Chapter 4

Application of Spatial Bayesian Hierarchical Models to fMRI Data

Kuo-Jung Lee

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/64823

Abstract

Bayesian modelling has attracted great interest in cognitive science and offered a flexible and interpretable way to study cognitive processes using functional magnetic resonance imaging data. In this chapter, a spatial Bayesian hierarchical model is applied to an event-related fMRI study of cognitive control using the Simon test. We consider a sparse spatial generalized linear mixed-effects model to capture the spatial dependence among activated voxels and temporal parameters and to benefit computationally by reducing dimensionality. We demonstrate that the proposed model has the capability of identification of the brain areas related to cognitive tasks. Moreover, the reduction in the false positive rate is observed in the simulation study, and the relevant brain regions involved in processing cognitive control are clearly detected in a real-life fMRI example.

Keywords: Bayesian, functional magnetic resonance imaging, Markov chain Monte Carlo, spatial generalized linear mixed-effects model

1. Introduction

Functional magnetic resonance imaging (fMRI) has increasingly become an important and popular modality that allows researchers to investigate brain activity resulting from a particular stimulus [1]. In an fMRI experiment, a subject is asked to perform a task by responding to a series of stimuli that may involve a motor, sensory or cognitive task, then the MR machine records the changes in the blood oxygen level dependent (BOLD) of the brain across different time points, resulting in three-dimensional fMRI time-series images. Numerous statistical models have been proposed to allow researchers to detect localized regions activated during a task, to describe the networks required for a particular brain function or to assess physical

INTECH
open science | open minds

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

characteristics elicited by cognitive processes in the brain. However, the growth in the complexity, the significant size and the hierarchical structure of fMRI data make it difficult to comprise a fully efficient computationally feasible statistical model to accurately explain the temporal and spatial characteristics of the data.

The standard approach used on fMRI data is known as statistical parametric mapping (SPM) [2], which applied either a voxel-wise *t*-test or *F*-test statistics. In order to obtain an activation map of the brain, the next step is to threshold the test statistics at a given overall error rate that leads to a major multiplicity problem. The most common way of solving this problem is to use Gaussian random field theory [3]. This technique is based on the assumption of a stationary Gaussian random field, which may not be satisfied in fMRI settings. Another limitation is that most current methods ignore at least one of the spatial or temporal relationships between observations. Ignoring either spatial or temporal correlations in the model leads to seriously biased conclusions [4].

This article introduces a novel Bayesian modelling approach to fMRI data analysis. Bayesian approaches have great potential in applications because they allow a flexible modelling of spatial and temporal correlations in data [5]. We consider a Bayesian spatiotemporal model in a computationally feasible manner to detect brain regions that are activated by the external stimulus in fMRI. Accurate and powerful single-subject task-related activation models are required in order to develop effective imaging biomarkers, which constitutes the primary scientific problem. Moreover, the Bayesian paradigm provides an attractive inferential framework that can directly incorporate the physical characteristics of an experiment. Our goal is to put forward a model and inferential framework by which to investigate task-specific changes in the BOLD signal. Clearly, the model must account for the spatial relationships between the voxels, but there are other possible sources of variation that should not be ignored. Ultimately, we develop a hierarchical Bayesian model, which not only takes into account the spatiotemporal and temporal drift relationships in the data under consideration, but also easily investigates the role of specific regions that integrate brain activity to coordinate cognition and behaviour.

The applied Bayesian hierarchical approach contains several characteristics [4, 6]. Latent binary variables are introduced to indicate activation/inactivation of voxels. The spatial generalized linear mixed model (SGLMM) [7, 8] is considered to capture the spatial dependence of the latent binary variables. In addition, the autoregression (AR) model is used to model the temporal dependence of signal changes. Several studies have found that spatial dependence also appears in the temporal parameters in AR models [4, 9]. Thus, neglecting the spatial dependence of temporal correlations moderates the computational intensity, but the simplification may produce a biased estimation of the temporal coefficients and consequently may result in the spurious detection of brain activities [10]. Therefore, we also consider the spatial linear mixed-effects model to spatially regularize the AR parameters.

In fMRI data, the posterior inferences are based on the estimations of parameters; however, the posterior distribution is typically extremely large, and is unavailable in analytic form. Hence, we employ Markov chain Monte Carlo (MCMC) [11] sampling techniques that combine Metropolis-Hastings [12] schemes to generate samples from the posterior distribution for the

http://dx.doi.org/10.5772/64823

purpose of performing inferential tasks. To reduce the computational burden and accelerate the sampling procedure, we parcel the brain so that sampling procedures can be performed in parallel. In addition, prior information on activation in the form of spatially informative variables, e.g. the grey and white areas of the brain, can be incorporated into the model.

2. Statistical modelling

In this section, we introduce the proposed statistical model in the analysis of fMRI data. To make inferences about task-related change in underlying neuronal activity, a general linear model is used to model BOLD signal changes for each voxel. In addition to the essential temporal dependence of BOLD signals in a voxel itself, the BOLD signal changes show spatially contiguous and locally homogenous among voxels [11]. Shmuel et al. [13] in the visual fMRI study have demonstrated BOLD response seems to be well approximated by separable spatiotemporal model. Thus, for computational convenience, we consider a Bayesian separable spatiotemporal model for BODL signal changes to simultaneously account for the temporal dependence in nearby time points and spatial dependence in local neighbouring voxels.

