

# A Bayesian Multiplicative Markov Random Field Model for Determining Activated Areas in Functional Magnetic Resonance Imaging Studies

John Kornak

School of Mathematical Sciences, University of Nottingham, and  
MRC Institute of Hearing Research, Nottingham

Anthony O'Hagan

Department of Probability and Statistics, University of Sheffield

Mark Haggard

MRC Institute of Hearing Research, Nottingham

April 18, 2005

John Kornak (corresponding author), Assistant Professor, Department of Radiology and Department of Epidemiology and Biostatistics, University of California, San Francisco, VAMC MRI Unit, 4150 Clement Street (114M), San Francisco, CA 94121 (email: kornak@itsa.ucsf.edu); Anthony O'Hagan, Professor, Department of Probability and Statistics, University of Sheffield, Sheffield, S3 7RH, UK (email: a.ohagan@sheffield.ac.uk); and Mark P. Haggard, Professor, MRC ESS Team in Children's Middle-ear Disease, Elseworth House, Box 58, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 2QQ, UK (email: mark.haggard@mrc-cbu.cam.ac.uk)

## Abstract

A Bayesian spatial model for detecting brain activation in functional neuroimaging (here focusing on functional magnetic resonance imaging [fMRI]) is developed that inherently trades the local functional response level with its spatial extent. The model makes no assumptions about the shape or form of activated areas, except that they are elevated above non-active areas and form contiguous regions as characterized by Markov random field prior distributions. The robustness of the model is tested under a variety of simulated conditions and the model is subsequently applied to real data from an auditory experiment. An observed secondary low-level response that occurs globally and is synchronized with the experimental stimulus necessitates an extension to the model in the form of an additional additive Markov random field.

Functional magnetic resonance imaging has been playing a major role in neuroscience for more than a decade. A whole range of statistical methods have been developed and adapted in order to analyze the highly multivariate datasets that are obtained from fMRI experiments. Initially, these approaches were adapted from existing ones for detecting activated areas in Positron Emission Tomography experiments. However, subsequent methods have been developed that are tailored specifically for fMRI. Most of these methods tend to focus on performing a separate analysis (e.g. linear model or parametric curve fitting) at each voxel (3D pixel) in order to determine the existence or not of a localized brain response to an experimental stimulus. The subsequent spatial stage of these analysis methods tend to focus on determining a level at which to threshold a 3D map of, for example estimated magnitudes, such that the large number of multiple comparisons is accounted for. A spatial map showing above-threshold voxels is then usually displayed as a map of activation.

Instead of level thresholding the map of statistics, we develop a Bayesian product Markov random field model that places a priori expected constraints on the smoothness of activation patterns. The presence of local activation is then dependent not only on the level of a statistic at a particular voxel, but also on both the level of the statistics at nearby voxels and the classification as either “active” or “inactive” of those same voxels. The main benefit of this approach is that the Bayesian spatial prior distributions provide a framework for detecting active regions much as a neurologist might; based on posterior evidence over a wide range of spatial scales, simultaneously considering the level of the voxel magnitudes along with the size of the elevated area.

Keywords and phrases: Bayesian image analysis, functional Magnetic Resonance Imaging, Markov chain Monte Carlo, Markov random field.

## 1 Introduction

Neuro-imaging techniques such as functional magnetic resonance imaging (fMRI) enable *in-vivo* measurement and localization of cortical functions in the brain, and the determination of interactions between various cortical regions. Until the beginning of the previous decade, only methods with poor spatial localization (e.g. electroencephalography) existed

to measure cortical activity. The advent of positron emission tomography (PET) in the 1980s provided an indirect way of localizing cortical activity through regional blood flow measurement. However, PET has very poor temporal resolution and requires subjects to be injected with small amounts of radioactive material, limiting the general applicability to research neuroscience.

The subsequent non-invasive technique of functional magnetic resonance imaging (fMRI) developed rapidly in the early-mid 1990s, providing greater convenience and safety as well as improved spatial and temporal resolution relative to PET in the detection of cortical activity. In fMRI, neural activation leads (via neural control of the vasculature) to increased local (oxygenated) blood flow, so as to replace locally consumed oxygen. Since oxygenated blood has different magnetic properties from de-oxygenated blood, the (over-)compensating blood flow leads to a measurable change in the MR signal. The response time course is referred to as the blood oxygenation level dependent (BOLD) hemodynamic response function (HDR) and this serves as the raw signal in fMRI datasets.

The most widespread statistical analyses for fMRI first estimates response magnitude parameters at the voxel (3D pixel) level and subsequently inserts these parameter estimates into a spatial method for hypothesis testing (in effect detecting activation). The dominant method for spatially analyzing parameter maps thresholds the image using the theory of Gaussian random fields (Worsley, 1994) to determine the voxels that can be considered active. Little consideration is given in these thresholding methods to the spatial properties of activation patterns except that the threshold is sometimes adjusted to either only consider voxel groups of at least  $n$  contiguous voxels, or a threshold is arbitrarily set and the size of the maximally sized supra-threshold region is considered as a test statistic (Friston, Worsely, Frackowiack, Mazziotta and Evans 1994). We here build on the basis of voxel-level analysis, but offer a spatial modeling methodology that goes beyond thresholding techniques for detecting activated areas, by considering the *a priori* expected spatial characteristics of activation patterns.

