

*Appl. Statist.* (2019)  
68, Part 3, pp. 521–541

# Bayesian analysis of functional magnetic resonance imaging data with spatially varying auto-regressive orders

Ming Teng,

*University of Michigan, Ann Arbor, USA*

Farouk S. Nathoo

*University of Victoria, Canada*

and Timothy D. Johnson

*University of Michigan, Ann Arbor, USA*

[Received August 2017. Revised September 2018]

**Summary.** Statistical modelling of functional magnetic resonance imaging data is challenging as the data are both spatially and temporally correlated. Spatially, measurements are taken at thousands of contiguous regions, called voxels, and temporally measurements are taken at hundreds of time points at each voxel. Recent advances in Bayesian hierarchical modelling have addressed the challenges of spatiotemporal structure in functional magnetic resonance imaging data with models incorporating both spatial and temporal priors for signal and noise. Whereas there has been extensive research on modelling the functional magnetic resonance imaging signal (i.e. the convolution of the experimental design with the functional choice for the haemodynamic response function) and its spatial variability, less attention has been paid to realistic modelling of the temporal dependence that typically exists within the functional magnetic resonance imaging noise, where a low order auto-regressive process is typically adopted. Furthermore, the auto-regressive order is held constant across voxels (e.g. AR(1) at each voxel). Motivated by an event-related functional magnetic resonance imaging experiment, we propose a novel hierarchical Bayesian model with automatic selection of the auto-regressive orders of the noise process that vary spatially over the brain. With simulation studies we show that our model is more statistically efficient and we apply it to our motivating example.

**Keywords:** Auto-regression; Bayesian hierarchical modelling; Functional magnetic resonance imaging; Imaging; Markov chain Monte Carlo methods

## 1. Introduction

In the analysis of functional magnetic resonance imaging (MRI) data a key challenge is dealing with spatial and temporal correlation. Temporal correlation can arise from many sources, including MRI scanner drift at very low frequencies, slow vascular or metabolic oscillations that are of moderate to low frequency and other sources of noise such as breathing and heart-beat. Simply ignoring these sources of temporal correlation may lead to increased false positive discoveries (Makni *et al.*, 2006). To deal with these issues, a variety of approaches have been proposed. One commonly used approach, namely prewhitening, works by estimating the

*Address for correspondence:* Timothy D. Johnson, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, USA.  
E-mail: [tdjtdj@umich.edu](mailto:tdjtdj@umich.edu)

temporal correlation and then decorrelating the noise by using the estimates (Bullmore *et al.*, 1996; Locascio *et al.*, 1997). Besides these stationary time series models, non-stationary  $1/f$ -models have also been proposed (Zarahn *et al.*, 1997; Bullmore *et al.*, 2004). According to Friston *et al.* (2000), prewhitening can produce an extraneous source of bias. Alternatively, a band-pass filtering technique known as precolouring can be applied to the data first, followed by statistical modelling that deals with the auto-correlation in the coloured data. For a review and discussion of these approaches the reader is referred to Woolrich *et al.* (2001). Whereas high pass filtering has proven to be beneficial in increasing the power of the statistical analysis, the low pass filtering involved in colouring is considered controversial in that it tends to add auto-correlation into the data (Skudlarski *et al.*, 1999; Della-Maggiore *et al.*, 2002).

Although accurate temporal modelling is important for estimation of the functional MRI signal and, especially, its standard error, traditional approaches apply a temporal model at each voxel, independently, i.e. they ignore spatial correlation. More specifically, this mass univariate approach, which is considered to be the classical approach to the analysis of functional MRI data, includes a smoothing step involving a spatial Gaussian filter that is applied to the data first (Friston *et al.*, 1995), followed by model estimation at each voxel, and then statistical inference is based on random-field theory (Worsley and Friston, 1995) which is applied to adjust for multiplicity in the spatial domain. Although this approach remains the most common approach for analysing functional MRI data it has been criticized on several grounds. For example, the Gaussian kernel that is used to smooth the data must be prespecified and introduces artificial spatial correlation into the data, similarly to how precolouring adds artificial temporal correlation. In addition, this approach does not directly account for spatial correlation in the model.

Partly as a result of these criticisms, Bayesian models with spatially structured priors have been proposed that allow for the calculation of posterior probability maps (PPMs) for activation. This Bayesian approach to inference is based on an explicit spatial model and does not require smoothing the data with a Gaussian kernel; nor does it require the use of random-field theory-based adjustments for multiplicity. A variety of spatiotemporal Bayesian models have been proposed. One model that is widely used and implemented within the SPM software (<https://www.fil.ion.ucl.ac.uk/spm/>) is the general linear model–auto-regression (GLM–AR) model (Penny *et al.*, 2003, 2005, 2007). The GLM–AR model assumes that the data can be decomposed into two sources of variability. The first source is the product of the design matrix for the functional MRI experiment convolved with a haemodynamic response function and experimental factors, and the second source represents temporally correlated noise that is modelled by using a low order AR structure. In addition, the regression coefficients and the AR coefficients vary across space and are assigned spatial smoothing priors. Gössl *et al.* (2001) proposed a model where the data are decomposed into three sources: a spatial stimulus, a deterministic trend and a white noise process. However, this modelling approach may not account for some higher frequency stochastic noise components. Woolrich *et al.* (2004b) assumed that the temporal noise arises from both large-scale and small-scale variation, and built a space–time simultaneous AR model that accounts for both scales of variation. Methods focusing on spatial variable selection have also been proposed (see, for example, Bezener *et al.* (2016), Lee *et al.* (2014) and Musgrove *et al.* (2016)), whereas Kim *et al.* (2010) proposed a mixture-of-experts model to represent spatial activation clusters. Although these models have some different characteristics which make the approaches unique, most of them commonly assume a homogeneous, low order AR or AR moving average process for the temporal noise. By homogeneous, we mean that the order of the AR or AR moving average process is assumed constant across all voxels. This assumption was also made in Penny *et al.* (2003). However, as we demonstrate by using a simple empirical example in the

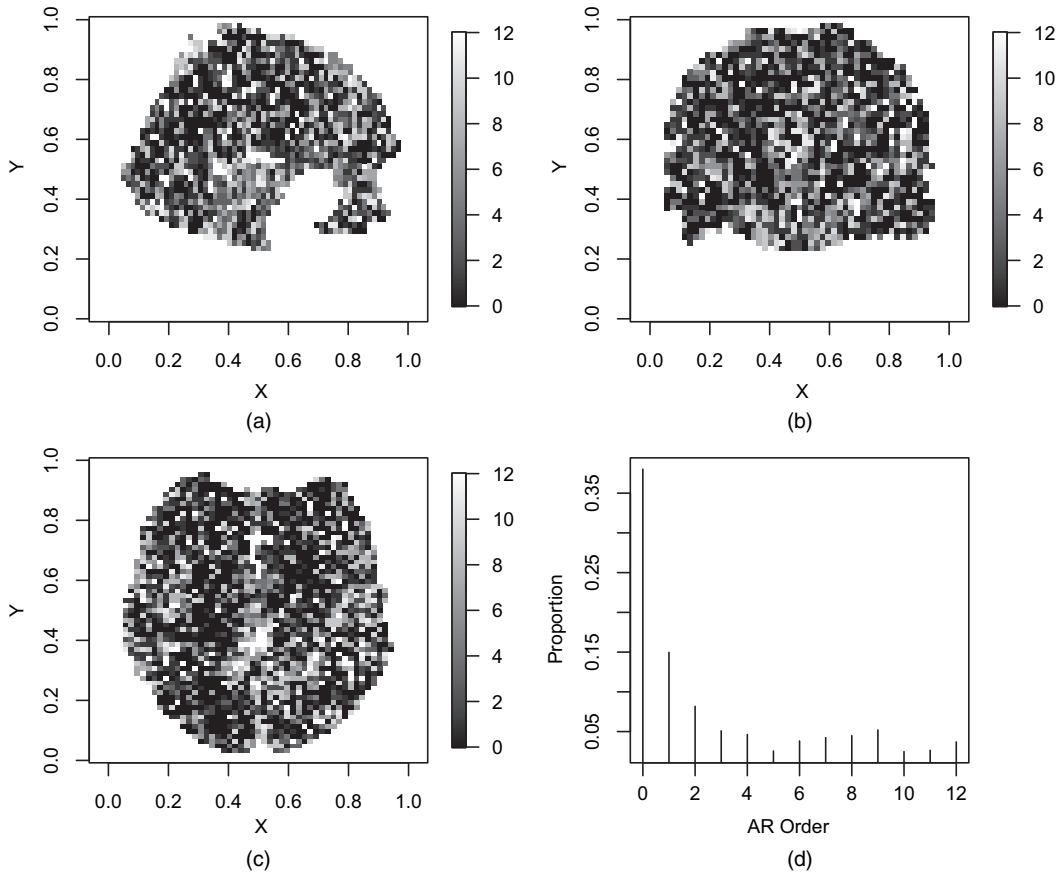
next section, this homogeneous AR order assumption may be violated in real functional MRI data.

Instead of formulating the model at each voxel and then adopting spatial smoothing priors for parameters across voxels, another branch of research is based on vector AR processes; see Harrison *et al.* (2003). This approach enables time-lagged dependence across voxels and spatial–temporal interaction but fitting these models across a large number of voxels is computationally intractable and low rank approximations must be used. These models are also useful for studying effective brain connectivity, where one time course is used to predict the other (Castruccio *et al.*, 2016; Chang and Glover, 2010). Another line of work chooses to model the temporal noise as a  $1/f$  long memory process and applies discrete wavelet transforms towards fitting the model (see, for example, Jeong *et al.* (2013), Bullmore *et al.* (2004), Fadili and Bullmore (2002) and Meyer (2003)). Although this approach seems promising, our focus in this paper will be with modelling short-term memory by using the classical AR process and spatial priors. The reason that we choose to work with the AR process is because of its mathematical amenability and simplicity, and its wide use in various areas of science. A novel aspect of our work is that we allow the data to determine the order of the AR process at each voxel, borrowing strength from neighbouring voxels. The AR order is determined by using ideas from Bayesian spatial variable selection as we expect neighbouring voxels to have similar temporal structure.

