Bayesian Statistics and Its Applications
Edited by S.K. Upadhyay, U. Singh and D.K. Dey
Anamaya Publishers, New Delhi, India

Bayesian Hierarchical Spatio-Temporal Analysis of fMRI Data: A Case Study

Sandeep Basak[1] **and Sumitra Purkayastha**[2]

[1]Upper Ground Floor, Bay # 4, GE Towers, Sector Road
Sector 53, DLF Phase 5, Gurgaon-122 002, India

[2]Theoretical Statistics and Mathematics Unit, Indian Statistical Institute
203 B.T. Road, Kolkata-700 108, India

Abstract

Analysis of fMRI data has traditionally proceeded along the direction of deciding separately if each voxel is activated by a stimulus. Linear model formulation of this problem takes into account the temporal dependence that is likely to be present between fMR observations. However, associated spatial and spatio-temporal dependence have been receiving substantial attention for the last few years. Hierarchical Bayesian modeling appears to be quite useful in such studies. Some relevant references, by no means exhaustive, are Genovese (1999, 2000), Gössl et al. (2000, 2001a), Penny et al. (2005), Woolrich et al. (2004a, b, 2005). The approach proposed in Gössl et al. (2000, 2001a) appear to be relatively simpler than those proposed in the others, even though it must be said that there is considerable flexibility in them (in Genovese, 1999, for example). The work reported in this paper has been inspired by Gössl et al. (2000, 2001a). We have employed some of the models proposed in Gössl et al. (2001a) on a data-set studied by Worsley (Worsley et al., 2002; Worsley; 2003, http://www.math.mcgill.ca/keith/fmristat). This has been pursued with a view to seeing how well do the models proposed in Gössl et al. (2001a) perform when employed on a data-set where non-Bayesian method has already been tried. The data-set studied by Prof. Worsley came in handy for this purpose. We wanted to see to what extent do the results obtained by employing some of the models in Gössl et al. (2001a) agree to those obtained by Prof. Worsley and also wanted to have some understanding about the computational aspects, the running time, in particular. We have used the results obtained by Prof. Worsley as the benchmark. Our results agree to a large extent with those obtained by Prof. Worsley. Our computations reveal certain important aspects. In particular, for reducing computing time, introducing some simplifications seemed unavoidable.

1. Introduction

Functional magnetic resonance imaging (fMRI) is a technique to determine which parts of the brain are activated by different types of physical sensation or activity, such as sight, sound or the movement

AMS 2000 Subject Classifications: 62F15; 62P10; 62M99.

of a subject's fingers. In functional brain imaging studies, the signal changes of interest are caused by the neuronal activity, but such electrical activity is not directly detectable by fMRI. The signal being measured in fMRI experiments is called the *blood oxygenation level dependent* (BOLD) response, which is a consequence of the hemodynamic changes (including local changes in blood flow, volume and oxygenation level) occurring within a few seconds of changes in neuronal activity induced by external stimuli. The BOLD signal is usually used as a proxy for the underlying neuronal activity. Most fMRI studies concern the detection of sites of activation ('hot-spots') in the brain and their relationship to the experimental stimulation and we refer to these studies as the *activation* studies. Interested readers can refer to Clare (1997), Frackowiak et al. (2004), and Jezzard et al. (2001) for the details on the theory and application of fMRI in neurology, neuroscience, psychology and psychiatry.

Analysis of fMRI data usually proceeds by modelling the BOLD signal as the dependent variable in a linear model where suitable transform(s) of the external stimulus (stimuli), obtained by convolving the stimulus (stimuli) with the hemodynamic response function (HRF) (see, e.g., Glover, 1999), has (have) played the role of the independent variable(s). The associated error distribution is assumed to be an autoregressive process of an appropriate order. The question whether a voxel is activated by one of the stimuli being employed is posed as one of testing whether the corresponding regression coefficient is greater than zero against being equal to zero (see, e.g., Worsley et al. (2002), Worsley (2003)). Some important references on statistical analysis of fMRI data and related issues are Bullmore et al. (1996), Friston et al. (1995), Lange (1996), Lazar et al. (2001), Worsley and Friston (1995) etc. and this list is by no means exhaustive.

The approach described above takes into account the temporal dependence that is likely to be present between the BOLD signals. However, this does not take into account the spatial dependence that is likely to be present between the BOLD signals. Possible spatio-temporal dependence is also ignored. On the other hand, it is only natural to demand that our analysis should take these aspects into account. Motivated by these, several authors have tried to incorporate spatial and spatio-temporal aspects both into modeling of fMRI data and subsequent stage of inference. Some relevant references are Genovese (1999, 2000), Gössl et al. (2000, 2001a), Penny et al. (2005), Woolrich et al. (2004a, 2005). All these works are based on a Bayesian formulation. Moreover, hierarchical modeling is seen to be quite useful in this problem. For a detailed discussion on hierarchical modeling and associated Bayesian methods, we refer the reader to Banerjee et al. (2004). It needs to be mentioned here that apart from the purpose with which a Bayesian formulation has been employed, several researchers have also employed Bayesian formulation for the purpose of estimating the HRF. Some relevant references are Gössl et al. (2001b), Marrelec et al. (2003), Woolrich et al. (2004b). We would like to add here that Bayesian approach has received substantial attention in analysis of fMRI data and related issues and the references cited above are by no means exhaustive. Some relevant non-Bayesian references are Polzehl and Spokoiny (2001), Purdon et al. (2001) and Solo et al. (2001).

The work reported in this paper has been inspired by Gössl et al. (2001a) (henceforth, abbreviated to GAF) and also by Gössl et al. (2000). The approach proposed in GAF appears to be relatively simpler than those proposed in the others, even though it must be said that there is considerable flexibility in them (in Genovese, 1999, for example). We have employed some of the models proposed in GAF on a data-set studied by Worsley (Worsley, et al. 2002; Worsley, 2003; http://www.math.mcgill.ca/keith/fmristat) with a view to seeing to what extent do the results obtained by employing some of the models in GAF agree to those obtained by Prof. Worsley and

also to have some understanding about the computational aspects, about the running time, in particular. We describe the experiment we have studied and the data-set we have analysed in the next two paragraphs.

The aim of the experiment has been to identify areas related to pain perception. We shall see that the experiment allows one to locate separately areas activated by painful heat stimulus and warm stimulus also. During the experiment, the subject was exposed to the following two stimuli: (a) a painful heat stimulus (49°C) to the left forearm and (b) a warm stimulus (35°C). The experiment began with a period of rest for 9 secs which was followed by each stimulus being alternately administered for 9 secs, intercepted by 9 secs of rest. The whole cycle was repeated 10 times for 6 mins. in total with TR = 3 secs. In other words, each type of scan lasted for 3 secs., and was replicated thrice.

The data-set consists of 13 slices of 128 × 128 images, for each scan. For every scan, the slices are interleaved, i.e. first images of the odd-numbered slices are scanned and then those of the even-numbered slices are obtained. Hence, there are 120 images corresponding to each slice. The work reported in this paper has been pursued with only the *third* slice.

Section 2 discusses the models of GAF we have considered and employed in our context. Section 3 describes our goals and also the methods we have employed. Section 4 pertains to certain issues associated with implementation of the methods. Section 5 presents the results and offers some relevant discussion. Section 6 concludes with some discussion on issues worth exploring in future.

2. Our Models

Suppose, for ease in notation, the *hot* stimulus is the first one and the *warm* stimulus is the second one. Denote by $s_j(t)$, the function representing the jth ($j = 1, 2$) stimulus. This function is usually taken to be a zero-one function (a *box-car* function). Fig. 1 shows the function $s_1(t)$. Its definition

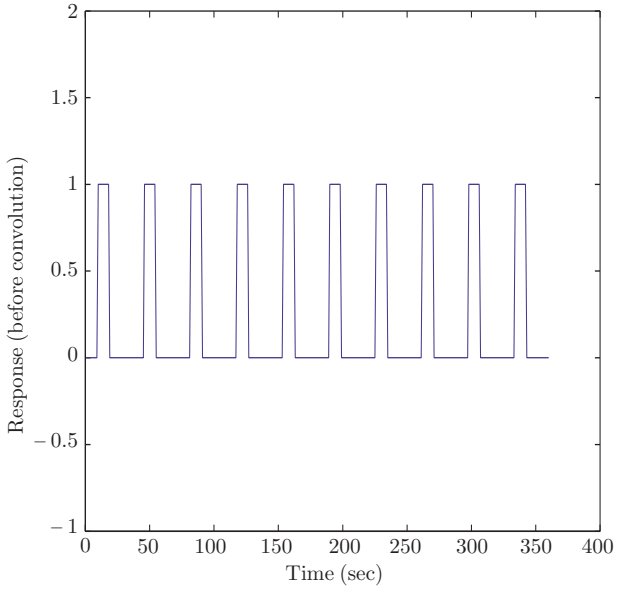

Fig. 1. Box-car function $s_1(t)$ representing the first stimulus.

requires onset time, duration, height (= 1, usually). The experiment we are studying involves two stimuli. In this situation, the box-car function $s_2(t)$ for the second stimulus is a *suitable* translate of $s_1(t)$.