Let $y_v=(y_{v1}, \ldots, y_{vT})'$, be a $T \times 1$ column vector and denote the observed BOLD signals from a voxel $v=1, \ldots, V$ at time t, $t=1, \ldots, T$. Following [14], we model the BOLD response for a particular voxel v with a linear regression model defined as

$$y_v = X_v\beta_v + L_v\rho_v + \varepsilon_v; \quad \varepsilon_v \sim N_T(0, \sigma^2 I_T), \tag{1}$$

where X_v is the design matrix, each column of which consists of values obtained from an impulse stimulus function [15, 16] with respect to a task convolved with the hemodynamic response function (HRF); β_v, is a $p \times 1$ vector and corresponds to the effect of stimuli on the BOLD signal changes. The temporal correlation is modelled by ρ_v, being an $r \times 1$ autoregression coefficient vector, with L_v, a $T \times r$ matrix of lagged prediction errors [2, 14]. We assume that the error terms ϵ_v in (1) are independently normally distributed $N_T(0, \sigma^2 I_T)$ with a mean vector 0 of length T and a covariance matrix $\sigma^2 I_T$ across voxels, where I_T is a $T \times T$ identity matrix [14]. In the Bayesian framework, the parameters are hierarchically assigned and the corresponding priors are defined, including spatial prior being introduced to capture the spatial dependence of brain activities among voxels.

For the purpose of detecting the activation of a voxel, a vector of binary random variables $\gamma_v=(\gamma_{v1}, \ldots, \gamma_{vp})'$ is introduced to indicate whether the voxel v is in response to a sequence of input stimuli. The voxel v is considered active to the stimulus j if $\gamma_{vj}=1$ and, on the other hand, inactive if $\gamma_{vj}=0$. Given that γ_{vj}, we assume β_{vj} has a spike and slab mixture prior of two normal distributions [17] given by

$$\beta_{vj} \mid \gamma_{vj} \sim \gamma_{vj} N\left(0, c_{vj}^2 \tau_{vj}^2\right) + \left(1-\gamma_{vj}\right) N\left(0, \tau_{vj}^2\right) \tag{2}$$

where $N(\mu, \sigma^2)$ is denoted as a normal distribution with a mean μ and a variance σ^2. In Eq. (2), we let c_{vj}^2 being fixed and assume τ_{vj}^2 to have an inverse gamma distribution $\mathcal{IG}\left(\frac{a_\tau}{2}, \frac{b_\tau}{2}\right)$ [4]. We consider that $X \sim \mathcal{G}(a, b)$ if the probability density function of X is defined by just below Eq. (3)

$$f\left(x|a,b\right) = \frac{b^a}{\Gamma\left(a\right)} x^{-a-1} e^{-b/x}, x > 0, \tag{3}$$

where $\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy$. We set c_{vj}^2 to be large resulting in the nonzero estimate of β_{vj} so that the stimulus j is considered to activate the voxel v. As George et al. [17] suggested, c_{vj}^2 should be taken less than 10^4 to avoid the computational problems and we find that $c_{vj}^2 = 10$ is a reasonable choice in our simulation studies and real fMRI example.

Since the response at a particular voxel is likely to be consistent to the responses of neighbouring voxels, we apply SGLMM [7, 35] to capture the spatial relationship. We assume $\gamma_j = (\gamma_{1j}, \ldots, \gamma_{Vj})'$ are independently distributed in accordance to the Bernoulli distribution $\gamma_{vj} \mid \eta_{vj} \sim Ber(\eta_{vj})$ with a logistic link $\text{logit}(\eta_{vj}) = \alpha_{vj} + m'_v \phi_j$, where $\text{logit}(\eta_{vj}) = \ln\left(\frac{\eta_{vj}}{1-\eta_{vj}}\right)$ and α_{vj} is a constant intended to incorporate the expert knowledge or anatomical information and ϕ_j is a vector of spatial random effects. The spatial dependence between the binary variables is implicitly captured by ϕ_j assuming to have

$$\phi_j | \kappa_j \sim N\left(0, \left(\kappa_j M' Q M\right)^{-1}\right), \tag{4}$$

where m'_v is v_{th} row of M, an $V \times q$ matrix consisting of multi-resolutional spatial basis vectors that are able to explain spatial variation sources. The columns of M, consist of the q principal eigenvectors with respect to the first q largest eigenvalues from the adjacency matrix A of voxels, an $V \times V$ matrix with the $(v, s)_{\text{th}}$ element in Eq. (5), defined as

$$A\left(v,s\right) = \begin{cases} 0 & v = s \\ 1 & v \sim s \end{cases}, \tag{5}$$

where $v \sim s$ indicates s and v are neighbours. In the three-dimensional fMRI data analysis, a 26-adjacent neighbourhood is considered [8]. Typically, q is less than $V/2$ or equal to the number of eigenvalues greater than 0.05. This choice can reduce the dimensionality significantly but still maintains the spatial structure of the data. The graph Laplacian matrix $Q = \text{diag}(A\mathbf{1}) - A$ relates spatial basis vectors to represent the image data [18] and $\mathbf{1}$ is a $q \times 1$ column vector of ones. We assume a conjugate prior for the smoothing parameter given by

$$\kappa_j \sim \mathcal{G}\left(\frac{a_\kappa}{2}, \frac{b_\kappa}{2}\right). \tag{6}$$

To avoid generating artefactual spatial structure in the posterior distribution, [19] suggested $a_\kappa = 1$ and $b_\kappa = 4000$ in Eq. (6). We refer $X \sim \mathcal{G}(a, b)$ if the probability density function of X is defined in just below Eq. (7)

$$f(x|a,b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}, x > 0. \tag{7}$$