Computation is an important issue when considering Bayesian spatial–temporal models for functional MRI data. Although the main focus of this paper lies with the development of a new model, another aspect of this work is the comparison of fully Bayesian and approximate Bayesian computation methods. Because of the computational burden that is associated with fitting models to high dimensional brain imaging data, approximate Bayesian methods have received considerable attention in the neuroimaging literature. One such method is variational Bayes (VB) inference (Penny *et al.*, 2003, 2007; Woolrich *et al.*, 2004a). As there are currently no theoretical results quantifying the accuracy of VB methods (in contrast with Markov chain Monte Carlo (MCMC) sampling which is justified by large sample theory of stationary Markov chains), the evaluation of VB models must be performed case by case. In some cases, the performance of VB models can be quite good and in other cases it can be quite poor. In addition to the implementation of our new model based on a suitably designed MCMC sampler, we also develop an MCMC algorithm to sample from the posterior of the original GLM–AR model. We then compare our model with both the VB implementation of the GLM–AR model (using SPM code) and our MCMC implementation of the GLM–AR model. Our studies indicate that under a low signal-to-noise (SNR) ratio the accuracy of MCMC sampling outperforms VB methods according to several criteria.

### 1.1. Motivating example

Our motivating example comes from a single subject in a functional MRI experiment examining a face repetition stimulus. The experiment involves the presentation of either famous faces or non-famous faces with each type of face presented two times. Convolving this experiment design with the canonical haemodynamic response function and its time and dispersion derivatives leads to a design matrix with 12 columns plus one extra column for an intercept term in the regression model. After performing the necessary preprocessing steps as described in Penny *et al.* (2005), we fit a simple linear regression at each of the voxels. After obtaining the residuals from each voxel-specific fit, we fit an AR process up to order 12 for each voxel by using the `ar` function in R (R Core Team, 2018). The Akaike information criterion AIC is used to estimate the true AR orders. Fig. 1 is a pictorial representation of the results.



**Fig. 1** AIC-estimated AR orders (the upper AR order bound is set to 12): (a) sagittal view; (b) coronal view; (c) axial view; (d) AR order

Fig. 1 shows considerable variability in the estimated AR order across voxels. Although most of the estimated AR orders are 4 or less, higher orders up to 12 are selected at some of the voxels. Furthermore, these estimated AR orders tend to show some extent of spatial clustering. If, as is often done, we simply model the data by using a homogeneous low order AR process, then the voxels with higher AR orders will be incorrectly modelled, and this inaccuracy in the modelling of temporal noise will have an effect on the inference on the covariates of interest (via underestimated standard errors), resulting in potentially false inferences about brain activation. To address this issue, we propose a spatially varying AR order (SVARO) model, where the AR orders vary spatially across the brain. This is made possible by adopting a spike-and-slab prior with a stochastic search variable selection scheme. Spatial clustering of AR orders is incorporated by imposing an Ising prior (Ising, 1925) as the latent indicator for the spike-and-slab prior. We update the latent indicators by using the Swendsen–Wang algorithm (Swendsen and Wang, 1987) alternating with Gibbs sampling in our MCMC algorithm. To avoid the phase transition problem that is associated with the Ising model, we derive theoretical bounds as in Li *et al.* (2015) and use these bounds to prevent critical slowing of the algorithm. We compare our model with the GLM–AR model of Penny *et al.* (2007) (implemented under two schemes: our self-written MCMC sampler and the VB algorithm that is available in the SPM software)

in terms of mean-squared error (MSE) and sensitivity. We conduct these comparisons by using two simulation studies and then compare results on the motivating data set.

The rest of the paper is organized as follows: in Section 2 we present our model and MCMC sampling scheme. We present results from our simulation studies in Section 3, followed by the analysis of the face repetition data set in Section 4. Lastly, we provide a discussion and outline some possible directions for future work in Section 5.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

<https://rss.onlinelibrary.wiley.com/hub/journal/14679876/series-c-datasets>

## 2. The spatially varying auto-regressive order model

First, we shall define the model likelihood and then specify the spatial and temporal priors. Then we discuss our posterior sampling scheme, including the construction of bounds for the hyperpriors that we use for the Ising priors that are used for variable selection; lastly we discuss inference based on posterior probability maps.

### 2.1. Model likelihood

We let  $P$  denote the maximum allowable order of the AR process at every voxel and let  $K$  denote the number of regression coefficients at each voxel. Using similar notation to that in Teng *et al.* (2016), for voxel  $n$  ( $n = 1, \dots, N$ ), we let  $\mathbf{y}_n$  denote the observed time series of length  $T$ . For simplicity, our model is specified conditionally on the first  $P$  observations at each voxel so that the likelihood function is constructed on the basis of the model for the remaining  $T - P$  observations in the time series. We let  $\mathbf{X}$  denote the  $(T - P) \times K$  design matrix,  $\mathbf{w}_n$  denote the  $K$ -dimensional vector of regression coefficients at voxel  $n$  and  $\mathbf{e}_n$  denotes the corresponding error term. Define the vector  $\mathbf{y}_n \equiv \mathbf{y}_{1:T,n}$ , the entire time series observed at voxel  $n$ . The hierarchical model is specified in several stages. The first stage is a general linear model:

$$\mathbf{y}_{P+1:T,n} = \mathbf{X}\mathbf{w}_n + \mathbf{e}_n, \tag{1}$$

where we emphasize again the implicit conditioning on  $\mathbf{y}_{1:P,n}$  ( $n = 1, \dots, N$ ). Let  $\tilde{\mathbf{E}}_n$  denote the embedded error (or lagged prediction) matrix of dimension  $(T - P) \times P$ , with  $t, p$  element  $(\mathbf{y}_{P+1:T,n} - \mathbf{X}\mathbf{w}_n)_{[t-p]}$  where the notation  $[i]$  denotes the  $i$ th index of the vector. Let  $\mathbf{z}_n \equiv \mathbf{z}_{P+1:T,n}$  denote a vector of independent and identically distributed mean 0 Gaussian random variables with precision  $\lambda_n$ . The second stage is then an AR model at each voxel:

$$\mathbf{e}_n = \tilde{\mathbf{E}}_n \mathbf{a}_n + \mathbf{z}_n \tag{2}$$

where  $\mathbf{a}_n$  is a vector of AR coefficients for the time series at voxel  $n$ .

Letting  $c$  denote a constant term, the log-likelihood for voxel  $n$  is

$$l_n = -\frac{\lambda_n}{2} \sum_{t=P+1}^T (y_{tn} - \mathbf{x}_t \mathbf{w}_n - \tilde{\mathbf{e}}_{tn} \mathbf{a}_n)^2 + \frac{T-P}{2} \log(\lambda_n) + c, \tag{3}$$

where  $\mathbf{x}_t$  is the  $(t - P)$ th row of the design matrix  $\mathbf{X}$  and  $\tilde{\mathbf{e}}_{tn}$  is the  $(t - P)$ th row of  $\tilde{\mathbf{E}}_n$ .

Summing this log-likelihood over the number of voxels,  $n$ , we obtain the overall log-likelihood:

$$l = \sum_{n=1}^N \left\{ -\frac{\lambda_n}{2} \sum_{t=P+1}^T (y_{tn} - \mathbf{x}_t \mathbf{w}_n - \tilde{\mathbf{c}}_{tn} \mathbf{a}_n)^2 + \frac{T-P}{2} \log(\lambda_n) + c \right\}. \tag{4}$$

2.2. Spatial prior

At the next level of the model we specify a spatial smoothing prior for the regression coefficients  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_N)$ , a  $K \times N$  matrix, with  $k$ th row  $\mathbf{W}_{k,\cdot}$ . Following Penny *et al.* (2005), we assume that the prior for  $\mathbf{W}$  takes the form

$$\pi(\mathbf{W}) = \prod_{k=1}^K \pi(\mathbf{W}_{k,\cdot}), \tag{5}$$

$$\mathbf{W}_{k,\cdot} \sim N\{\mathbf{0}, \alpha_k^{-1} (\mathbf{S}^T \mathbf{S})^{-1}\}, \quad k = 1, \dots, K. \tag{6}$$

*A priori* the regression coefficients within voxels are independent (5) whereas, spatially, the  $k$ th,  $k = 1, \dots, K$ , regression coefficients (across voxels) are modelled dependently through an  $N$ -dimensional multivariate normal distribution (6). Here  $\mathbf{S}$  is known as a Laplacian matrix. The  $n$ th diagonal term of this matrix is equal to the corresponding number of first-order neighbours of the voxel  $n$ . All off-diagonal terms are 0 except for  $-1$  in off-diagonal elements  $(n, j)$  and  $(j, n)$  if voxel  $j$  is a neighbour of voxel  $n$ , for  $n = 1, \dots, N$ . This form for the prior accommodates spatial smoothing while also being sparse and convenient to work with computationally. Also, note that each regression coefficient is smoothed independently of one another and this smoothness is controlled by  $\alpha_k$ . In the SPM12 software this prior is referred to as the low resolution tomography prior. Ultimately, what is of primary interest in studies of brain activation is a posterior probability of some function of these regression coefficients, and this posterior probability is computed at each voxel to produce PPMs (see Penny *et al.* (2005)). The precision parameter in distribution (6),  $\alpha_k$ , is assigned a conditionally conjugate hyperprior distribution:

$$\alpha_k \stackrel{\text{IID}}{\sim} \text{gamma}(q_1, q_2), \quad k = 1, \dots, K. \tag{7}$$