The corresponding fMRI response is not instantaneous; there is a blurring and a delay of the peak response by about 6 secs. A standard way of capturing this is to assume that the noise-free fMRI response depends on the external stimulus by convolution with a hemodynamic response function (HRF) $h(t)$. Thus, for each of the two stimuli, the corresponding noise-free fMRI response is given by

$$z_t^{(j)} = \int_0^\infty h(u) s_j(t-u)\, du, \quad j = 1, 2.$$

The function $h(t)$ we shall work with is the one due to Glover (1999):

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} \exp\left(-\frac{t-d_1}{b_1}\right) - c\left(\frac{t}{d_2}\right)^{a_2} \exp\left(-\frac{t-d_2}{b_2}\right),$$

where t is time in secs, $d_j = a_j b_j$ is the time to the peak, and $a_1 = 6, a_2 = 12, b_1 = b_2 = 0.9$ secs, and $c = 0.35$ (Fig. 2).

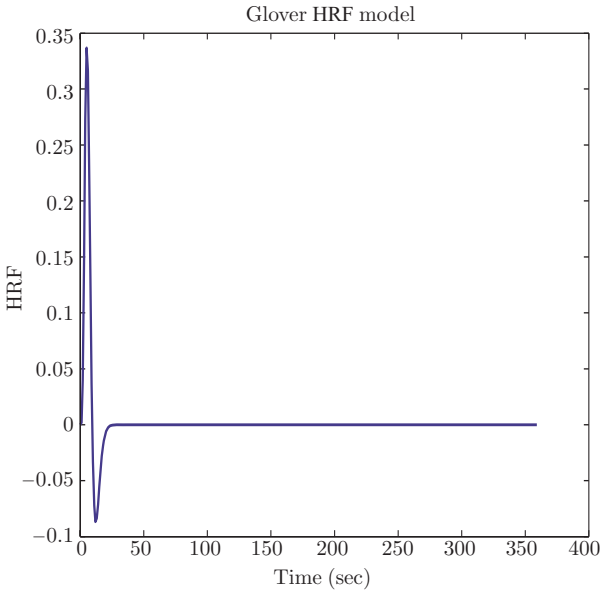

Fig. 2. Glover's HRF $h(t)$.

The function $z_t^{(1)}$, obtained after convolution, is shown in Fig. 3. The function $z_t^{(2)}$ will obviously be a translate of $z_t^{(1)}$. Both the functions $z_t^{(1)}$ and $z_t^{(2)}$ are subsampled at $T = 120$ scan acquisition times. These quantities have been computed by using the MATLAB function **fmridesign.m**, which is one module of the software **fmristat**, available at http://www.math.mcgill.ca/keith/fmristat.

To describe the model let I be the number of voxels in the slice under consideration. In our problem, $I = 128 \times 128$. For the ith voxel, $i = 1, \ldots, I$, the time series $\{y_{it} : t = 1, \ldots, T\}$ of fMR signals is related to the quantities $\{z_t^{(1)} : t = 1, \ldots, T\}$ and $\{z_t^{(2)} : t = 1, \ldots, T\}$ as follows (GAF):

$$y_{it} = w_t' a_i + z_t^{(1)} b_{it} + z_t^{(2)} c_{it} + \epsilon_{it}, \tag{1.1}$$

$$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_i^2), \tag{1.2}$$

where $w_t' a_i$ is the drift, w_t a known design vector that models parametrically the drift, a_i the multiplicative constant (a vector) associated with the drift, b_{it} the multiplicative constant associated with the *hot* stimulus, c_{it} the multiplicative constant associated with the *warm* stimulus and ϵ_{it} the noise.

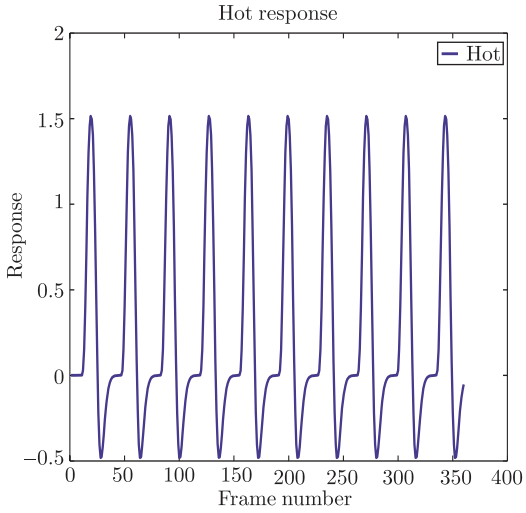

Fig. 3. Convolution $z_t^{(1)}$ of stimulus $s_1(t)$ and Glover's HRF $h(t)$.

Our observation model is given by (1.1)-(1.2). Note that this model ignores temporal dependence present between fMRI observations. Keeping in mind this issue and also the issues of spatial and spatio-temporal dependence, GAF has replaced this model by those described in the following.

2.1 Model for Spatial Dependence

The model for the observations is assumed to be

$$\begin{aligned} y_{it} &= \quad a_i + z_t^{(1)} b_i + z_t^{(2)} c_i + \epsilon_{it}, \\ \epsilon_{it} &\overset{\text{i.i.d.}}{\sim} N(0, \sigma_i^2). \end{aligned} \tag{2}$$

The prior for

$$\boldsymbol{b} = (b_1, \ldots, b_I)'$$

is assumed to be an intrinsic autoregressive prior, also called pairwise difference prior (Besag et al., 1991; Banerjee et al. 2004, Section 3.3). For every voxel i, a neighbourhood denoted by ∂i is assumed to consist of n_i nearest neighbours. This number n_i equals 8 for all but the voxels remaining on the boundary of the slice, where this number equals 3 or 5 according as it is one of the four corners or not. The neighborhood relation between i and j is denoted by $i \sim j$. In other words, for every voxel i, ∂i is the 3×3 window around i without the voxel i. The prior is given by

$$p(\boldsymbol{b}|\lambda) \propto \exp\left\{ -\frac{\lambda}{2} \sum_{\substack{i \sim j \\ i > j}} (b_i - b_j)^2 \right\}. \tag{3.1}$$

It can also be written as

$$p(\boldsymbol{b}|\lambda) \propto \exp\left(-\frac{\lambda}{2}\boldsymbol{b}'Q^{b,\text{space}}\boldsymbol{b}\right), \tag{3.2}$$

where the precision matrix $Q_{b,\text{space}}$ (i.e., the inverse of the variance-covariance matrix) is of the form

$$Q_{ij}^{b,\text{space}} = \begin{cases} n_i & \text{if } i = j, \\ -1 & \text{if } i \sim j, \\ 0 & \text{otherwise.} \end{cases}$$

This model has the following conditional distributions:

$$b_i|b_{j\neq i}, \lambda \sim N\left(\frac{1}{n_i}\sum_{j\in\partial i} b_j, \frac{1}{\lambda n_i}\right).$$

Priors $p(\boldsymbol{a}|\lambda)$, $p(\boldsymbol{c}|\lambda)$ for

$$\boldsymbol{a} = (a_1, \ldots, a_I)' \text{ and } \boldsymbol{c} = (c_1, \ldots, c_I)'$$

are similar. All these priors are assumed to be independent.

In the last stage of the hierarchy, we assume, following GAF, Gamma$(1, 1)$ priors for the unknown hyperparameters, i.e., precision parameter λ and inverse variances $1/\sigma_i^2$ of the observation errors.

Notice that the model (2) is a special case of the one given by (1.1) and (1.2), with the activation coefficients b_i and c_i depending only on the voxel i. The prior given by (3.1) (or (3.2)) allows us to introduce spatial correlation. Compared to pixelwise parametric modeling, estimates are spatially smoothed.

