Spatial Statistics 21 (2017) 96–113

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Spatial Bayesian hierarchical model with variable selection to fMRI data

Kuo-Jung Lee [a,*], Shulan Hsieh [b,c,d], Tanya Wen [c,e]

[a] Department of Statistics, National Cheng-Kung University, Tainan, Taiwan
[b] Cognitive Electrophysiology Lab: Control, Aging, Sleep, and Emotion, National Cheng Kung University, Tainan, Taiwan
[c] Department of Psychology, National Cheng Kung University, Tainan, Taiwan
[d] Institute of Allied Health Sciences, Department of Public Health, National Cheng Kung University, Tainan, Taiwan
[e] MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 21 January 2016
Accepted 6 June 2017
Available online 23 June 2017

Keywords:
Bayesian variable selection
fMRI
MCMC
Spatial random mixed-effects model

ABSTRACT

A spatial Bayesian hierarchical model is proposed to analyze functional magnetic resonance imaging data for complex spatial and temporal structures. Several studies found that spatial dependence not only appears in signal changes but also in temporal correlations among voxels; however, current statistical approaches ignore the spatial dependence of temporal correlations, thereby keeping computational efficiency. We incorporated the spatial random effects model to simultaneously consider spatial dependence arising from both signal changes and temporal correlations. We conducted simulation studies to demonstrate that the proposed approach increases the accuracy of the detection of brain activities while remaining computationally feasible. A real event-related fMRI data is provided to further illustrate the application and usefulness of the proposed model.

1. Introduction

Functional magnetic resonance imaging (fMRI) allows researchers to investigate brain activity – resulting from a stimulus – by observing blood oxygenation level dependent (BOLD) signals

* Corresponding author.
E-mail addresses: kuojunglee@mail.ncku.edu.tw (K. Lee), psyhsl@mail.ncku.edu.tw (S. Hsieh), tanya.wen@mrc-cbu.cam.ac.uk (T. Wen).

http://dx.doi.org/10.1016/j.spasta.2017.06.002
2211-6753/

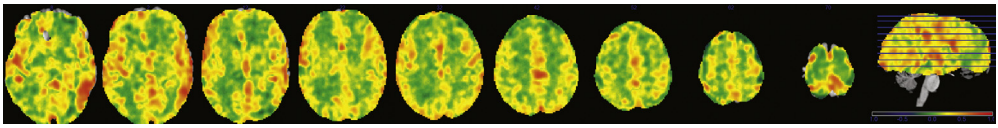


Fig. 1. An example of the distribution of temporal coefficients in AR(1) taken from (Lee et al., 2014). A red–blue color scale was used, where red represents correlation +1, blue represents correlation −1, and green represents correlations around 0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Detre and Floyd, 2001). The data is derived from a subject performing a task in response to a stimulus, while three-dimensional images containing BOLD signals are collected every 1–3 seconds (s). Numerous statistical models have been proposed (to allow researchers) to detect localized regions activated during a task. The broader goal is to identify the networks required for a particular brain function, or to assess physical characteristics of the brain elicited by cognitive processes, see Friston et al. (2007) and references therein. From a statistical point of view, the Bayesian paradigm provides an (attractive) inferential framework (Zhang et al., 2015). In general, the brain activation structure of a group of subjects is analyzed, and the conclusions are generalized to the entire population. The efficiency of group inferences relies on the accuracy of single subject data analysis (Bowman et al., 2008; Chen et al., 2012). We thus propose a new Bayesian modeling approach to substantially improve the sensitivity of detection for single subject fMRI sessions.

Non-Bayesian approaches apply a Gaussian random field (GRF) theorem to minimize the error introduced by spatial normalization. However, fMRI data exhibits a complicated spatial and temporal structure. As recently published by Eklund et al. (2016) the invalid assumption of GRF may significantly inflate false-positive rates in neuroimaging studies. A variety of Bayesian approaches were introduced to model temporal and spatial dependences. They also help integrate the neighboring information for an activation indicator with temporal data, see Woolrich et al. (2004). Smith et al. (2003) and Smith and Fahrmeir (2007) introduced a latent binary variable to indicate whether a voxel is activated by a task. Lee et al. (2014) extended the approach by applying spatial Bayesian variable selection (SBVS) with Ising distribution to capture the spatial dependence between brain activations, and used an auto-regression model to depict temporal correlations in signal changes. In addition to the mentioned Ising distribution, Gaussian Markov random field (GRMF) models (Quirós et al., 2010) and spatial probit models (Kalus et al., 2014) are also widely used. However, all current methodologies are computationally intensive.

We propose a model with consideration of both spatial and temporal dependences in a computationally feasible manner. The spatial correlation in signal is modeled via latent indicator variables showing activation/inactivation of voxels, while the temporal dependence in signal is captured via auto-correlation parameters. The description of spatial dependence among latent indicator variables is based on the spatial generalized linear mixed-effects model (SGLMM) (Hughes and Haran, 2013; Musgrove et al., 2016). In addition, studies have found that the temporal autocorrelations appear similar in nearby voxels. Fig. 1 from Lee et al. (2014) shows the spatial distribution of the maximum likelihood estimate for temporal coefficient in the auto-regression model, AR(1), in each voxel. Nearby voxels in some domains have similar temporal correlation values, and hence the map may contain spatial correlation. Not considering the spatial dependence of temporal correlations also has an impact of the parameter estimation on detecting the brain activation partially due to the inaccurate estimate of temporal coefficients (Arbabshirani et al., 2014). We aim to improve the estimates of brain activity for each voxel by accurately estimating the temporal parameters in the auto-regression models. Therefore, we proposed a spatial random effect model to include the spatial dependence of temporal characteristics. This way, both the model estimation and the accuracy of the detection (of brain activity) will be improved. We will assert functional connectivity in the brain from the similar patterns of activation in different brain areas.

Our modeling strategy combines several ideas from the literature that we summarized above. In the time dimension, we employ an auto-regression model for correlated errors, and describe the patterns of temporal coefficients via a spatial random effect model. In the spatial dimension, we capture dependence of brain activities among nearby voxels via the SGLMM. To remain computationally

feasible in the model estimation, the parcellation of a brain image enables us to quickly update the activation maps using a Markov chain Monte Carlo (MCMC) algorithm (Musgrove et al., 2016). The proposed model can incorporate prior information on activation in the form of spatially informative variables.

The remainder of the paper is organized as follows. In Section 2, we introduce the Bayesian formulation of our statistical model, including the spatial dependence and the priors of the related parameters. Section 3 provides simulations to validate the model. In Section 4, we apply our methods to a data set from an event-related fMRI experiment realizing the Simon task. In Section 5, we summarize our findings and discuss future research directions.

2. Statistical modeling

Bayesian methods in neuroimaging studies have received recent attention, see (Bowman et al., 2008; Caffo et al., 2011; Genovese, 2000; Goldsmith et al., 2014; Smith et al., 2003; Smith and Fahrmeir, 2007; Woolrich et al., 2004; Xia et al., 2009). They can directly incorporate the physical characteristics of the experiment. Our model not only identifies the voxels that have a response to a task, but also detects the areas that have more activity in response to some stimulus in an fMRI experiment.

The basic idea is now described, while the completion of details of the proposed model is given in Section 2.1. Let y_{vt} be the observed BOLD signal from a voxel $v = 1, \ldots, N$ at time $t = 1, \ldots, T_v$. Then let x_{vtj} be the corresponding transformed jth stimulus by means of the convolution of a stimulus function s_{jt} with a parametric hemodynamic response function (HRF) h_{vt}. Typically, s_{jt} corresponds to a 'boxcar' function indicating if the stimulus is active/inactive at time t, while the HRF h_{vt} consists of a linear combination of two gamma functions (Friston et al., 1998; Glover, 1999; Gössl et al., 2001; Smith and Fahrmeir, 2007). In hemodynamics we deal with short-duration stimuli, for which the BOLD signal fits in a pattern. This includes an increase above the baseline 2 s after the onset of neuronal activity, a peak at about 5–8 s, and a fall below the baseline (the so called undershoot) for a period of 10 s (Aguirre et al., 1997). In addition, the change of BOLD signal intensity is partly due to stimulus-independent effects, such as movement, blood flow, slow drift, etc., which are modeled by covariates $z_{tk}, k = 1, \ldots, q$. Thus, we model the response y_{vt} with

$$y_{vt} = \sum_{k=1}^{q} z_{tk}\delta_{vk} + \sum_{j=1}^{p} x_{vtj}\beta_{vj} + \varepsilon_{vt},$$

where β_{vj} is the so called magnitude of activation that we measure by the BOLD signal level for the jth stimulus, while δ_v denotes the total magnitude contributed by the mentioned stimulus-independent effects. Finally, ε_{vt} is the measurement error. Modeling the inherent spatial and temporal characteristics is accomplished by making appropriate distributional assumptions for β_{vj} and ε_{vt}, and through the choice of prior distributions for the parameters.

A voxel v has a neural activation induced by the jth stimulus $\boldsymbol{x}_{vj}$ whenever the corresponding magnitude β_{vj} is nonzero. Detecting whether a voxel is activated by a stimulus is equivalent to detecting whether the variable $\boldsymbol{x}_{vj}$ has an effect on the BOLD signal y_{vt}. In the literature, this approach is called the variable selection problem (George and McCulloch, 1993, 1997). Several studies (Smith et al., 2003; Smith and Fahrmeir, 2007; Lee et al., 2014) from a Bayesian perspective, have dealt with variable selection problems. However, a naive fully Bayesian approach, which incorporates both spatial and temporal dependences, may encounter posteriors that are computationally intractable, even with the most sophisticated methods. In Section 2.1, we further develop our model to also incorporate spatial dependence in the temporal characteristics, pay careful attention to the prior specifications with the dual goal that we accurately reflect the nature of the experiment while mitigating the computational burden.

