


Fig. 4. Simulated data with block design: (a) posterior clustering map—different colors correspond to different clustering allocations; (b) posterior mean map of ψ ; (c) posterior mean map of α .

Table 2

Simulated data: accuracy, precision, FPR, and FNR for different choice of true HRFs for block design (B) and event-related design (E).

	True HRF	Poisson	Canonical	Inverse logit
B	Accuracy	95.67	95.78	95.22
	Precision	100.00	100.00	100.00
	FPR	0.00	0.00	0.00
	FNR	13.18	12.84	14.53
E	Accuracy	95.33	91.33	90.22
	Precision	100.00	100.00	100.00
	FPR	0.00	0.00	0.00
	FNR	14.19	26.35	29.73

Wager (2007) in addition to the Poisson HRF. The canonical HRF is defined as the difference of two Gamma densities,

$$h(t) = (t/d_1)^{a_1} \exp(-(t-d_1)/b_1) - c(t/d_2)^{a_2} \exp(-(t-d_2)/b_2), \quad (14)$$

with $a_1 = 6, a_2 = 12, b_1 = 0.9, b_2 = 0.9, c = 0.35, d_1 = a_1 * b_1, d_2 = a_2 * b_2$ (Glover, 1999). The inverse logit HRF is generated as a superposition of three separate inverse logit functions, which describe, respectively, the rise following activation, the subsequent decrease and undershoot, and the stabilization of the HRF,

$$h(t) = \alpha_1 L((t-T_1)/D_1) + \alpha_2 L((t-T_2)/D_2) + \alpha_3 L((t-T_3)/D_3), \quad (15)$$

with $L(x) = 1/(1 + e^{-x}), \alpha_1 = 1, T_1 = 15, D_1 = 1.33, \alpha_2 = -1.3, T_2 = 27, D_2 = 2.5, \alpha_3 = 0.3, T_3 = 66, D_3 = 2$. For the Poisson HRF, we set the delay parameter $\lambda = 4$ as in Jeong et al. (2013). The data are generated so to obtain comparable SNRs, despite the three HRFs have different response amplitudes. Table 2 shows an overall good performance of our model even when the true HRFs are different from the assumed Poisson HRF.

Sensitivity analysis

We performed a sensitivity analysis to study how different prior specifications affect the estimation. Tables 3–5 show results for parameters $\tau, d,$ and e , for both block and event-related design. In each table, accuracy, precision, FPR and FNR are reported for one simulated data set. Table 3 shows that setting larger values of the variance parameter τ in the prior distribution for the β 's may moderately impact accuracy and lead to a relatively higher FNR. We did not notice sensitivity to moderate changes of the prior specifications for the parameters ψ and α of the error term, see Table 1 for example. As for the MRF prior on the γ 's, its role is to encourage activations to take place in regions of neighboring voxels, penalizing detections in isolated voxels. As discussed in the Priors section, the parameter d controls the sparsity of the model, whereas e regulates the smoothness of the distribution, with higher values encouraging the selection of voxels with neighbors already selected as active. As expected, Tables 4 and 5 show sensitivity of the results to the values of those parameters. More specifically, larger values

Table 3

Simulated data: sensitivity analysis on the parameter τ (keeping fixed $\eta = 1, d = -2.5, e = 0.3, a_0 = 3, b_0 = 2, a_1 = b_1 = 1$), for block design (B) and event-related design (E). Results are in percentages.

		$\tau = 1$	$\tau = 5$	$\tau = 10$
B	Accuracy	94.33	93.78	92.67
	Precision	100.00	100.00	100.00
	FPR	0.00	0.00	0.00
	FNR	17.23	18.92	22.30
E	Accuracy	92.33	91.00	90.56
	Precision	100.00	100.00	100.00
	FPR	0.00	0.00	0.00
	FNR	23.31	27.36	28.72

Table 4

Simulated data: sensitivity analysis on the parameter d (keeping fixed $\eta = 1, e = .3, \tau = 5, a_0 = 3, b_0 = 2, a_1 = b_1 = 1$), for block design (B) and event-related design (E). Results are in percentages.

		$d = -4.0$	$d = -2.5$	$d = -1.5$	$d = -1.3$
B	Accuracy	91.78	93.78	94.44	94.22
	Precision	100.00	100.00	100.00	98.80
	FPR	0.00	0.00	0.00	0.50
	FNR	25.00	18.92	16.89	16.55
E	Accuracy	89.44	91.00	92.67	92.44
	Precision	100.00	100.00	100.00	98.31
	FPR	0.00	0.00	0.00	0.66
	FNR	32.09	27.36	22.30	21.62

of d or e generally correspond to a lower FNR, at the expense of a higher FPR and a lower precision. We note that setting $d = -2.5$ implies that the prior probability of selection is less than 10% when a voxel has no neighbors and that $e = 0.3$ is a value that should allow a reasonable power in detecting true activations, and far from the phase transition limit. Some sensitivity to η , the parameter of the DP prior, was also observed, with large values generally inducing a larger number of clusters (result not shown).

Synthetic data

In this section we simulate a synthetic fMRI data set using data from the study of the Wellcome Department of Imaging Neuroscience that we present in the section A case study for fMRI data. Our simulation approach is similar in spirit to the one of Quirós et al. (2010). More specifically, we considered a slice captured under the “no attention” condition and then obtained synthetic data as the sum of two components, $Y_{syn} = Y + w$, where Y denotes the component simulated from our model as described below and w is the selected slice from the real fMRI study (slice 27). The use of the synthetic data allows us to carry out comparisons of the results from our model with the popular software SPM8, available at <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>.

In order to simulate Y , we first obtained the design matrix X for a block design as before. Then, we convolved X with a Poisson hemodynamic function and applied the wavelet transform. The true activation map defines three rectangular active regions, obtained by setting $\gamma = 1$, as shown in Fig. 5(a). The true values of the β 's in these regions were sampled from a Uniform($-80, 80$). We set the values of (ψ_i, α_i) at $(10, 0.8), (5, 0.5),$ and $(1, 0.2)$, leading to three different clusters of voxels. Finally, we obtained Y as $Y = W^T X$. Our final synthetic dataset comprised time series data for 256 scans of 64×64 voxels, mirroring a real block-design fMRI study. Fig. 5(b) shows the first scan.