It is also found that the temporal correlations between voxels tend similar [4]. Such spatial dependence is modelled by a spatial linear mixed-effects model. For computational convenience and simplicity, we make a transformation of ρ_{vr} such that $\tilde{\rho}_{vr} = \log\frac{1+\rho_{vr}}{1-\rho_{vr}}$ and assume $\tilde{\rho}_{vr} \sim N(m'_v \varphi_r, \lambda_r^2)$. For the spatial random effect for the temporal parameter $\tilde{\rho}_s$ is assumed to be

$$\varphi_r | \omega_r \sim (0, (\omega_r M' Q M)^{-1}), \tag{8}$$

where φ_r is the spatial random effect for the rth order of the temporal parameters, and M and Q in Eq. (8) are the same as those listed in Eq. (4). Finally, we assume that $\omega_r \sim \mathcal{G}\left(\frac{a_\omega}{2}, \frac{b_\omega}{2}\right)$, $\lambda_r^2 \sim \mathcal{IG}\left(\frac{a_\lambda}{2}, \frac{b_\lambda}{2}\right)$ and $\sigma^2 \sim \mathcal{IG}\left(\frac{a_\sigma}{2}, \frac{b_\sigma}{2}\right)$. The values of a's and b's in the gamma or inverse gamma distributions are determined by the user to reflect the strength of one's prior belief before observing the data.

3. Posterior inference

To explore parallel computation, which allows us to substantially speed-up expensive operations in the MCMC iteration, we partition the brain into non-overlapping areas such as rectangular three-dimensional lattices or Brodmann areas. We then carry out the statistical model in Section 2 to analyse the partitioned data. In this chapter, for simplicity, we use Brodmann's map to parcel the brain into distinct regions before implementing the proposed model to the real fMRI data. In addition, separate regions divided by a data-drive segmentation procedure using functional clustering can be considered. Next, we introduce the likelihood function and the priors combining the parcellation procedure to form the posterior distribution for inference.

Suppose that a brain can be divided into G parcels, where the gth parcel contains V_g voxels, $g = 1, \ldots, G$ and where we denote the voxel-level parameters as $\theta_v = (\beta_v, \rho_v, \gamma_v, \sigma_v^2)$ and the

parcel-level parameters as $\Theta_g=(\varphi, \kappa, \tau^2, \lambda^2, \phi, \omega)$. The posterior distribution is obtained by combining the priors $\pi(\theta_v, \Theta_g)$ and the likelihood $L(\theta_v \mid y_v)$, which are defined in Eqs. (9) and (10) as follows [8, 20]:

$$\pi\left(\theta_v,\Theta_g\right)= \prod_{v=1}^{V_g}\pi(\rho_v)\pi\left(\sigma_v^2\right)\pi(\beta_v \mid \gamma_v)\prod_{j=1}^{p}\pi\left(\gamma_j \mid \phi_j\right)\pi\left(\phi_j \mid \kappa_j\right)\pi\left(\kappa_j\right)\pi\left(\tau_j^2\right)\prod_{k=1}^{r}\pi(\lambda_k)\pi(\varphi_k \mid \omega_k)\pi(\omega_k) \tag{9}$$

and

$$L(\theta_v \mid y_v)=\left(\frac{1}{2\pi\sigma_v^2}\right)^{-T/2}\exp\left\{-\frac{\left(y_v-X_v\beta_v-L_v\rho_v\right)'\left(y_v-X_v\beta_v-L_v\rho_v\right)}{2\sigma_v^2}\right\}. \tag{10}$$

Due to the intractability of the posterior, Gibbs and Metropolis-Hastings updates [20, 34] are applied to sample the posterior distribution for estimation and inference of the model parameters.

The posterior quantities of interest are $P(\gamma_{vj}=1 \mid y)$ for the activation map and $E(\beta_{vj} \mid y)$ for the magnitude of the effect caused by the stimulus j on the BOLD signal changes for voxel v. They can be directly estimated based on MCMC samples by

$$\hat{P}\left(\gamma_{vj}=1 \mid y\right)=\frac{1}{M}\sum_{m=1}^{M}\gamma_{vj}^{(m)} \text{ and } \hat{E}\left(\beta_{vj} \mid y\right)=\frac{1}{M}\sum_{m=1}^{M}\beta_{vj}^{(m)}, \tag{11}$$

where $\gamma_{vj}^{(m)}$ and $\beta_{vj}^{(m)}$ are the mth sample from a total M MCMC samples for γ_{vj} and β_{vj}, respectively. The distribution of $\hat{P}(\gamma_{vj}=1 \mid y)$ in a map provides a way to visually inspect brain regions with peak, high, low and practically no activation. In addition, $\hat{E}(\beta_{vj} \mid y)$ in Eq. (11) offers researchers to view the strength of response in the brain to the stimulus.

The construction of the binary activation map is obtained by thresholding the posterior probability of $\gamma_{vj}=1$. A threshold value c needs to be used in the detection of brain activity, that is, a voxel v is defined as active to a stimulus j if $\hat{P}(\gamma_{vj}=1 \mid y)>c$. However, threshold determination lacks agreement among investigators. Under certain conditions, a median selection criterion, $c=0.5$, results in the minimum prediction risk [21]. Smith et al. [6] and Lee et al. [4] defined a threshold by matching a Bayes factor approximation to a likelihood ratio test for activation such that $c=0.8722$ at a 5% level of significance. Additional to both, Kalus et al. [22] applied the false discovery rate (FDR) [23] to determine a threshold. In practice, it would be reasonable to construct activation maps for a grid of thresholds and to evaluate elicited results

with neuroscience experts. In our experience, activation maps seem to be comparably robust against an exact choice of the threshold.

4. Simulation studies

We conduct stimulation studies to investigate the performance of the proposed model on the detection of brain activity. We measure the accuracy of the classification of voxels as either active/inactive and show the proposed model to be robust with regard to different spatial structures. Finally, we illustrate the implementation of the proposed model to a simulated fMRI data that mimics a study in face repetition effects in memory tests [24].