2.2 Model for Temporal Dependence

The model for the observations is assumed to be the following (Gössl et al., 2000, 2001a):

$$\begin{aligned} y_{it} &= a_{it} + z_t^{(1)} b_{it} + z_t^{(2)} c_{it} + \epsilon_{it}, \\ \epsilon_{it} &\overset{\text{i.i.d.}}{\sim} N(0, \sigma_i^2). \end{aligned} \tag{4}$$

The priors for $\boldsymbol{a}_i = (a_{i,1}, \ldots, a_{i,T})'$, $\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,T})'$, $\boldsymbol{c}_i = (c_{i,1}, \ldots, c_{i,T})'$ are as follows:

We assume that for $t = 3, 4, \ldots, T$,

$$a_{it} = 2a_{i,t-1} - a_{i,t-2} + \zeta_{it}, \quad \zeta_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_{\zeta,i}^2), \tag{5.1}$$

$$b_{it} = 2b_{i,t-1} - b_{i,t-2} + \eta_{it}, \quad \eta_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_{\eta,i}^2), \tag{5.2}$$

$$c_{it} = 2c_{i,t-1} - c_{i,t-2} + \xi_{it}, \quad \xi_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_{\xi,i}^2), \tag{5.3}$$

and assume appropriate initial priors for $b_{i,1}, b_{i,2}$ (and also for $a_{i,1}, a_{i,2}$ and $c_{i,1}, c_{i,2}$) so that the prior for

$$\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,T})'$$

can be written as

$$p(\boldsymbol{b}_i|\lambda_{\eta,i}) \propto \exp\left(-\frac{1}{2}\lambda_{\eta,i}\boldsymbol{b}_i'\boldsymbol{Q}^{b,\text{temp}}\boldsymbol{b}_i\right), \tag{6}$$

where $$\boldsymbol{Q}^{b,\text{temp}} = \begin{bmatrix} 2 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -2 & 6 & -4 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & \cdots & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & \cdots & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}_{T\times T},$$

and $\lambda_{\eta,i}$ is the precision parameter of the prior, given by

$$\lambda_{\eta,i} \stackrel{\text{def}}{=} \frac{1}{\sigma^2_{\eta,i}}.$$

The expression for the prior given by (6) will be derived later. We need to mention only that the priors for $b_{i,1}$ and $b_{i,2}$ are assummed to be independent $N(0, \sigma^2_{\eta,i})$ variables. Priors for $\boldsymbol{a}_i = (a_{i,1}, \ldots, a_{i,T})'$ and $\boldsymbol{c}_i = (c_{i,1}, \ldots, c_{i,T})'$ are similar. All these priors are assumed to be independent.

In the last stage of the hierarchy, we assume that

$$\sigma_{\eta,1} = \sigma_{\eta,2} = \cdots = \sigma_{\eta,I} = \sigma_{\xi,1} = \sigma_{\xi,2} = \cdots = \sigma_{\xi,T} = \sigma_{\zeta,1} = \sigma_{\zeta,2} = \cdots = \sigma_{\zeta,T} = \sigma, \tag{7.1}$$

say. This implies

$$\lambda_{\eta,1} = \lambda_{\eta,2} = \cdots = \lambda_{\eta,I} = \lambda_{\xi,1} = \lambda_{\xi,2} = \cdots = \lambda_{\xi,T} = \lambda_{\zeta,1} = \lambda_{\zeta,2} = \cdots = \lambda_{\zeta,T} = \lambda, \tag{7.2}$$

say. Also, we assume, following GAF, Gamma$(1, 1)$ priors for the unknown hyperparameters, i.e. precision parameter λ and inverse variances $1/\sigma_i^2$ of the observation errors.

Equations in (5.1)-(5.3) enforce smoothness, by imposing Gaussian prior restrictions, on the second-order differences

$$\zeta_{it} = a_{it} - 2a_{i,t-1} + a_{i,t-2}, \;\; \eta_{it} = b_{it} - 2b_{i,t-1} + b_{i,t-2}, \;\; \xi_{it} = c_{it} - 2c_{i,t-1} + c_{i,t-2}.$$

These differences may be interpreted as penalties. They penalize deviations from a straight line for which the penalty is zero. Thus, b_{it}'s (and similarly a_{it}'s and c_{it}'s) follow a second-order random walk.

One specific reason for extending (1.1)-(1.2) to (4), (5.1)-(5.3) is to allow flexibility in the form of baseline trend as opposed to a particular form of baseline trend (like a polynomial of degree 3) and also to allow time-dependent effects of covariates.

2.3 An Additive Spatio-Temporal Model

The model for the observations is assumed to be

$$\begin{aligned} y_{it} = \;& a_{it} + z_t^{(1)} b_{it} + z_t^{(2)} c_{it} + \epsilon_{it}, \\ \epsilon_{it} \stackrel{\text{i.i.d.}}{\sim} & N(0, \sigma_i^2). \end{aligned}$$

The previous two models took care, respectively, of the spatial aspect or the temporal aspect of the data separately by introducing appropriate smoothness priors. To obtain models that simultaneously consider these two aspects, a combination of the above properties is necessary. This can be done by different types of space-time interactions. Here we consider an additive model as a reasonable compromise between computational complexity and model complexity. This is obtained by splitting the activation effects for the *hot* and *warm* stimulus b_{it} and c_{it}, and the baseline trend a_{it} as follows:

$$a_{it} = \alpha_i + \beta_{it}, \;\; b_{it} = \gamma_i + \delta_{it}, \;\; c_{it} = \varphi_i + \theta_{it}.$$

We define

$$\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{iT})', \;\; \boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{iT})', \;\; \boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iT})',$$

and assume that each of the vectors $\boldsymbol{\beta}_i$, $\boldsymbol{\delta}_i$, $\boldsymbol{\theta}_i$ are centered at zero. This is to ensure identifiability.

We assume a spatial smoothness prior for the time-constant part γ_i (cf. (3.1) or (3.2)) and a voxelwise temporal second-order random walk prior (cf. (6)) for the time-varying effects δ_{it}. Similar priors are assumed for α_i, $\boldsymbol{\beta}_i$, φ_i and $\boldsymbol{\theta}_i$.

The priors for the precisions are assumed to be Gamma$(1, 1)$. Also, as before, all the priors are assumed to be independent.

3. Our Goal and Methods

We have mentioned earlier that the problem of deciding if a voxel is activated by a specific stimulus is formulated as one of testing if the relevant activation coefficient is zero. This is usually answered within the frequentist framework. Such an approach does not take into account the spatial or spatio-temporal dependence among the fMRI observations which are natural issues in our context. We are addressing such issues within the Bayesian framework.

3.1 Spatial Model

We are interested in the following probabilities:

$$\text{(i) } P(b_i > 0|\boldsymbol{Y}), \;\; \text{(ii) } P(c_i > 0|\boldsymbol{Y}), \;\; \text{(iii) } P(b_i > c_i|\boldsymbol{Y}), \; i = 1, \ldots, I,$$

where $\boldsymbol{Y}$ is the vector of all $(= I \cdot T)$ observations. We shall plot the above values of the probabilities for all the pixels to get activation maps as before. These probabilities are obtained by employing Gibbs sampling (Gelfand and Smith, 1990). The final results are given by images of the posterior probabilities as above. Some initial calculations leading to the full conditionals for the steps of Gibbs sampling are presented below.

To begin with, we work out a more precise expression for $p(\boldsymbol{b}|\lambda)$. The total number of terms in the sum

$$\sum_{\substack{i \sim j \\ i > j}} (b_i - b_j)^2 \text{ is } A/2,$$

where $A = 4 \cdot 3 + (\sqrt{I} - 2) \cdot 4 \cdot 5 + (\sqrt{I} - 2)^2 \cdot 8 = 8I - 12\sqrt{I} + 4$.