The spatial model presented can intuitively be thought of as estimating the probability that an elevated area of blood oxygenation is genuine activation in response to a stimulus by considering the evidence for activation to be a trade-off between the response level (integrated over the activated area) and its spatial connectedness. This trade-off is incorporated by using a prior that is the product of a continuous and a binary Markov random field (MRF) prior, favoring the spatial contiguity of activated regions. (The multiplication is performed on a voxel by voxel basis.) Using Bayes' Theorem, this prior is then combined with information in the data via the likelihood, to make posterior inferences about activation patterns. This has an advantage over many other Bayesian spatial techniques considered thus far, in that it does not make strong assumptions about the shape of activated areas. We demonstrate the approach in a fully Bayesian setting to spatially analyze fMRI parameter estimates that have been obtained by the fitting procedure of Kornak, Haggard and O'Hagan (1999) which is outlined in Section 2.1. However, the technique presented is general enough to be applied to any kind of parameter estimates e.g.  $t$ -statistics (Frackowiack et al., 1997) or the parameter estimates of Genovese (2000).

During the time we have developed our approach, there have been many other attempts to spatially model fMRI voxel-level responses via a Bayesian approach e.g. Descombes,

Kruggel and von Cramon (1998), Hartvig (2002), Gössl, Auer and Fahrmeir (2000), Taskinen (1999), Rajapakse and Piyaratna (2001), Friston et al. (2003b) and Friston et al. (2003a). While some of these other approaches implicitly consider spatial extent versus overall magnitude through either MRF modeling or object recognition analysis, they do not have the flexibility and simplicity of the present product Markov random field representation. The flexibility of the model proposed here comes from having no model constraints on the shape of activations (apart from contiguity) and the model’s simplicity is derived from the intuitive representation of a binary activation indicator map multiplied on an element by element basis with a response level map.

The Bayesian model developed here considers the detection of activation in a way that more closely resembles the approach a neuroscientist might take when looking at a map of magnitude responses in the brain. Although the neuroscientist will intuitively search for regions where there is a large change in magnitude of BOLD signal between stimulus conditions, they will temper their decision as to whether it is likely to represent genuine activation by whether the area covered by this high signal change is large or not. Response areas consisting of just a few voxels are considered much less likely to be genuine activation than a large area of contiguous voxels with the same signal change. The use of MRF priors to increase the a priori preference for smoother and wider regions of activation mimics this decision process of trading the response magnitude with the spatial extent covered when deciding which regions define genuine activity.

In Section 2 we describe the modeling of the temporal biological process for obtaining response level estimates on a voxel-by-voxel basis. Section 3 develops an initial spatial fMRI model to analyze the response level estimates. In Section 4, we give the initial model’s posterior distribution and outline the inference procedure based on that posterior. Section 5 discusses a simulation study based on the initial model. Section 6 applies the model to real data while Section 7 applies an extended spatial model to the same real data that allows for secondary global activation. Finally, Section 8 gives some discussion and summary.

## 2 The hemodynamic response function

The archetypal BOLD HDR response to activation is displayed in Figure 1. There is a small initial dip followed by a large rise that over-compensates for the previous oxygen depletion brought about by increased blood flow, reaching a peak over the order of 4 to 8 seconds. The HDR then falls at a lower rate than the rise and subsequently dips below the baseline for a while (the ‘negative overshoot’, usually mis-named ‘undershoot’), before gradually returning to baseline. This delayed response is characterized by the BOLD HDR function  $h(\tau)$ , where  $\tau$  is the time from the onset of activity.

---

Figure 1 about here

---

## 2.1 Parameter extraction

An efficient way to extract signals is to model their archetypal form with a function that contains few parameters, but that is able to adequately describe the main features in the signal. The function we adopt to model the BOLD HDR is the Poisson function, a scaled and shifted form of the mathematical function for the Poisson density equation:

$$h_s(\tau) = \begin{cases} y_s \frac{\lambda_s^{\tau-1} e^{-\lambda_s}}{(\tau-1)!} & \tau = 1, 2, \dots, T-1 \\ 0 & \tau = 0 \end{cases} \quad (1)$$

where  $y_s \in [-\infty, \infty]$  and  $\lambda_s \in [0, \infty]$ . This function was fitted to averaged response cycles by least squares estimation. The voxel response magnitude parameters (scaling factors  $y_s$  for each voxel  $s$ ) were combined to form spatial maps that were analyzed for coherent spatial variation within the Bayesian model. We only present an outline of the approach here in order to provide a particular set of parameters for the Bayesian spatial model. For further details about this non-linear HDR modeling approach and to compare with alternative functional forms that we have fitted to the HDR, see Kornak et al. (1999) or Kornak, Haggard and Hall (2000). In order to stabilize parameter estimation, data were smoothed spatially with a 5 mm full width at half maximum (fwhm) Gaussian kernel; fwhm is the width of the kernel at half of the height of the maximum.