2.1. Bayesian formulation

Let y_v, a $T_v \times 1$ vector, denote a time series data for voxel v. A general linear model is considered for y_v as follows

$$\boldsymbol{y}_v = Z_v\boldsymbol{\delta}_v + X_v\boldsymbol{\beta}_v + L_v\boldsymbol{\rho}_v + \boldsymbol{\epsilon}_v; \quad \boldsymbol{\epsilon}_v \sim N(\mathbf{0}, \sigma_v^2 I), \tag{1}$$

where Z_v is a $T_v \times k$ matrix used to account for long memory trends, hardware related low frequency drift, residual movement effects, and aliased physiological noise, such as respiration and cardiac pulsation, with the corresponding effects, $\boldsymbol{\delta}_v$ (Friston et al., 2007; Lund et al., 2006). The $T_v \times p$ matrix X_v is the design matrix, each column of which consists of values obtained from a stimulus function with respect to a task, (Rajapakse et al., 1998; Lindquist et al., 2009) convolved with the canonical HRF (Friston et al., 2007), and $\boldsymbol{\beta}_v$, a $p \times 1$ vector, corresponding to the effects of stimuli. Temporal correlation is modeled by $\boldsymbol{\rho}_v$, an $r \times 1$ vector of auto-regression coefficients, with L_v a $T_v \times r$ matrix of lagged prediction errors (Penny et al., 2005, 2003). Finally, we assume that the error terms $\boldsymbol{\epsilon}_v$ are independently and normally distributed across voxels, with a mean vector $\mathbf{0}$ and a covariance matrix $\sigma^2 I$.

One important goal in this study is to detect neural activation in a voxel, that is, whether one has a response to a stimulus. For this purpose, we introduce a vector of binary random variables, $\boldsymbol{\gamma}_v = (\gamma_{v1}, \ldots, \gamma_{vp})^T$. The voxel is considered active to the stimulus j if and only if $\gamma_{vj} = 1$. It was found in brain studies that along with a voxel, its neighboring voxels behave similarly, they are active or inactive in a group. To account for the spatial relationship, Lee et al. (2014) developed a two-stage spatio-temporal model using the Ising prior for $\boldsymbol{\gamma}$ and AR(1) for temporal correlation between time points. However, there appeared computational difficulties arising from the simultaneous use of spatial and temporal parameters. Namely, at each MCMC iteration, a normalizing constant and the inverse of the covariance matrix need to be evaluated. This increases the computational burden on the MCMC sampler. Although the authors proposed the so called path sampling approach (Gelman and Meng, 1998) to compute the normalizing constant, there is still a thorny issue about the inversion of the $T_v \times T_v$ covariance matrix at each MCMC iteration.

To avoid the computational difficulty in evaluating the normalizing constant when applying Ising distribution to model the spatial dependence of binary random variables, we adopted an alternative Markov random field (MRF) prior, SGLMM (Musgrove et al., 2016). *In addition, we considered a parcellation technique* (Musgrove et al., 2016) *to partition the brain for computational efficiency. Thereby, a parcel-dependent model has been established. Within each parcel, an individual model is applied and fitted independently. This may simplify the spatial dependence of fMRI BOLD signal changes; however, by doing so, model estimations can be done in parallel, reducing analysis time, considerably. In addition, different regions of the brain may have different signal-to-noise ratios, so our model can more accurately capture distinct local characteristics.* Next, we introduce the priors of parameters in the model and the posterior distribution.

2.2. Priors

Suppose that a brain is partitioned into G non-overlapping parcels, where the gth parcel contains n_g voxels, $g = 1, \ldots, G$. We then fit the proposed model into (1) independently for each parcel. In the Bayesian framework, each parameter is necessarily assigned a prior.

Instead of the Zellner's prior used for β_{vj}, we consider a spike-and-slab mixture prior (George and McCulloch, 1997)

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

where c_{vj}^2 is fixed and τ_{vj}^2 is assumed to have an inverse gamma distribution, $\mathcal{IG}\left(\frac{a_\tau}{2}, \frac{b_\tau}{2}\right)$. We set c_{vj}^2 large enough resulting in a nonzero estimate for β_{vj}, that is, stimulus j activates voxel v. As George and McCulloch (1997) suggested, c_{vj}^2 should be less than 10^4 to avoid computational problems. We found that $c_{vj}^2 = 10$ is a reasonable choice for our simulations and the real fMRI example. Instead of a mixture prior with a Dirac delta spike and a normal slab or g-prior, our choice is useful to create

credible intervals for the β's of interest. Furthermore, it allows straightforward posterior estimates, due to the option of Rao-Blackwellization.

We assume that $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{n_g j})'$ are independent, Bernoulli distributed conditionals on η_{vj}, where $\eta_{vj} = P\left(\gamma_{vj} = 1 | s_{vj}, \alpha_{vj}\right)$; that is,

$$\gamma_{vj}|\eta_{vj} \overset{ind}{\sim} Ber(\eta_{vj}). \tag{2}$$

We let

$$\text{logit}\,(\eta_{vj}) = \alpha_{vj} + s_{vj}, \tag{3}$$

with $\text{logit}\,(\eta_{vj}) = \log\left(\frac{\eta_{vj}}{1-\eta_{vj}}\right)$, where s_{vj} is a random effect and α_{vj} is a fixed constant.

The α_{vj} is used to incorporate expert knowledge and anatomical information, while s_{vj} carries the spatial random effect to characterize the spatial dependence of $\boldsymbol{\gamma}_j$. Let $S_j = \left(s_{1j}, \ldots, s_{n_g j}\right)$ constitute a GMRF. That is, $S_j \sim N\left(0, (\kappa_j Q_g)^{-1}\right)$, where κ_j is a smoothing parameter, $Q_g = \text{diag}(A_g \mathbf{1}) - A_g$ is the precision matrix, $\mathbf{1}$ denotes a column vector of 1 s, and A_g is the underlying spatial structure of the random field. We assume that the spatial weights $a_{uv} = a_{vu} \geq 0$ for each pair of s_u and s_v in A_g satisfying $a_{uu} = 0$ for any u. To speed up computation, Hughes and Haran (2013) proposed a dimension reduction approach to the random field via introducing $s_{vj} = \boldsymbol{m}_v' \boldsymbol{\phi}_j$, where $\boldsymbol{\phi}_j$ is the vector of spatial random effects, and $\boldsymbol{m}_v'$ is the vector consisting of the spatial characteristics of data from the row vector M_g, an $n_g \times q$ matrix. The columns of the latter matrix are the q eigenvectors corresponding to the largest eigenvalues of the adjacency matrix A_g of all voxels in the gth parcel. The eigenvectors are multi-resolutional spatial basis vectors that describe spatial variation in the parcel. In general, brain activity in one voxel tends to be similar to that of nearby voxels. In accordance, only positive spatial dependence will be considered in the analysis, that is the columns of M_g are restricted to eigenvectors of A_g with positive eigenvalues and typically $q \leqslant n_g/2$ (Hughes and Haran, 2013).

Therefore, the model in (3) can be rewritten as $\text{logit}\,(\eta_{vj}) = \alpha_{vj} + \boldsymbol{m}_v' \boldsymbol{\phi}_j$ and the corresponding prior for the spatial random effect vector $\boldsymbol{\phi}_j$ is assigned as follows (Hughes and Haran, 2013)

$$\boldsymbol{\phi}_j|\kappa_j \sim N(0, (\kappa_j \tilde{Q}_g)^{-1}) \tag{4}$$

where $\tilde{Q}_g = M_g' Q_g M_g$. Note that $\tilde{Q}_g$ is a $q \times q$ matrix and $q \ll n_g$ so the operational cost is significantly reduced for the MCMC procedure. The prior for the spatial smoothing parameter κ_j is

$$\kappa_j \sim \Gamma\left(\frac{a_\kappa}{2}, \frac{b_\kappa}{2}\right).$$

It was found in Lee et al. (2014) that the temporal correlations between voxels also tend to behave similarly for neighboring voxels. We model the spatial dependence between temporal correlations as follows. We assume

$$\rho_{vr} \sim N(\boldsymbol{m}_v' \boldsymbol{\varphi}_r, \lambda_r^2), \tag{5}$$

where $\boldsymbol{\varphi}_r$ is the spatial random effect for the rth order of temporal correlations and

$$\lambda_r^2 \sim \mathcal{IG}\left(\frac{a_\lambda}{2}, \frac{b_\lambda}{2}\right).$$

We assume that the priors of $\boldsymbol{\varphi}_r$ have a hierarchical structure given by

$$\varphi_r|\omega_r \sim N(0, (\omega_r \tilde{Q}_g)^{-1});$$
$$\omega_r \sim \Gamma\left(\frac{a_\omega}{2}, \frac{b_\omega}{2}\right),$$

where ω_r is a spatial smoothing parameter. Finally, we assume that δ_{vk} are independent normal distributions with mean 0 and variance $s_{v\delta}^2$, that is,

$$\delta_{vq} \overset{ind.}{\sim} N(0, s_{v\delta}^2)$$

for $k = 1, \ldots, q$ and

$$\sigma_v^2 \sim \mathcal{IG}\left(\frac{a_\sigma}{2}, \frac{b_\sigma}{2}\right).$$

The values of a's and b's in the gamma and inverse gamma distributions are determined by the user. They incorporate one's prior assumptions before observing the data. The values of these hyperparameters in the priors of smoothing parameters κ_j and ω_r are chosen as $a_\kappa = a_\omega = 0.5$ and $b_\kappa = b_\omega = 1/2000$. Their purpose is not to bring artifacts into the spatial structure of the posterior density (Kelshall and Wakefield, 1999). In the inverse gamma prior for σ's and λ's, the nearly non-informative priors are used, that is, $a_\sigma = a_\sigma = a_\lambda = b_\lambda = 1$.