4.1. Benchmark example

Consider a 30×30 binary activation image γ_v, v 1, …, 900, generated independently from

$$\gamma_v \mid \eta_v \sim Ber(\eta_v), \tag{12}$$

where $\text{logit}(\eta_v)=\alpha_v + m'_v\phi$ and

$$\phi \mid \kappa \sim N\left(0,(\kappa M'QM)^{-1}\right). \tag{13}$$

We let $\alpha_v=0$, $\kappa=0.5$, where M is a 300×900 matrix whose columns are $q=300$ principal eigenvectors of the adjacency matrix, A, for the image corresponding to the first $q=300$ largest eigenvalues. Moreover, m'_v is the v_{th} row of M, and $Q=\text{diag}(A\mathbf{1})-A$.

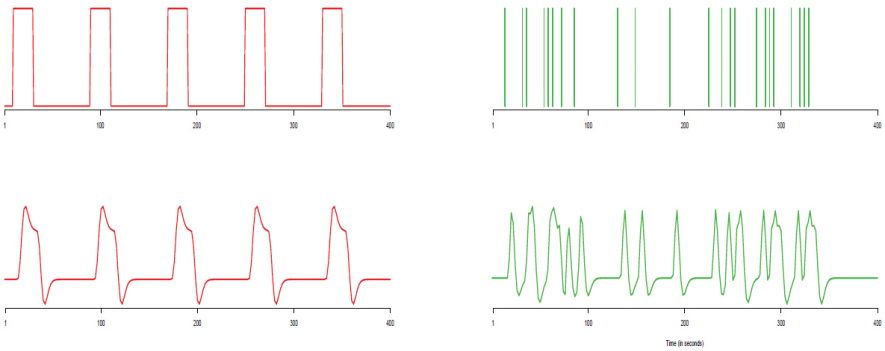
We consider an AR(1) temporal correlation between time points within voxels. For each voxel v, let $\tilde{\rho}_v=\log\frac{1+\rho_v}{1-\rho_v}$ and $\tilde{\rho}_v$ are generated from

$$\tilde{\rho}_v \sim N(m'_v\varphi, \lambda^2) \text{ and } \varphi|\omega \sim N(0,(\omega M'QM)^{-1}). \tag{14}$$

We let $\lambda^2=0.1$ and $\omega=2$.

We consider a stimulus and the block and event-related designs with a total number of time points spanning 400 and repetition time (TR) equal to 2s. In the block design, the duration time is 20s. In the event-related design, the stimulus is randomly assigned. The stimulus function is convolved with a HRF modelled by a double gamma function [15] to create the design matrix

X_v. The stimulus functions and the corresponding design matrices for the block and event-related designs are shown in **Figure 1(a)** and **(b)**, respectively.



(a) Block design. (b) Event-related design.

Figure 1. The stimulus functions and the predicted BOLD signal changes used in the Benchmark example.

Given X_v, γ_v, ρ_v, the BOLD signal changes y_v are generated from a normal distribution with a covariance $\sigma_v^2\Lambda_v$, where $\sigma_v^2=1$ and the $(u, v)_{\text{th}}$ element of Λ_v is $\rho^{|u-v|}$, and a mean of $X_v\beta_v$, where

$$\beta_v = \begin{cases} 3 & \text{if} \quad \gamma_v = 1; \\ 0 & \text{if} \quad \gamma_v = 0. \end{cases} \tag{15}$$

We then apply the proposed model to detect the activation areas, that is, to identify which voxel has $\gamma_v=1$. We ran MCMC chains with 100,000 iterations in order to ensure the largest Monte Carlo standard error (MCSE) [25] of all posterior probabilities of $\gamma_v=1$, less than 0.01.

We obtained posterior activation maps by setting the posterior probability threshold at 0.8722, that is, the voxel is categorized as active if $\hat{P}(\gamma_v=1 \mid y)>0.8722$, and it is categorized as inactive otherwise [26]. To measure the performance of our proposed model, we calculate the true classification rate (TCR), the true positive rate (TPR) and the false positive rate (FPR), which are defined as

$$\text{TCR} = \frac{\text{number of correctly classified voxels}}{\text{number of voxels}}; \tag{16}$$

$$\text{TPR} = \frac{\text{number of active voxels correctly claimed as active}}{\text{number of active voxels}}; \tag{17}$$

http://dx.doi.org/10.5772/64823

$$\text{FPR} = \frac{\text{number of inactive voxels correctly claimed as active}}{\text{number of inactive voxels}}. \quad (18)$$

Table 1 shows the proposed method performs well with regard to detecting the active voxels where only a small number of active voxels are falsely identified as inactive in both designs. Over 10 replications, the model with consideration of spatial dependence performs better on the detection of activations. Especially, FPR reduces around 50%.

(a)		
	Block design (%)	Event-related design (%)
TCR	92.44 (90.28–96.67)	91.12 (91.82–96.00)
TPR	99.97 (99.94–100)	99.97 (99.94–100)
FPR	6.98 (3.45–8.75)	6.54 (3.54–8.66)
(b)		
	Block design (%)	Event-related design (%)
TCR	96.33 (95.89–97.56)	95.78 (95.12–97.50)
TPR	99.99 (99.97–100)	99.99 (99.97–100)
FPR	3.42 (1.12–5.27)	3.22 (1.32–4.78)