Hence,

$$p(\boldsymbol{b}|\lambda) \propto (\sqrt{\lambda})^{4I - 6\sqrt{I} + 2} \cdot \exp\left\{ -\frac{\lambda}{2} \sum_{\substack{i \sim j \\ i > j}} (b_i - b_j)^2 \right\}.$$

The prior for λ is Gamma (a, b), with $a = b = 1$. Hence, it is given by

$$p(\lambda) = \exp(-\lambda), \;\; \lambda > 0. \tag{8}$$

The prior for $1/\sigma_i^2$ is given by $\sigma_i^2 \sim IG(a, b)$ with $a = b = 1$. Write $t = 1/\sigma_i^2$. Then

$$p(t) = \exp(-t), \;\; t > 0.$$

From this, we obtain

$$p(\sigma_i^2) = \exp\left(-\frac{1}{\sigma_i^2}\right) \cdot \left(\frac{1}{\sigma_i^2}\right)^2, \;\; \sigma_i^2 > 0. \tag{9.1}$$

Define

$$\boldsymbol{\sigma} = (\sigma_1^2 \; \sigma_2^2 \; \ldots \; \sigma_I^2)^t.$$

Then

$$p(\boldsymbol{\sigma}) = \prod_{i=1}^{I} p(\sigma_i^2). \tag{9.2}$$

In view of (2), the likelihood is given by

$$l(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\sigma}|\boldsymbol{Y}) \propto \prod_{i=1}^{I} \frac{1}{\sigma_i^T} \exp\left(-\frac{1}{2\sigma_i^2} \sum_{t=1}^{T} (y_{it} - a_i - z_t^{(1)} b_i - z_t^{(2)} c_i)^2\right),$$

where $\boldsymbol{Y}$ is the vector of all $(= I \cdot T)$ observations. Posterior of the parameters is given by

$$p(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\sigma}, \lambda|\boldsymbol{Y}) \propto l(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\sigma}|\boldsymbol{Y}) \cdot p(\boldsymbol{b}|\lambda) \cdot p(\boldsymbol{c}|\lambda) \cdot p(\boldsymbol{a}|\lambda) \cdot p(\lambda) \cdot p(\boldsymbol{\sigma}).$$

The steps of Gibbs sampling is obtained from the expression of the posterior. In Appendix A.1, we present the expressions for the full conditionals.

3.2 Temporal Model

We are interested in the following probabilities

$$\text{(i) } P(b_{it} > 0|\boldsymbol{Y}), \;\; \text{(ii) } P(c_{it} > 0|\boldsymbol{Y}), \;\; \text{(iii) } P(b_{it} > c_{it}|\boldsymbol{Y}), \; i = 1, \ldots, I, \, t = 1, \ldots, T,$$

where $\boldsymbol{Y}$ is the vector of all $(= I \cdot T)$ observations. For a given t, we can plot the above values of the probabilities for all the pixels to get activation maps as before. These probabilities are obtained by employing Gibbs sampling. In what follows, we present some initial calculation leading to the full conditionals for the steps of Gibbs sampling.

To begin with, we derive the expression for the prior for $\boldsymbol{b}_i$ given by (6). Refer to (5.2), (7.1) and (7.2). Let us define

$$\begin{aligned} \eta_{it} &= b_{it} - 2b_{i,t-1} + b_{i,t-2}, \quad t = 3, \ldots, T, \\ &= b_{it}, \qquad\qquad\qquad\quad t = 1, 2. \end{aligned}$$

We assume that η_{it}, $t = 1, \ldots, T$, are i.i.d. $N(0, \sigma_{\eta,i}^2)$. Define now

$$\boldsymbol{\eta}_i \stackrel{\text{def}}{=} (\eta_{i,1}, \ldots, \eta_{i,T})',$$

and recall that

$$\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,T})'.$$

Observe now that

$$\boldsymbol{\eta}_i = \boldsymbol{A}\boldsymbol{b}_i,$$

where $$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}_{T\times T}$$

The pdf of $\boldsymbol{\eta}_i$ is given by

$$p(\boldsymbol{\eta}_i) \propto \lambda^{T/2} \exp\left(-\frac{1}{2}\lambda \boldsymbol{\eta}_i' \boldsymbol{\eta}_i\right).$$

Hence, the pdf of $\boldsymbol{b}_i$ is given by

$$p(\boldsymbol{b}_i|\lambda) \propto \lambda^{T/2} \exp\left(-\frac{1}{2}\lambda \boldsymbol{b}_i' \boldsymbol{A}' \boldsymbol{A} \boldsymbol{b}_i\right).$$

Let us write $\boldsymbol{Q}^{b,\text{temp}} = \boldsymbol{A}'\boldsymbol{A}$. It is easy to verify that

$$\boldsymbol{Q}^{b,\text{temp}} = \begin{bmatrix} 2 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -2 & 6 & -4 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & \cdots & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & \cdots & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}_{T\times T}.$$

Notice that the algebraic calculation above implies the following:

$$\boldsymbol{Q}^{b,\text{temp}} = \boldsymbol{Q}^{c,\text{temp}} = \boldsymbol{Q}^{a,\text{temp}} = \boldsymbol{Q}, \text{ say.}$$

The prior for the common precision parameter λ (cf. (7.1) and (7.2)) is same as in (8).

The prior for

$$\boldsymbol{\sigma} = (\sigma_1^2\ \sigma_2^2\ \ldots\ \sigma_I^2)'$$

is same as in (9.1) and (9.2).

The likelihood is given by

$$l(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{a}_i,\ i = 1, \ldots, I, \boldsymbol{\sigma}|\boldsymbol{Y}) \propto \prod_{i=1}^{I} \frac{1}{\sigma_i^T} \exp\left(-\frac{1}{2\sigma_i^2}\sum_{t=1}^{T}(y_{it} - a_{it} - z_t^{(1)} b_{it} - z_t^{(2)} c_{it})^2\right),$$

where $\boldsymbol{Y}$ is the vector of all $(= I \cdot T)$ observations. Posterior of the parameters is given by

$$p(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{a}_i,\ i = 1, \ldots, I, \boldsymbol{\sigma}, \lambda|\boldsymbol{Y}) \propto l(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{a}_i,\ i = 1, \ldots, I, \boldsymbol{\sigma}|\boldsymbol{Y})$$

$$\times \prod_{i=1}^{I} p(\boldsymbol{b}_i|\lambda) \cdot \prod_{i=1}^{I} p(\boldsymbol{c}_i|\lambda) \cdot \prod_{i=1}^{I} p(\boldsymbol{a}_i|\lambda) \cdot p(\lambda) \cdot p(\boldsymbol{\sigma}).$$

The full conditionals associated with the Gibbs sampling are obtained from the expression of the posterior. They appear in Appendix A.2.

3.3 Spatio-Temporal Model

We are interested in the following probabilities:

$$\text{(i) } P(b_{it} > 0|\boldsymbol{Y}), \quad \text{(ii) } P(c_{it} > 0|\boldsymbol{Y}), \quad \text{(iii) } P(b_{it} > c_{it}|\boldsymbol{Y}), \; i = 1, \ldots, I, \; t = 1, \ldots, T,$$

where $\boldsymbol{Y}$ is the vector of all $(= I \cdot T)$ observations. For a given t, we can plot the above values of the probabilities for all the pixels to get activation maps as before. These probabilities are obtained by employing Gibbs sampling. We skip the details. Interested readers can contact the authors to obtain these information.

4. Issues of Implementation

All our computations have been done on a typical 1.8 GHz Pentium IV desktop PC with 128MB of RAM. There are several simplifications we have introduced with the purpose of reducing running time. The most important is the following: We simulated λ and σ_i^2s directly from their prior distributions, instead of their *full conditional distributions*. Moreover, to reduce the running time of the Gibbs sampler to a reasonable time frame, and to reduce the working memory load in the computer, we have worked with only part of the data-set, i.e., we have taken only first 60 time points for the temporal model and first 40 time points for the spatio-temporal model (instead of the full 120 available to us in our data-set). This data reduction was done only for the models for temporal dependence and spatio-temporal dependence in which the large number of parameters forced us to apply the above step to make our program run within reasonable time limits. For the spatial model, however, we worked with the full data-set.

The Gibbs sampler and the codes for the calculation of the posterior probabilities, were all written in MATHEMATICA while the final plots of the probabilities (i.e., the activation maps) were prepared using MATLAB. For preserving independence among successive samples, we have selected every 5^{th} value generated by the Gibbs sampler, thus discarding all but m of the $5m$ (see Raftery and Lewis, 1992) random variates it generates.

Table 1 gives the number of iterations and approximate running time of the Gibbs sampler.

Table 1. Approximate running times

	Iterations		
Model	Burn-in	Convergence	Time taken (approx.)
Spatial	1000	5000	2 hr 20 min
Temporal	1000	1000	5 hr 50 min
Spatio-temporal	500	500	15 hr 10 min

In order to make a model assessment, we carried out a preliminary analysis of the residuals (for a few random pixels) to check for heteroscedascity and non-normality. All the results obtained validated our assumptions.

5. Results and Discussion

For describing how we present our results, let us restrict our attention to the model for spatial dependence. We want following probabilities:

$$\text{(i) } P(b_i > 0|\boldsymbol{Y}), \quad \text{(ii) } P(c_i > 0|\boldsymbol{Y}), \quad \text{(iii) } P(b_i > c_i|\boldsymbol{Y}), \; i = 1, \ldots, I,$$

where $\boldsymbol{Y}$ is the vector of all $(= I \cdot T)$ observations. For every voxel, denoted by i, of the slice under consideration, we present these quantities in plots. Fig. 4 (a, b) gives the first and third of these plots, for the model for spatial dependence. Such plots are known as *posterior probability maps*. Voxels with high probabilities constitute the region(s) of activation due to the stimulus.