2.3. Posterior and Monte Carlo estimates

Denote the voxel-level parameters by $\boldsymbol{\theta}_v = (\boldsymbol{\beta}_v, \boldsymbol{\delta}_v, \boldsymbol{\rho}_v, \boldsymbol{\gamma}_v, \sigma_v^2)$ and the parcel-level parameters by $\boldsymbol{\Theta}_g = (\boldsymbol{\phi}, \kappa, \tau^2, \lambda^2, \boldsymbol{\varphi}, \omega)$. The posterior distribution is obtained by combining the priors $\pi(\boldsymbol{\theta}_v, \boldsymbol{\Theta}_g)$ and the likelihoods $L(\boldsymbol{\theta}_v|\boldsymbol{y}_v)$, which are defined by

$$\begin{aligned}\pi(\boldsymbol{\theta}_v, \boldsymbol{\Theta}_g) = &\prod_{v=1}^{n_g} \pi(\boldsymbol{\delta}_v)\pi(\boldsymbol{\rho}_v)\pi(\sigma_v^2)\pi(\boldsymbol{\beta}_v|\boldsymbol{\gamma}_v)\prod_{j=1}^{p} \pi(\boldsymbol{\gamma}_j|\boldsymbol{\phi}_j)\pi(\boldsymbol{\phi}_j|\kappa_j)\pi(\kappa_j)\pi(\tau_j^2)\\ &\times \prod_{k=1}^{r} \pi(\lambda_k)\pi(\boldsymbol{\varphi}_k|\omega_k)\pi(\omega_k)\end{aligned}$$

and

$$L(\boldsymbol{\theta}_v|\boldsymbol{y}_v) = \left(\frac{1}{2\pi\sigma_v^2}\right)^{-T/2} \exp\left\{-\frac{(\boldsymbol{y}_v - Z_v\boldsymbol{\delta}_v - X_v\boldsymbol{\beta}_v - L_v\boldsymbol{\rho}_v)'(\boldsymbol{y}_v - Z_v\boldsymbol{\delta}_v - X_v\boldsymbol{\beta}_v - L_v\boldsymbol{\rho}_v)}{2\sigma_v^2}\right\}.$$

The posterior based on our prior specifications is typically of extremely large dimension, and is unavailable in closed form. We depend on the use of MCMC throughout, to perform the required inferential tasks. Due to the intractability of the posteriors, Gibbs updates and Metropolis–Hastings updates were used to sample the posterior distribution. The details of derivation of all conditionals are given in Appendix.

The MCMC procedure produces samples from the joint posterior distribution of all the model parameters, allowing for their estimation and inference. Although the parametrization is complex in the proposed hierarchical model, the computation is efficient. Estimation of the single-subject fMRI data set can be done in parallel and considerable time saving can be achieved via the SGLMM as well.

The posterior probability $P(\gamma_{vj} = 1|\boldsymbol{y})$ for the activation map is directly calculated from MCMC samples by

$$\hat{P}(\gamma_{vj} = 1 \mid \boldsymbol{y}) = \frac{1}{M}\sum_{k=1}^{M} \gamma_{vj}^{(m)},$$

where $\gamma_{vj}^{(m)}$ is the mth sample among M pieces of MCMC samples for γ_{vj}. The distribution of $\hat{P}(\gamma_{vj} = 1|\boldsymbol{y})$ in a map provides a way to visualize brain regions with peak, high, low, and practically no activation. Here, voxels are classified active if their posterior activation probabilities are greater than a threshold. Similarly, the posterior estimation of $E(\beta_{vj} \mid \boldsymbol{y})$ is the average of all MCMC samples for β_{vj}.

2.4. Activation classification

The construction of the binary activation map is done by thresholding the posterior probability of $\gamma_{vj} = 1$. This means that a voxel v is considered active to a stimulus j if $\hat{P}(\gamma_{vj} = 1|\boldsymbol{y}) > c$. Threshold determination lacks agreement between researchers. Often, $c = 0.5$ results in minimal prediction risk (Barbieri and Berger, 2004). Smith and Fahrmeir (2007) and Lee et al. (2014) defined

the threshold as $c = 0.8722$ for activation by matching a Bayes factor approximation to a likelihood ratio test under 5% significance level.

For threshold determination, our preference was to minimize the false discovery rate (FDR) (Benjamini and Hochberg, 1995). It is the number of voxels falsely/incorrectly classified active divided by the number D of all voxels classified active. Our algorithm design followed Kalus et al. (2014). Let ζ denote the binary decision to classify a voxel active ($\zeta_v = 1$) or inactive ($\zeta_v = 0$). FDR is thus defined as

$$\text{FDR} = \frac{1}{D}\sum_{v}^{N}(1-\gamma_v)\zeta_v.$$

In a Bayesian framework, the threshold is determined by controlling the posterior expectation value of FDR (Newton et al., 2004),

$$E(\text{FDR} \mid \boldsymbol{y}) = \frac{1}{D}\sum_{v}^{N}\left(1 - P(\gamma_v = 1 \mid \boldsymbol{y})\right)\zeta_v.$$

Putting $\zeta_v = I\left(\hat{P}(\gamma_v = 1 \mid \boldsymbol{y} > t)\right)$ and replacing $P(\gamma = 1 \mid \boldsymbol{y})$ by its estimation, we can evaluate $E(\text{FDR} \mid \boldsymbol{y})$ over a grid after choosing a threshold to match a specified level, say 0.05. In our experience, activation maps seem to be robust against the choice of the threshold.

3. Simulation studies

There were two different simulation scenarios, block and event-related. We also compared the accuracy of classification with and without consideration of spatial dependence of temporal correlations in Section 3.1. Next, we design a simulation to investigate if the proposed approach is robust to different spatial and temporal dependence structures presented in Section 3.2.

3.1. Benchmark example

In this simulation study, our goal is not only to demonstrate that the proposed model is able to identify the active areas but also to compare performance with and without consideration of spatial dependence of temporal correlations while the dependence really appears in the data.

Our designs have a total number of 400 time points and a repetition time (TR) of 2 s. In the block design, the duration of each block is 20 s. In the event-related design, the stimulus is assigned randomly. The stimulus function is convolved with an HRF modeled by a double gamma function to create the design matrix X_v.

A 30×30 binary image $\boldsymbol{\gamma}$ was generated from (2) using $\boldsymbol{\phi}$ simulated from (4). We let $\alpha_v = 0$, $\kappa = 0.5$, M be a 300×900 matrix whose columns are the $q = 300$ principal eigenvectors of A corresponding to the $q = 300$ largest eigenvalues. We consider first order spatial dependence, that is, only 4 of the nearest voxels are considered neighbors in 2-dimensional data. Moreover, $\boldsymbol{m}_v'$ is the vth row of M, and $Q = \text{diag}(A\mathbf{1}) - A$. We consider an AR(1) for the temporal noise within voxels. For each voxel v, ρ_v is hierarchically generated by (5). We let $\lambda^2 = 0.1$ and $\omega = 2$.

We generate $\beta_v \sim U(1, 5)$ when $\gamma_v = 1$, otherwise $\beta_v \sim U(-0.1, 0.1)$. We set $\sigma_v^2 = 1$. Given X_v, β_v, γ_v, ρ_v, and σ_v^2, the BOLD signal $\boldsymbol{y}_v$ is generated by

$$\boldsymbol{y}_v \sim N\left(X_v\beta_v, \sigma_v^2\Lambda_v\right), \tag{6}$$

where the (s, v)th element of Λ_v is $\rho_v^{|s-v|}$. We considered two threshold values, one was the fixed value of 0.8722, and the other one was determined by FDR. To evaluate the detection performance of our proposed method, we considered the following measures: True Classification Rate (TCR), True

Table 1
The median of TCR, TPR, and FPR and the corresponding 5th and 95th quantiles (within the parentheses) over 1000 replications. MSE(β) and MSE(ρ) are the mean square errors on the estimates of βs and ρs. With/Without: do/do not consider the spatial dependence in the temporal coefficients.