2.3. Temporal prior

The key difference between our model and the model of Penny *et al.* (2007) lies in our modelling of the temporal noise. Rather than assuming that AR orders are homogeneous across the brain (we refer the readers to Teng *et al.* (2016) and Penny *et al.* (2007) for model details), we allow for variability in the order of the AR processes across voxels. In addition, we adopt a spatial prior for this variability under the assumption that the order of the AR processes of neighbouring voxels will be similar and that their coefficients will also be similar. Specifically, for each voxel  $n$  and AR order  $p$ ,  $p = 1, \dots, P$ , we assign the latent indicator variable  $\gamma_{pn}$  to the  $p$ th AR coefficient  $a_{pn}$ , such that, given  $\gamma_{pn}$  ( $p = 1, \dots, P; n = 1, \dots, N$ ),  $a_{pn}$  will be conditionally independent.  $\gamma_{pn}$  will take value 1 if the  $p$ th AR coefficient is non-zero at voxel  $n$  and will take the value 0 otherwise. Conditionally on  $\gamma_{pn}$ ,  $a_{pn}$  will either have a normal distribution or unit mass at 0. This is commonly referred to as the spike-and-slab prior (George and McCulloch, 1993; Mitchell and Beauchamp, 1988), though we note that our formulation is a *spatial* spike-and-slab prior and that this prior is assigned to the coefficients of the AR process governing the temporal noise:

$$\pi(\mathbf{a} | \boldsymbol{\gamma}) = \prod_n \prod_p \pi(\mathbf{a}_{pn} | \gamma_{pn}),$$

$$\pi(a_{pn} | \gamma_{pn}) = \gamma_{pn} \phi(a_{pn}; 0, \tau_p^{-1}) + (1 - \gamma_{pn}) I_0(a_{pn}).$$

Here,  $\phi(\cdot; a, b)$  is the probability density function of a normal distribution with mean  $a$  and variance  $b$  and  $I_0(\cdot)$  is the indicator function that its argument equals 0, and  $\gamma_{pn}$  is the binary indicator.  $\tau_p$  is the precision of the normal component and is again given a gamma prior  $\tau_p \sim \text{gamma}(r_1, r_2)$ .

The advantages of introducing such a prior are threefold: first, the coefficients in the AR process at each voxel that lack support from the data can be effectively removed from the model. This enables us to infer which AR coefficients are present in which voxels. Second, the number of voxels with large AR orders is non-zero but expected to be small, which is an aspect of this prior that can be controlled by tuning the hyperparameters. Third, for some of the voxels it is possible that some AR coefficients are 0 whereas others are non-zero. For example if the AR order at a particular voxel is 5, the second and third coefficient may be 0 whereas the fourth and fifth are non-zero. The model proposed is sufficiently flexible to allow for this behaviour, since we have a total of  $P$  independent Ising processes: one for each possible  $p \in \{1, \dots, P\}$ . Note that this enables voxels to have independent temporal error structures when  $\gamma_{pn} = 0$  for all  $p$ .

There are of course other model selection techniques that could have been considered. For example a type of Bayesian lasso could have been used as an alternative to the spike-and-slab prior. Wang *et al.* (2007) have applied the lasso to the selection of AR processes, and for the Bayesian lasso we refer to Schmidt and Makalic (2013). A recent alternative prior known as the non-local prior for variable selection has been proposed by Johnson and Rossell (2012) that has desirable consistency properties and yields smaller prediction errors in large sample settings. A review of Bayesian priors that can be employed for model selection is presented in O’Hara and Sillanpää (2009).

We assume that the indicator processes are independent across the set  $\{1, \dots, P\}$ :  $\pi(\gamma) = \prod_{p=1}^P \pi(\gamma_p)$ , where  $\gamma_p = (\gamma_{p1}, \dots, \gamma_{pN})^T$ . The simplest variable selection model would assume that  $\gamma_{pn}$  follows a Bernoulli distribution (George and McCulloch, 1993). Here, to borrow information across neighbours, we choose to use the Ising prior (Ising, 1925) independently for each  $p = 1, \dots, P$ :

$$\pi(\gamma_p) \propto \exp\left\{\beta_{0p} \sum_n \gamma_{pn} + \beta_{1p} \sum_{n_1 \sim n_2} I(\gamma_{pn_1} = \gamma_{pn_2})\right\}, \tag{8}$$

where  $\beta_{0p}$  and  $\beta_{1p}$  are two hyperparameters controlling the sparsity and smoothness of the binary latent field respectively. Larger values of  $\beta_{0p}$  result in less sparsity and larger values of  $\beta_{1p}$  result in more smoothness. One issue with the Ising model that requires some care is the choice of hyperparameters. When  $\beta_{1p}$  take values near the phase transition boundary, the mixing of an MCMC sampler will suffer from a critical slowdown (Stanley *et al.*, 1987). To avoid the phase transition boundary, we adopt an analytical approach similar to Li *et al.* (2015) to quantify the value for the bounds of both  $\beta_{0p}$  and  $\beta_{1p}$ . An outline of the bound derivation is given below in Section 2.5.

### 2.4. Posterior sampling scheme

Most parameter updates that are related to the posterior sampling of our model can be accomplished via Gibbs sampling. One exception is the update to the latent indicators  $\gamma_p$ ,  $p = 1, \dots, P$ . For each  $\gamma_p$  we use the Swendsen–Wang algorithm alternating with Gibbs sampling (Johnson *et al.*, 2013). This strategy, i.e. mixing Swendsen–Wang updates with Gibbs updates for  $\gamma_p$ , has proved successful in improving the mixing of the Markov chain sampler and results in faster block updates in various studies (Higdon, 1998).

2.5. Bound construction for the Ising model hyperparameters

The hyperparameters in the Ising priors play a vital role in posterior estimation. Without careful selection, we are faced with mixing issues that are associated with ‘phase transition’ (Stanley *et al.*, 1987). There are various approaches to sampling such hyperparameters. Johnson *et al.* (2013) estimated them by using path sampling and Gelman and Meng (1998) and Shu *et al.* (2015) proposed a Monte Carlo expectation–maximization algorithm to obtain a point estimate of the hyperparameters, but these procedures are too time consuming for functional MRI data and our model, considering that we need to estimate  $P$  independent Ising fields and thousands of voxels. Smith and Fahrmeir (2007) proposed to update the hyperparameters and binary indicators together, but this approach still suffers from the potential possibility of sampling over the phase transition boundary. Here, we adopt an approach that is similar to that considered in Li *et al.* (2015) and construct some theoretical bounds to prevent phase transition. The resulting hyperparameter values are then chosen as fixed values within the estimated bounds. This procedure turns out to work well in our analysis and studies.

To construct the bounds, we first write out the posterior conditional density with respect to  $\gamma_p$ :

$$\pi(\gamma_p | \cdot) \propto \exp \left\{ \beta_{0p} \sum_{n=1}^N \gamma_{pn} + \beta_{1p} \sum_{n_1 \sim n_2} I(\gamma_{pn_1} = \gamma_{pn_2}) + \sum_{n=1}^N \sum_{t=1}^T \frac{-\lambda_n}{2} \left( e_m - \sum_{j=1}^P \tilde{e}_{mja_{jn}} \right)^2 \right\}. \quad (9)$$

In our model, where multiple AR orders exist across space, it is natural to assume that

- (a) there are relatively few time series with large AR order and
- (b)  $\pi(\gamma_p \neq \mathbf{0} | \cdot) > \pi(\gamma_p = \mathbf{0} | \cdot)$  for small  $p$ .

On a three-dimensional lattice, selected voxels give rise to the greatest number of neighbouring pairs when they form a cube. Therefore, following Li *et al.* (2015), the approximate locations of the voxels selected are obtained by using cubes. Let  $\pi_p$  denote those voxels whose  $p$ th AR coefficient is non-zero. Let  $V_p = \lceil (\pi_p N)^{1/3} \rceil$  where  $\lceil x \rceil$  is the greatest integer less than  $x$ . Then, for a cube containing  $V_p^3$  voxels, there are  $3V_p^2(V_p - 1)$  neighbouring pairs. On the basis of this we derive

$$\beta_{0p} \sum_{n=1}^N \gamma_{pn} + \beta_{1p} \sum_{n_1 \sim n_2} I(\gamma_{pn_1} = \gamma_{pn_2}) = \beta_{0p} V_p^3 + 3\beta_{1p} V_p^2 (V_p - 1), \quad (10)$$

for the  $p$ th latent Ising image. According to assumption (a) for large  $p$  (typically  $p > 8$ )  $\beta_{0p} + 3\beta_{1p} < 0$ . According to assumption (b) for small  $p$  (typically  $p < 4$ ),

$$\begin{aligned} \sum_{n=1}^N \sum_{t=1}^T -0.5\lambda_n \left( e_m - \sum_{j=1, j \neq p}^P \tilde{e}_{mja_{jn}} \right)^2 &\leq \sum_{n=1}^N \sum_{t=1}^T -0.5\lambda_n \left( e_m - \sum_{j=1}^P \tilde{e}_{mja_{jn}} \right)^2 \\ &\quad + \beta_{0p} \sum_{n=1}^N \gamma_{pn} + \beta_{1p} I(\gamma_{pn_1} = \gamma_{pn_2}). \end{aligned} \quad (11)$$

Reorganizing this by moving the first term on the right-hand side to the left-hand side produces

$$\begin{aligned} \sum_{n=1}^N \sum_{t=1}^T -0.5\lambda_n \left\{ \left( e_m - \sum_{j=1, j \neq p}^P \tilde{e}_{mja_{jn}} \right)^2 - \left( e_m - \sum_{j=1}^P \tilde{e}_{mja_{jn}} \right)^2 \right\} \\ \leq \beta_{0p} \sum_{n=1}^N \gamma_{pn} + \beta_{1p} I(\gamma_{pn_1} = \gamma_{pn_2}). \end{aligned} \quad (12)$$



The two terms in the curly brackets on the left-hand side can be considered as one with and without  $\tilde{e}_{mp}a_{pn}$ . Thus, it can be approximately considered as the residual sum of squares of an AR when the  $p$ th coefficient  $a_{pn}$  is included or excluded in the model. Let  $R_{pn}^2$  denote this coefficient of determination; then we have

$$\beta_{0p} \sum_{n=1}^N \gamma_{pn} + \beta_{1p} \sum_{n_1 \sim n_2} I(\gamma_{pn_1} = \gamma_{pn_2}) \geq -0.5 \sum_{n=1}^N \sum_{t=1}^T \frac{R_{pn}^2}{1 - R_{pn}^2}. \tag{13}$$

Combined with equation (10), we have

$$\beta_{0p} V_p^3 + 3\beta_{1p} V_p^2 (V_p - 1) \geq -0.5\pi_p NT \frac{R_{pn}^2}{1 - R_{pn}^2}. \tag{14}$$

As an example, let the number of voxels equal 56526: that of the real data that we analyse below. Suppose, for example, for the  $p$ th AR coefficient we set  $\pi_p = 0.1$  and that 5% of the total variation can be explained by the inclusion of the  $p$ th AR coefficient,  $a_{pn}$ . From inequality (14) we then have the bound  $\beta_{0p} + 2.83\beta_{1p} \geq -9.26$ .