The HDR fitting procedure has four steps:

1. Estimate the baseline at each voxel using two minutes of resting MRI signal before and after the experiment. Note that intermediate drift of overall signal is removed at the pre-processing stage by de-trending (high-pass filtering using low-frequency Fourier basis functions) in the fMRI statistical package SPM (see e.g. Frakowiack et al., 1997).
2. Average the data across cycles to produce a ‘mean response’ cycle  $h_s(\tau)$  for each voxel  $s$ .
3. Fit the Poisson curve to each voxel’s mean response cycle using non-linear least squares estimation.
4. Create a summary map of the parameter magnitude estimates.

These summary map estimates are then used as input into the spatial model.

## 3 The spatial fMRI parameter model

The objective of the spatial model is to delineate regions of activity responding to the stimulus and to optimally map the level of responses based on the magnitude parameter

estimates. We initially propose the following model:

$$y = z \odot x + \epsilon \tag{2}$$

where  $y$  is the observed magnitude response map,  $z$  is a (unobserved) binary response indicator map of activation delineating the pattern of activation,  $x$  is the activity level pattern (only relevant in regions that are activated, i.e. where  $z = 1$ ),  $\epsilon$  is the residual error field and the operator  $\odot$  represents element-by-element multiplication, i.e. for each voxel  $s$ ,  $y_s = z_s x_s + \epsilon_s$ . Note that this model may prove adequate for many data sets. However, the dataset reported in this paper demonstrated a small scale stimulus-locked response in addition to the standard functionally localized activation. This effect is accounted for in Section 7 by incorporating a secondary response MRF ( $v$ ).

The model is applied to the data on a slice-by-slice basis rather than in three dimensions and the discussion below refers to maps of a single slice plane with  $K = m \times n$  voxels. Slices may be in any orientation within 3-D space. The choice of slice direction only depends on convenience of registration and visualization. In magnetic resonance imaging, other things being equal, the signal-to-noise ratio (SNR) depends on neural activation and voxel volume, because volume determines the strength of the magnetic signal in relation to magnetic background noise. Theoretically available resolution at  $3 \times 3 \times 3$  mm may in practice be set to  $4 \times 4 \times 4$  or  $3 \times 3 \times 8$  (anisotropic), in order to increase SNR. In a given imaging plane, between-slice resolution equivalent to within-slice resolution (isotropy) is not always required. It was considered to be more important to first test the expression of constraints of the Bayesian model near to the theoretical resolution limit for a 3 Tesla (T) scanner, i.e. at  $3 \times 3$  mm but with reasonable SNR, than to demonstrate three-dimensional operation.

When inter-slice distance and slice-thickness are chosen such that voxels form an isotropic three-dimensional lattice, extension to three dimensions should in principle not pose any extra difficulties.

### 3.1 Markov random field prior distributions

In order to perform Bayesian inference in image reconstruction problems, it is necessary to formulate a prior distribution for the ‘true’ image of interest, in this case  $z \odot x$ . Markov random fields (MRFs) provide a useful class of models for this purpose. Guyon (1995) provides an introductory reference to MRFs on lattice networks and Winkler (1995) discusses their applications within Bayesian image analysis.

In a MRF, the distribution of values that can be taken by a particular pixel in the field is constrained by local neighborhood probability structure. This leads to a relatively simple conditional probability distribution at each pixel that only depends (probabilistically) on neighboring pixel values, and possibly on some global parameters (e.g. global mean). The neighborhood of a site  $s$  is denoted by  $\partial s$ , where each site is defined as a neighbor (or not) of site  $s$ , depending on how ‘close’ under some measure it is to site  $s$ .

When MRFs are employed as prior distributions in image analysis, they are generally unable to fully quantify prior expectations of the ‘true image’, but describe *a priori* expected characteristics of the image such as the expected smoothness of the true image

(although they can be used to impose negative correlation too). These priors act like stochastic kernel smoothing filters, Besag, York and Mollié (1991).

Formally, a random field  $x$  is Markov with respect to a neighborhood system  $\partial$  if

$$\pi(x) > 0, \quad \text{for all } x \in \mathcal{X} \quad (3)$$

$$\text{and } \pi(x_s|x_{-s}) = \pi(x_s|x_{\partial s}), \quad \text{for all } s \in S \quad (4)$$

where  $-s$  denotes the set of all pixels other than  $s$ . Condition (3) is known as the positivity condition (Besag, 1974), and is satisfied for all applications here, although Kaiser and Cressie (2000) point out that this is a stronger condition than is actually necessary. Condition 4 states that the conditional density for any site  $s$ , given the realization of all other sites, depends only on the site's neighbors ( $\partial s$ ).

The Hammersley-Clifford Theorem (see Besag 1974 or Cressie 1993) provides an approach to constructing MRFs, by providing conditions for an MRF which are equivalent to the standard definition:

$x$  is an MRF with respect to a neighborhood system  $\partial$  if and only if

$$\pi(x) > 0, \quad \text{for all } x \in \mathcal{X} \quad (5)$$

$$\text{and } \pi(x) = \frac{1}{K} \exp \left\{ -\frac{1}{T} \sum_{C \in \mathcal{C}} V_C(x) \right\}, \quad \text{for all } x \in \mathcal{X} \quad (6)$$

where  $K$  is the normalizing constant and  $T$  is known as the *temperature* parameter, so named because of the role it plays in statistical physics. The function  $V_C$  is the *clique potential* of the clique  $C$  and depends only on pixel values  $x_s$  such that  $s \in C$ . (A *clique* is a subset of pixels such that every pixel is a neighbor of every other pixel in the subset.) Condition 6 states that  $x$  has a *Gibbs distribution*, i.e. it is the product of exponential functions of clique potentials and therefore, the theorem states that every Gibbs distribution is an MRF and vice-versa, provided that the positivity condition holds.