	Block design		Event-related design	
	With	Without	With	Without
Fixed threshold: 0.8722				
TCR	96.36 (90.25, 98.34)%	90.20 (88.01, 94.25)%	95.87 (95.25, 97.35)%	92.24 (91.98, 96.02)%
TPR	94.78 (90.01, 97.60)%	88.22 (85.86, 92.15)%	94.45 (90.01, 96.23)%	88.07 (86.01, 92.34)%
FPR	2.32 (1.26, 5.25)%	5.67 (4.40, 10.89)%	2.76 (1.47, 5.65)%	6.01 (4.43, 10.72)%
FDR				
TCR	94.05 (87.90, 96.72)%	91.00 (87.92, 94.24)%	96.87 (95.95, 98.10)%	93.01 (92.01, 96.01)%
TPR	97.70 (91.62, 100)%	88.03 (84.87, 95.09)%	96.14 (90.23, 96.24)%	88.00 (85.10, 92.37)%
FPR	2.85 (1.52, 5.78)%	9.43 (4.59, 11.99)%	4.50 (3.84, 7.92)%	10.01 (4.45, 12.20)%
MSE(β) $\times 10^{-2}$	7.93 (2.55, 32.10)	13.23 (2.71, 55.30)	8.15 (2.72, 40.45)	12.73 (2.90, 58.50)
MSE(ρ) $\times 10^{-3}$	4.81 (4.02, 5.35)	5.44 (4.24, 8.36)	5.22 (3.80, 8.87)	10.13 (7.72, 20.50)

Positive Rate (TPR), and False Positive Rate (FPR), which are defined as follows

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

$$\text{TPR} = \frac{\text{number of active voxels correctly classified}}{\text{number of active voxels}};$$

$$\text{FPR} = \frac{\text{number of inactive voxels incorrectly classified}}{\text{number of inactive voxels}}.$$

TCR is an overall index of accuracy in the identification of active and inactive voxels. TPR is the rate of active voxels correctly identified, and is used to measure the power of the method. FPR is the rate of inactive voxels that are claimed as active; it can be considered the type I error rate of the method. Among these three criteria, it is preferred to have larger values for TCR and TPR, and a smaller value for FPR. Moreover, to evaluate the accuracy of the model estimation, we calculate the mean squared errors (MSE) of estimates on βs and ρs defined by

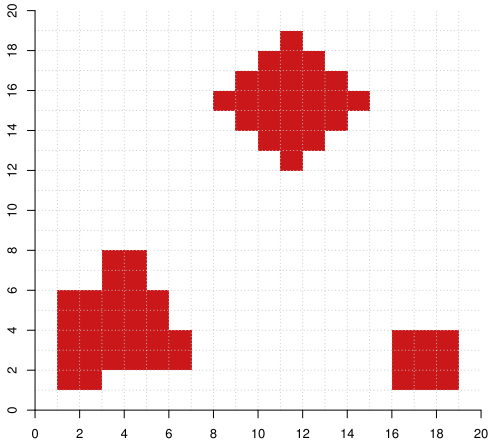
$$\text{MSE}\,(\beta) = \frac{1}{N}\sum_{v=1}^{N}(\beta_v - \hat{\beta}_v)^2 \quad \text{and} \quad \text{MSE}\,(\rho) = \frac{1}{N}\sum_{v=1}^{N}(\rho_v - \hat{\rho}_v)^2.$$

Smaller values for MSE indicate better performance than larger values.

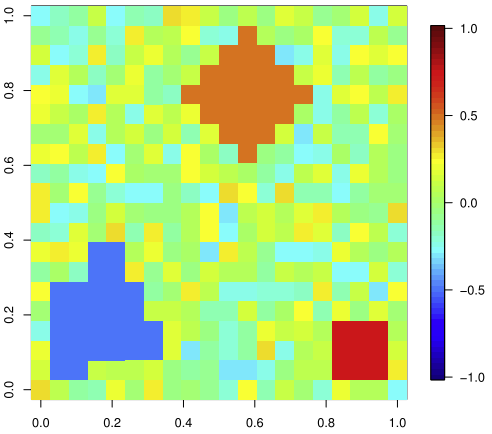
We compared the results over 1000 replications with/without considering the spatial dependence of temporal correlations in the model for different experimental designs. The consideration of spatial dependence of temporal correlations allowed our model to outperform the model without the same consideration, see Table 1. We found that the FPR was reduced by 50% when considering spatial dependence. It seemed that 0.8722 is too conservative, that is, some active voxels were not correctly detected. In contrast, a more accurate classification of voxels and a higher detection of active voxels using FDR is needed to determine the threshold.

3.2. Different spatial dependence structures for activations and temporal coefficients

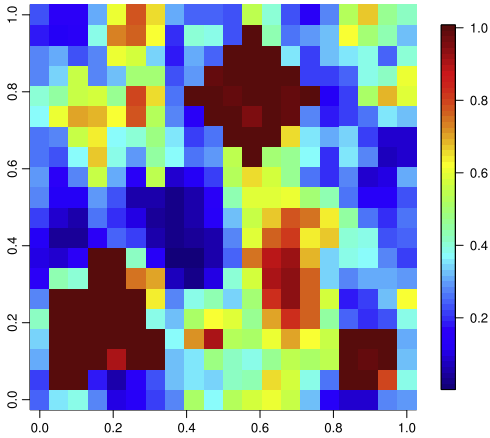
In this section, we demonstrate that the proposed model is applicable for the detection of active–inactive (γ) and temporal (ρ) images using different spatial dependence structures. We considered three scenarios to assess the performance. For a specific case, a 20 $\times$ 20 active–inactive image is shown in Fig. 2a where the areas in red color indicate active voxels (Bezener et al., 2017). In the temporal image, ρ is randomly assigned a value from $U(-0.3, 0.3)$ for inactive areas, and from -0.5,



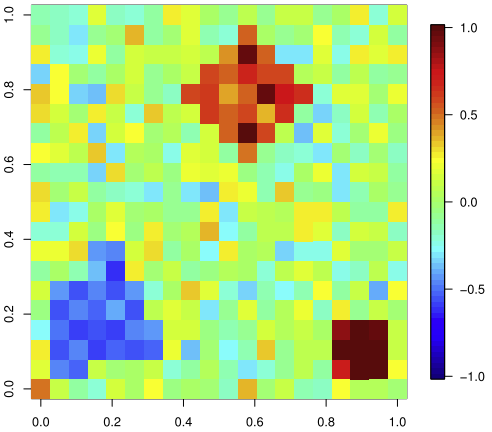
(a) Simulated active–inactive image. The areas in red indicate that the voxels are active, $\gamma_v = 1$, and otherwise, they are not, $\gamma_v = 0$.



(b) Simulated ρ image.



(c) Estimated posterior probability of $\gamma = 1$.



(d) Posterior estimate of ρ.

Fig. 2. Simulated and posterior estimated images for activation and ρ. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

0.5, and 0.75 for active areas, respectively, as shown in Fig. 2b. Additionally, we considered the active–inactive images generated from Ising and auto-logistic models, and the temporal image from a CAR model. We employed both block and event-related designs with one stimulus here. Given γ, X, β, σ, and ρ, the BOLD signal changes $\boldsymbol{y}_v$ were simulated by (6).

We carry out the proposed approach to detecting the activation and estimating the parameters of interest. Fig. 2 displays the simulated active–inactive and ρ-images and the corresponding posterior activation probabilities and estimated ρ images, respectively, for one simulated data set from a specific case. Table 2 shows the median with 5th and 95th quantiles within the parentheses of TCR, TPR, and FPR over 1000 replications. The results of block and event-related designs are similar in the simulation, so we only present the results from the event-related design in Table 2. The result shows an overall good performance of our model even for different spatial dependence structures of the active–inactive and temporal images.

Table 2
The median of TCR, TPR, and FPR and the corresponding 5th and 95th quantiles (within the parentheses) for different spatial dependence structures over 1000 simulations. The threshold is determined by FDR. Ising, Auto-logistic, Specific_Binary refer to the spatial structure for the activation map generated by Ising, auto-logistic models, and Fig. 2a, respectively. Specific_ρ and CAR_ρ denote the spatial structure for the ρ map generated by a specific case and a CAR model, respectively.

	Specific_Binary	Ising	Auto-logistic
Specific_ρ			
TCR	96.30 (96.00, 99.75)%	95.72 (94.50, 98.75)%	96.10 (96.00, 98.85)%
TPR	97.45 (93.55, 100)%	97.80 (93.55, 100.00)%	96.15 (91.94, 98.38)%
FPR	3.42 (1.47, 5.92)%	3.35 (1.18, 6.80)%	3.54 (1.47, 7.10)%
CAR_ρ			
TCR	96.20 (95.00, 99.75)%	96.02 (94.00, 98.75)%	95.10 (95.00, 98.85)%
TPR	97.50 (95.16, 98.38)%	96.80 (93.55, 100.00)%	95.15 (91.94, 98.38)%
FPR	3.45 (1.47, 5.62)%	3.30 (1.18, 6.50)%	3.65 (1.78, 6.80)%

4. Application

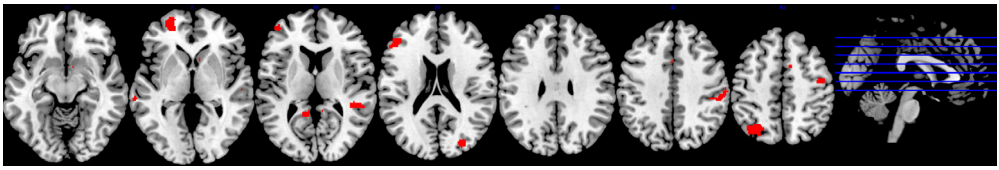
We considered an application to an fMRI study of the Simon effect (Liu et al., 2004; Wen and Hsieh, 2015), a typical paradigm for investigating cognitive control. In this Simon task, participants responded to one color with the right hand, and to the other color with the left hand. Congruence in this aspect means that a color appears on the default side (congruent condition). Obviously, in-congruence depicts appearance on the opposite side (in-congruent condition).