We ran a MCMC for model fitting, performing 20,000 iterations and discarding the first 10,000 iterations as burn-in. Again, we assessed convergence by the Raftery–Lewis diagnostic test implemented in the R package “coda”. We assigned the same non-informative priors as in the section Simulation from the model to all the parameters of the

Table 5

Simulated data: sensitivity analysis on the parameter e (keeping fixed $\eta = 1, d = -2.5, a_0 = 3, b_0 = 2, a_1 = b_1 = 1$), for block design (B) and event-related design (E). Results are in percentages.

		$e = 0.10$	$e = 0.30$	$e = 0.50$	$e = 0.65$
B	Accuracy	92.78	93.78	94.33	94.56
	Precision	100.00	100.00	100.00	99.60
	FPR	0.00	0.00	0.00	0.17
	FNR	21.96	18.92	17.23	16.22
E	Accuracy	90.22	91.00	91.89	93.67
	Precision	100.00	100.00	100.00	100.00
	FPR	0.00	0.00	0.00	0.00
	FNR	29.73	27.36	24.66	19.26

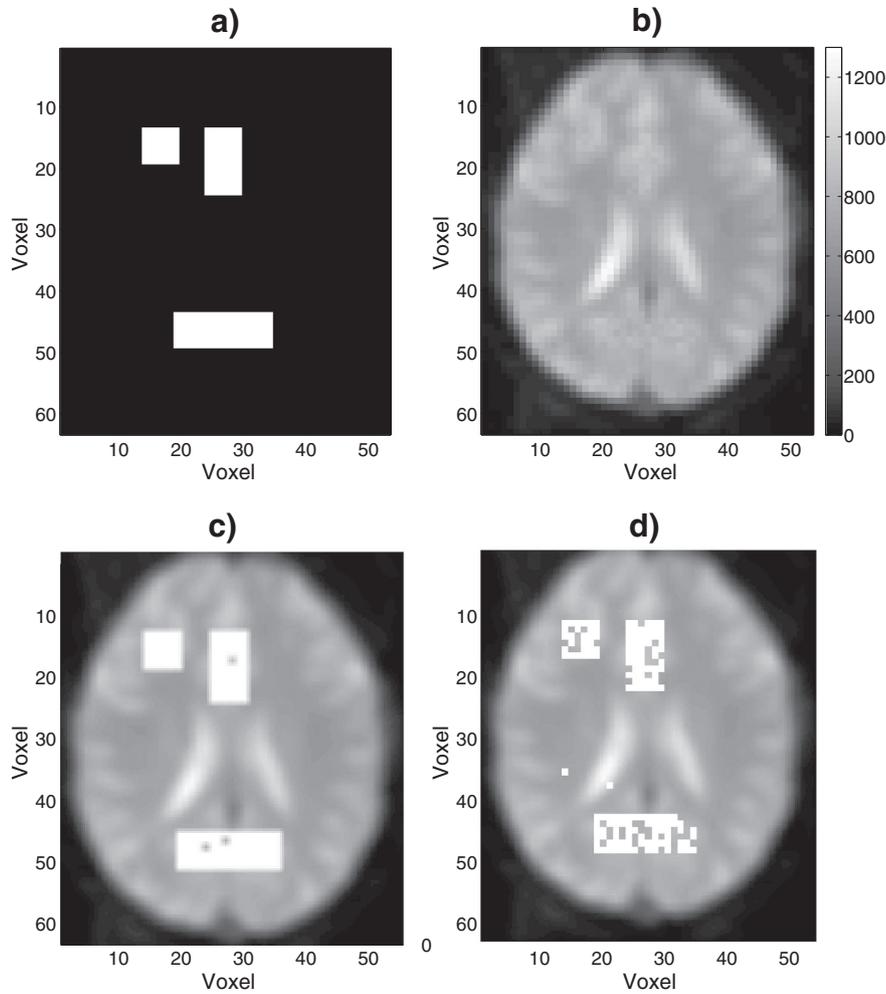


Fig. 5. Synthetic data: (a) true map of γ ; (b) first scan of our synthetic data. (c) Overlay on slice 27 of the posterior activation map obtained by assigning value 1 to those voxels with $p(\gamma_v = 1 | y) > 0.8$ and value 0 otherwise. (d) Result from SPM8; overlay on slice 27 of the posterior probability map of activation obtained with a Bayesian analysis, as described in the text and Eq. (16), and showing voxels with posterior probability greater than 0.999.

model and applied, at each iteration, discrete wavelet transforms using Daubechies minimum phase wavelets with 4 vanishing moments. Fig. 5(c) shows the posterior activation map, obtained by assigning value 1 to those voxels with $p(\gamma_v = 1 | y) > 0.8$ and value 0 otherwise. The accuracy of our detection is 99.86%. For comparison, we looked into the estimation results obtained from SPM8 by performing a Bayesian analysis, which detects regions where the response is different under the active and rest conditions, whether positive or negative. More specifically, we fit a GLM model with a design matrix with one regressor, convolved with a canonical HRF, and with a Gaussian MRF prior on the model coefficients (Penny et al., 2005). Fig. 5(d) shows the posterior probability map of activation produced by the software SPM8 following the method of Friston and Penny (2003), which calculates the posterior probability that a particular effect exceeds a threshold κ as

$$p = 1 - \Phi \left(\frac{\kappa - w^T M_{\beta|y}}{\sqrt{w^T C_{\beta|y} w}} \right), \quad (16)$$

with $M_{\beta|y}$ and $C_{\beta|y}$ the posterior mean and covariance of the parameter β . In particular, we obtained Fig. 5(d) by applying an F-contrast with contrast weight vector $w = [1, 0]^T$ to the estimation of the regression coefficients, and using a threshold of 0.999. We note that in SPM the detection of the activated voxels is based on the posterior probabilities of the beta coefficients, whereas in our model it is based on the posterior

probabilities of the γ coefficients, as described in the section [Simulation from the model](#). Table 6 shows comparisons of our method with SPM8 in terms of accuracy, FPR, and FNR. Our method outperforms SPM8, showing a higher accuracy, higher precision, lower FPR and lower FNR. Furthermore, Fig. 6 reports the posterior mean estimates, calculated from the MCMC samples, and the scatter plots of posterior mean estimates versus true values for the parameters β , λ , showing the correctness of our estimation for these parameters.

Fig. 7 shows the posterior clustering map of the voxels, obtained as described in the previous simulation, and the posterior mean maps of the estimates of the parameters ψ and α . As in the previous simulated scenario, we are able to recover the true clustering of the voxels, with all maps showing three large clusters and a few misclassified voxels, and to well estimate the true values of the parameters.