The purpose of the smoothing MRF priors used in the spatial fMRI parameter model has an emphasis slightly different from that typically seen in image reconstruction. For the fMRI analysis, as in image modeling generally, the objective is to split the voxel parameter magnitudes into separate constituent parts, i.e. signal (here activation) and noise. The role of the smoothing priors here, is to describe the smoothness characteristics of these individual parts and aid separation on that basis. Therefore the location of activation is determined from two sources of evidence: i) response magnitude level and ii) smoothness or connectedness properties. This is because activation is expected to give a higher magnitude signal than noise and also to have broader spatial frequency characteristics. The smoothing properties of the priors are not fully exploited in their usual sense, because the original data were spatially smoothed to aid parameter estimation (see Section 2.1). However, if the parameter estimation procedure is incorporated within a fully spatio-temporal framework, then the need for pre-smoothing would be removed, since spatial smoothness constraints on the parameters would be incorporated within their prior distributions.

The MRF models used here are mainly based on the *conditional auto-regressive (CAR)* or *auto-Gaussian* process, which is a generalization into two dimensions of the conditional definition for the standard auto-regression time series process and for the auto-Gaussian

case used here, the conditions of the Hammersley-Clifford Theorem have been shown to be satisfied in Besag (1974). Cressie (1993), Ripley (1981, 1988) and Besag (1974) give further details on the derivation of two-dimensional CAR processes.

In the rest of this Section we detail the individual components of the spatial model along with their corresponding prior distributions (where required).

### The $y$ map

Let  $y = \{y_s, s = 1, \dots, K\}$  be the map of voxel BOLD hemodynamic response parameters to the stimulation paradigm, where  $y = (y_1, \dots, y_K)^T$  is the corresponding column vector. For our purposes,  $y$  consists of the estimated Poisson magnitude parameters as obtained in Section 2.1.

### The $z$ field and its prior distribution

The field  $z = \{z_s, s = 1, \dots, K\}$  is a hidden binary random field on the same set of sites (voxels) as the set  $y$ , with voxel locations being assigned the values 1 or 0 to indicate the presence or absence of activity respectively. The  $z$ -field provides a summary of location of activation, which is of interest in determining the existence, location, extent and pattern shift of activity and acts as a ‘mask’ or ‘switch’ for the  $x$  ‘response level’ map (described later). A similar approach used in texture classification called *Double MRF*, for switching between MRFs (as opposed to switching a field on and off, as here) is described in Melas and Wilson (1997).

We adopt an alternative binary model to the often used auto-logistic model (e.g. Besag 1974) as a prior distribution for  $z$ , to impose the required *a priori* spatial constraint of ‘connectedness’ on the regions of activity. The model incorporates the attribute of spatial connectedness through a latent continuous random field  $w$ , which is then thresholded at  $w = 0$ . Values for  $z$  at each voxel of 0 or 1 are allocated depending on whether  $w$  is below or above threshold at the voxel, i.e.  $z_s = I_{w_s > 0}, s = 1, \dots, K$  where  $I_{w_s > 0}$  is the indicator map for where  $w > 0$ . Note that  $z|w$  is wholly deterministic and hence the prior for  $z$  is fully defined through the prior for  $w$ . This approach has been used by De Oliveira (2000) in the reverse problem, where the desire is to predict an underlying un-observable Gaussian random field from observed binary data, and by Weir and Pettitt (1999) to estimate the strength of directional interaction parameters (and hence auto-correlation) in fully observed binary spatial lattice processes.

In contrast to the approaches taken in the works cited,  $w$  is modeled as an (improper) intrinsic Gaussian Markov random field (GMRF) of first order:

$$\pi(w) \propto \exp \left\{ -\frac{1}{2} \sum_{s \sim t} (w_s - w_t)^2 \right\} \quad (7)$$

This distribution for  $w$  is the (improper) limiting case of the CAR model corresponding to a generalization of the random walk in 2D. Strictly speaking, the thresholding of an intrinsic GMRF field has no meaning, since the impropriety of an intrinsic GMRF leads to it having an improper flat prior in respect of the level of  $w$ , i.e. only the differences

are defined and the whole field is free to shift up and down. However, when combined with the information in the data through the product of the prior and likelihood, the level of  $w$  becomes defined (i.e. the marginal distribution for each  $w_s$  becomes proper, Besag, Green, Higdon and Mengersen, 1995) and it becomes meaningful to talk about thresholding  $w$ .

The main advantage of using an intrinsic form for  $w$  is that there is no need to specify a prior for the mean level of  $w$ , and this freedom is desirable if no prior constraint on the proportion of activated voxels is required. Consequentially, there is an arbitrariness of the threshold level which implies that the variance parameter of  $w$  is also arbitrary (we fix  $\text{var}(w) = 1$ ).