In the fMRI experiment, a fixation cross appeared for 500 milliseconds (ms) at the beginning of each trial. This was followed by the target displayed on the screen for up to 1400 ms before a response was given. Participants responded to the colored circles by pressing the corresponding buttons on their response pad. Afterwards, a blank screen (lasting from 0–8 s) followed before the start of the next trial. One run consisted of 160 trials (with 40 trials for each type: Right-congruent, Right-in-congruent, Left-congruent, Left-in-congruent), lasting for approximately 9.5 min. The order and the timing of each trial were pseudo-randomized using optseq2 (Dale, 1999) to maintain an optimum jitter level during each sequence. Stimulus providing was controlled via E-PRIME 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA).

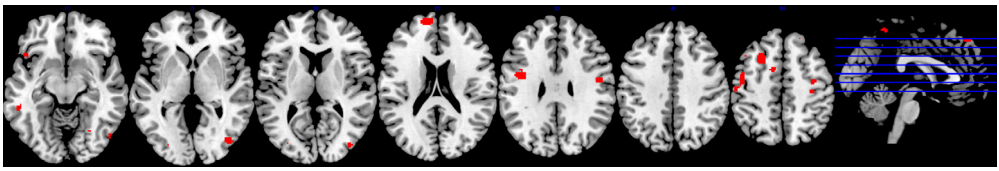
One participant's brain image data was selected for illustration. The fMRI BOLD signal data was preprocessed by FSL (FMRIB, 2015) including motion correction, realignment, slice timing correction, spatial smoothing, and high-pass filtering. The incorrect trials were removed and not modeled (so the incorrect trials were part of the implicit baseline). This is a common procedure for cognitive tasks with responses of high accuracy. The subject got 2 out of 160 trials incorrect; that is, the error rate was only 1.25%. Therefore the effects due to few incorrect trials would be mostly noise and can be ignored. But none of the volumes corresponding to incorrect trials were removed since it is difficult and not practical given the low temporal resolution of the BOLD signal. The four conditions in this experiment were convolved with a double gamma function.

We applied two partition schemes: one based on Brodmann and the other one on a simple cubic lattice. We subdivided the preprocessed data into 52 Brodmann areas. The areal voxel counts were between 209 and 833. In the case of simple cubic lattice, we partitioned the brain into 40 regions, each containing at least 500 voxels. Both schemes yielded similar activation patterns. In this analysis, we present the output obtained with Brodmann areas. The columns of M for modeling spatial dependence were chosen with the constraint that the eigenvalues be greater than 0.05. Dimensionality was dramatically reduced with such procedure, but the data still maintained the spatial characteristics.

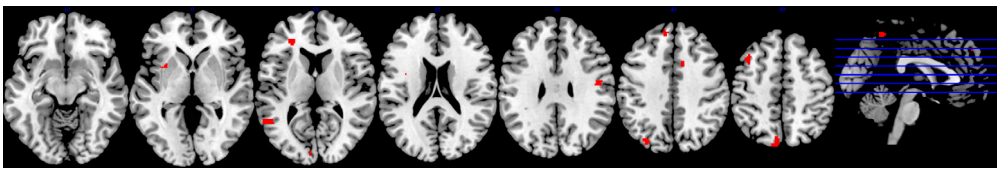
Then 100,000 MCMC samples were used to perform the statistical inferences regarding activation detection. The statistics of interest were registered into standard templates like MNI152, for ease of visualization. We compared the two activation maps: with and without consideration of spatial dependences between temporal correlations. Although 2nd order was considered for temporal auto-correlation, the result was barely different from that with 1st order. Therefore, we used 1st order temporal auto-correlation throughout the real example analysis.



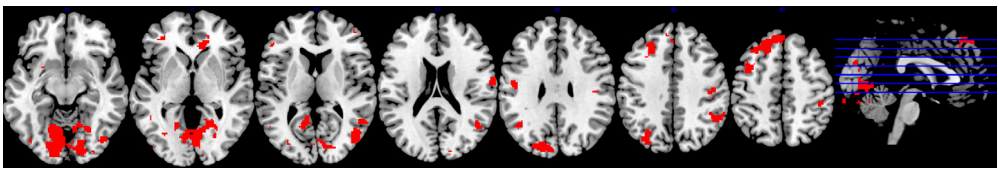
(a) Right congruence.



(b) Right in-congruence.



(c) Left congruence.



(d) Left in-congruence.

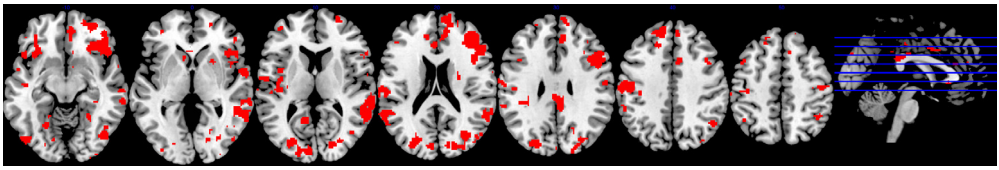
Fig. 3. The activation areas corresponding to the posterior probability greater than a threshold determined by FDR shown in red for different tasks. Spatial dependence of temporal correlations is considered in the model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Activation detection

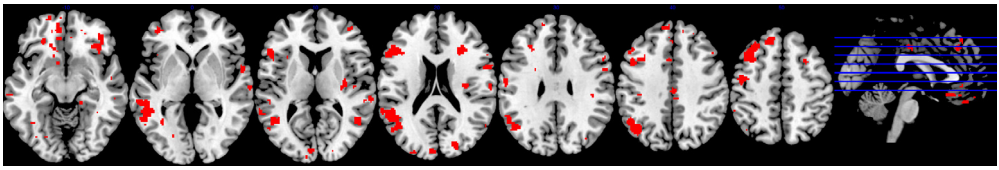
In this section, we present the posterior activation maps for different stimuli. A voxel is considered to be activated by the stimulus j when the posterior probability of γ_{vj} is greater than c. Here c is either a deterministic value (0.8722) (Smith and Fahrmeir, 2007; Lee et al., 2014) or it is determined by FDR (0.05) (Kalus et al., 2014). However, in the event-related fMRI simulation, 0.8722 performed in a conservative manner, giving way to 0.05. The activation maps for different tasks are shown in Fig. 3 with consideration of the spatial dependence of temporal correlations; and in Fig. 4 without consideration of the spatial dependence. More activated areas – especially for the task of "Left Incongruence" – are present without the consideration of spatial dependence of temporal correlations in Fig. 4(d), but this might be false positive. Possibly the statistical model failed to catch some spatial dependence existing among temporal correlations.

4.2. Contrast effect

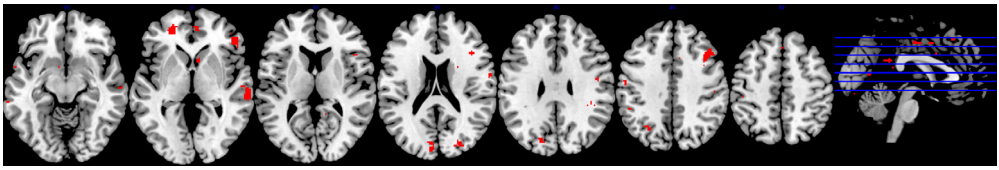
One interesting question about the standard Simon fMRI study is, which brain areas are related to the inhibitory control that resolves response conflict between multiple concurrently activated



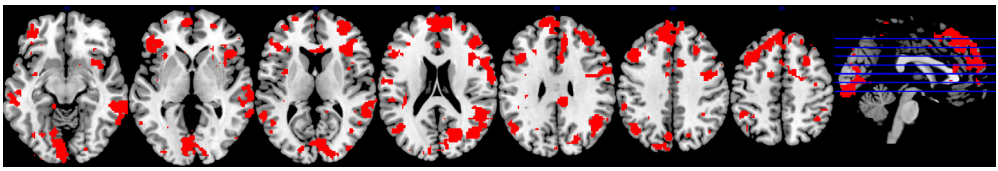
(a) Right congruence.



(b) Right in-congruence.



(c) Left congruence.



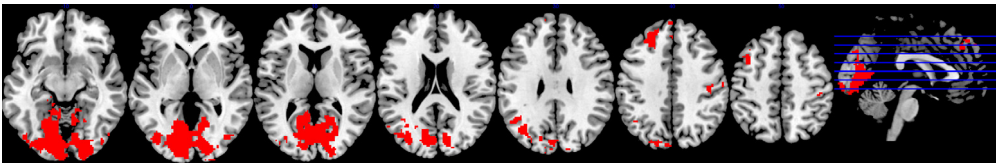
(d) Left in-congruence.