Fig. 4 (a) shows some very highly activated regions (colored white) and a substantial amount of high activation regions (colored red) due to the *hot* stimulus. Due to the presence of such a high amount of activation in different regions of the brain, it is quite difficult to distinguish where exactly the primary activation takes place due to the *hot* stimulus. However, Fig. 4(b) very clearly distinguishes between the activated and non-activated regions. It is seen that the primary activation is located in the central part of the brain (colored white). Additionally small activated regions can be seen in the lower part of the brain. This figure essentially tells us which parts of the brain are responsible for perception of pain finding which is the aim of the experiment. Also, as a consequence, this has motivated us to concentrate on 'such' type of maps in the other models also.

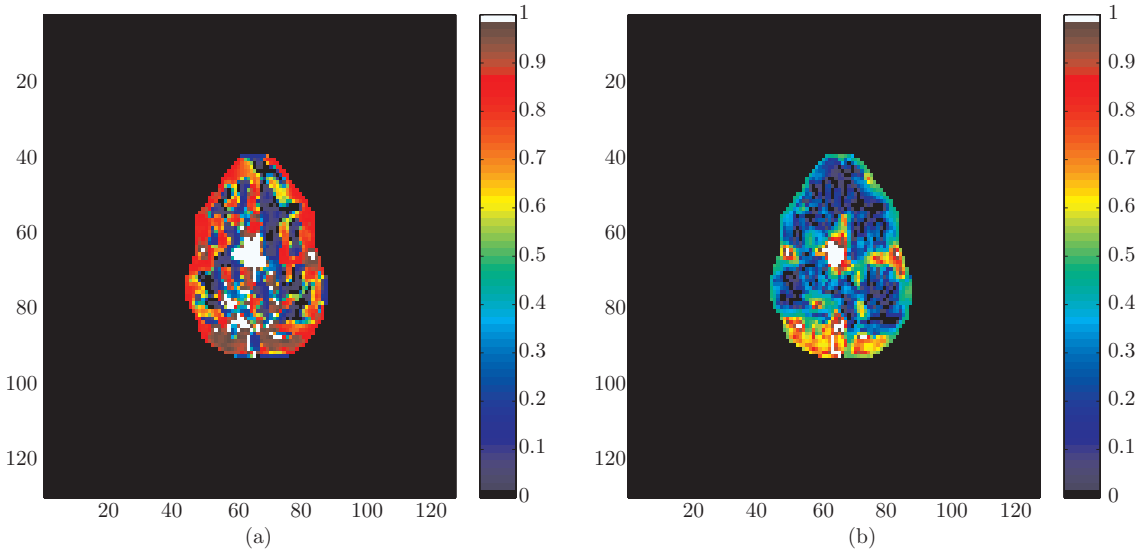

Fig. 4. Plots of (a) $\boldsymbol{P}(b_i > 0|Y)$ and (b) $\boldsymbol{P}(b_i > c_i|Y)$ for the model for spatial dependence.

We now consider the models for temporal and spatio-temporal dependence. Let us recall that we are interested in the following probabilities:

$$\text{(i) } P(b_{it} > 0|\boldsymbol{Y}), \quad \text{(ii) } P(c_{it} > 0|\boldsymbol{Y}), \quad \text{(iii) } P(b_{it} > c_{it}|\boldsymbol{Y}), \; i = 1, \ldots, I, \; t = 1, \ldots, T,$$

where $\boldsymbol{Y}$ is the vector of all $(= I \cdot T)$ observations. For a given t, we can plot the above values of the probabilities for all the pixels to get posterior probability maps as before.

The figures given next and pertaining to (iii) above were obtained from the application of the temporal model. Being obtained from the temporal model, each such map corresponds to a specific

time point of the experiment. Fig. 5 (a) corresponds to the 22$^{\text{nd}}$ time point of the experiment, while Fig. 5 (b) corresponds to the 27$^{\text{th}}$ time point of the experiment. Both these maps were thresholded at the probability level 1-10^{-3}, meaning that all those pixels for which the corresponding probabilities were greater than our threshold were assigned values above 0.8 in the colorbar. This sort of thresholding is necessary here to bring out the activated regions. (What we are doing here is what is known in image processing as *Contrast Stretching*. See Appendix A.3 for details.)

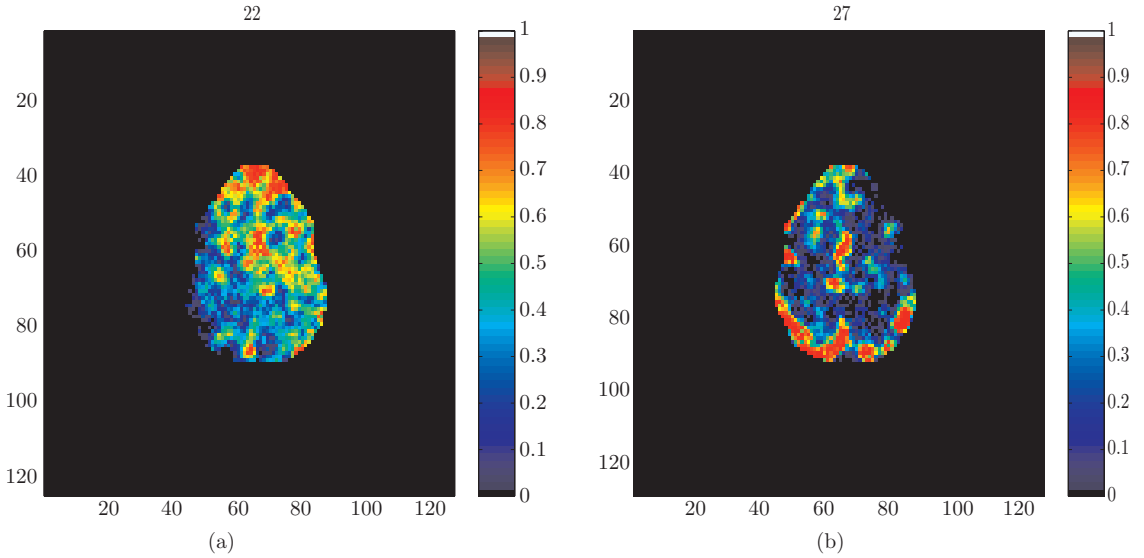

Fig. 5. Plots of $P(b_{it} > c_{it}|Y)$ at (a) $t = 22$ and (b) $t = 27$ for the model for temporal dependence.

What is readily seen to be *common* among the two plots is the activation around the central region of the brain. However, in Fig. 5 (a) we see significant activation in some parts of the top of the brain. Also, in Fig. 5 (b) parts of the lower left portion and lower right portion (near the boundary) show high activation. These observations suggest that there are temporal fluctuations in the activated regions of the brain which our temporal model can catch. This is because we had designed our priors in such a way so that we had a temporally varying activation profile (β_{it}). On the other hand, the spatial model, which assumes a time constant activation α_i, can detect only the primary activation areas (areas which get activated more or less throughout the whole duration of the experiment), but is unable to detect the temporally varying activation regions. However, it can be seen that the plots from the temporal model does not assign very high probabilities to the primary activation regions, as was the case for the spatial model. Also the activation is spread over a large part of the brain, instead of some specific regions. This may be due to the large number of parameters associated with the temporal model which may lead to a splitting of the information (thus widening the Bayesian credibility regions). Also to be noted is the fact that this analysis was carried out using half the data-set (as was mentioned earlier). So the inferences from this model cannot be as precise as those from the spatial model.

Finally, the plots given next and pertaining to (iii) above were obtained from the additive spatio-temporal model. As before, each such plot corresponds to a particular time point of the experiment. Fig. 6 (a) corresponds to the 10$^{\text{th}}$ time point of the experiment, while Fig. 6 (b) corresponds to the 18$^{\text{th}}$ time point of the experiment. Both these maps were thresholded at 1-10^{-6} level. *It should be*

noted that as the number of parameters in our model increase, we have to increase our thresholding limits in order to bring out the activated regions. Thresholding at lower levels do not give satisfactory results, i.e., we do not get pictures/images where we can clearly distinguish the activated regions from the non-activated ones. If we threshold at lower levels, then we get large parts of the brain as activated regions. So, unless we threshold at higher probability levels, we are unable to decide which are the areas of the brain that actually stand out as activated compared to the other surrounding areas.