Table 6

Synthetic data: comparison of accuracy, FPR, and FNR of our model with SPM8. Results are in percentages.

	Our model	SPM (0.99)	SPM (0.999)
Accuracy	99.86	96.35	97.53
Precision	100.00	85.16	98.76
FPR	0.00	1.41	0.10
FNR	1.52	25.25	25.25

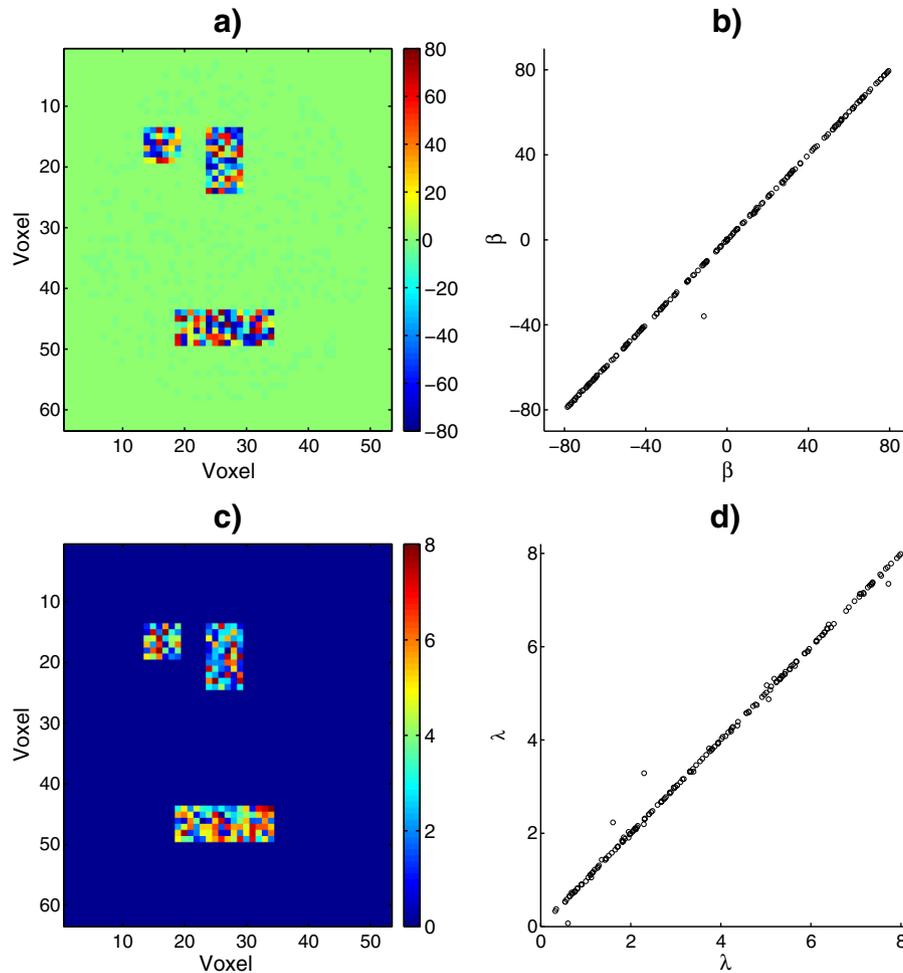


Fig. 6. Synthetic data: (a) posterior mean map of β ; (b) scatter plot of posterior mean estimates vs. true values of β ; (c) same as (a) for λ ; (d) same as (b) for λ .

A case study for fMRI data

Here we apply our model to the data set provided by the Wellcome Department of Imaging Neuroscience, available at <http://www.fil.ion.ucl.ac.uk/spm/data/attention/>. The data was original collected by Büchel and Friston (1997). The experiment was performed on a single subject, under 4 different conditions, and using a 2 Telsa Magnet

VISION (Siemens, Erlangen) whole body MRI system equipped with a head volume coil. The subject was scanned during 4 runs, with each run lasting 5 min and 22 s, producing 100 images for each run. This resulted in 400 contiguous multi-slice (32 slices, slice thickness 3 mm, giving 9.6 cm vertical field of view) T_2 -weighted fMRI images ($T_E = 40$ ms, 90 ms/image, $T_R = 3.22$ s, 64×64 pixels [19.2×19.2 cm]). The image volume obtained covered the whole brain, except for the

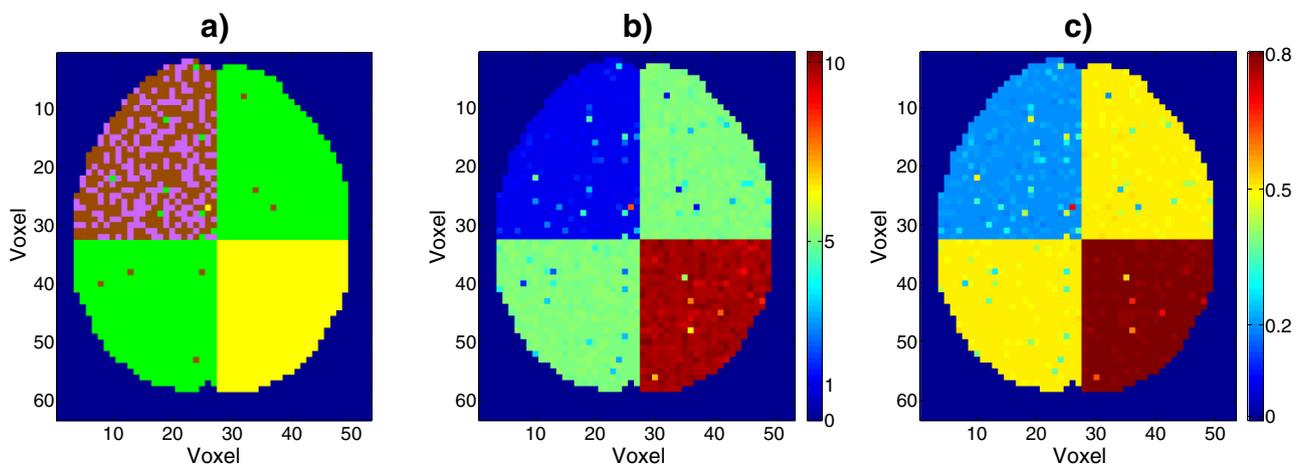


Fig. 7. Synthetic data: (a) posterior clustering map—different colors correspond to different clusters; (b) posterior mean map of ψ ; (c) posterior mean map of α .