Note that inequality (14) just gives a range of values for the hyperparameters, rather than providing the values directly. In practice, the exact values of hyperparameters are largely determined by the researcher, which should be combined with one’s prior experience and an initial analysis of the data. We suggest obtaining such values on the basis of some exploratory *ad hoc* approaches, e.g. a linear regression at each voxel followed by fitting an AR process. Then the estimated AR orders can be used as a reference when determining the hyperparameters in the Ising model. This method has turned out to work well empirically as we demonstrate in Section 4.

### 2.6. Posterior probability maps

A primary emphasis on functional MRI data analysis is inference for activation. For Bayesian modelling of functional MRI data this is typically achieved via PPMs. Let  $\mathbf{c}^T \mathbf{w}_n$  denote a contrast of interest of the regression coefficients. A PPM is a map of the posterior probability of activation for each voxel:  $\Pr(\mathbf{c}^T \mathbf{w}_n > \delta_e | \mathbf{y})$ . Here  $\delta_e$  is a prespecified activation threshold, e.g. a value that corresponds to 1% of the global mean value. Thus, a PPM looks at the probability of the contrast  $\mathbf{c}^T \mathbf{w}_n$  being greater than activation threshold  $\delta_e$ , given the data.

To determine activation in the brain formally, we can look at a thresholded PPM. This is obtained by exerting a second threshold, namely a probability threshold  $\delta_p$ , onto the original PPM. Thus, a voxel is considered activated if  $\Pr(\mathbf{c}^T \mathbf{w}_n > \delta_e | \mathbf{y}) > \delta_p$ . This  $\delta_p$  reflects the confidence of the inference and usually takes a value above 0.9 (e.g. 0.95 or 0.99). This process discretizes the PPM into null and activated voxels and is commonly used in summarizing a Bayesian analysis for brain activation.

### 3. Simulation study

To evaluate the performance of our model, we make comparisons with the standard GLM–AR spatial model. One implementation of this model that we make comparisons with is the VB method that is available in the SPM12 software. Another implementation is our self-written MCMC sampler for the same model. Although the accuracy of VB models has been verified in a setting with high SNR by Teng *et al.* (2016), under low SNR, MCMC sampling outperforms VB models according to certain metrics. This will be illustrated presently. Henceforth, we shall refer to the VB implementation and MCMC implementation of the GLM–AR model as PVB and PMCMC respectively.

### 3.1. Simulation design

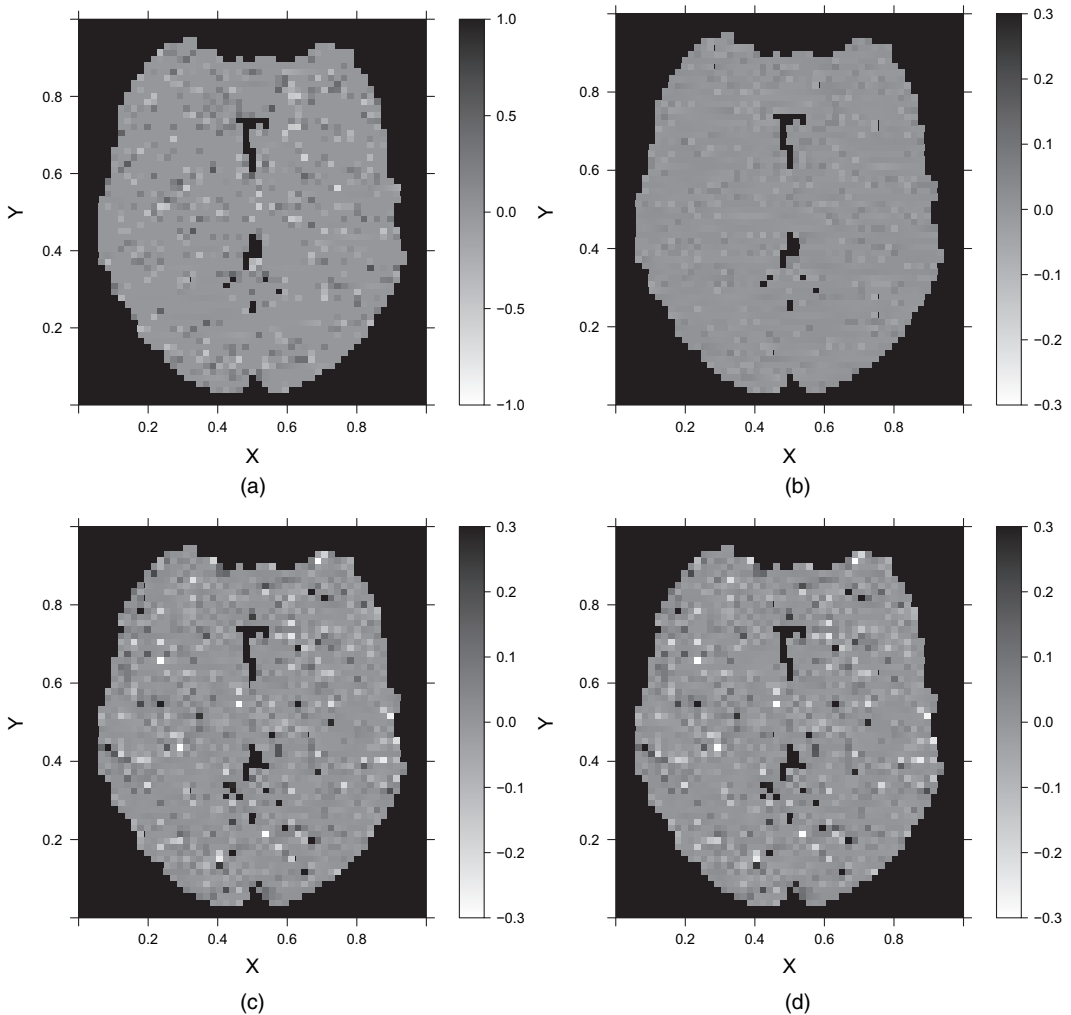
Our design matrix consists of two columns ( $K = 2$ ); the first column is the experimental design (fashioned after the face repetition data set) convolved with the canonical haemodynamic response function and the second column is the intercept. The parameters of interest, corresponding to the experimental design (one at each voxel), are generated under a mean 0 multivariate normal distribution,  $\mathbf{W}_{1,\cdot} \sim N\{0, (10\mathbf{S}^T\mathbf{S})^{-1/2}\}$ , whereas the intercepts are generated under a mean 100 multivariate normal distribution,  $\mathbf{W}_{2,\cdot} \sim N\{100, (10\mathbf{S}^T\mathbf{S}^T)^{-1/2}\}$ . The model noise will have a precision of  $\lambda_n = 0.1$  ( $n = 1, \dots, N$ ). This corresponds to a fairly small SNR, where the temporal noise will play a greater role in the data. In what follows, we shall carry out two simulations. In the first case the data will be simulated under our model, and in the second case the data will be generated under the standard spatial GLM-AR model. In these two simulations, we investigate the estimation accuracy of the slopes,  $\mathbf{W}_{1,\cdot}$ , intercepts,  $\mathbf{W}_{2,\cdot}$ , and AR coefficients ( $\mathbf{a}_p, p = 1, \dots, P$ ), and we also examine whether the difference in inference for these coefficients will lead to a possible difference in the final inference on brain activation. All simulations are based on 100 replicate data sets, and we perform the simulations on a two-dimensional axial slice of the brain.

### 3.2. Simulation 1

The purpose of this first simulation study is twofold. First, since we are generating the data under our model, we are testing our model under an ideal situation to determine its statistical properties and to ensure that the algorithm is coded correctly. Second, we compare our model with the standard spatial GLM-AR model under the situation where the AR order and coefficients vary spatially as we have seen empirically (see Fig. 1). For the GLM-AR model we use the commonly used assumption that the AR order is small and set it to 1. Here we simulate the AR parameters under our SVARO model, i.e. from the Ising priors. When simulating data we assume that the largest possible AR order at each voxel is 8; however, it could be as small as 0. The precision parameters are set as  $\tau_p = 400$  ( $p = 1, \dots, P$ ). For simplicity, we assume that all AR orders are generated spatially according to the same values for the hyperparameters of the Ising model, i.e.  $\beta_{0p} = -0.2$  and  $\beta_{1p} = 0.3$ . The AR order in PMCMC and PVB are set to  $P = 1$  as is standard practice. We note here that the GLM-AR model is misspecified and we expect that its performance will suffer regardless of which posterior sampling algorithm is used. We could have set the AR order to 8 for the GLM-AR model, but this is never done in practice and assumes knowledge of the true maximum value. For our SVARO model, we allow the order to be as large as 12 and let the algorithm determine spatially which of the 12 AR coefficients are set to 0 and which enter the model via the variable selection mechanism.

We compare the SVARO with the PVB and PMCMC models in estimating the first AR coefficient. As shown in Fig. 2, the SVARO model shows little error compared with the truth, indicating that our model has captured the AR parameter quite well. In contrast, PMCMC and PVB exhibit more bias, indicating a lack of fit for the temporal noise. Note that we are displaying the SVARO estimate only for the first AR coefficient for simplicity and direct comparison; the other AR coefficients are well estimated as well.