Fig. 4. The activation areas corresponding to the posterior probability greater than a threshold determined by FDR shown in red for different tasks. Spatial dependence of temporal correlations is not considered in the model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

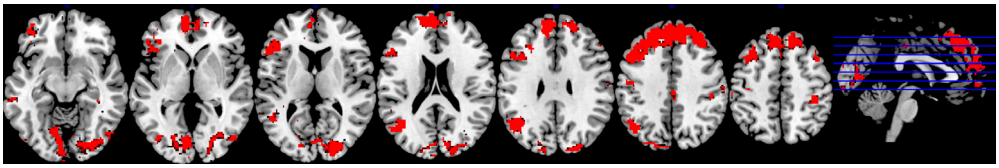
responses. To answer it, we have to determine which brain regions are significantly more active in in-congruent than in congruent trials. This corresponds to determine if the posterior probability

$$p\left(\beta_{\text{right-incon}} + \beta_{\text{left-incon}} > \beta_{\text{right-con}} + \beta_{\text{left-con}} \mid \boldsymbol{y}\right) \tag{7}$$

is greater than a threshold. The posterior probability images obtained by (7) with and without considering spatial dependence of temporal coefficients are shown in Figs. 5a and 5b, respectively. The active regions include frontal and occipital lobes detected by the models with and without considering spatial dependence of temporal correlations. However, there seems to be more activity in the medial prefrontal cortex in Fig. 5b without considering spatial dependence . For the Simon task, it is expected that the lateral prefrontal cortex be activated as part of the cognitive control network and the multiple demand network. However, the medial prefrontal cortex is part of the default mode network (Niendam et al., 2012). Therefore, it is possible that the activation in the medial part belongs to false positives. This is an important example that illustrates an advantage of our approach. Namely, considering spatial dependence of temporal correlations improves the posterior activation by reducing spurious detections.



(a) Posterior activation when the in-congruent conditions are greater than congruent conditions *with* consideration of spatial dependence in temporal parameters.



(b) Posterior activation when the in-congruent conditions are greater than congruent conditions *without* consideration of spatial dependence in temporal parameters.

Fig. 5. Areas show more activity in in-congruent than congruent condition. The threshold is determined by FDR.

5. Discussion and conclusions

In this work we applied a new approach for performing Bayesian variable selection with consideration of spatial dependence among activation coefficients and among 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 spatial structures, parcellations, and experimental designs. In addition, the proposed approach decreased the false positive rate and incorporated the prior information about brain functions for subject-level inference.

In addition to local spatial dependence of temporal correlations, we can also use nonparametric Bayesian variable selection (Zhang et al., 2014) to extend the proposed model to detect neural activation in response to a stimulus, and to investigate the associations or clustering—of spatially remote voxels exhibiting fMRI time series behavior with similar characteristics. For completeness, we will also consider a voxel-dependent HRF in the model, and use SGLMM to capture the dependence among nearby voxels. To achieve the inferred association between voxel time courses, we plan to group the voxels by applying a hierarchical Dirichlet process (HDP) prior on the parameters of temporal correlations. In practice, this may be more suitable to model the spatial pattern of temporal correlations and may further increase the accuracy of brain activation detection.

Acknowledgments

Research was supported by the Ministry of Science and Technology (MOST), Taiwan (MOST 102-2420-H-006-006-MY2, 104-2118-M-006-003-). In addition, 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). In the end, we thank the Mind Research and Imaging Center (MRIC) at NCKU for consultation and instrument availability. MRIC is supported by MOST.

Appendix. Posterior distribution and full conditionals

For the gth parcel, the temporal correlation of signal changes over time is modeled by an autoregression model with order r. For a given voxel v, the signal change at time t is assumed to follow a regression model as follows

$$y_{vt} = \boldsymbol{z}_{vt}\boldsymbol{\delta}_v + \boldsymbol{x}_{vt}\boldsymbol{\beta}_v + e_{vt}\,;$$
$$e_{vt} = \rho_1 e_{t-1} + \cdots + \rho_r e_{t-r} + \epsilon_{vt}\,,$$

where $\boldsymbol{\beta}_v = (\beta_1, \ldots, \beta_p)'$, $\boldsymbol{\delta}_v = (\delta_1, \ldots, \delta_q)'$, $\boldsymbol{x}_{vt} = (x_{t1}, \ldots, x_{tp})$, $\boldsymbol{z}_{vt} = (z_{t1}, \ldots, z_{tq})$, and ϵ_{vt} is normally distributed with a mean 0 and variance σ_v^2. Since $e_{vt} = y_{vt} - \boldsymbol{z}_{vt}\boldsymbol{\delta}_v - \boldsymbol{x}_{vt}\boldsymbol{\beta}_v$, we can rewrite the model as

$$\begin{aligned} y_{vt} &= \boldsymbol{z}_{vt}\boldsymbol{\delta}_v + \boldsymbol{x}_{vt}\boldsymbol{\beta}_v + \rho_1(y_{v,t-1} - \boldsymbol{z}_{v,t-1}\boldsymbol{\delta}_v - \boldsymbol{x}_{v,t-1}\boldsymbol{\beta}_v) + \cdots \\ &\quad + \rho_r(y_{v,t-r} - \boldsymbol{z}_{v,t-r}\boldsymbol{\delta}_v - \boldsymbol{x}_{v,t-r}\boldsymbol{\beta}_v) + \epsilon_{vt} \\ &= \boldsymbol{z}_{vt}\boldsymbol{\delta}_v + \boldsymbol{x}_{vt}\boldsymbol{\beta}_v + L_v\boldsymbol{\rho}_v + \epsilon_{vt}, \end{aligned}$$

where $L_v = (y_{v,t-1} - \boldsymbol{z}_{v,t-1}\boldsymbol{\delta}_v + \boldsymbol{x}_{v,t-1}\boldsymbol{\beta}_v, \ldots, y_{v,t-r} - \boldsymbol{z}_{v,t-r}\boldsymbol{\delta}_v - \boldsymbol{x}_{v,t-r}\boldsymbol{\beta}_v)$ and $\boldsymbol{\rho}_v = (\rho_1, \ldots, \rho_r)'$.

Given the representation of the proposed model above, the posterior distribution of parameters can be written as

$$p(\boldsymbol{\theta}, \Theta_g|\boldsymbol{y}_v) \propto L(\boldsymbol{\theta}_v|\boldsymbol{y}_v)\pi(\boldsymbol{\theta}_v, \Theta_g),$$

where $\boldsymbol{y}_v$ is the time series data of the voxel v , $\boldsymbol{\theta}_v = (\boldsymbol{\beta}_v, \boldsymbol{\delta}_v, \boldsymbol{\rho}_v, \boldsymbol{\gamma}_v, \sigma_v^2)$ lists the voxel-level, and the parcel-level parameters are $\Theta_g = (\boldsymbol{\phi}, \kappa, \tau^2, \lambda^2, \boldsymbol{\varphi}, \omega)$. Within the gth parcel with n_g voxels, we assume

$$\begin{aligned} \pi(\boldsymbol{\theta}_v, \Theta_g) = &\prod_{v=1}^{n_g} \pi(\boldsymbol{\delta}_v)\pi(\boldsymbol{\rho}_v)\pi(\sigma_v^2)\pi(\boldsymbol{\beta}_v|\boldsymbol{\gamma}_v)\prod_{j=1}^{p} \pi(\boldsymbol{\gamma}_j|\boldsymbol{\phi}_j)\pi(\boldsymbol{\phi}_j|\kappa_j)\pi(\kappa_j)\pi(\tau_j^2) \\ &\times \prod_{k=1}^{r} \pi(\lambda_k)\pi(\boldsymbol{\varphi}_k|\omega_k)\pi(\omega_k). \end{aligned}$$

The likelihood for the vth voxel is given by

$$L(\boldsymbol{\theta}_v|\boldsymbol{y}_v) = \left(\frac{1}{2\pi\sigma_v^2}\right)^{-T/2} \exp\left\{-\frac{(\boldsymbol{y}_v - Z_v\boldsymbol{\delta}_v - X_v\boldsymbol{\beta}_v - L_v\boldsymbol{\rho}_v)'(\boldsymbol{y}_v - Z_v\boldsymbol{\delta}_v - X_v\boldsymbol{\beta}_v - L_v\boldsymbol{\rho}_v)}{2\sigma_v^2}\right\}.$$

Given the posterior distribution, we derive all conditional distributions of the parameters by Gibbs sampling. Note that in order to alleviate the notational burden, we drop v unless necessary.

- Now we begin with the derivation of the conditional distribution of β_j.

$$\begin{aligned} p(\beta_j|\boldsymbol{y}) &\propto \exp\left\{-\frac{\sum_t(y_t - \boldsymbol{z}_t\boldsymbol{\delta} - \boldsymbol{x}_t\boldsymbol{\beta} - \boldsymbol{l}_t\boldsymbol{\rho})^2}{2\sigma^2}\right\} \exp\left\{-\frac{\beta_j^2}{2c_j^{2r_j}\tau_j^2}\right\} \\ &\propto \exp\left\{-\frac{\sum_t(y_t - \boldsymbol{z}_t\boldsymbol{\delta} - \boldsymbol{x}_{t(j)}\boldsymbol{\beta}_{(j)} - x_{tj}\beta_j - \boldsymbol{l}_t\boldsymbol{\rho})^2}{2\sigma^2}\right\} \exp\left\{-\frac{\beta_j^2}{2c_j^{2r_j}\tau_j^2}\right\} \\ &\propto \exp\left\{-\frac{\sum_t(y_t - \boldsymbol{z}_t\boldsymbol{\delta} - \boldsymbol{x}_{t(j)}\boldsymbol{\beta}_{(j)} - \tilde{x}_{tj}\beta_j - \boldsymbol{l}_{\beta t(j)}\boldsymbol{\rho})^2}{2\sigma^2}\right\} \exp\left\{-\frac{\beta_j^2}{2c_j^{2r_j}\tau_j^2}\right\} \\ &\propto \exp\left\{-\frac{\sum_t(d_{tj} - \tilde{x}_{tj}\beta_j)^2}{2\sigma^2}\right\} \exp\left\{-\frac{\beta_j^2}{2c_j^{2r_j}\tau_j^2}\right\} \\ &\propto \exp\left\{-\frac{\beta_j^2(c_j^{2r_j}\tau_j^2\sum_t\tilde{x}_{tj}^2 + \sigma^2) - 2c_j^{2r_j}\tau_j^2\beta_j\sum_t d_{tj}\tilde{x}_{tj}}{2\sigma^2 c_j^{2r_j}\tau_j^2}\right\}, \end{aligned}$$