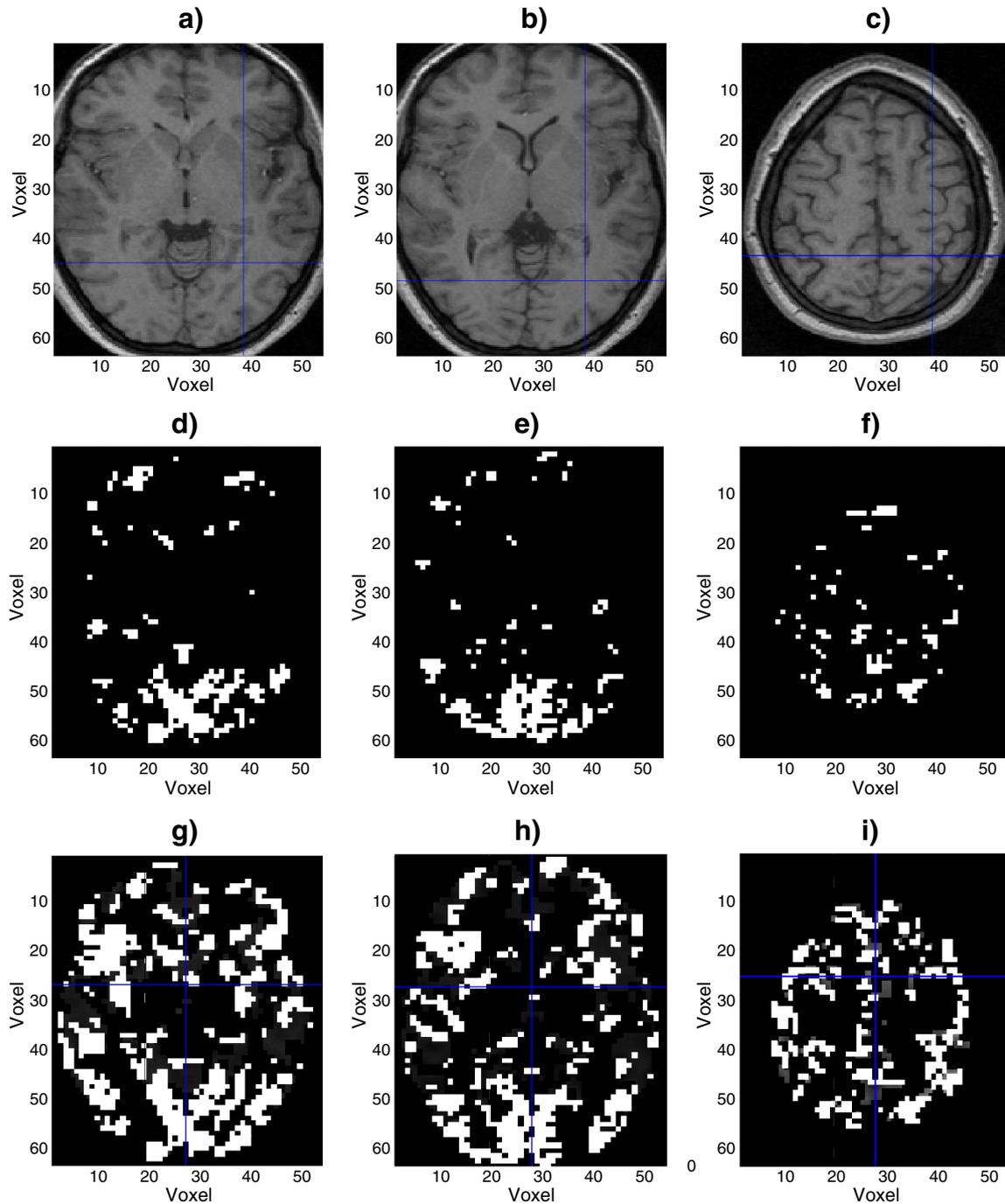


Fig. 8. Real fMRI data: First row: structural MRI images for (a) slice 17 (V1), containing the primary visual cortex; (b) slice 19 (V5), containing the motion-selective cortical area; (c) slice 36 (PP), containing the posterior parietal cortex. Second row: posterior activation maps, obtained by assigning value 1 to those voxels with $p(\gamma_v = 1 | y) > 0.9$, and value 0 otherwise. Third row: overlay of PPMs obtained by a Bayesian analysis with SPM8, as described in the text and Eq. (16), showing voxels with posterior probability greater than 0.999.

lower half of the cerebellum and the most inferior part of the temporal lobes. The first 10 scans in each run were discarded in order to eliminate magnetic saturation effects. The four conditions were “Fixation”, “Attention”, “No Attention”, and “Stationary”. Each condition lasted 32.2 s, giving 10 multi-slice volumes. The subject was asked to look at a fixed point (size 0.3°) in the middle of a transparent screen. During visual motion conditions, 250 white dots (size 0.1°) moved radially from the fixation point in random directions at a constant speed (4.7/s) toward the border of the screen where they vanished. During “Attention” and “No Attention” conditions, the subject fixated centrally while the white dots emerged from the fixation point toward the edge of the screen. During the “Attention” condition, the subject was asked to

detect changes of speed, and during the “No Attention” condition the subject was asked to just look at the moving points. During “Fixation” condition the subject saw only dark screen except for the visible fixation dot. During “Stationary” condition, the fixation point and 250 stationary dots were presented to the subject.

The data set comprises smoothed, spatially normalized, realigned, slice-time corrected images, on which we applied our wavelet-based Bayesian nonparametric model. We excluded the images obtained during “Stationary” condition, and grouped the other three conditions into two categories: a vector was defined with elements set to 1 for the images obtained during the “Attention” and “No Attention” conditions and to 0 during “Fixation”. This vector, representing the stimulus pattern