The standard pairwise interaction auto-logistic model as a prior for  $z$  is foregone here for two reasons:

Firstly, there are computational difficulties in obtaining the normalizing function of the probability density for  $\delta|(z$  and all other model parameters), where  $\delta$  is the neighbor interaction parameter, in that it is an unknown function of  $\delta$ . This function would need to be either approximated at each iteration of the Markov chain Monte Carlo sampling algorithm (see Section 4) or fixed in advance. Apart from approximations being undesirable in general for obvious reasons, the estimates would have to be made by some approximation procedure which may have relatively high computational burden (see Green (1996) for a discussion and Weir (1997) for an approximation procedure).

Secondly, if the pairwise interactions of the auto-logistic model are strong enough to prevent typical realizations of the field from having local noise, then high correlation over longer ranges sets in, leading to realizations consisting wholly of 1 or 0 values with occasional isolated voxels of the opposite type (Tjelmeland and Besag 1998; Descombes and Kruggel 1999). Therefore, any posterior inferences (e.g. the number of regions of activity), based on the auto-logistic model prior for  $z$ , may be considerably biased. This problem has not been completely solved by the model suggested for  $z$  here which leads to *a priori* realizations which are too noisy (see the simulation example in Appendix A). Models do exist that produce realizations with the required properties e.g. Tjelmeland and Besag (1998), Descombes and Kruggel (1999). However, there is a considerable increase in model complexity, which would add too high a cost to the computational burden here.

### The $x$ -field and its prior distribution

The purpose of the  $x$ -field is to model the spatially varying response level of regions that are activated. The values that  $x$  takes in non-activated areas are irrelevant in terms of posterior inference, since in Equation (2),  $z \odot x = 0$  wherever  $z = 0$ , so that  $y = \epsilon$ . In activated voxels,  $z = 1$  so that  $z$  is now ‘switching’ the activity on, leading to  $y = x + \epsilon$ . The  $z$  map effectively masks  $x$  outside activated areas. The  $x$ -field is modeled as the natural logarithm of a proper GMRF of the (first-order) CAR form with constant mean  $\mu$ , i.e.  $\log(x) \sim MVN(\mu\mathbf{1}, \kappa^2(I - \beta N)^{-1})$ , see Cressie (1993) Chapters 6 and 7 where it is referred to as the conditional Gaussian or auto-Gaussian model.

The initial intention was to model  $x$  directly as a CAR process, since the CAR process

provides a simpler model than taking  $x$  as the exponential of a CAR process. In real data there do occur from time to time regions which respond negatively, i.e. apparently de-activate reliably although the physiological basis for this is not understood. As a consequence, posterior realizations of  $x$  cross the  $x = 0$  plane in order to accommodate the negative regions. Permitting  $x$  to cross the  $x = 0$  plane creates a difficulty, however, because  $z$  becomes non-identifiable in regions where  $x$  is approximately 0. Hence,  $\log x$  was subsequently modeled as a CAR process to prevent negative  $x$  regions.

$$\pi(x_s|x_{-s}, \mu, \kappa^2, \beta) = \frac{1}{\sqrt{2\pi\kappa^2}} \exp \left\{ -\frac{1}{2\kappa^2} \left[ (\log x_s - \mu) - \beta \sum_{r \in \partial s} (\log x_r - \mu) \right]^2 \right\} \quad (8)$$

This ‘log-CAR’ model has the added advantage that the positive tail of the log Gaussian distribution is ‘wider’ than that of the standard CAR. Therefore, the log-CAR model is more amenable to fitting any extremely high, but possibly relatively small (in terms of spatial extent), saturated regions of activity. This appears desirable for neuro-imaging where stimuli and tasks are arranged that are likely to stimulate highly specific brain systems.

The mean parameter  $\mu$  controls the mean level of the response parameters and  $\kappa^2$  controls the spatial ‘smoothness’ of the response level. We give examples of the estimation of these parameters through a fully Bayesian approach in subsequent sections, but it is beyond the scope of this paper to offer a study of optimal prior distributions for these parameters for determining activation. In any case, these priors will depend on the particular scanner, stimulus and hypothesis.

Various methods of considering edge effects were tried. However, because edges of the image were all far outside of the brain, the choice was found to have negligible effect on inferences regarding activated regions. The torus approach, where voxels at the edge of the image are considered to be neighbors of voxels at the opposing edge, was therefore adopted for its relatively simple code.

### Hyper-priors for $\mu$ , $\kappa^2$ and $\beta$

The hyper-parameters  $\mu$ ,  $\kappa^2$  and  $\beta$  control characteristics of the log CAR MRF that is used as a prior for  $x$ . The variance parameter  $\kappa^2$  influences the level of smoothness in the underlying CAR process, and  $\beta$  controls the strength of influence that neighboring pixel values have on each other relative to that of the field’s global mean level  $\mu$  (in the intrinsic case ( $\beta = 0.25$ ) all influence is placed on the neighboring voxel values). The forms of the hyper-prior distributions for  $\mu$  and  $\kappa^2$  are chosen primarily on grounds of plausibility, given that little is known about the distribution of these parameters. Since a range of plausible distributional forms exist, the choice is made so that they take a conjugate form where the corresponding posterior conditional distributions take a known form and are hence amenable to the Gibbs sampler (although this is not necessary since the Metropolis-Hastings algorithm can be used even when non-conjugate priors are used and the normalizing constants of the full conditionals are unknown, Besag *et al.* 1995).