Table 1 summarizes the average MSE, MSE, for various parameters. These summaries are obtained by averaging the MSE of the corresponding parameters across all the voxels and over simulation replicates. It is clear that the SVARO model has the smallest MSE for these three parameters. In addition, PMCMC outperforms PVB in estimating the coefficient of the haemodynamic response  $\mathbf{W}_{1,\cdot}$ , which is the primary parameter on which inference is based. This finding is in line with our previous findings in Teng *et al.* (2016), where we found that a low



**Fig. 2** (a) True first AR coefficients, and difference maps between the true AR coefficients and the estimated posterior mean from the (b) SVARO, (c) PMCMC and (d) PVB models: the posterior means are averaged over the 100 replicated simulation data sets

SNR is one setting where MCMC sampling outperforms VB models for this particular model (GLM-AR). We also calculate and present the log-pseudomarginal likelihood LPML, averaged over simulations (Gelfand and Dey, 1994) in Table 1. The SVARO model has a larger LPML than the MCMC implementation of GLM-AR, indicating a better model fit. Note that LPML cannot be obtained from the VB algorithm. In terms of timing, the SVARO algorithm takes 108 min with 10000 iterations following 10000 burn-in iterations, PMCMC takes 11 min with the same number of iterations and PVB is the fastest at 1 min computation time.

We next investigate how the differences that are observed for the individual parameters will impact the overall inference of interest. A sensitivity plot is presented in Fig. 3, which is obtained by plotting the average sensitivity against a range of marginal posterior probability thresholds from 0.9 to 1. We choose this range because it covers those values that are most often used in practice.

**Table 1.**  $\overline{MSE}$ , log-pseudomarginal likelihood LPML and timing for the three models†

Method	$\overline{MSE}$			LPML	Timing (min)
	$W_{1..}$	$W_{2..}$	$a_1$		
SVARO	0.478	0.030	0.001	-1842902	108
PMCMC	113%	135%	509%	-1926620	11
PVB	199%	138%	510%	— ‡	1

† $\overline{MSE}$  is calculated by averaging the MSE in each voxel and over simulation replicates. The MSE values for PMCMC and PVB are relative to those in the SVARO model.  
‡Not applicable.

In terms of the underlying activation threshold, we use two thresholds: the true value of the contrast that corresponds to the top 10% and top 5% of all voxels. Thus, corresponding to a certain activation threshold and a certain probability threshold, the higher the sensitivity, the better the model is in terms of capturing activation. Again, a notable difference is observed when comparing the three methods, with the SVARO model giving the uniformly highest sensitivity across the entire range of probability thresholds and PVB resulting in the lowest sensitivity. PMCMC is better than PVB but still underperforms relatively to SVARO.

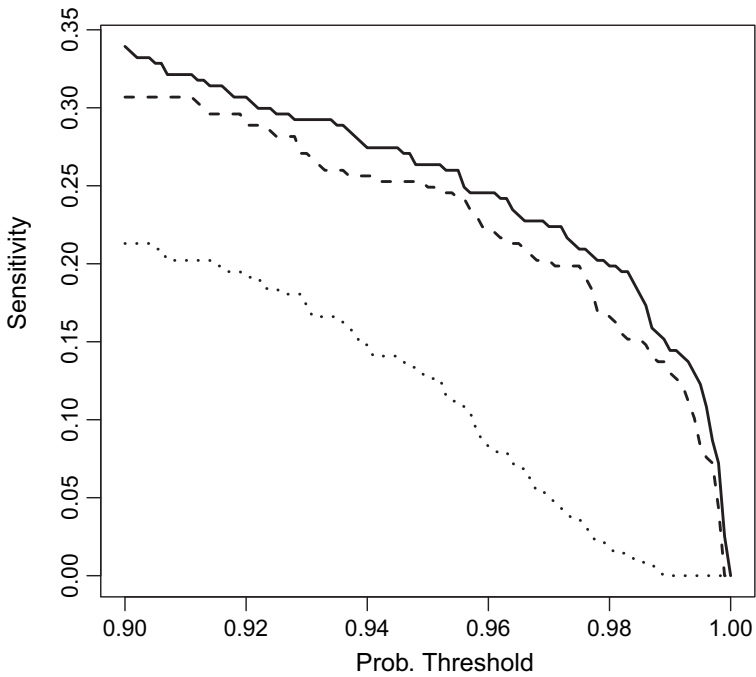
We plot the PPMs in Fig. 4, which depicts the locations of the true activations and the PPMs from the SVARO model. In addition, differences in the probability maps comparing the SVARO with the PMCMC and PVB methods are also depicted. Again, the SVARO model appears to perform the best in producing the highest posterior probabilities for regions that are truly activated. PMCMC is similar to the SVARO model but its probabilities on those activated regions are slightly lower than those from the SVARO model, especially near the boundary of activated regions. PVB underperforms compared with the other two approaches by providing greater posterior probability on null locations while providing smaller posterior probability on activated locations.

### 3.3. Simulation 2

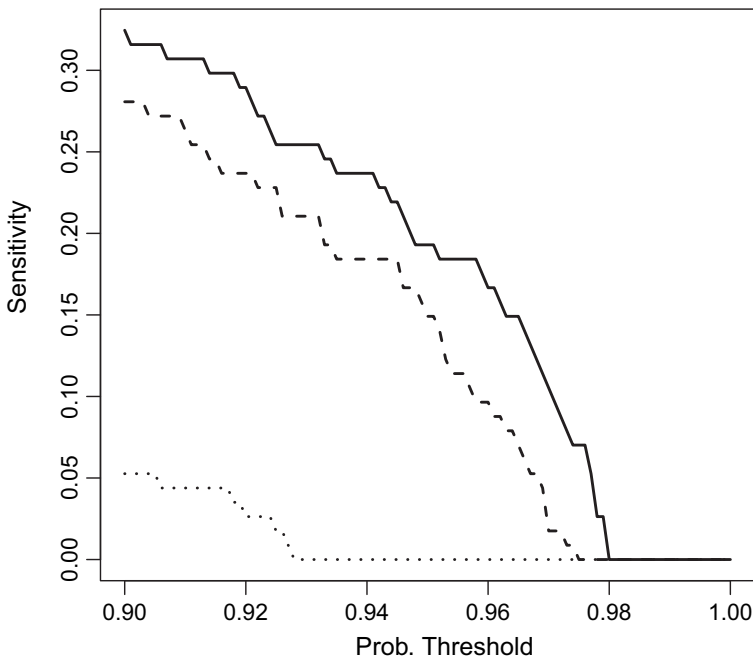
In this simulation study we compare the performance of our model under model misspecification against the correctly specified GLM-AR model by using the two algorithms PMCMC and PVB. Here we simulate data under the competing GLM-AR model. The AR coefficients are simulated by using the low resolution tomography prior and the AR order is set to 1 for every voxel, which is a typical assumption that is used when analysing functional MRI data. The prior precision,  $\tau_{a_1}$  is set at 400. The maximum AR order that is allowed for our model is set at 12.

Table 2 shows the average MSE summaries of the estimators. When data are simulated under the competing model, the SVARO model still exhibits good performance in terms of average MSE for the two regression parameters, suffering only a slight loss of efficiency compared with the GLM-AR model using the PMCMC algorithm. Nevertheless, the average LPML indicates that, overall, the SVARO model fits the simulated data better than does the GLM-AR model. It is worth mentioning that PVB again underperforms relatively to PMCMC in terms of the haemodynamic response parameter and the AR coefficient.

Fig. 5 presents the sensitivity curves. Although the data were simulated under a constant order AR assumption, the SVARO method demonstrates similar sensitivity to that of PMCMC. The

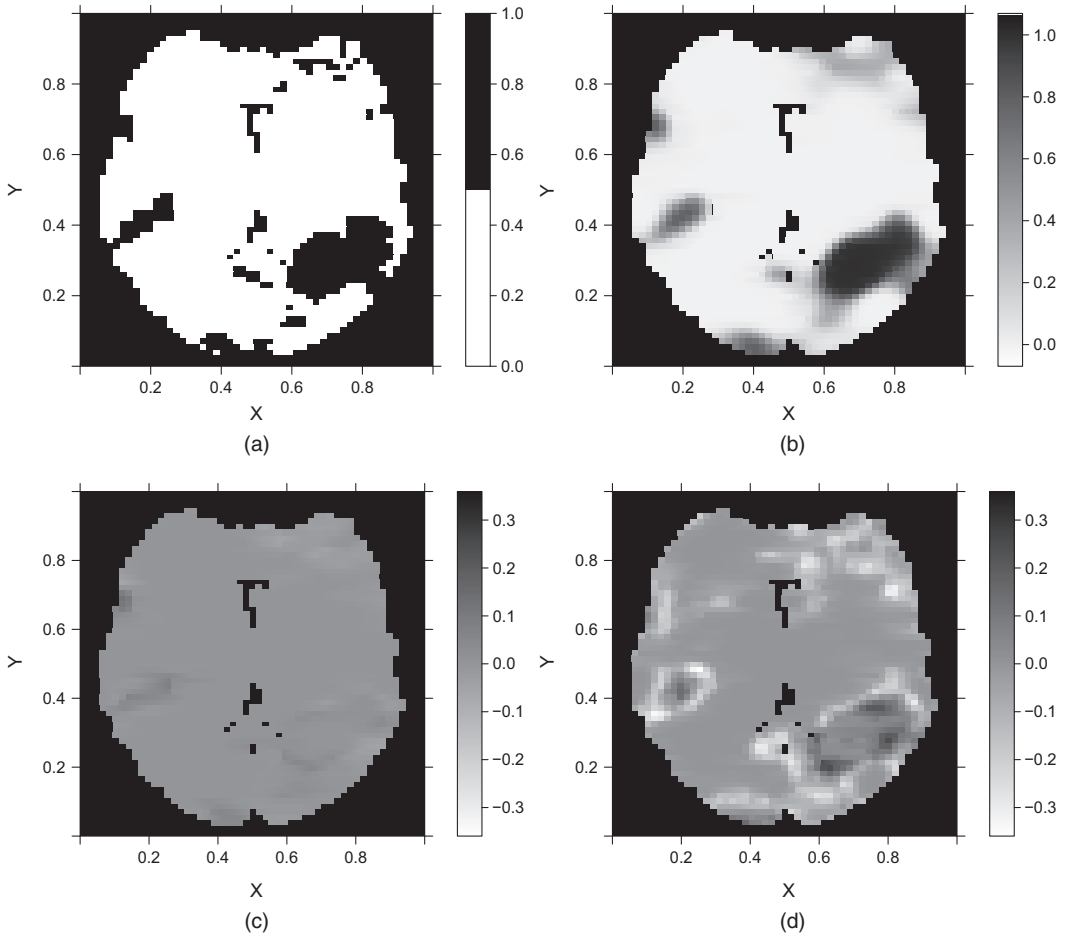


(a)



(b)

**Fig. 3** Thresholded sensitivity curves for the three methods SVARO (—), PMCMC (---) and PVB (.....) with two activation thresholds: (a) activation threshold corresponds to the top 10% of the parameter estimates; (b) top 5%



**Fig. 4** (a) True activation map (■, activated voxels), and PPMs of activation obtained by using (b) the SVARO, (c) SVARO-PMCMC and (d) SVARO-PVB models: (c) and (d) reflect the difference of the two alternative approaches relative to the SVARO model

sensitivity curve for PVB is uniformly smaller than the other two because of the inaccurate estimation of the haemodynamic response parameter.