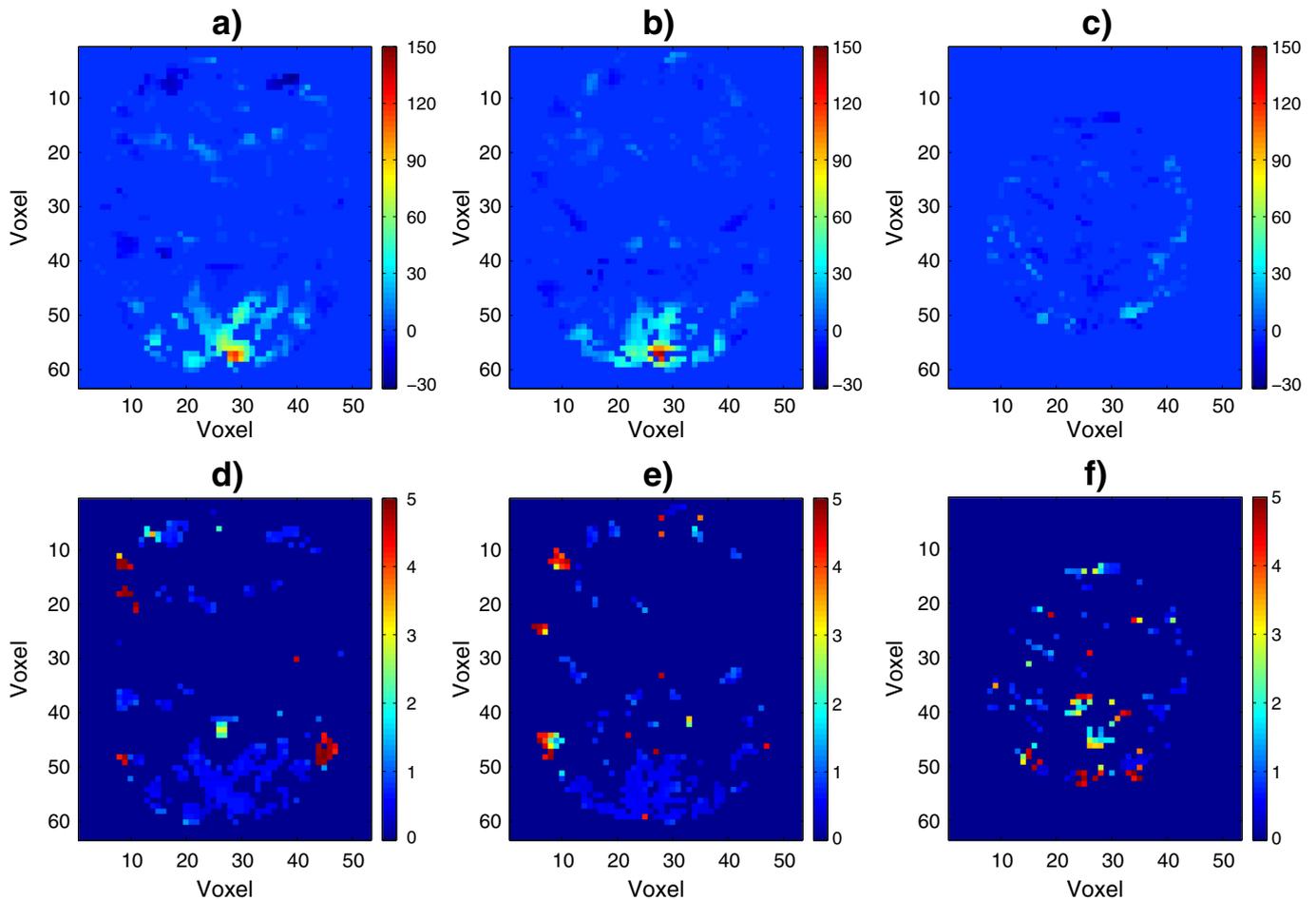


Fig. 9. Real fMRI data: First row: posterior mean maps of β for (a) V1, (b) V5 and (c) PP. Second row: posterior mean maps of λ for (d) V1, (e) V5 and (f) PP.

over time, was convolved with a Poisson hemodynamic function with voxel-dependent parameter λ_v , according to our model (1), to obtain the covariate X_v . We set $\tau = 100$, and $\eta = 1$, $a_1 = 1$, $b_1 = 1$, $a_0 = 3$, $b_0 = 2$, $d = -2.5$, $e = 0.3$, as in the simulation studies. We set $u_1 = 0$, $u_2 = 5$ in the prior for λ , to model the on-set delay of the HRF in accordance to the characteristics of the specific experimental design. We ran the MCMC with 20,000 iterations, discarding the first 10,000 as burn-in. At each iteration the Add, Delete, Swap steps for activation detection were repeated 12 times. As in Büchel and Friston (1997), we show results for slices containing the primary visual cortex (V1), the motion-selective cortical area (V5) and the posterior parietal cortex (PP), since neuroimaging studies have shown that these brain regions are activated by a stimulus with visual motion (Bushnell et al., 1981; Mountcastle et al., 1981; Treue and Maunsell, 1996). Structural MRI images of the slices are shown in the first row of Fig. 8. We have obtained automatic gray matter segmentation on the structural images for V1, V5 and PP with SPM8, and applied our model only on the resulting images, comprising voxels likely to correspond to grey matter. Our matlab code performed 1000 MCMC iterations in 64 min for V1, 61 min for V5, and 30 min for PP, on a desktop computer with CPU 3.10 GHz, and 4 GB of RAM.

Fig. 8 shows the posterior activation maps, obtained with our method, for the three slices under consideration. Fig. 9 shows the posterior mean estimates of the β and λ parameters. In both figures, images in the first column refer to slice 17 (V1), containing the primary visual cortex, those in the second column to slice 19 (V5), containing the motion-selective cortical area and those in the third column to slice 36 (PP), containing the posterior parietal cortex. Brighter regions in the images

denote higher probabilities, that is regions that are activated in response to the given “Attention” and “No Attention” visual stimulus. Activations appear to emerge mainly in the occipital lobe, which is known to be involved in the perception of visual motion. The positive values of the β coefficients are consistent with the hypothesis that the BOLD response in that region should increase upon presentation of a visual stimulus. Fig. 8 reports, for comparison, results from SPM8 obtained by fitting a GLM with an AR(1) noise, in which the regressor is convolved with a canonical HRF, and by carrying out the estimation via a two-sided Bayesian analysis with a GMRF prior. All methods appear to detect relevant activations in the designated areas, with activations that are mostly localized. We notice however that the detection is much sharper in our Bayesian maps, whereas the Bayesian analysis with SPM8 identifies wider areas as active.

Fig. 10 shows the posterior clustering maps of the three slices under study and the posterior mean maps of the parameters ψ and α . These maps capture association of neural activity over time as a spatio-temporal clustering of distant voxels. Patterns of similar neuronal dynamics are clearly visible. A close inspection of the results suggests that the use of the long memory process to model the error term and the clustering induced by the DP prior improve the deconvolution of the signal and the noise and explain, at least partially, the sparser activation observed in our model with respect to the SPM8 fit. In addition, the estimates of the variance parameter ψ and the decay parameter α in the middle and bottom row of Fig. 10 indicate that active voxels are characterized by relatively lower variance and increased auto-correlation of the time series. On the contrary, non-active voxels appear to be characterized by increased noise and are less dependent over time.