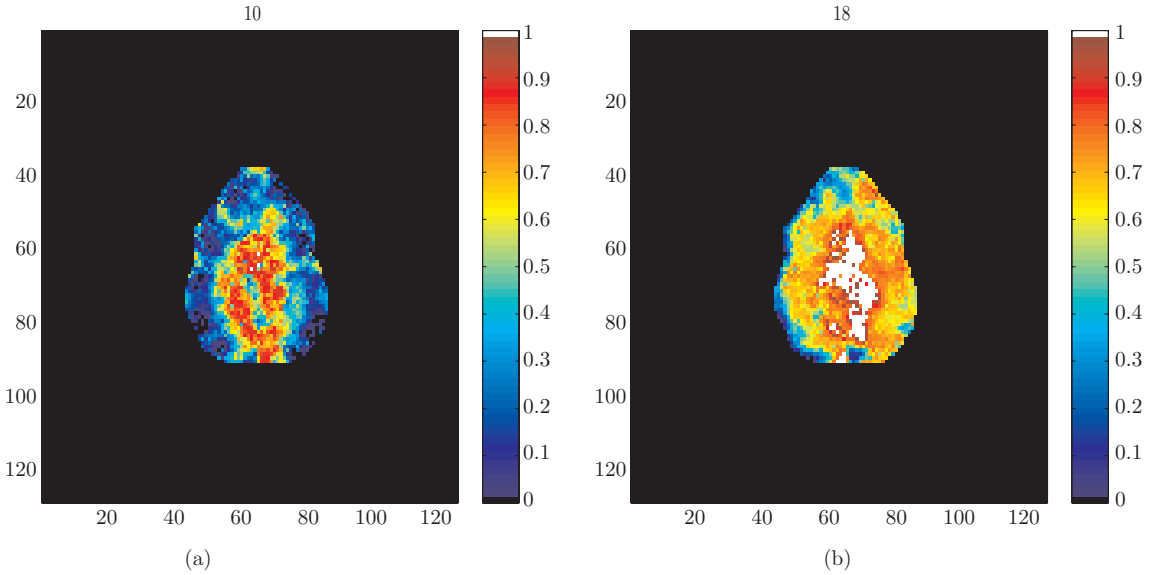

Fig. 6. Plots of $P(b_{it} > c_{it}|Y)$ at (a) $t = 10$ and (b) $t = 18$ for the model for spatio-temporal dependence.

Fig. 6 (a, b) shows quite a substantial part of the brain activated including the central part which we had already identified in the results of the spatial and temporal models. However, a comparison between these two plots shows an interesting feature. In Fig. 6 (a) the regions which are quite highly activated (deep red regions), a large part of those regions show very high activation (white regions) in Fig. 6 (b). Also those regions which show low levels of activation (blue and green regions) in Fig. 6 (a) show increased activation in Fig. 6 (b) (yellow and orange regions). This suggests that regions of the brain do not get activated all at once, the levels of activation vary smoothly over time. This is what one would expect, as BOLD signal depends on the blood flow inside the brain which is expected to vary smoothly over time. However, as has been said about results for the temporal model, the results obtained from the spatio-temporal model show activation over a large part of the brain (including the suspected central region). The large number of parameters and the small portion of the data-set we used for this model ($1/3^{\text{rd}}$ part of the original data-set) may have caused this. Hence, the inferences drawn from the results of this model will obviously not be as precise as those of the spatial model. However, the results show that this model is able to capture the primary activated regions as well as capture the smooth temporal variation of different parts of the brain in response to the stimulus.

We conclude this section by observing that the region(s) activated by pain perception and picked up by the methods we have employed agree to a large extent with the one obtained in Worsley

(Worsley et al., 2002; Worsley, 2003; http://www.math.mcgill.ca/keith/fmristat) by application of non-Bayesian method. However, we get activation over other regions and larger regions.

6. Future Directions

In order to simplify our calculations, we have assumed that λ_i's are all equal (cf. (7.2)). We have also simplified calculations by simulating directly from the prior distributions of the hyperparameters λ's and σ_i^2's. We should do the whole exercise without these simplifications. However, in view of our experience, we may add that we shall require *some* simplifications for reducing computing time. We may also try with different hyperprior specifications, say by changing the values of γ_a and γ_b, to see how our results vary.

To reduce the computation time of our programs, we have worked with only half of our original data-set in case of the temporal model and one third of our data-set was used for fitting the spatio-temporal model. Besides this, our results and inference were obtained by running the analysis on a single slice. In future one might try running the analysis on all the slices and the full data-set, to draw more accurate inferences. Here also we have to address separately the issue of reducing computing time.

A non-additive spatio-temporal model has been suggested in GAF that takes into account the temporal variations in adjacent pixels, unlike the *additive* spatio-temporal model. They propose priors in which both the spatial and temporal dependencies are incorporated simultaneously. This is one direction in which some future work may be pursued. Also, models and methods proposed in Genovese (1999), Gössl et al. (2001b), Marrelec et al. (2003), Woolrich et al. (2004a) may be tried. Carlin and Banerjee (2003) also may be helpful in this context.

The referee has asked if there is any motivation for choosing ∂i, for every voxel i, the 3×3 window around i without the voxel i. In the present context, this choice has been driven by requirement of simplicity and consequential reduction in computing time. However, we also wish to address this issue by varying the window size and making a decision by employing some tool of model selection. In particular, we may try an empirical Bayes approach using ML-II method (Berger, 1985, pp. 99–101) in this context. Purdon et al. (2001) and Solo et al. (2001) also are relevant references here.

The analysis presented in GAF is preceded by application of suitable algorithm for correcting for the subject's motion. Usually an image registration algorithm is applied for this purpose. This is a common preprocessing step and is quite effective in practice (Genovese, 2000). We may also think of some sort of image processing on the scanned images, possibly to reduce noise and increase sharpness. Such techniques can be applied to our images. We have ignored these issues as our focus was more on successful implementation of some of the models proposed in GAF.

It has to be noted that we are employing several models on the same data-set and have also kept several others in mind. Issues of model selection need to be addressed here. Also, in addition to looking at certain posterior probabilities we need to develop and work out the appropriate Bayes tests for activation studies in analysis of fMRI data. See Genovese (1999) in this context.

A final extension might be the incorporation of further substantial prior information, e.g., anatomical priors obtained from structural images to constrain estimates of activity to be in or near the grey matter. This would both lend biological validity and restrict the number of voxels to be analysed and so speed up computation—both important issues.

Acknowledgements

The work reported in this paper has been supported partially by fund available through the project "Analysis of fMRI Data and Human Brain Mapping" of Division of Theoretical Statistics and

Mathematics, Indian Statistical Institute. This project was undertaken with a view to initiating research in statistical brain imaging at Indian Statistical Institute. Prof. J.K. Ghosh introduced Sumitra Purkayastha to this area and also encouraged him to pursue research in this area. Sumitra Purkayastha is grateful to him for these. The authors are also grateful to him for several helpful discussions and to Professors N.A. Lazar and W.D. Penny for their encouraging comments on an earlier version of this paper. Some of their suggestions have been incorporated in the section on future directions of work. They are also grateful to Prof. K.J. Worsley for bringing some important references to their notice. Finally, the authors wish to record their sincere gratitude to an anonymous referee for going through an earlier version of this paper and offering several helpful suggestions.

References

Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical modeling and analysis for spatial data*. Boca Raton: Chapman & Hall/CRC.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.

Bullmore, E., Brammer, M., Williams, S.C.R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R. and Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, **35**, 261-277.

Carlin, B.P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. In *Bayesian Statistics* **7**, eds. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West. Oxford: Oxford University Press, 45-64.

Clare, S. (1997). *Functional Magnetic Resonance Imaging: Methods and Applications*, Unpublished Ph.D. thesis. [http://www.fmrib.ox.ac.uk/∼stuart]

Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Price, C.J., Zeki, S., Ashburner, J. and Penny, W. (2004). *Human brain function* (2nd ed.). Elsevier, Academic Press: Amsterdam.

Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J. and Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage*, **2**, 45-53.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

Genovese, C.R. (1999). Imaging and spatio-temporal inference (with discussion). In *Bayesian Statistics* **6**, eds. J. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford University Press, 255-274.

Genovese, C.R. (2000). A Bayesian time-course model for functional magnetic resonance imaging data (with discussion). *Journal of American statistical association*, **95**, 691-719.

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

Gössl, C., Auer, D.P. and Fahrmeir, L. (2000). Dynamic models in fMRI. *Magnetic Resonance in Medicine* **43**, 72-81.

Gössl, C., Auer, D.P. and Fahrmeir, L. (2001a). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, **57**, 554-562.