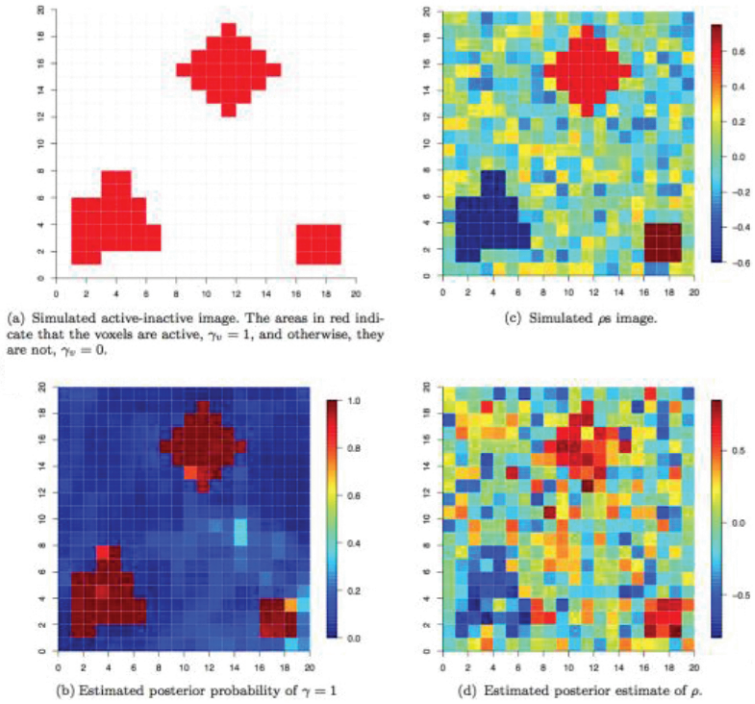
The values within the parentheses are the range of different rates over 10 replications.

Table 1. The percentage of correct classification of voxels with a threshold at 0.8722 for considering the spatial dependence in (a) and without considering spatial dependence in (b) for temporal parameters in the model.

4.2. Structures with spatial dependence

To demonstrate that the proposed model is able to handle the different binary and temporal image spatial dependencies, we generate a dataset that is the same as the benchmark example in the paper of [27]. We create a 20×20 binary image as shown in **Figure 2(a)**, where the areas in red indicate that the voxels are active, and otherwise, they are not. Moreover, the values of ρs are generated from a uniform distribution between –0.3 and 0.3 for the non-active areas. However, we assign the fixed values –0.5, 0.5, and 0.75 to ρ in each active area, respectively, as shown in **Figure 2(b)**. This is designed to investigate the patterns of temporal dependencies within the different areas. Similar to the settings discussed in Section 3.1, we consider both block and event-related designs with one stimulus. First, β_v's generated from $U(2, 5)$ and σ_v^2 from $U(1, 3)$. The design matrix X_v and temporal correlation matrix Λ_v are defined to be the same as in Section 3.1. Given X_v, γ_v, ρ_v, β_v, and σ_v^2 the BOLD signal changes are generated from a normal distribution with a mean $X_v\beta_v$ and a covariance $\sigma_v^2\Lambda_v$. We carry out the proposed approach to detect the activation and estimate the parameters of interest. For ease of visualization of the results, we provide one of the estimated binary images without thresholding and the corresponding estimated image, as shown in **Figure 2(c)**. By visually inspecting estimated values of ρs in **Figure 2(d)** compared to the simulated ones in **Figure 2(b)**, we are confident that the spatial dependences also can be captured by the model.

After thresholding by a value determined by controlling FDR as 0.05, the accuracy of the classification of voxels measured by TCR is 99.25%, and the FPR is 4.84% with consideration of the spatial dependence of the temporal correlations over 10 replications. On the other hand, when the threshold is taken to be equal to 0.8722, TCR is 99.17% and FPR is 5.01%. Therefore, the proposed model can be applied to detect the activation of brain image data even for different spatial dependence structures.



(a) Simulated active-inactive image. The areas in red indicate that the voxels are active, $\gamma_v = 1$, and otherwise, they are not, $\gamma_v = 0$.

(c) Simulated ρs image.

(b) Estimated posterior probability of $\gamma = 1$

(d) Estimated posterior estimate of ρ.

Figure 2. Simulated and estimated binary and ρ images.

4.3. Parcellation effect

In the previous simulations, image data was analysed together. In this simulation, we partition the data and then apply the proposed model to analyse each unit of the partitioned data. Our goal is to investigate the effect of the parcellation of an image on the estimation of active voxels. Given the binary image γ shown in **Figure 3(a)**, a unit of data is generated with the parameters being the same as the settings discussed in Section 3.1.

We consider four different parcellations, as shown in **Figure 3(a)**, **(b)** and **(d)**. **Table 2** shows the values of different measurements over 10 replications. Little difference can be found on the measures between different parcellations. However, the computational time is dramati-

http://dx.doi.org/10.5772/64823

cally reduced for case (d), where each parcellation contains fewer voxels. The same results in block and event-related fMRI experiments.

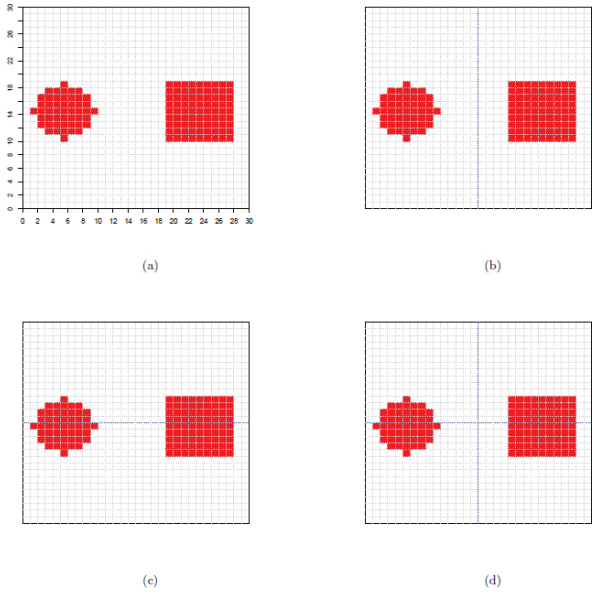


Figure 3. Four different parcellation schemes. The areas in red are active areas corresponding to $\gamma_v = 1$. The blue dashed line is used to partition the area into non-overlapping parts.