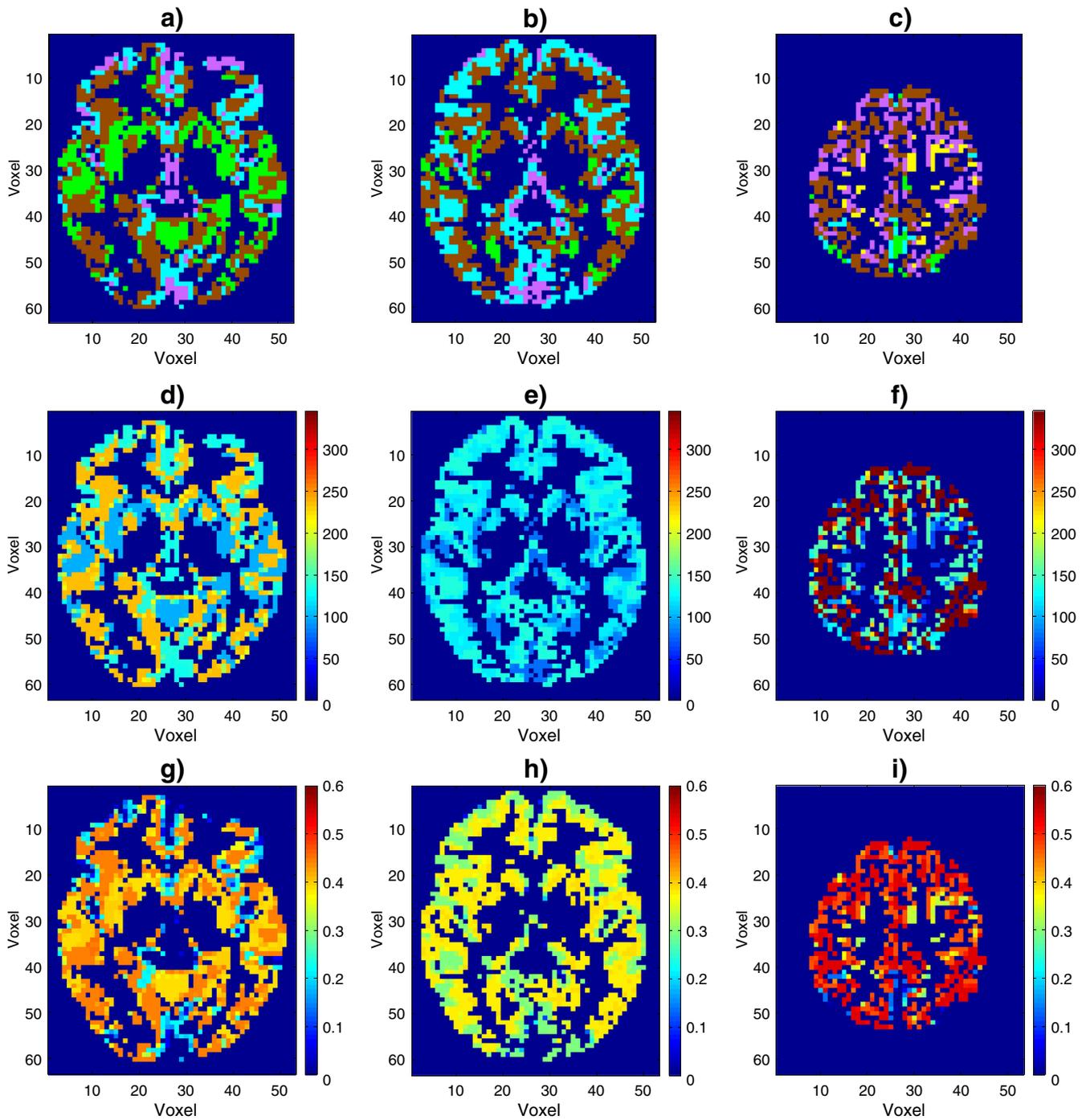


Fig. 10. Real fMRI data: posterior clustering maps of (a) V1, (b) V5, and (c) PP—different colors correspond to different clusters. Posterior mean maps of ψ for (d) V1, (e) V5 and (f) PP; posterior mean maps of α on (g) V1, (h) V5 and (i) PP.

Discussion

In this paper we have introduced a novel wavelet-based Bayesian nonparametric regression model for the analysis of fMRI data. Our inferential strategy detects regions of the brain which exhibit neuronal activity in response to a stimulus, while simultaneously inferring the association, or clustering, of spatially remote voxels that exhibit fMRI time series with similar characteristics. Our model combines several features into a single framework. In the time dimension, we allow for correlated errors and then employ wavelet transforms. We also account for the estimation of a voxel dependent delay parameter of the hemodynamic response function. Furthermore, we describe patterns of similar

neuronal dynamics via the clustering of the voxel time courses induced by a Dirichlet process prior model. In the spatial dimension, we capture dependence among nearby voxels via a Markov random field prior. Our results on simulated data have shown that the proposed modeling strategy works well for both block and event-related experimental design.

Our results on real fMRI data have confirmed the autocorrelated nature of the data, as many of the cerebral responses have shown small estimated values of the long memory parameter α . Other authors have compared estimation performances of models for fMRI data that make different assumptions on the error structure and have pointed out that models ignoring autocorrelation in the errors result in standard errors of the estimates that are inflated, see for example [Friston and](#)

Penny (2003). In our approach, we have exploited the decorrelation properties of discrete wavelet transforms, well studied in the literature, to aid the inferential task. In all applications in the paper we have applied discrete wavelet transforms with Daubechies minimum phase wavelets with 4 vanishing moments. Ko and Vannucci (2006) and Jeong et al. (2013) have shown that wavelets with higher degrees of regularity produce slightly better estimates of the long memory parameter, as they ensure wavelet coefficients approximately uncorrelated. On the other hand, the support of the wavelets increases with the regularity and boundary effects may arise in the DWT, so a trade-off is often necessary.

Our model has employed a MRF prior distribution on the detection indicator, as a way to encourage activations to take place in regions of neighboring voxels, penalizing detections in isolated voxels. MRF priors have been successfully used in a variety of applications, particularly for the analysis of genomic and imaging data. Care is needed in the specification of such priors, as allowing the smoothness parameter to vary can lead to the so-called *phase transition*, a situation in which the expected number of variables equal to 1 increases massively for small increments of the parameter, as described for example by Stingo et al. (2011). Phase transition has consequences such as the loss of model sparsity, and consequently a critical slow down of the MCMC, together with a higher number of falsely detected activations.