Overall, these simulation studies show that our SVARO model is more efficient from a statistical point of view than the GLM-AR model when the data display heterogeneous AR orders across the image—a situation that we believe is more common than not. They also show that our model does not suffer much in terms of statistical efficiency when the AR order is small and homogeneous across the image.

#### 4. Analysis of face perception functional magnetic resonance imaging data

We turn our focus back to the face repetition data set that originally motivated our model development and compare results from the two models and three algorithms. In this analysis we use the complete experimental design consisting of famous faces, repeated twice, and unfamiliar faces, repeated twice. These four design vectors are then convolved with the canonical haemodynamic response function as well as its time and dispersion derivatives. An intercept term is

**Table 2.**  $\overline{MSE}$ , LPML and computation time for the three methods†

Method	$\overline{MSE}$			LPML	Timing (min)
	$W_{1..}$	$W_{2..}$	$a_1$		
SVARO	0.502	0.031	0.003	-1817287	206
PMCMC	99%	97%	54%	-1875900	11
PVB	167%	98%	49%	—‡	1

† $\overline{MSE}$  is calculated by averaging the MSE over voxels and simulation replicates. The  $\overline{MSE}$ -values for PMCMC and PVB are relative to those in SVARO.

‡Not applicable.

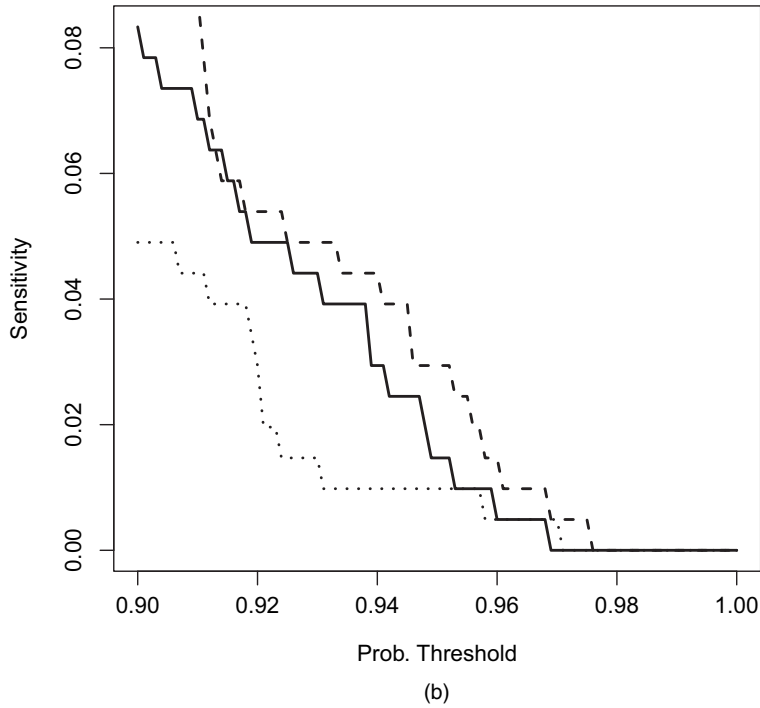
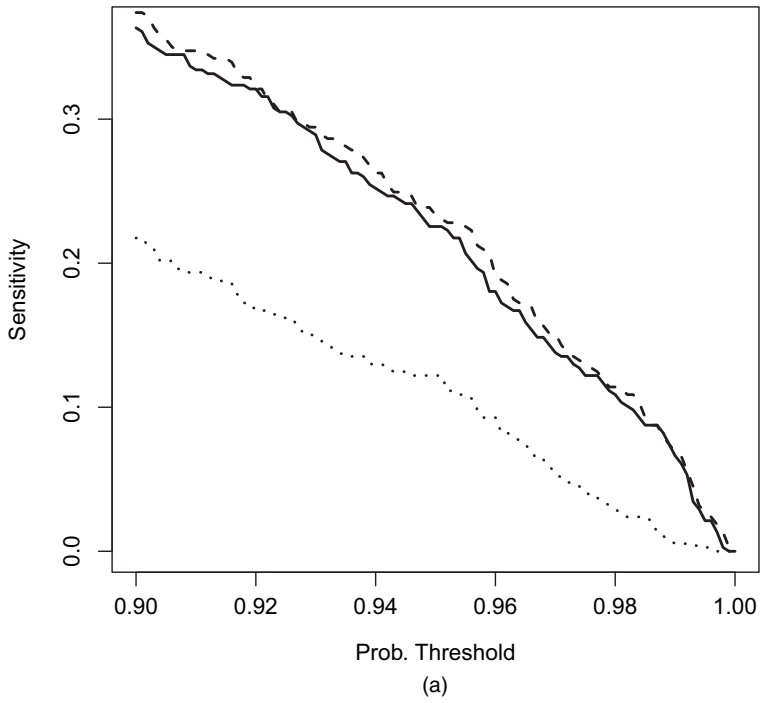
also added to the final design matrix for a total of 13 covariates. We allow an AR order up to a maximum of  $P = 12$  when fitting the SVARO model, and an AR(3) model for the GLM-AR model using both the PMCMC and the PVB algorithms. Although the choice of an AR(3) model for the last two approaches may seem arbitrary, this is exactly the justification for the use of the SVARO model where such an arbitrary assumption need not be made. We consider data that are collected on a single subject in what follows.

Preprocessing steps are applied to the data prior to fitting the Bayesian models. All functional images are aligned to the first image by using a six-parameter rigid body transformation. Then slice timing correction is performed to set the standard acquisition time as the 12th slice. Images are spatially normalized to a standard echo planar imaging image. The global mean  $g$  is computed and each time series is divided by  $100/g$  to represent a percentage of  $g$ . Finally, a high pass filter with cut-off frequency of  $1/128$  Hz is applied to the data and design matrix to remove low frequency signals that arise through MRI scanner drift.

Fig. 6 presents the distribution of estimated posterior modal AR orders from our model across voxels. The most frequent order is the zero order, or no auto-correlation in the time series, accounting for approximately 35% of the voxels. Interestingly, the next highest is order 8, with 9.4% of the voxels. Overall, roughly 51% of the voxels exhibit an AR order that is greater than 3. The existence of these higher orders and the variability in the orders is in general agreement with our exploratory analysis of the face repetition data set (see Section 1.1) and indicates the necessity of our proposed model.

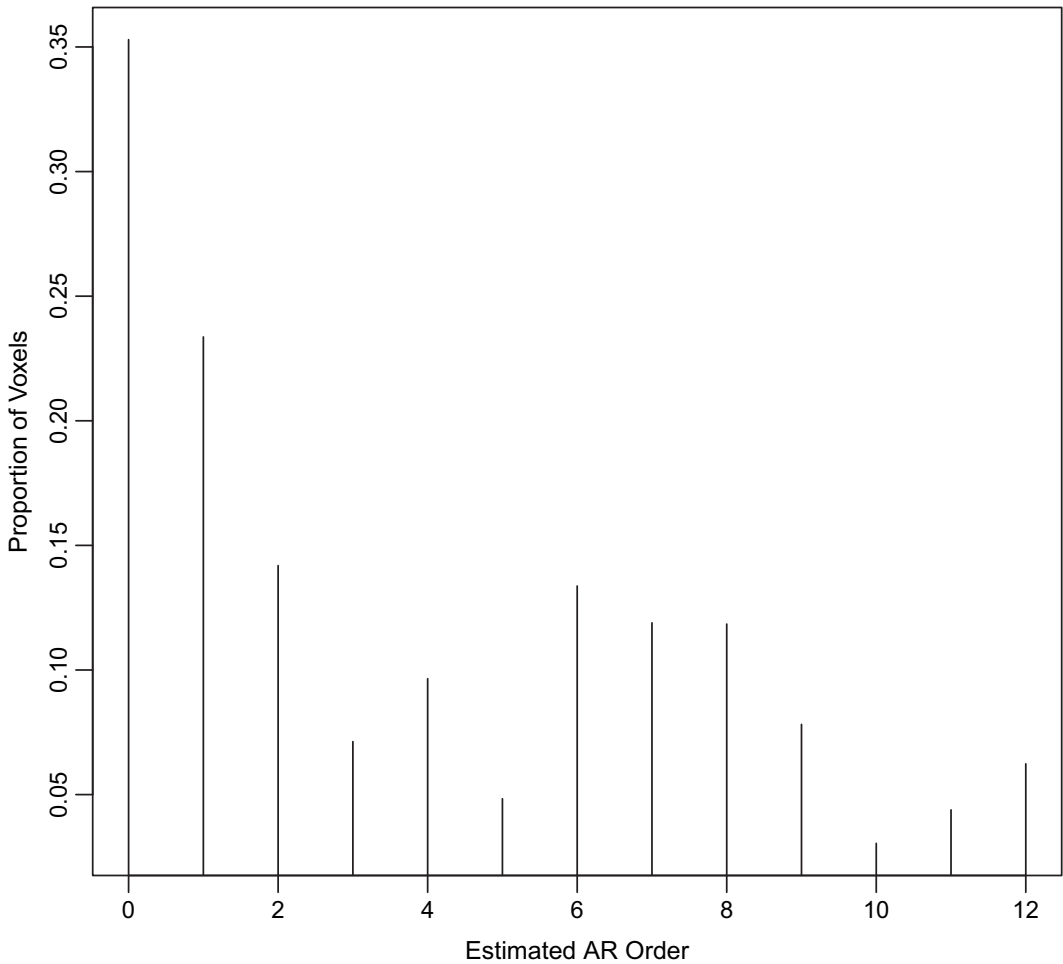
Next, we compare models or algorithms with respect to a particular contrast of interest, namely the effect of fame, i.e. famous faces *versus* unfamiliar faces. The estimated posterior mean and standard deviation (SD) maps for the fame contrast are displayed in Fig. 7. Although the posterior SD estimated from the SVARO and PMCMC models are very close, the estimated posterior means are different between the two approaches. Also, the estimated SD that is obtained from PVB shows apparent discretization. This is due to a graph partitioning that is incorporated in the algorithm for computational speed (Penny *et al.*, 2007). It is clear that the boundaries of these graph-partitioned regions have substantially larger estimated SD than do the interior locations. The posterior mean of PVB also seems to exhibit artefacts at the partition boundaries, though the effect is not as pronounced. The estimated LPML is  $-4.66 \times 10^7$  under our proposed model and is  $-4.82 \times 10^7$  under the GLM-AR model with MCMC sampling. According to this model selection criterion, our proposed model is preferred.