Parcellation Scheme	(a)	(b)	(c)	(d)
TCR	99.12 (98.61–99.84) %	99.07 (98.68–99.56) %	98.94 (98.03–99.64) %	99.08 (98.76–99.51) %
TPR	100%	100%	100%	100%
FPR	6.35 (1.51–9.97) %	6.68 (3.67–9.36) %	7.45 (3.12–13.87) %	5.82 (2.83–9.44) %
Running time ratio	8.75	2.75	2.75	1

The values within the parentheses are the ranges of different rates over 10 replications.

Table 2. The average percentage of correctly classified voxels (TCR), true positives (TPR), and false positives (FPR) over 10 replications for different image parcellations corresponding to **Figure 3(a)–(d)**.

4.4. A real-life simulation example

To further demonstrate the capability of the proposed model to accurately identify the activation regions, we simulate an fMRI dataset from an **R** package: **neuRosim** [28]. The simulated data follows the example of event-related fMRI study [24]. We consider four different tasks denoted by N1, N2, F1 and F2, which are presented randomly in the experiment. The onsets for each condition are schematically shown in **Figure 4**. We then specify three areas

that are activated by the tasks shown in **Figure 5**. The simulated four-dimensional fMRI data consists of 351 scans, each containing three-dimensional image of size 53×63×46 and being collected in every 2 s resulting the total time for the experiment is 11 min and 42 s.

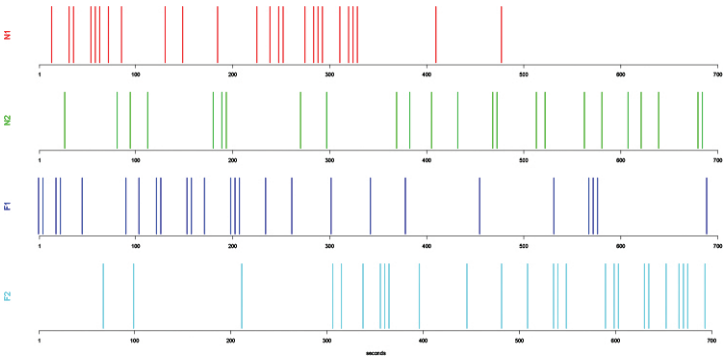


Figure 4. Depiction of temporal occurrences of four different tasks.

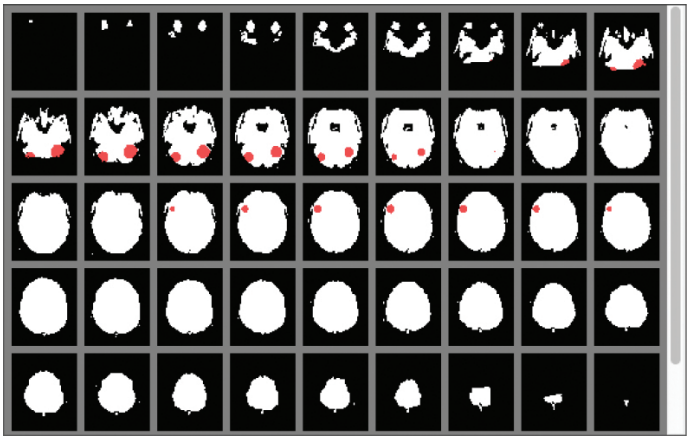


Figure 5. Activated regions upon representation of different tasks.

Now we apply the proposed model to analyse the simulated data. We split the array into disjointed arrays based on the Brodmann level regions. The voxel for all regions ranges from 236 to 969. We collect 1,000,000 samples to estimate the posterior probability of activation in each voxel. We consider the following probability:

$$P\left(\min_{j} \gamma_{vj} = 1 | y\right) \tag{19}$$

http://dx.doi.org/10.5772/64823

which is used to denote whether the voxel v is activated by any of applied stimuli. Once the posterior probability is greater than 0.8722, then the corresponding voxel is considered as activated by a stimulus. The activation maps registered into MNI 152 template (a reference brain map from the Montreal Neurological Institute) and the areas considered active are in yellow as shown in **Figure 6**. In addition to comparing specified activation regions in **Figure 5** to those detected in **Figure 6**, we also calculate the three measures, TCR, TPR and FPR which are 97.71%, 99.34% and 3.25%, respectively. The results show that proposed model performs well on classifying the active and inactive voxels.

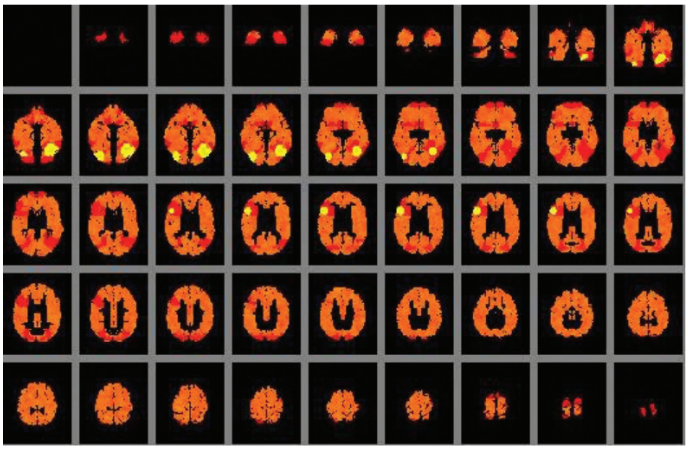


Figure 6. Estimated probability of activation in response to any face task with activation areas in yellow.