All the analyses have been conducted fitting our model on each separate slice, thus assuming a two-dimensional lattice. Slice-by-slice analyses are known to suffer from several limitations. For example, interpretation might be hard as it may be difficult to combine the results from separate slices as well as relate the slices to meaningful regions of the brain. Furthermore, the magnitude of the estimates might vary largely across slides. In real studies, the voxels overlay a three-dimensional lattice. However, the use of a 3D Ising model presents itself with further modeling and computational challenges, in terms of hyper-parameter selection and posterior computation. See, for example, Risser et al. (2009) and Vincent et al. (2010). Future work may consider extending our framework to the 3D analysis of a few specific brain regions. Further computational benefits may result by considering scalable approximate Bayesian inference, for example by resorting to Variational Bayes approaches as in Penny et al. (2003) and Roussos et al. (2012), and by exploring parallel computing, in particular by taking advantage of the Matlab built-in support for GPU computation, which will allow us to substantially speed-up expensive operations within a single MCMC iteration. Similarly, our approach can be extended to consider multiple signal coefficients per voxel by using priors that promote sparsity and allow to couple the regression parameters in space and/or time (see, for example, Babacan et al., 2014; Hernandez-Lobato et al., 2013; van Gerven et al., 2010).

A novel feature of our model is that we allow clustering of the time course responses of distant brain regions via a Dirichlet process (DP) prior. In our approach, the DP prior model induces a clustering of the voxels that exhibit time series signals with similar variance and long-memory behavior, as a way of capturing statistical association of neural activity over time. An interesting future direction is to explore extensions of our modeling strategy to multi-subject data, by using hierarchical Dirichlet process priors that capture the correlation of voxel time courses across multiple subjects, therefore linking the same anatomical regions across subjects (see, for example, Sanyal and Ferreira, 2012).

Acknowledgement

MG and FV were partially supported by the University of Texas MD Anderson Cancer Center's Cancer Center Support Grant (NIH CA016672).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2014.03.024>.

References

- Aguirre, G., Zarahn, E., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics II. *NeuroImage* 5 (3), 199–212.
- Babacan, S.D., Nakajima, S., Do, M.N., 2014. Bayesian group-sparse modeling and variational inference. *IEEE Trans. Signal Process.* (in press).
- Bogdan, M., Ghosh, J., Tokdar, S., 2008. A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing. In: Balakrishnan, N., Peña, E., Silvapulle, M. (Eds.), *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*. Vol. 1 of IMS Collections. Institute of Mathematical Statistics, Beachwood, Ohio, USA, pp. 211–230.
- Bowman, F., Caffo, B., Bassett, S., Kilts, C., 2008. A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* 39 (1), 146–156.
- Brown, P.J., Vannucci, M., Fearn, T., 1998. Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B (Stat Methodol.)* 60 (3), 627–641.
- Büchel, C., Friston, K., 1997. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cereb. Cortex* 7 (8), 768–778.
- Bullmore, E., Fadili, J., Maxim, V., Sendur, L., Whitcher, B., Suckling, J., Brammer, M., Breakspear, M., 2004. Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage* 23, 234–249.
- Bushnell, M., Goldberg, M., Robinson, D., 1981. Behavioral enhancement of visual responses in monkey cerebral in monkey cerebral cortex. I. modulation in posterior parietal cortex related to selective visual attention. *J. Neurophysiol.* 46, 755–772.
- Buxton, R., Frank, L., 1997. A model for the coupling between cerebral blood flow and oxygenation metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* 17 (1), 64–72.
- Craigmile, P.F., Percival, D.B., 2005. Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Trans. Inform. Theory* 51 (3), 1039–1048.
- Daubechies, I., 1992. Ten lectures on wavelets. SIAM, CBMS-NSF Conference Series, vol. 61.
- Escobar, M., West, M., 1995. Bayesian density estimation and inference using mixtures. *J. Am. Stat. Assoc.* 90 (430), 577–588.
- Fadili, M., Bullmore, E., 2002. Wavelet-generalised least squares: a new BLU estimator of linear regression models with $1/f$ errors. *NeuroImage* 15, 217–232.
- Ferguson, T., 1973. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* 1 (2), 209–230.
- Flandin, G., Penny, W., 2007. Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage* 34 (3), 1108–1125.
- Friston, K., 2011. Functional and effective connectivity: a review. *Brain Connect.* 1 (1), 13–36.
- Friston, K., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. *Hum. Brain Mapp.* 1 (2), 153–171.
- Friston, K., Penny, W., 2003. Posterior probability maps and SPMs. *NeuroImage* 19 (3), 1240–1249.
- Friston, K.J., 1994. Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* 2, 56–78.
- George, E., McCulloch, R., 1993. Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* 88 (423), 881–889.
- George, E., McCulloch, R., 1997. Approaches for Bayesian variable selection. *Stat. Sin.* 7 (2), 339–373.
- Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* 9, 416–429.
- Gössl, C., Auer, D., Fahrmeir, L., 2001. Bayesian spatio-temporal inference in functional magnetic resonance imaging. *Biometrics* 57 (2), 554–562.
- Guindani, M., Müller, P., Zhang, S., 2009. A Bayesian discovery procedure. *J. R. Stat. Soc. Ser. B (Stat Methodol.)* 71 (5), 905–925.
- Harrison, L., Green, G., 2010. A Bayesian spatiotemporal model for very large data sets. *NeuroImage* 50 (3), 1126–1141.
- Hernandez-Lobato, D., Hernandez-Lobato, J.M., Dupont, P., 2013. Generalized spike-and-slab priors for Bayesian group feature selection using expectation propagation. *J. Mach. Learn. Res.* 14, 1891–1945.
- Hjort, N., Holmes, C., Müller, P., Walker, S.G., 2010. *Bayesian Nonparametrics*. Cambridge University Press.
- Jbabdi, S., Woolrich, M., Behrens, T., 2009. Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. *NeuroImage* 44 (2), 373–384.
- Jeong, J., Vannucci, M., Ko, K., 2013. A wavelet-based Bayesian approach to regression models with long memory errors and its application to fMRI data. *Biometrics* 69, 184–196.
- Kalus, S., Sämann, P., Fahrmeir, L., 2013. Classification of brain activation via spatial Bayesian variable selection in fMRI regression. *ADAC* 1–21.
- Ko, K., Vannucci, M., 2006. Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes. *J. Stat. Plann. Infer.* 136 (10), 3415–3434.
- Lazar, N.A., 2008. *The statistical analysis of functional MRI data*. Statistics for Biology and Health/Springer, New York.
- Li, F., Zhang, N., 2010. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *J. Am. Stat. Assoc.* 105 (491), 1202–1214.
- Lindquist, M., 2008. The statistical analysis of fMRI data. *Stat. Sci.* 23 (4), 439–464.
- Lindquist, M., Wager, T., 2007. Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Hum. Brain Mapp.* 28 (8), 764–784.
- Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (7), 674–693.
- Meyer, F., 2003. Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series. *IEEE Trans. Med. Imaging* 22 (3), 315–322.
- Mountcastle, V., Anderson, R., Motter, B., 1981. The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* 1 (11), 1218–1225.