Finally, we look at the effect of fame by using thresholded PPMs. The activation threshold



**Fig. 5** Thresholded sensitivity curves for the three methods SVARO (—), PMCMC (---) and PVB (.....) with two activation thresholds: (a) activation threshold corresponds to the top 10% of the parameter estimates; (b) top 5%

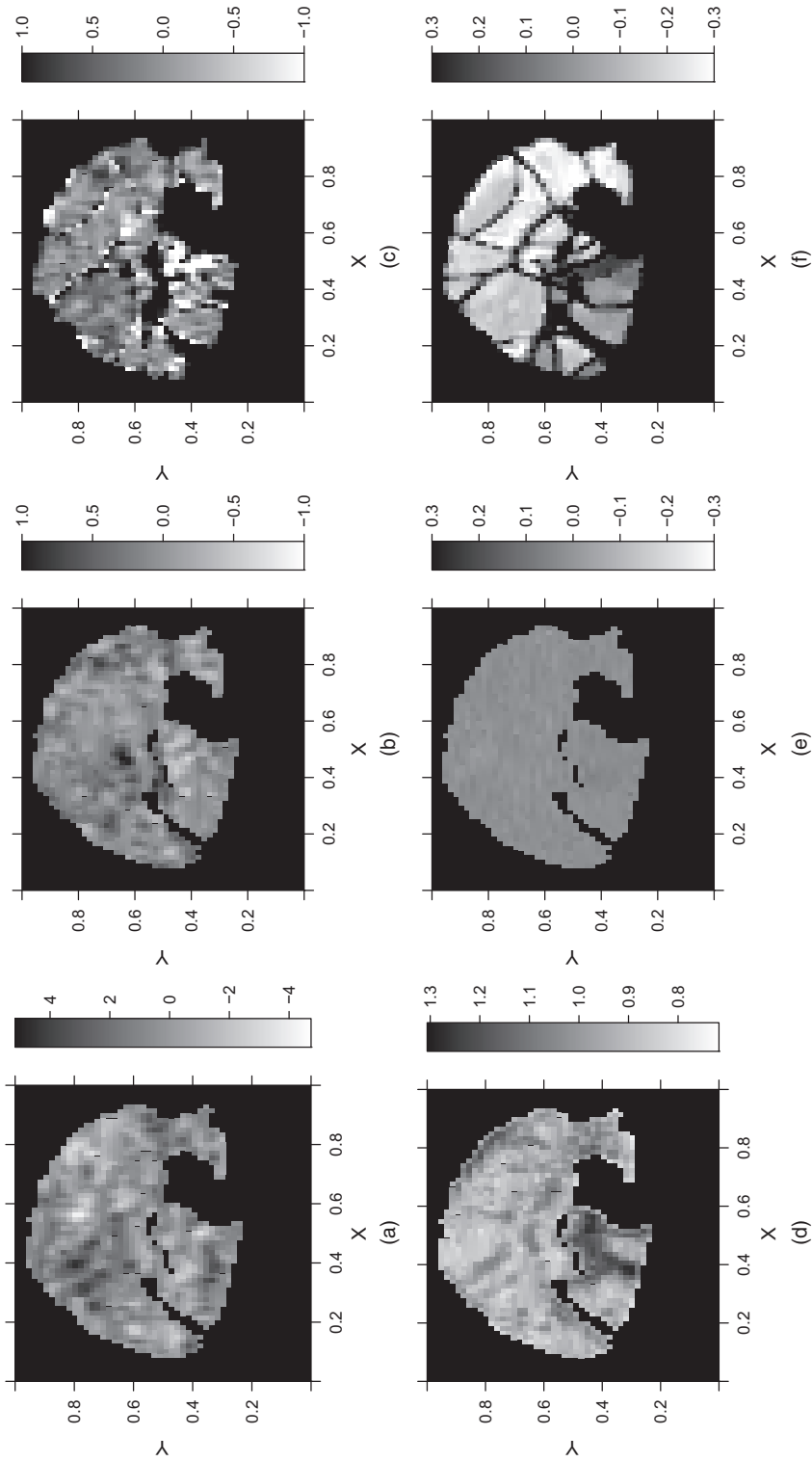




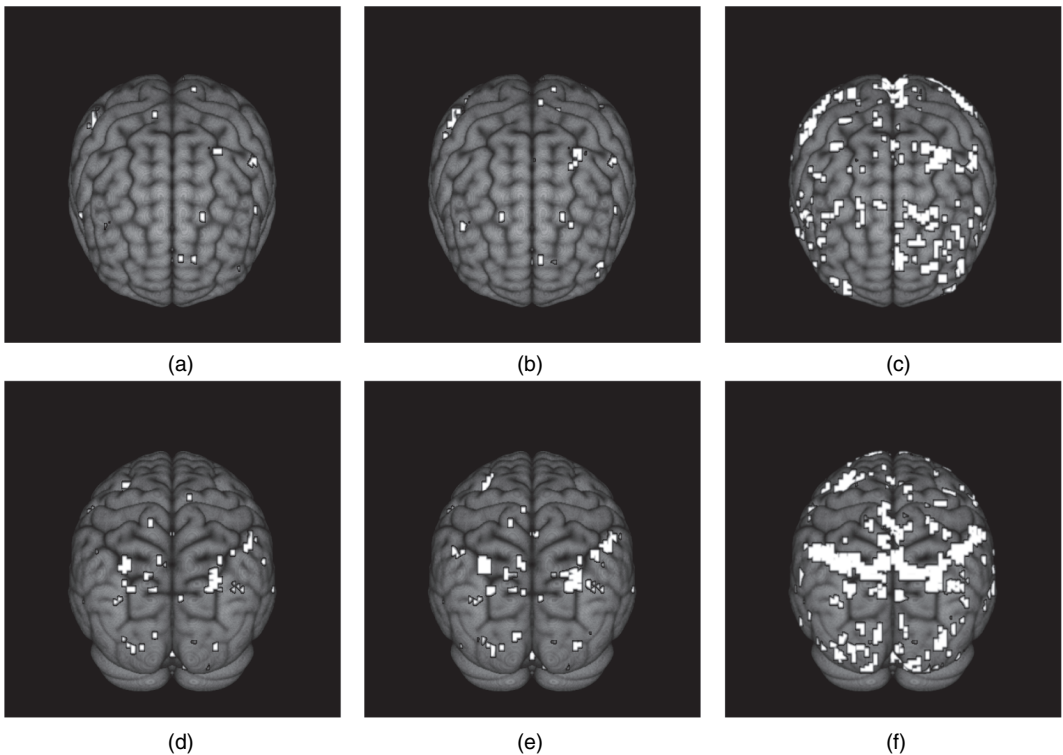
**Fig. 6** Histogram of the estimated posterior modal AR order from the SVARO model across all voxels for the face repetition data set

is set to 0.2% of the global mean value, and the probability threshold is set to 0.95. Fig. 8 shows the activation regions projected onto the cerebral cortex surface. We can see that there is a match in terms of a majority of activation regions inferred from the SVARO and PMCMC models. A closer look reveals that PMCMC tends to make more scattered predictions across the posterior regions of the brain. The numbers of activation regions from PVB are far greater than the number that is obtained from the other two approaches and are more widely dispersed across the brain. From our simulation study results, we suspect that these scattered activated regions are likely to be false positive results.

The main parameters of interest are the regression parameters corresponding to the haemodynamic response function for the four experimental tasks. We use Geweke (1984) diagnostics along with false discovery rate correction for dependent variables (Benjamini and Yekutieli, 2001) to determine convergence of our model. Although there is a multivariate version of the Geweke diagnostic, our parameter space is too large for this method. Out of the four parameters for each of the 56526 voxels (a total of 226104 parameters), 99.89% of the parameters converged



**Fig. 7** Posterior estimates from the middle (27/53) slice of the brain, sagittal view: (a)–(c) posterior mean of fame; (d)–(f) posterior standard deviation of fame; (a), (d) results from the SVARO model; (b), (e) difference map, SVARO-PMCMC; (c), (f) difference map, SVARO-PVB



**Fig. 8** Activation maps for fame contrast: (a)–(c) anterior view; (d)–(f) posterior view; (a), (d) activation (white) from the SVARO model; (b), (e) activation from PMCMC; (c), (f) activation from PVB

to their posterior distributions. In terms of computation time, PVB took about 1 h, PMCMC took 1 day and the SVARO model took about 1 week of computation. Much of our MCMC algorithm is amenable to parallel programming, which is an avenue for further development.