Gössl, C., Fahrmeir, L. and Auer, D.P. (2001b). Bayesian modeling of the hemodynamic response function in BOLD fMRI. *NeuroImage*, **14**, 140-148.

Jezzard, P., Matthews, P.M. and Smith, S.M. (2001). (Eds.) *Functional MRI: An introduction to methods*. New York: Oxford University Press.

Lange, N. (1996). Statistical approaches to human brain mapping by functional magnetic resonance imaging. *Statistics in Medicine*, **15**, 389-428.

Lazar, N.A., Eddy, W.F., Genovese, C.R. and Welling, J. (2001). Statistical issues in fMRI brain imaging. *International Statistical Review*, **69**, 105-127.

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

Penny, W.D., Trujillo-Barreto, N.J. and Friston, K.J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, **24**, 350-362.

Polzehl, J. and Spokoiny, V.G. (2001). Functional and dynamic magnetic resonance imaging using vector adaptive weights smoothing. *Journal Royal Statist. Soc. C,* **50**, 485-501.

Purdon, P.L., Solo, V., Weisskoff, R.M. and Brown, E.N. (2001). Locally regularized spatiotemporal modeling and model comparison for functional MRI. *NeuroImage*, **14**, 912-923.

Raftery, A.E. and Lewis, S.M. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics* **4**, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, 255-274. Oxford: Oxford University Press.

Solo, V., Purdon, P.L., Weisskoff, R.M. and Brown, E.N. (2001). A signal estimation approach to functional MRI. *IEEE Transactions on Medical Imaging*, **20**, 26-35.

Woolrich, M.W., Jenkinson, M., Brady, J.M. and Smith, S.M. (2004a). Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging*, **23**, 213-231.

Woolrich, M., Behrens, T.E.J. and Smith, S.M. (2004b). Constrained linear basis set for HRF modelling using variational Bayes. *NeuroImage*, **21**, 1748-1761.

Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F. and Smith, S. M. (2005). Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Transactions on Medical Imaging*, **24**, 1-11.

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

Worsley, K. J. (2003). Detecting activation in fMRI data. *Statistical Methods in Medical Research*, **12**, 401-418.

Worsley, K.J. and Friston, K.J. (1995). Analysis of fMRI time series revisited—again. *NeuroImage*, **2**, 173-181.

Appendix A.1 Full Conditionals for the Steps of Gibbs Sampling for the Model for Spatial Dependence

- *Full conditional of each* b_i: Note that

$$p(b_i|\boldsymbol{b}_{-i}, \boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\sigma}, \lambda, \boldsymbol{Y}) \propto \exp\left(-\frac{1}{2\sigma_i^2}\sum_{t=1}^{T}(y_{it} - a_i - z_t^{(1)}b_i - z_t^{(2)}c_i)^2 - \frac{\lambda}{2}(n_i b_i^2 - 2b_i \sum_{j\in\partial i} b_j)\right).$$

From this, it is possible to show that

$$p(b_i|\boldsymbol{b}_{-i}, \boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\sigma}, \lambda, \boldsymbol{Y}) \sim N(\mu_1, \tau_1^2),$$

where $\mu_1 = \dfrac{\dfrac{\sigma_i^2}{n_i}\displaystyle\sum_{j\in\partial i} b_j + \dfrac{1}{n_i\lambda}\sum_{t=1}^{T} z_t^{(1)}(y_{it} - a_i - z_t^{(2)}c_i)}{\sigma_i^2 + \dfrac{\displaystyle\sum_{t=1}^{T}(z_t^{(1)})^2}{n_i\lambda}}$, $\tau_1^2 = \dfrac{\dfrac{\sigma_i^2}{n_i\lambda}}{\sigma_i^2 + \dfrac{\displaystyle\sum_{t=1}^{T}(z_t^{(1)})^2}{n_i\lambda}}$.

- *Full conditional of each c_i*: Proceeding as above, we can show that

$$p(c_i|\boldsymbol{c}_{-i}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{\sigma}, \lambda, \boldsymbol{Y}) \sim N(\mu_2, \tau_2^2),$$

where $$\mu_2 = \frac{\frac{\sigma_i^2}{n_i}\sum_{j\in\partial i} c_j + \frac{1}{n_i\lambda}\sum_{t=1}^{T} z_t^{(2)}(y_{it} - a_i - z_t^{(1)}b_i)}{\sigma_i^2 + \frac{\sum_{t=1}^{T}(z_t^{(2)})^2}{n_i\lambda}}, \quad \tau_2^2 = \frac{\frac{\sigma_i^2}{n_i\lambda}}{\sigma_i^2 + \frac{\sum_{t=1}^{T}(z_t^{(2)})^2}{n_i\lambda}}.$$

- *Full conditional of each a_i*: It is possible to show that

$$p(a_i|\boldsymbol{a}_{-i}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{\sigma}, \lambda, \boldsymbol{Y}) \sim N(\mu_3, \tau_3^2),$$

where $$\mu_3 = \frac{\frac{\sigma_i^2}{n_i}\sum_{j\in\partial i} a_j + \frac{1}{n_i\lambda}\sum_{t=1}^{T}(y_{it} - z_t^{(1)}b_i - z_t^{(2)}c_i)}{\sigma_i^2 + \frac{T}{n_i\lambda}}, \quad \tau_3^2 = \frac{\frac{\sigma_i^2}{n_i\lambda}}{\sigma_i^2 + \frac{T}{n_i\lambda}}.$$

- *Full conditional of each σ_i^2*:

$$p(\sigma_i^2|\boldsymbol{\sigma}_{-i}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \lambda, \boldsymbol{Y}) \propto \left\{\frac{1}{\sigma_i^T}\exp\left(-\frac{1}{2\sigma_i^2}\sum_{t=1}^{T}(y_{it} - a_i - z_t^{(1)}b_i - z_t^{(2)}c_i)^2\right)\right\} \times \exp\left(-\frac{1}{\sigma_i^2}\right) \times (\frac{1}{\sigma_i^2})^2.$$

  Therefore,

$$p(\sigma_i^2|\boldsymbol{\sigma}_{-i}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \lambda, \boldsymbol{Y}) \sim IG\left(1 + \frac{1}{2}\sum_{t=1}^{T}(y_{it} - a_i - z_t^{(1)}b_i - z_t^{(2)}c_i)^2, \frac{T+2}{2}\right).$$

- *Full conditional of λ*: It is given by

$$p(\lambda|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{\sigma}, \boldsymbol{Y}) \sim \text{Gamma}\left(a^\star, b^\star\right),$$

  where

$$a^\star = a + \frac{1}{2}\left\{\sum_{i=1}^{I} n_i a_i^2 + \sum_{i=1}^{I} n_i b_i^2 + \sum_{i=1}^{I} n_i c_i^2\right\} - \left(\sum_{i\sim j} a_i a_j + \sum_{i\sim j} b_i b_j + \sum_{i\sim j} c_i c_j\right),$$
$$b^\star = b + \frac{3}{2}(4I - 6\sqrt{I} + 2).$$

We need to mention here that for reducing computing time, at each step of iteration we have drawn sample from the priors of the σ_i^2's and also λ. However, for the other parameters, the samples are drawn from the full posteriors.

Appendix A.2 Full Conditionals for the Steps of Gibbs Sampling for the Model for Temporal Dependence

- *Full conditional of each* $\boldsymbol{b}_i$:

$$p(\boldsymbol{b}_i|\boldsymbol{b}_{j\neq i}, \boldsymbol{c}_i, \boldsymbol{a}_i,\, i=1,\ldots,I, \boldsymbol{\sigma}, \lambda, \boldsymbol{Y}) \propto \left\{\frac{1}{\sigma_i^T}\exp\left(-\frac{1}{2\sigma_i^2}\sum_{t=1}^{T}(y_{it}-a_{it}-z_t^{(1)}b_{it}-z_t^{(2)}c_{it})^2\right)\right\} \times \exp\left(-\frac{1}{2}\lambda \boldsymbol{b}_i'\boldsymbol{Q}\boldsymbol{b}_i\right).$$

Let us write

$$\begin{aligned}
&\boldsymbol{y}_i = (y_{i1},\ldots,y_{iT})',\\
&\boldsymbol{\mu}_i = (a_{i1}+z_1^{(1)}b_{i1}+z_1^{(2)}c_{i1},\ldots,a_{iT}+z_T^{(1)}b_{iT}+z_T^{(2)}c_{iT})',\\
&\boldsymbol{Z}^{(1)} = \mathrm{diag}(z_1^{(1)},\ldots,z_T^{(1)}),\\
&\boldsymbol{Z}^{(2)} = \mathrm{diag}(z_1^{(2)},\ldots,z_T^{(2)}).
\end{aligned}$$