The prior distribution for  $\mu$  is taken to be Gaussian with mean  $\mu_1$  and variance  $\sigma_1^2$ , i.e.  $\mu \sim N(\mu_1, \sigma_1^2)$ .

The prior for  $\kappa^2$  is taken to be inverse gamma ( $IG(b, a)$ ), where  $\kappa^2 \sim IG(b, a)$  if

$$\pi(\kappa^2) \propto \frac{1}{(\kappa^2)^{a+1}} \exp\left\{-\frac{b}{\kappa^2}\right\} \quad (9)$$

The prior distribution for  $\beta$  is chosen such that  $4\beta$  is distributed as a beta distribution,  $Be(p, q)$

$$\pi(\beta) \propto (4\beta)^{p-1} (1 - (4\beta))^{q-1} \quad (10)$$

The range of plausible values considered for  $\beta$  is  $[0, 0.25]$ , although in general  $\beta$  can take negative values in a CAR process (implying negative first-order spatial correlation).

The choice of  $\mu_1, \sigma_1^2, b, a, p$  and  $q$ , will depend on equipment specification, experimental signal-to-noise ratio (SNR), and empirical information gathered from past experiments and analysis. Also note that, for significant spatial auto-correlation to exist in  $x$ ,  $\beta$  must lie very close to the intrinsic limit of 0.25, Besag and Kooperberg (1995). Therefore, the choice of  $p$  and  $q$  will need to reflect this. However, in practice, this can cause problems by applying too much prior weight close to the intrinsic limit especially if one is interested in estimation of the MRF parameters; this is not typically the case in fMRI where the main requirement is the localization and description of regions of activation (see Section 5).

### 3.2 Formulation of the model likelihood and the $\epsilon$ map

The likelihood  $\pi(y|x, z, \sigma^2)$  in the present model is assumed to take the form

$\prod_{s=1}^K \pi(y_s|z_s, x_s, \sigma^2)$ , i.e. the  $y_s$  are assumed to be conditionally independent given  $z$  and  $x$  and  $\pi(y_s|x, z, \sigma^2) = \pi(y_s|x_s, z_s, \sigma^2)$ . The conditional distributions  $\pi(y_s|x_s, z_s, \sigma^2)$  are taken as i.i.d. Gaussian,  $N(z_s x_s, \sigma^2)$ . That is, the error in the parameter estimates is assumed to be i.i.d. zero mean Gaussian noise with constant (unknown) variance. The noise is generated by both the noise in the original voxel time series and the parameter estimation procedure.

#### Prior for $\sigma^2$

The choice of prior distribution for  $\sigma^2$  is taken as  $IG(b_1, a_1)$ . The choice of  $a_1$  and  $b_1$  will depend on many factors, but will become more precise as more data sets are acquired, under similar conditions.

## 4 The posterior distribution and inference

The priors, hyper-priors and the likelihood are combined through Bayes' theorem. All hyper-parameters and  $\sigma^2$  are assumed to be *a priori* mutually independent of each other and of  $w$  and hence  $z$ . Also note that  $z|w = I_{w>0}$  is deterministic and  $x$  and  $w$  are also assumed *a priori* independent. The ensuing full posterior distribution is:

$$\begin{aligned}
& \pi(x, z, w, \mu, \beta, \kappa^2, \sigma^2 | y) = \\
& \pi(y | x, z, \sigma^2) \pi(x | \mu, \beta, \kappa^2) \pi(w) \pi(\mu) \pi(\beta) \pi(\kappa^2) \pi(\sigma^2) \\
& \propto \frac{1}{(2\pi\sigma^2)^{\frac{K}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} (x \odot I_{w>0} - y)^T (x \odot I_{w>0} - y) \right\} \\
& \frac{[\det(I - \beta H)]^{\frac{1}{2}}}{(2\pi\kappa^2)^{\frac{K}{2}} \prod_{s=1}^K x_s} \exp \left\{ -\frac{1}{2\kappa^2} (\log x - \mu \mathbf{1})^T (I - \beta H) (\log x - \mu \mathbf{1}) \right\} \\
& \exp \left\{ -\frac{1}{2} \sum_{\langle s,t \rangle} (w_s - w_t)^2 \right\} \pi(\mu) \pi(\beta) \pi(\kappa^2) \pi(\sigma^2) \\
& \propto \frac{1}{(\sigma^2)^{\frac{K}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{s=1}^K (x_s I_{w_s>0} - y_s)^2 \right\} \\
& \frac{[\det(I - \beta H)]^{\frac{1}{2}}}{(\kappa^2)^{\frac{K}{2}} \prod_{s=1}^K x_s} \exp \left\{ -\frac{1}{2\kappa^2} \left[ \sum_{s=1}^K (\log x_s - \mu)^2 - 2\beta \sum_{\langle s,t \rangle} (\log x_s - \mu)(\log x_t - \mu) \right] \right\} \\
& \exp \left\{ -\frac{1}{2} \sum_{\langle s,t \rangle} (w_s - w_t)^2 \right\} \pi(\mu) \pi(\beta) \pi(\kappa^2) \pi(\sigma^2) \tag{11}
\end{aligned}$$

The term  $\prod_{s=1}^K x_s$  in the prior for  $x$  arises from the Jacobian of the transformation from  $x$  to  $\log x$ .