## 5. Discussion

In this paper, we have developed a Bayesian hierarchical model, SVARO, that allows the AR order to vary spatially across the brain, with the orders themselves displaying a certain level of spatial clustering based on an Ising model. We compared our proposed model with a self-written MCMC sampler for the standard GLM–AR model and the VB implementation for the same model. The results are interesting; under a low SNR ratio, the VB method seems to suffer from variance overestimation, leading to a larger MSE than those of the other two methods. It is likely that, as temporal noise increases, a more vital role is played by the AR correlation that increases the posterior correlation between different parameters and this makes the mean field assumption underlying the VB approximation less accurate.

We have shown that our model outperforms the GLM–AR model not only in terms of accuracy and sensitivity, but also according to formal model selection based on the LPML-criterion. Through an application of our proposed model and through an exploratory analysis, we have shown that the typical constant low order AR assumption can be violated in real functional MRI data. It is very likely that this issue, seen in the face repetition data set, is also present in other functional MRI data sets.

There is a computational price to be paid for gaining the flexibility that we have proposed in our model. Our model takes a longer time to run than either implementation of the GLM–AR algorithm. This is mostly due to the estimation of the varying AR orders and the associated greater number of parameters to estimate. However, as previously mentioned, there are elements of our MCMC algorithm that are amenable to parallel programming. This will be investigated in future work.

Whereas we have based our model specification on a set of independent Ising processes, one for each possible order of the AR process, another approach would be to assume a Potts model for the orders of the AR coefficients. A Potts model, combined with a Dirichlet process prior for parameters, has been investigated for selecting covariates of interest in brain imaging (Johnson *et al.*, 2013). Here we can apply it also to the selection of AR orders to yield a more parsimonious, yet still flexible, model. Investigation of hyperparameter estimation in the Ising model and the use of alternative spatial models is also of interest, as is increasing the scope of our comparison of methods to include wavelet approaches that focus on long memory errors, or vector AR models (Harrison *et al.*, 2003).

## References

- Benjamini, Y. and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, **29**, 1165–1188.
- Bezener, M., Eberly, L. E., Hughes, J., Jones, G. and Musgrove, D. R. (2016) Bayesian spatiotemporal modeling for detecting neuronal activation via functional magnetic resonance imaging. In *Handbook of Big Data Analytics* (eds W. K. Härdle, S. H. Lu and X. Shen). New York: Springer.
- Bullmore, E., Brammer, M., Williams, S. C., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R. and Sham, P. (1996) Statistical methods of estimation and inference for functional MR image analysis. *Magn. Resonance Med.*, **35**, 261–277.
- Bullmore, E., Fadili, J., Maxim, V., Şendur, L., Whitcher, B., Suckling, J., Brammer, M. and Breakspear, M. (2004) Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage*, **23**, suppl., S234–S249.
- Castruccio, S., Ombao, H. and Genton, M. G. (2016) A multi-resolution spatio-temporal model for brain activation and connectivity in fMRI data. *Preprint arXiv:1602.02435*. School of Mathematics and Statistics, University of Newcastle, Newcastle upon Tyne.
- Chang, C. and Glover, G. H. (2010) Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage*, **50**, 81–98.
- Della-Maggiore, V., Chau, W., Peres-Neto, P. R. and McIntosh, A. R. (2002) An empirical comparison of SPM preprocessing parameters to the analysis of fMRI data. *NeuroImage*, **17**, 19–28.
- Fadili, M. and Bullmore, E. (2002) Wavelet-generalized least squares: a new BLU estimator of linear regression models with 1/f errors. *NeuroImage*, **15**, 217–232.
- Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S. and Turner, R. (1995) Analysis of fMRI time-series revisited. *NeuroImage*, **2**, 45–53.
- Friston, K., Josephs, O., Zarahn, E., Holmes, A., Rouquette, S. and Poline, J.-B. (2000) To smooth or not to smooth?: Bias and efficiency in fMRI time-series analysis. *NeuroImage*, **12**, 196–208.
- Gelfand, A. E. and Dey, D. K. (1994) Bayesian model choice: asymptotics and exact calculations. *J. R. Statist. Soc. B*, **56**, 501–514.
- Gelman, A. and Meng, X.-L. (1998) Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.*, **13**, 163–185.
- George, E. I. and McCulloch, R. E. (1993) Variable selection via Gibbs sampling. *J. Am. Statist. Ass.*, **88**, 881–889.
- Geweke, J. F. (1984) Measures of conditional linear dependence and feedback between time series. *J. Am. Statist. Ass.*, **79**, 907–915.
- Gössl, C., Auer, D. P. and Fahrmeir, L. (2001) Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, **57**, 554–562.
- Harrison, L., Penny, W. D. and Friston, K. (2003) Multivariate autoregressive modeling of fMRI time series. *NeuroImage*, **19**, 1477–1491.
- Higdon, D. M. (1998) Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Am. Statist. Ass.*, **93**, 585–595.
- Ising, E. (1925) Beitrag zur Theorie des Ferromagnetismus. *Zeits. Phys.*, **31**, 253–258.
- Jeong, J., Vannucci, M. and 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.

- Johnson, T. D., Liu, Z., Bartsch, A. J. and Nichols, T. E. (2013) A bayesian non-parametric Potts model with application to pre-surgical fMRI data. *Statist. Meth. Med. Res.*, **22**, 364–381.
- Johnson, V. E. and Rossell, D. (2012) Bayesian model selection in high-dimensional settings. *J. Am. Statist. Ass.*, **107**, 649–660.
- Kim, S., Smyth, P. and Stern, H. (2010) A bayesian mixture approach to modeling spatial activation patterns in multisite fMRI data. *IEEE Trans. Med. Imngng*, **29**, 1260–1274.
- Lee, K.-J., Jones, G. L., Caffo, B. S. and Bassett, S. S. (2014) Spatial bayesian variable selection models on functional magnetic resonance imaging time-series data. *Baysn Anal.*, **9**, 699–732.
- Li, F., Zhang, T., Wang, Q., Gonzalez, M. Z., Maresh, E. L. and Coan, J. A. (2015) Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. *Ann. Appl. Statist.*, **9**, 687–713.
- Locascio, J. J., Jennings, P. J., Moore, C. I. and Corkin, S. (1997) Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. *Hum. Brain Mappng*, **5**, 168–193.
- Makni, S., Ciuciu, P., Idier, J. and Poline, J.-B. (2006) Joint detection-estimation of brain activity in fMRI using an autoregressive noise model. In *Proc. 3rd Int Symp. Biomedical Imaging: Nano to Macro*, pp. 1048–1051. New York: Institute of Electrical and Electronics Engineers.
- Meyer, F. G. (2003) Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series. *IEEE Trans. Med. Imngng*, **22**, 315–322.
- Mitchell, T. J. and Beauchamp, J. J. (1988) Bayesian variable selection in linear regression. *J. Am. Statist. Ass.*, **83**, 1023–1032.
- Musgrove, D. R., Hughes, J. and Eberly, L. E. (2016) Fast, fully bayesian spatiotemporal inference for fMRI data. *Biostatistics*, **17**, 291–303.
- O’Hara, R. B. and Sillanpää, M. J. (2009) A review of Bayesian variable selection methods: what, how and which. *Baysn Anal.*, **4**, 85–117.
- Penny, W., Flandin, G. and Trujillo-Barreto, N. (2007) Bayesian comparison of spatially regularised general linear models. *Hum. Brain Mappng*, **28**, 275–293.
- Penny, W., Kiebel, S. and Friston, K. (2003) Variational Bayesian inference for fMRI time series. *NeuroImage*, **19**, 727–741.
- Penny, W. D., Trujillo-Barreto, N. J. and Friston, K. J. (2005) Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, **24**, 350–362.
- R Core Team (2018) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Schmidt, D. F. and Makalic, E. (2013) Estimation of stationary autoregressive models with the Bayesian LASSO. *J. Time Ser. Anal.*, **34**, 517–531.
- Shu, H., Nan, B. and Koeppe, R. (2015) Multiple testing for neuroimaging via hidden Markov random field. *Biometrics*, **71**, 741–750.
- Skudlarski, P., Constable, R. T. and Gore, J. C. (1999) ROC analysis of statistical methods used in functional MRI: individual subjects. *NeuroImage*, **9**, 311–329.
- Smith, M. and Fahrmeir, L. (2007) Spatial bayesian variable selection with application to functional magnetic resonance imaging. *J. Am. Statist. Ass.*, **102**, 417–431.
- Stanley, H. E., Stauffer, D., Kertesz, J. and Herrmann, H. J. (1987) Dynamics of spreading phenomena in two-dimensional Ising models. *Phys. Rev. Lett.*, **59**, 2326–2328.
- Swendsen, R. H. and Wang, J.-S. (1987) Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, **58**, 86–88.
- Teng, M., Johnson, T. and Nathoo, F. (2016) A comparison of variational Bayes and Hamiltonian Monte Carlo for Bayesian fMRI time series analysis with spatial priors. *Preprint arXiv:1609.02123*. Department of Statistics, University of Michigan, Ann Arbor.
- Wang, H., Li, G. and Tsai, C.-L. (2007) Regression coefficient and autoregressive order shrinkage and selection via the lasso. *J. R. Statist. Soc. B*, **69**, 63–78.
- Woolrich, M. W., Behrens, T. E. and Smith, S. M. (2004a) Constrained linear basis sets for HRF modelling using variational bayes. *NeuroImage*, **21**, 1748–1761.
- Woolrich, M. W., Jenkinson, M., Brady, J. M. and Smith, S. M. (2004b) Fully bayesian spatio-temporal modeling of fMRI data. *IEEE Trans. Med. Imngng*, **23**, 213–231.
- Woolrich, M. W., Ripley, B. D., Brady, M. and Smith, S. M. (2001) Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage*, **14**, 1370–1386.
- Worsley, K. J. and Friston, K. J. (1995) Analysis of fMRI time-series revisited again. *NeuroImage*, **2**, 173–181.
- Zarahn, E., Aguirre, G. K. and D’Esposito, M. (1997) Empirical analyses of bold fMRI statistics. *NeuroImage*, **5**, 179–197.