- Müller, P., Parmigiani, G., Rice, K., 2007. FDR and Bayesian multiple comparisons rules. In: Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., West, M. (Eds.), *Bayesian Statistics*, 8. Oxford University Press, Oxford, UK.
- Müller, P., Parmigiani, G., Robert, C., Rousseau, J., 2004. Optimal sample size for multiple testing: the case of gene expression microarrays. *J. Am. Stat. Assoc.* 99 (468), 990–1001.
- Neal, R., 2000. Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Stat.* 9 (2), 249–265.
- Penny, W., Kiebel, S., Friston, K., 2003. Variational Bayesian inference for fMRI time series. *NeuroImage* 19 (3), 727–741.
- Penny, W., Trujillo-Barreto, N., Friston, K., 2005. Bayesian fMRI time series analysis with spatial priors. *NeuroImage* 24 (2), 350–362.
- Poldrack, R., Mumford, J., Nichols, T., 2011. *Handbook of Functional MRI Data Analysis*. Cambridge University Press.
- Propp, J.G., Wilson, D.B., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Struct. Algorith.* 9 (1), 223–252.
- Quirós, A., Díez, R., Gamerman, D., 2010. Bayesian spatiotemporal model of fMRI data. *NeuroImage* 49 (1), 442–456.
- Raftery, A., Lewis, S., 1992. One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Stat. Sci.* 7 (4), 493–497.
- Risser, L., Vincent, T., Ciuciu, P., Idier, J., 2009. Robust extrapolation scheme for fast estimation of 3D Ising field partition functions: application to within-subject fMRI data analysis. *Med. Image Comput. Assist. Interv.* 12 (Pt 1), 975–983.
- Rodríguez, A., Müller, P., 2013. Nonparametric Bayesian inference. Vol. 9 of NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Mathematical Statistics and American Statistical Association.
- Roussos, E., Roberts, S., Daubechies, I., 2012. Variational Bayesian learning of sparse representations and its application in functional neuroimaging. *Machine learning and interpretation in neuroimaging*. *Lect. Notes Comput. Sci* 7263, 218–225.
- Sanyal, N., Ferreira, M., 2012. Bayesian hierarchical multi-subject multiscale analysis of functional MRI data. *NeuroImage* 63 (3), 1519–1531.
- Savitsky, T., Vannucci, M., Sha, N., 2011. Variable selection for nonparametric Gaussian process priors: models and computational strategies. *Stat. Sci.* 26 (1), 130–149.
- Sha, N., Vannucci, M., Tadesse, M.G., Brown, P., Dragoni, I., Davies, N., Roberts, T., Contestabile, A., Salmon, M., Buckley, C., Falciani, F., 2004. Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. *Bio-metrics* 60 (3), 812–819.
- Smith, A., Lewis, B., Ruttimann, U., Ye, F., Sinwell, T., Yang, Y., Duyn, J., Frank, J., 1999. Investigation of low frequency drift in fMRI signal. *NeuroImage* 9 (5), 526–533.
- Smith, M., Fahrmeir, L., 2007. Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *J. Am. Stat. Assoc.* 102 (478), 417–431.
- Stephens, M., 2000. Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B (Stat Methodol.)* 62 (4), 795–809.
- Stingo, F., Chen, Y., Tadesse, M., Vannucci, M., 2011. Incorporating biological information into linear models: a Bayesian approach to the selection of pathways and genes. *Ann. Appl. Stat.* 5 (3), 1978–2002.
- Strehl, A., Ghosh, J., 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* 3, 583–617.
- Stingo, F., Vannucci, M., Downey, G., 2012. Bayesian wavelet-based curve classification via discriminant analysis with Markov random tree priors. *Stat. Sin.* 22 (2), 465–488.
- Tadesse, M., Sha, N., Vannucci, M., 2005. Bayesian variable selection in clustering high-dimensional data. *J. Am. Stat. Assoc.* 100 (470), 602–617.
- Tewfik, A., Kim, M., 1992. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inf. Theory* 38 (2), 904–909.
- Thirion, B., Tucholka, A., Keller, M., Pinel, P., Roche, A., Mangin, J., Poline, J., 2007. High level group analysis of fMRI data based on Dirichlet process mixture models. *Inf. Process. Med. Imaging* 20, 482–494.
- Treue, S., Maunsell, J., 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382, 539–541.
- Turkheimer, F., Aston, J., Banati, R., Riddell, C., Cunningham, V., 2003. A linear wavelet filter for parametric imaging with dynamic pet. *IEEE Trans. Med. Imaging* 22 (3), 289–301.
- van Gerven, M.A.J., Cseke, B., Heskes, T., 2010. Efficient Bayesian multivariate fMRI analysis using a sparsifying spatio-temporal prior. *NeuroImage* 50 (1), 150–161.
- Vannucci, M., Corradi, F., 1999. Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. *J. R. Stat. Soc. Ser. B (Stat Methodol.)* 61 (4), 971–986.
- Vincent, T., Risser, L., Ciuciu, P., 2010. Spatially adaptive mixture modeling for analysis of fMRI time series. *IEEE Trans. Med. Imaging* 29 (4), 1059–1074.
- Woolrich, M., Jenkinson, M., Brady, J., Smith, S., 2004. Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE Trans. Med. Imaging* 23 (2), 213–231.
- Wornell, G., Oppenheim, A., 1992. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Process.* 40 (3), 611–623.
- Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. *NeuroImage* 15 (1), 1–15.
- Zarahn, E., Aguirre, G., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics I. *NeuroImage* 5 (3), 179–197.