For the present model, Markov chain Monte Carlo (MCMC) sampling is used to sample each random variable in turn from the full posterior distribution. (Note that the random variables include the individual site location variables of each random field.) We employ the Gibbs sampler wherever possible and use the Metropolis-Hastings algorithm for variables for which the full conditional distribution is not easily obtained.

The field  $w$  and random variables,  $\sigma^2$ ,  $\mu$  and  $\kappa^2$ , are sampled using the Gibbs sampler with single site updating. Derivation of these full posterior conditional distributions is relatively straight-forward for these variables except in the case of  $w$ . Sampling from the conditional distribution of  $w_s$  is relatively tricky because the probability density function is discontinuous at  $w_s = 0$  where  $z_s$  changes its binary value.

$$\begin{aligned}
& \pi(w_s | w_{-s}, y, x, \mu, \beta, \kappa^2, \sigma^2) = \\
& A \exp \left\{ -\frac{1}{2\sigma^2} z_s (x_s^2 - 2y_s x_s) - 2(w_s - \bar{w}_s)^2 \right\} \\
& = \begin{cases} A \exp \{ -2(w_s - \bar{w}_s)^2 \} & w_s \leq 0 \\ A \exp \left\{ -\frac{1}{2\sigma^2} (x_s^2 - 2y_s x_s) - 2(w_s - \bar{w}_s)^2 \right\} & w_s > 0 \end{cases} \tag{12}
\end{aligned}$$

We give the derivation of the full conditional distribution for  $w_s$  in Appendix B and describe an algorithm to sample from it.

Since the conditional distributions of  $x$  and  $\beta$  are of unknown form Gibbs sampling can not be used and Metropolis-Hastings is consequently employed to sample from the

conditional distributions. We use random walk Metropolis proposal steps for individual  $x_s$  sites and for  $\beta$  (e.g. Besag et al. 1995) with appropriate adjustments at boundary values.

We used a fully Bayesian approach and based posterior inferences for all parameters on output from the MCMC algorithm (thereby incorporating all sources of uncertainty).

## 5 Simulation experiment

We performed a simulation experiment in order to ascertain the level of sensitivity in the spatial model (2) to hyper-parameter values, hyper-prior distribution choices and signal-to-noise ratio (SNR). This was assessed under two criteria: i) the quality of parameter estimates and ii) the plausibility of posterior map estimates with respect to the separation of  $z \odot x$  and  $\epsilon$ .

These simulations were performed with  $x$  as a CAR process (as opposed to  $\log x$ , which was adopted later, for the reasons stated in Section 3.1). Since  $x$  takes the form of a CAR process, it was possible to use the Gibbs sampler to sample  $x$  from the posterior distribution.

For this simulation experiment, a CAR process for  $x$  was simulated on a  $20 \times 20$  lattice with a toroidal edge process, i.e. voxels at the edge of the lattice are considered to be neighbors of pixels at the opposing edge. The parameter values chosen were  $\beta = 0.249$ ,  $\kappa^2 = 1.0$ ,  $\mu = 5$ . A mask with a disk of radius 7.5 voxels about the lattice's center was applied to the simulated data. All the masked voxels outside the disk, were set to 0, and the remainder of the CAR process was retained to represent the region of activity  $z \odot x$ , giving an approximate disk of 172 activated voxels. Three levels of Gaussian i.i.d. noise  $\sigma^2 = 0.01, 0.5, 0.1$  (high, medium and low SNR respectively) were added and the results of the simulation experiment were compared for the different SNR cases.

Non-informative improper priors were employed for  $\kappa^2$  and  $\sigma^2$ ,  $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$  and  $\pi(\kappa^2) \propto \frac{1}{\kappa^2}$ . A vague prior,  $N(0, 100000)$  was assigned to  $\mu$  and a variety of prior distributions were employed for  $4\beta$ :  $Be(0, 0)$  (improper),  $Be(1, 1)$ , (flat/uniform), and  $Be(400, 1.5)$  (strong). A burn-in sample of 1000 was discarded and a subsequent 3000 samples were used for posterior inference.