where $\boldsymbol{l}_t$ is the tth row of L, $\boldsymbol{\beta}_{(j)}$ is vector $\boldsymbol{\beta}$ excluding β_j, $\boldsymbol{x}_{t(j)}$ is vector $\boldsymbol{x}_t$ excluding x_{tj}, $\boldsymbol{l}_{\beta t(j)} = (y_{t-1} - \boldsymbol{z}_{t-1}\boldsymbol{\delta} - \boldsymbol{x}_{t-1,(j)}\boldsymbol{\beta}_{(j)}, \ldots, y_{t-r} - \boldsymbol{z}_{t-r}\boldsymbol{\delta} - \boldsymbol{x}_{t-r,(j)}\boldsymbol{\beta}_{(j)})$, $d_{tj} = y_t - \boldsymbol{z}_t\boldsymbol{\delta} - \boldsymbol{x}_{t(j)}\boldsymbol{\beta}_{(j)} - \boldsymbol{l}_{\beta t(j)}(r)\boldsymbol{\rho}$, and $\tilde{x}_{tj} = x_{tj} - \rho_1 x_{t-1,j} - \cdots - \rho_r x_{t-r,j}$. Therefore, the conditional distribution of β_j is

$$\beta_j \sim N\left(\hat{\beta}_j, \hat{\sigma}_{\beta_j}^2\right),$$

where

$$\hat{\beta}_j = \frac{c_j^{2r_j}\tau_j^2 \sum_t d_{tj}\tilde{x}_{tj}}{c_j^{2r_j}\tau_j^2 \sum_t \tilde{x}_{tj}^2 + \sigma^2};$$
$$\hat{\sigma}^2_{\beta_j} = \frac{\sigma^2 c_j^{2r_j}\tau_j^2}{c_j^{2r_j}\tau_j^2 \sum_t \tilde{x}_{tj}^2 + \sigma^2}.$$

- Now we derive the conditional distribution of δ_q as

$$\begin{aligned}
p(\delta_q|y) &\propto \exp\left\{-\frac{\sum_t (y_t - \boldsymbol{z}_t\boldsymbol{\delta} - \boldsymbol{x}_t\boldsymbol{\beta} - \boldsymbol{l}_t\boldsymbol{\rho})^2}{2\sigma^2}\right\}\exp\left\{-\frac{\delta_j^2}{2s_\delta^2}\right\} \\
&\propto \exp\left\{-\frac{-\sum_t (y_t - \boldsymbol{z}_{t(j)}\boldsymbol{\delta}_{(j)} - z_{tj}\delta_j - \boldsymbol{x}_t\boldsymbol{\beta} - \boldsymbol{l}_t\boldsymbol{\rho})^2}{2\sigma^2}\right\}\exp\left\{-\frac{\delta_j^2}{2s_\delta^2}\right\} \\
&\propto \exp\left\{-\frac{-\sum_t (y_t - \boldsymbol{z}_{t(j)}\boldsymbol{\delta}_{(j)} - \tilde{z}_{tj}\delta_j - \boldsymbol{x}_t\boldsymbol{\beta} - \boldsymbol{l}_{\delta t(j)}\boldsymbol{\rho})^2}{2\sigma^2}\right\}\exp\left\{-\frac{\delta_j^2}{2s_\delta^2}\right\} \\
&\propto \exp\left\{-\frac{-\sum_t (d_{\delta tj} - \tilde{z}_{tj}\delta_j)^2}{2\sigma^2}\right\}\exp\left\{-\frac{\delta_j^2}{2s_\delta^2}\right\} \\
&\propto \exp\left\{-\frac{\delta_j^2(s_\delta^2 \sum_t \tilde{z}_{tj}^2 + \sigma^2) - 2\delta_j s_\delta^2 \sum_t d_{\delta tj}\tilde{z}_{tj}}{2\sigma^2 s_\delta^2}\right\},
\end{aligned}$$

where $\boldsymbol{\delta}_{(q)}$ is vector $\boldsymbol{\delta}$ excluding δ_q, $\boldsymbol{z}_{t(q)}$ is vector $\boldsymbol{z}_t$ excluding z_{tq}, $\boldsymbol{l}_{\delta t(q)} = (y_{t-1} - \boldsymbol{z}_{t-1(q)}\boldsymbol{\delta}_{(q)} - \boldsymbol{x}_{t-1}\boldsymbol{\beta}, \ldots, y_{t-r} - \boldsymbol{z}_{t-r(q)}\boldsymbol{\delta}_{(q)} - \boldsymbol{x}_{t-r}\boldsymbol{\beta})$, $d_{\delta tj} = y_t - \boldsymbol{z}_{t(j)}\boldsymbol{\delta}_{(q)} - \boldsymbol{x}_t\boldsymbol{\beta} - \boldsymbol{l}_{\delta t(q)}\boldsymbol{\rho}$, and $\tilde{z}_{tj} = z_{tj} - \rho_1 z_{t-1,j} - \cdots - \rho_r z_{t-r,j}$. Therefore, the conditional distribution of δ_j is

$$\delta_j \sim N\left(\hat{\delta}_j, \hat{\sigma}^2_{\delta_j}\right),$$

where

$$\hat{\delta}_j = \frac{s_\delta^2 \sum_t d_{\delta tj}\tilde{z}_{tj}}{s_\delta^2 \sum_t \tilde{z}_{tj}^2 + \sigma^2};$$
$$\hat{\sigma}^2_{\delta_j} = \frac{\sigma^2 s_\delta^2}{s_\delta^2 \sum_t \tilde{z}_{tj}^2 + \sigma^2}.$$

- The conditional distribution of σ^2 is

$$p(\sigma^2|y) \propto \left(\frac{1}{\sigma^2}\right)^{(T-r+a_\sigma)/2+1} \exp\left\{-\frac{\sum_t (y_t - \boldsymbol{z}_t\boldsymbol{\delta} - \boldsymbol{x}_t\boldsymbol{\beta} - \boldsymbol{l}_t\boldsymbol{\rho})^2 + b_\sigma}{2\sigma^2}\right\}$$

$$\sigma^2 \sim \mathcal{IG}\left(\frac{T-r+a_\sigma}{2}, \frac{\sum_t (y_t - \boldsymbol{z}_t\boldsymbol{\delta} - \boldsymbol{x}_t\boldsymbol{\beta} - \boldsymbol{l}_t\boldsymbol{\rho})^2 + b_\sigma}{2}\right).$$

- The conditional distribution of τ_j^2 is

$$p(\tau_j^2|y) \propto \left(\frac{1}{\tau_j^2}\right)^{(n_g+a_c)/2+1} \exp\left\{-\frac{1}{2\tau_j^2}\sum_v \frac{\beta_{vj}^2}{c_j^{2\gamma_{vj}}} - \frac{b_\tau}{2\tau_j^2}\right\}$$

$$\tau_j^2 \sim \mathcal{IG}\left(\frac{n_g + a_\tau}{2}, \frac{\sum_v \frac{\beta_{vj}^2}{c_j^{2\gamma_{vj}}} + b_\tau}{2}\right).$$

- The conditional distribution of κ_j is

$$\kappa_j \sim \Gamma\left(\frac{q_g + a_\kappa}{2}, \frac{\boldsymbol{\phi}_j' M_g' Q_g M_g \boldsymbol{\phi}_j + b_\kappa}{2}\right),$$

  where q_g is the rank of M_g.
- The conditional distribution of γ_j is

$$p(\gamma_j|y) \propto \eta_j^{\gamma_j}(1-\eta_j)^{1-\gamma_j}\left(\frac{1}{c_j^{2\gamma_j}\tau_j^2}\right)^{1/2} \exp\left\{-\frac{\beta_{vj}^2}{2c_j^{2\gamma_{vj}}\tau_j^2}\right\}.$$

  We have

$$p(\gamma_j = 1|y) \propto \eta_{vj}\left(\frac{1}{c_j^2}\right)^{1/2} \exp\left\{-\frac{\beta_{vj}^2}{2c_j^2\tau_j^2}\right\};$$

$$p(\gamma_j = 0|y) \propto (1-\eta_{vj}) \exp\left\{-\frac{\beta_{vj}^2}{2\tau_j^2}\right\}.$$

  Therefore,

$$p(\gamma_j = 1|y) = \frac{1}{1+h_j},$$

  where $h_j = \exp\left\{-\alpha_j - \boldsymbol{m}'\boldsymbol{\phi}_j + \frac{\log(c_j^2)}{2} - \frac{\beta_j^2(c_j^2-1)}{2c_j^2\tau_j^2}\right\}$.
- The conditional distribution of $\boldsymbol{\phi}_j$ is

$$p(\boldsymbol{\phi}_j|y) \propto \exp\left\{-\frac{\boldsymbol{\phi}_j'\kappa_j M_g' Q_g M_g \boldsymbol{\phi}_j}{2}\right\} \prod_v \eta_{vj}^{\gamma_{vj}}(1-\eta_j)^{1-\gamma_{vj}}$$

$$\propto \exp\left\{-\frac{\boldsymbol{\phi}_j'\kappa_j M_g' Q_g M_g \boldsymbol{\phi}_j}{2} + \sum_v \gamma_{vj}(\alpha_{vj} + \boldsymbol{m}_v'\boldsymbol{\phi}_j) - \sum_v \log(1+\exp\{\boldsymbol{m}_v'\boldsymbol{\phi}_j\})\right\}.$$