5. Application

We apply the spatial Bayesian variable selection approach discussed in Section 2 to real fMRI data, a study of the Simon effect [29]. In this Simon task, participants responded to one colour with the right hand, and to the other colour with the left hand. Congruence in this aspect means that a colour appears on the default side (congruent condition), or it may appear on the opposite side (incongruent condition). One participant's brain image data is selected to be analysed for the purpose of illustration. The four-dimensional fMRI BOLD image data was pre-processed using FSL from Oxford Centre for Functional MRI of the Brain [30], including motion correction, realignment, slice timing correction, spatial smoothing and high-pass filtering, before being analysed using our model. Incorrect trials were removed. The four tasks in this experiment were convolved with a double gamma function. We divided the pre-processed data into 48 Brodmann areas. The statistics of interest were registered into a standard template MNI152 for ease of visualization. We used a first-order temporal autocorrelation throughout the study.

We consider a voxel to be active when the posterior probability of γ_v is greater than c, which is either a deterministic value of 0.8722 [4] or is determined by an FDR equal to 0.05 [22]. However, in this event-related fMRI, 0.8722 seems too conservative, so the value of c is determined corresponding to an FDR equal to 0.05. The activation maps for different tasks are shown in **Figure 7(a)–(d)**.

One interesting question is to compare active domains corresponding to congruency. Posterior probabilities helped us find the following inequality between the incongruent and congruent tasks groups:

$$P\left(\beta_{\text{right-incon}} + \beta_{\text{left-incon}} > \beta_{\text{right-con}} + \beta_{\text{left-con}} | y\right), \tag{20}$$

where β_s is the magnitude of corresponding effect on the BOLD signal changes. The posterior activations are shown in **Figures 8** and **9** when posterior probabilities are thresholded by a critical value corresponding to FDR equal to 0.05. The active regions include frontal and occipital lobes detected by the models with and without consideration of spatial dependence of temporal correlations. These results are consistent with other studies [31, 32]. However, more active areas were detected by the model without consideration of spatial dependence of temporal correlations. For something like the Simon task, it is expected that the lateral prefrontal cortex to be activated as they are parts of the cognitive control network/multiple demand network, while the medial prefrontal cortex is part of the default mode network [33]. Therefore, it is possible that the activation in the medial part may be some spread of activity and therefore false positives. This is an important example that illustrates considering the

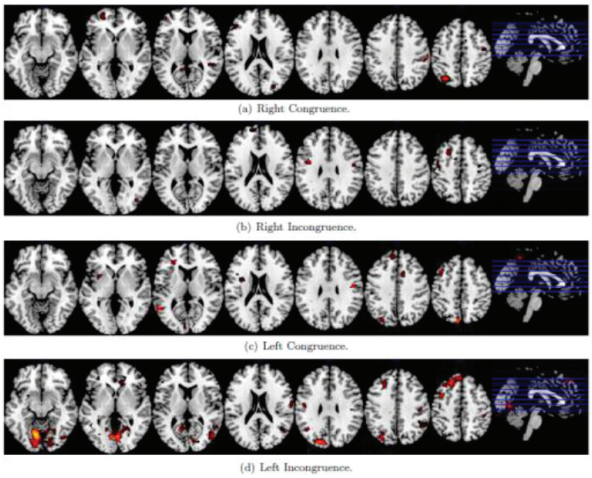


Figure 7. The activation areas with a posterior probability greater than c, corresponding to FDA = 0.05, shown in red for different tasks.

http://dx.doi.org/10.5772/64823

spatial dependence of temporal correlations may improve the estimation and reduce the some spurious detection of activation.

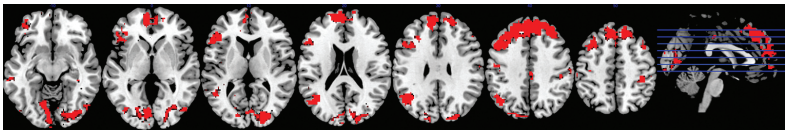


Figure 8. Areas shown more activity in incongruent than congruent task without consideration of spatial dependence of temporal correlations in the model.

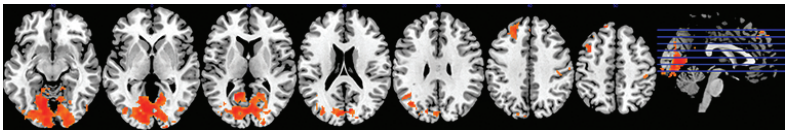


Figure 9. Areas shown more activity in incongruent than congruent task with consideration of spatial dependence of temporal correlations in the model.

6. Discussion and conclusions

In this work, we applied a new approach to performing Bayesian variable selection with consideration of spatial dependencies from both regression and temporal coefficients in single-subject event-related fMRI data. Through simulations, improvement in the detection of brain activity was observed for a wide range of different spatial structures, parcellations and experimental designs. We found that the proposed approach potentially decreases false positive rates as shown in **Table 1** in the simulation and **Figure 9** in the real example. In addition, prior information from brain structure and function for subject-level inference can be incorporated into the analysis.

It is worth noting that the proposed Bayesian approach partitions a brain into several regions before implementing the model to the fMRI data. In addition to Brodmann areas, several parcellations of the brain into distinct regions are available, such as the separate regions divided by a data-drive segmentation procedure using functional clustering. From the simulation study, we found the computational burden to be greatly reduced by the joint use of parcellation and an SGLMM. In particular, the probability of activation and activation magnitudes were readily computed without requiring an adjustment for multiple comparisons in a post-processing step. For broader goals, it would be of interest to extend the model to group studies. In addition, we are confident that Bayesian approaches represent an important direction in fMRI and in high-dimensional research in general.

Acknowledgements

This research was, in part, supported by the Ministry of Education, Taiwan, R.O.C. and National Cheng Kung University (the Aim for the Top University Project to National Cheng Kung University, NCKU). We thank the Mind Research and Imaging Center (MRIC) at NCKU for consultation and instrument availability.