It follows that

$$\boldsymbol{\mu}_i = \boldsymbol{a}_i + \boldsymbol{Z}^{(1)}\boldsymbol{b}_i + \boldsymbol{Z}^{(2)}\boldsymbol{c}_i.$$

Now note that

$$p(\boldsymbol{b}_i|\boldsymbol{b}_{j\neq i}, \boldsymbol{c}_i, \boldsymbol{a}_i,\, i=1,\ldots,I, \boldsymbol{\sigma}, \lambda, \boldsymbol{Y}) \propto \exp\left(-\frac{1}{2\sigma_i^2}(\boldsymbol{y}_i-\boldsymbol{\mu}_i)'(\boldsymbol{y}_i-\boldsymbol{\mu}_i)\right)\exp\left(-\frac{1}{2}\lambda\boldsymbol{b}_i'\boldsymbol{Q}\boldsymbol{b}_i\right).$$

Observe that

$$\begin{aligned}
&\frac{1}{2\sigma_i^2}(\boldsymbol{y}_i-\boldsymbol{\mu}_i)'(\boldsymbol{y}_i-\boldsymbol{\mu}_i)+\frac{1}{2}\lambda\boldsymbol{b}_i'\boldsymbol{Q}\boldsymbol{b}_i\\
&= \frac{1}{2\sigma_i^2}[\boldsymbol{Z}^{(1)}\boldsymbol{b}_i-(\boldsymbol{y}_i-\boldsymbol{a}_i-\boldsymbol{Z}^{(2)}\boldsymbol{c}_i)]'[\boldsymbol{Z}^{(1)}\boldsymbol{b}_i-(\boldsymbol{y}_i-\boldsymbol{a}_i-\boldsymbol{Z}^{(2)}\boldsymbol{c}_i)]+\frac{1}{2}\lambda\boldsymbol{b}_i'\boldsymbol{Q}\boldsymbol{b}_i\\
&= \boldsymbol{b}_i'\left[\frac{1}{2\sigma_i^2}\boldsymbol{Z}^{(1)\prime}\boldsymbol{Z}^{(1)}+\frac{\lambda}{2}\boldsymbol{Q}\right]\boldsymbol{b}_i-\frac{(\boldsymbol{y}_i-\boldsymbol{a}_i-\boldsymbol{Z}^{(2)}\boldsymbol{c}_i)'\boldsymbol{Z}^{(1)}\boldsymbol{b}_i}{\sigma_i^2}+\frac{(\boldsymbol{y}_i-\boldsymbol{a}_i-\boldsymbol{Z}^{(2)}\boldsymbol{c}_i)'(\boldsymbol{y}_i-\boldsymbol{a}_i-\boldsymbol{Z}^{(2)}\boldsymbol{c}_i)}{\sigma_i^2}
\end{aligned}$$

All these show that the conditional distribution to be found is a T-variate normal distribution with dispersion matrix and mean vector given by

$$\left[\frac{1}{\sigma_i^2}\boldsymbol{Z}^{(1)\prime}\boldsymbol{Z}^{(1)}+\lambda\boldsymbol{Q}\right]^{-1} \text{ and } \left[\frac{1}{\sigma_i^2}\boldsymbol{Z}^{(1)\prime}\boldsymbol{Z}^{(1)}+\lambda\boldsymbol{Q}\right]^{-1}\frac{\boldsymbol{Z}^{(1)}(\boldsymbol{y}_i-\boldsymbol{a}_i-\boldsymbol{Z}^{(2)}\boldsymbol{c}_i)}{\sigma_i^2}.$$

- *Full conditional of each* $\boldsymbol{c}_i$: It is given by a T-variate normal distribution with dispersion matrix and mean vector given by

$$\left[\frac{1}{\sigma_i^2}\boldsymbol{Z}^{(2)\prime}\boldsymbol{Z}^{(2)}+\lambda\boldsymbol{Q}\right]^{-1} \text{ and } \left[\frac{1}{\sigma_i^2}\boldsymbol{Z}^{(2)\prime}\boldsymbol{Z}^{(2)}+\lambda\boldsymbol{Q}\right]^{-1}\frac{\boldsymbol{Z}^{(2)}(\boldsymbol{y}_i-\boldsymbol{a}_i-\boldsymbol{Z}^{(1)}\boldsymbol{b}_i)}{\sigma_i^2}.$$

- *Full conditional of each* $\boldsymbol{a}_i$: It is given by a T-variate normal distribution with dispersion matrix and mean vector given by

$$\left[\frac{1}{\sigma_i^2}\boldsymbol{I}_T+\lambda\boldsymbol{Q}\right]^{-1} \text{ and } \left[\frac{1}{\sigma_i^2}\boldsymbol{I}_T+\lambda\boldsymbol{Q}\right]^{-1}\frac{(\boldsymbol{y}_i-\boldsymbol{Z}^{(1)}\boldsymbol{b}_i-\boldsymbol{Z}^{(2)}\boldsymbol{c}_i)}{\sigma_i^2}.$$

- *Full conditional of each* σ_i^2:

$$p(\sigma_i^2|\boldsymbol{\sigma}_{-i}, \boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{c}_i,\ i=1,\ldots,I,\ \lambda, \boldsymbol{Y})$$
$$\propto \left\{\frac{1}{\sigma_i^T}\exp\left(-\frac{1}{2\sigma_i^2}\sum_{t=1}^{T}(y_{it}-a_{it}-z_t^{(1)}b_{it}-z_t^{(2)}c_{it})^2\right)\right\}\times\exp\left(-\frac{1}{\sigma_i^2}\right)\times\left(\frac{1}{\sigma_i^2}\right)^2.$$

Therefore,

$$p(\sigma_i^2|\boldsymbol{\sigma}_{-i}, \boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{c}_i,\ i=1,\ldots,I,\ \lambda, \boldsymbol{Y}) \sim IG\left(1+\frac{1}{2}\sum_{t=1}^{T}(y_{it}-a_{it}-z_t^{(1)}b_{it}-z_t^{(2)}c_{it})^2, \frac{T+2}{2}\right).$$

- *Full conditional of* λ:

$$p(\lambda|\boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{c}_i,\ i=1,\ldots,I,\ \boldsymbol{\sigma}, \boldsymbol{Y})$$
$$\propto \prod_{i=1}^{I}\exp\left(-\frac{1}{2}\lambda\boldsymbol{a}_i'\boldsymbol{Q}\boldsymbol{a}_i-\frac{1}{2}\lambda\boldsymbol{b}_i'\boldsymbol{Q}\boldsymbol{b}_i-\frac{1}{2}\lambda\boldsymbol{c}_i'\boldsymbol{Q}\boldsymbol{c}_i\right)\cdot\exp(-\lambda)\cdot\lambda^{3T/2}.$$

Therefore, the required conditional distribution is given by

$$p(\lambda|\boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{c}_i,\ i=1,\ldots,I,\ \boldsymbol{\sigma}, \boldsymbol{Y}) \sim \text{Gamma}\,(a^\star, b^\star),$$

where $\quad a^\star = 1+\dfrac{\sum_{i=1}^{I}(\boldsymbol{a}_i'\boldsymbol{Q}\boldsymbol{a}_i+\boldsymbol{b}_i'\boldsymbol{Q}\boldsymbol{b}_i+\boldsymbol{c}_i'\boldsymbol{Q}\boldsymbol{c}_i)}{2},\ \ b^\star=\dfrac{3T+2}{2}.$

We need to mention here that for reducing computing time, at each step of iteration we have drawn sample from the priors of the σ_i^2's and also λ. However, for the other parameters, the samples are drawn from the full posteriors.

Appendix A.3 Details on Contrast Stretching

Denote 1-10^{-3} by T (threshold level). On the colorbar in MATLAB, we can assign values from 0 to 1, 0 representing black, 1 representing white. Note that each voxel value is a probability and hence is a number in $[0,1]$. However, we want to enhance those voxels having values (i.e., probabilities) greater than T. This is achieved by employing the following piecewise linear transformation $f\colon[0,1]\longrightarrow[0,1]$, characterized by $f(0)=0, f(T)=0.8, f(1)=1$, on each voxel value (i.e., probability) x. It is given by

$$f(x) = \frac{0.8\times x}{T} \qquad \text{if } 0\le x\le T,$$
$$= 0.8+\frac{x-T}{1-T}\times(1-0.8) \text{ if } T\le x\le 1.$$