- The conditional distribution of λ_r^2 is

$$\lambda_r^2 \sim \mathcal{IG}\left(\frac{n_g + a_\lambda}{2}, \frac{\sum_v(\tilde{\rho}_{vr} - \boldsymbol{m}_v'\boldsymbol{\varphi}_r)^2 + b_\lambda}{2}\right).$$

- The conditional distribution of ω_r is

$$\omega_r \sim \Gamma\left(\frac{q_g + a_\omega}{2}, \frac{\boldsymbol{\varphi}_r' M_g' Q_g M_g \boldsymbol{\varphi}_r + b_\omega}{2}\right).$$

- The conditional distribution of $\boldsymbol{\rho}_v$

$$\boldsymbol{\rho}_v \sim N\left((L_v'L_v + \sigma_v^2\Lambda_v^{-1})^{-1}(A_v'L_v + \sigma_v^2\tilde{\boldsymbol{\varphi}}_v\boldsymbol{m}_v\Lambda_v), \sigma_v^2(L_v'L_v + \sigma_v^2\Lambda_v^{-1})\right),$$

  where $A_v = \boldsymbol{y}_v - Z_v\boldsymbol{\delta}_v - X_v\boldsymbol{\beta}_v$, $\tilde{\boldsymbol{\varphi}}_v = \left(\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_r\right)'$, and $\boldsymbol{\rho}_v$ satisfies the condition that the roots of the polynomial $a^r - \sum_{i=1}^r \rho_i a^{r-i}$ lie within the unit circle for the stationary property of the $AR(r)$.

- The conditional distribution of $\boldsymbol{\varphi}_r$ is

$$p(\boldsymbol{\varphi}_r|y) \propto \exp\left\{-\frac{\boldsymbol{\varphi}_r'\omega_r M_g' Q_g M_g \boldsymbol{\varphi}_r}{2}\right\} \exp\left\{-\sum_v \frac{(\rho_{vr} - \boldsymbol{m}_v'\boldsymbol{\varphi}_r)^2}{2\lambda_r^2}\right\}.$$

That is,

$$\varphi_r \sim N\left[\left(\sum_v \boldsymbol{m}_v\boldsymbol{m}_v' + \lambda_r^2\omega_r M_g' Q_g M_g g\right)^{-1} \times \sum_v \rho_{vr}\boldsymbol{m}_v, \lambda^2\left(\sum_v \boldsymbol{m}_v\boldsymbol{m}_v' + \lambda_r^2\omega_r M_g' Q_g M_g\right)^{-1}\right].$$

References

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. NeuroImage 5, 199–212.

Arbabshirani, M.R., Damaraju, E., Phlypo, R., Plis, S., Allen, E., Ma, S., Mathalon, D., Preda, A., Vaidya, J.G., Adali, T., Calhoun, V.D., 2014. Impact of autocorrelation on functional connectivity. NeuroImage 102, Part 2, 294–308.

Barbieri, M., Berger, J.O., 2004. Optimal predictive model selection. Ann. Statist. 32, 870–897.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 289–300.

Bezener, M., Hughes, J., Jones, G.L., 2017. Bayesian spatiotemporal modeling using hierarchical spatial priors with applications to functional magnetic resonance imaging. In: Härdle, W.K., Lu, H.H., Shen, X. (Eds.), Handbook of Big Data Analytics. Springer International Publishing.

Bowman, F., Caffo, B., Bassett, S., Kilts, C., 2008. A Bayesian hierarchical framework for spatial modeling of fMRI data. NeuroImage 39, 146–156.

Caffo, B., Bowman, F., Eberly, L., Bassett, S., 2011. A Markov chain Monte Carlo based analysis of a multilevel model for functional MRI data. In: Brooks, S., Gelman, A., Jones, G., Meng, X.-L. (Eds.), Handbook of Markov Chain Monte Carlo. CRC Press, Boca Raton, FL.

Chen, G., Saad, Z.S., Nath, A.R., Beauchamp, M.S., Cox, R.W., 2012. FMRI group analysis combining effect estimates and their variances. NeuroImage 60, 747–765.

Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–114.

Detre, J.A., Floyd, T., 2001. Functional MRI and its applications to the clinical neurosciences. Neuroscientist 7, 64–79.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci..

FMRIB, 2015. Oxford, U. Fmrib Software Library.

Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D., 2007. Statistical Parametric Mapping: The Analysis of Funtional Brain Images. Elsevier/Academic Press.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Ruggb, M., Turner, R., 1998. Event-related fMRI: characterizing differential responses. Neuroimage 7, 30–40.

Gelman, A., Meng, X.-L., 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Statist. Sci. 13, 163–185.

Genovese, C., 2000. A Bayesian time-course model for functional magnetic resonance imaging data. J. Amer. Statist. Assoc. 95, 691–703.

George, E.I., McCulloch, R.E., 1993. Variable selection via Gibbs sampling. J. Amer. Statist. Assoc. 88, 881–889.

George, E.I., McCulloch, R.E., 1997. Approaches for Bayesian variable selection. Statist. Sinica 7, 339–374.

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

Goldsmith, J., Huang, L., Crainceanu, C.M., 2014. Smooth scalar-on-image regression via spatial Bayesian variable selection. J. Comput. Graph. Statist. 23, 46–64.

Gössl, C., Fahrmeir, L., Auer, D.P., 2001. Bayesian spatiotemporal inference in functional magnetic resonance images. Biometric 57, 554–562.

Hughes, J., Haran, M., 2013. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J. R. Stat. Soc. Ser. B Stat. Methodol. 75, 139–159.

Kalus, S., Sämann, P.G., Fahrmeir, L., 2014. Classification of brain activation via spatial Bayesian variable selection in fMRI regression. Adv. Data Anal. Classif. 8, 63–83.

Kelshall, J.E., Wakefield, J.C., 1999. Discussion of "Bayesian models for spatially correlated disease and exposure data". In: Best, N.G., Waller, L., Thomas, A., Conlon, E.M., Arnold, R., Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 6. Oxford University Press, pp. 147–150.

Lee, K., Jones, G.L., Caffo, B., Bassett, S., 2014. Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. Bayesian Anal. 9, 699–732.

Lindquist, M.A., Atlas, J.M.L.L.Y., Wager, T.D., 2009. Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. NeuroImage 45, S187–198.
Liu, X., Banich, M., Jacobson, B., Tanabe, J., 2004. Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. Neuroimage 22, 1097–1106.
Lund, T.E., Madsen, K.H., Sidaros, K., Luo, W.L., Nichols, T.E., 2006. Non-white noise in fMRI: Does modelling have an impact?. NeuroImage 29, 54–66.
Musgrove, D.R., Hughes, J., Eberly, L.E., 2016. Fast, fully Bayesian spatiotemporal inference for fMRI data. Biostatistics 17, 291–303.
Newton, M.A., Noueiry, A., Sarkar, D., Ahlquist, P., 2004. Detecting differential gene expression with a semiparametric hierarchical mixture method. Biostatistics 5, 155–176.
Niendam, T.A., Laird, A.R., Ray, K.L., Dean, Y.M., Glahn, D.C., Carter, C.S., 2012. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. Cogn. Affect. Behav. Neurosci. 12, 241–268.
Penny, W., Kiebel, S., Friston, K., 2003. Variational Bayesian inference for fMRI time series. NeuroImage 19, 727–741.
Penny, W.D., Trujillo-Barreto, N.J., Friston, K.J., 2005. Bayesian fMRI time series analysis with spatial priors. NeuroImage 24, 350–362.
Quirós, A., Diez, R.M., Gamerman, D., 2010. Bayesian spatiotemporal model of fMRI data. NeuroImage 49, 442–456.
Rajapakse, J.C., Kruggel, F., Cramon, J.M.M.D.Y., 1998. Modeling hemodynamic response for analysis of functional MRI time-series. Hum. Brain Mapp. 6, 283–300.
Smith, M., Fahrmeir, L., 2007. Spatial Bayesian variable selection with application to functional magnetic resonance imaging. J. Amer. Statist. Assoc. 102, 417–431.
Smith, M., Pütz, B., Auer, D., Fahrmeirc, L., 2003. Assessing brain activity through spatial Bayesian variable selection. NeuroImage 20, 802–815.
Wen, T., Hsieh, S., 2015. Neuroimaging of the joint Simon effect with believed biological and non-biological co-actors. Front. Hum. Neurosci. 9, 1–13.
Woolrich, M., Jenkinson, M., Brady, J., Smith, S., 2004. Fully Bayesian spatio-temporal modeling for fMRI data. IEEE Trans. Med. Imaging 23, 213–231.
Xia, J., Liang, F., Wang, Y.M., 2009. FMRI analysis through Bayesian variable selection with a spatial prior. In: IEEE Int. Symp. on Biomedical Imaging (ISBI), pp. 714–717.
Zhang, L., Guindani, M., Vannucci, M., 2014. A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. NeuroImage 95, 162–175.
Zhang, L., Guindani, M., Vannucci, M., 2015. Bayesian models for functional magnetic resonance imaging data analysis. Wiley Interdiscip. Rev. Comput. Stat. 7, 21–41.