Author details

Kuo-Jung Lee

Address all correspondence to: kuojunglee@mail.ncku.edu.tw

Department of Statistics, National Cheng-Kung University, Tainan, Taiwan

References

[1] J. A. Detre and T. Floyd. Functional MRI and its applications to the clinical neurosciences. The Neuroscientist. 2001;7:64–79.

[2] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, editors. Statistical Parametric Mapping: The Analysis of Functional Brain Images. London: Academic Press; 2006.

[3] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A unified statistical approach for determining significant voxels in images of cerebral activation. Human Brain Mapping. 1996;4:58–73.

[4] K. Lee, G. L. Jones, B. Caffo, and S. Bassett. Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. Bayesian Analysis. 2014;9:699–732.

[5] L. Zhang, M. Guindani, and M. Vannucci. Bayesian models for functional magnetic resonance imaging data analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2015;7:21–41.

[6] M. Smith and L. Fahrmeir. Spatial Bayesian variable selection with application to functional magnetic resonance imaging. Journal of the American Statistical Association. 2007;102:417–431.

[7] J. Hughes and M. Haran. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society Series B. 2013;75:139–159.

[8] D. R. Musgrove, J. Hughes, and L. E. Eberly. Fast, fully Bayesian spatiotemporal inference for fMRI data. Biostatistics. 2016;17:291–303.

[9] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith. Fully Bayesian spatio-temporal modeling for fMRI data. IEEE Transactions on Medical Imaging. 2004;23:213–231.

[10] A. Eklund, T. Nichols, M. Andersson, and H. Knutsson. Empirically investigating the statistical validity of SPM, FSL and AFNI for single subject fMRI analysis. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI); New York, NY; 2015; pp. 1376–1380.

[11] A. Quirós, R. M. Diez, and D. Gamerman. Bayesian spatiotemporal model of fMRI data. NeuroImage. 2010;49:442–456.

[12] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970;57:97–109.

[13] S. Brooks, A. Gelman, G. Jones, and X. Meng. Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC; 2011 Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487–2742.

[14] W. D. Penny, N. J. Trujillo-Barreto, and K. J. Friston. Bayesian fMRI time series analysis with spatial priors. NeuroImage. 2005;24:350–362.

[15] M. A. Lindquist, J. M. Loh, L. Y. Atlas, and T. D. Wager. Modeling the hemodynamic response function in fMRI efficiency, bias and mis-modeling. NeuroImage. 2009;45:S187–S198.

[16] J. C. Rajapakse, F. Kruggel, and J. M. M. D. Y. Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. Human Brain Mapping. 1998;6:283–300.

[17] E. I. George and R. E. McCulloch. Approaches for Bayesian variable selection. Statistica Sinica. 1997;7:339–374.

[18] D. M. Cvetkovć, M. Doob, and H. Sachs. Spectra of Graphs: Theory and Applications. 3rd ed. New York, NY: Wiley; 1998.

[19] N. G. Best, L. Waller, A. Thomas, E. M. Conlon, and R. Arnold. Discussion of Bayesian models for spatially correlated disease and exposure data. In: J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. Bayesian Statistics 6. Oxford University Press; 1999. pp. 147–150.

[20] A. Shmuel, E. Yacoub, D. Chaimow, N. K. Logothetis, and K. Ugurbil. Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. NeuroImage. Oxford 2007;35:539–552.

[21] M. Barbieri and J. O. Berger. Optimal predictive model selection. Annals of Statistics. 2004;32:870–897.

[22] S. Kalus, P. G. Sämann, and L. Fahrmeir. Classification of brain activation via spatial Bayesian variable selection in fMRI regression. Advances in Data Analysis and Classification. 2014;8:63–83.

[23] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B. 1995;57:289–300.

[24] R. Henson, T. Shallice, M. Gorno-Tempini, and R. Dolan. Face repetition effects in implicit and explicit memory tests as measured by fMRI. Cerebral Cortex. 2002;12:178–186.

[25] G. Jones, M. Haran, B. Caffo, and R. Neath. Fixed-width output analysis for Markov chain Monte Carlo. Journal of the American Statistical Association. 2006;101:1537–1547.

[26] O. Josephs and R. Henson. Event-related fMRI: modelling, inference and optimisation. The Royal Society: Philosophical Transactions B. 1999;354:1215–1228.

[27] M. Bezener, J. Hughes, and G. L. Jones. Bayesian spatiotemporal modeling using hierarchical spatial priors with applications to functional magnetic resonance imaging. Available from: http://users.stat.umn.edu/~galin/.

[28] M. Welvaert, J. Durnez, B. Moerkerke, G. Berdoolaege, and Y. Rosseel. neuRosim: An R package for generating fMRI data. Journal of Statistical Software. 2011;44:1–18.

[29] T. Wen and S. Hsieh. Neuroimaging of the joint Simon effect with believed biological and non-biological co-actors. Frontiers in Human Neuroscience. 2015;9:1–13.

[30] FMRIB, Oxford, UK. FMRIB Software Library [Internet]. Available from: http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/.

[31] X. Liu, M. Banich, M. Jacobson, and J. Tanabe. Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. NeuroImage. 2004;22:1097–1106.

[32] B.U. Forstmann, W. P. M van den Wildenberg, and K. R. Riderinkhof. Neural mechanisms, temporal dynamics, and individual differences in interference control. Neuroscience. 2008;20:1854–1865.

[33] T. A. Niendam, A. R. Laird, K. L. Ray, Y. M. Dean, D. C. Glahn, and C. S. Carter. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. Cognitive, Affective, and Behavioural Neuroscience. 2012;12:241–268.

[34] A. Gelman, J. C. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. Bayesian Data Analysis. 3rd ed. Chapman and Hall/CRC; 2013.

[35] C. Berrett and C. A. Calder. Bayesian spatial binary classification. Spatial Statistics. 2016;16:72–102.