---

Figure 2 about here

---

Figure 2 shows the resulting maps for the simulated model in the  $Be(0, 0)$  case with  $\sigma^2 = 0.01$ . The map of the simulated response data  $y$  is shown in panel (a), and the other panels show the MMSE posterior estimates of the various maps in the model. The  $x$ -map (b) picks up the pattern of the CAR process within the activated region. Moving away

from the activated area, the  $x$ -field tends toward the global mean since the data no longer play any direct role in influencing the  $x_s$  values. The  $z$ -map (c) accurately demarcates the region of activity, as expected, given the high SNR. The  $w$ -map (d) shows the hidden intrinsic GMRF which is thresholded to generate  $z$ . The  $z \odot x$ -map (e) is the voxel-wise product map of  $z$  and  $x$ . This gives the fitted response map and again gives a plausible fit. The final map of residuals  $\epsilon$  (f) is interesting, since the residuals within the area of activation are of lower magnitude than elsewhere. At first sight this may seem to be a problem practically, theoretically or both. However, there is no capacity in the model or in the MMSE estimation procedure to penalize spatially correlated residuals. Here, the loss function minimizes least squares error subject to the errors being (individually) normal realizations. Hence, it appears that within the activated area,  $x$  is fitting some of the noise as well as the CAR process.

The posterior maps when employing the  $Be(1, 1)$  and  $Be(400, 1.5)$  priors appear virtually identical to the  $Be(0, 0)$  case and are therefore omitted. This indicates that posterior map estimation (of primary interest in determining activation pattern), based on MMSE estimation, is not sensitive to the choice of prior for  $\beta$  given the current parameter settings and choice of true activation pattern.

---

Table 1 about here

---

Table 1 summarizes the effect of the different priors for  $\beta$  on hyper-parameter estimation in the high SNR case ( $\sigma^2 = 0.01$ ). The marginal posterior mean estimates derived from the MCMC output for each of the hyper-parameters  $\mu$ ,  $\beta$ ,  $\kappa^2$  and  $\sigma^2$  along with the residual sum of squares (RSS), i.e.  $\sum_{s=1}^K \epsilon_s$ . The marginal parameter estimates (based on the 3000 post burn-in samples) of  $\sigma^2$  and  $\kappa^2$  reflect the fact that  $x$  is picking up some of the noise, since  $\kappa^2$  is larger than its simulated value of 1.0 and  $\sigma^2$  is smaller than its value 0.01. The  $\kappa^2$  and  $\sigma^2$  estimates appear quite robust across the different beta priors and after testing over a range of start values, we found that the procedure was robust to different start values for these hyper-parameters except when  $\kappa^2$  was started close to 0. However, note that in the  $Be(0, 0)$  case the parameter estimates for  $\beta$  and  $\mu$  are inaccurate. In particular, notice that the variance for  $\mu$  is extremely high. The choice of prior for  $\beta$  had an interesting effect on the MCMC output for  $\mu$  and  $\beta$  which can be seen in the MCMC traces for  $\beta$  and  $\mu$  when  $4\beta \sim Be(0, 0)$  ( $\sigma^2 = 0.01$ ) in Figure 3. To understand the behavior of the chains, first note that the improper  $Be(0, 0)$  prior for  $4\beta$  has a singularity at  $\beta = 0.25$  giving a relative *a priori* attraction to this value. When the sample values of  $\beta$  approach 0.25, the model for  $x$  comes close to the intrinsic Gaussian form Besag and Kooperberg (1995) and the parameter  $\mu$  becomes non-identifiable; when  $\beta = 0.25$ , the conditional distribution for  $\mu$  depends only on its prior distribution. Since a vague prior has been chosen for  $\mu$ , the parameter's MCMC output tends to wander when

samples of  $\beta$  get close to 0.25. Only when  $\mu$  returns to a plausible value, is there a chance of  $\beta$  moving away from 0.25 again. It can be seen from the traces that the variability of  $\mu$  is high when  $\beta$  is close to 0.25.

---

Figure 3 about here

---

For the  $Be(1, 1)$  prior, the *a priori* mass for  $\beta$  toward the extremes has been reduced and hence there is less tendency for the chain to reach  $\beta = 0.25$ . The mean and median estimate for  $\beta$  are now of lower magnitude (underestimates as opposed to overestimates) and the MCMC traces (not shown) exhibit much less of the non-identifiability behavior, i.e.,  $\beta$  was not stuck for long periods at 0.25. The estimates for  $\mu$  in this case are slight overestimates. For the  $Be(400, 1.5)$  prior, the estimates for  $\beta$  were the closest to the true value of any of the models. This was to be expected because the mean of the prior for  $\beta$  had its mean at the correct value. Further, because this prior placed zero mass at  $\beta = 0.25$ , the non-identifiability problem was virtually non-existent.

The choice of prior for  $\beta$  is important in terms of obtaining sensible parameter estimates for  $\beta$  and  $\mu$ . However, the results of applying these different Beta priors show that the existence of some non-identifiability of  $\mu$  has little effect on the posterior estimates of the individual maps in the model (which are of primary interest), or on the other hyperparameter estimates,  $\kappa^2$  and  $\sigma^2$ . Clearly, the non-informative  $Be(0, 0)$  prior is unsuitable, given the extent of the non-identifiability problem for  $\mu$ . However, the  $Be(1, 1)$  prior appears to give a more practical approach to modeling vague prior knowledge about  $\beta$ .

---

Table 2 about here

---

Table 2 displays the MMSE posterior parameter estimates for the different levels of  $\sigma^2$ , i.e. 0.01, 0.5, 1.0 in the  $4\beta \sim Be(1, 1)$  case. For the higher levels of  $\sigma^2$ , these estimates are inaccurate for all but  $\mu$ . It appears that at lower SNR levels, the model is unable to distinguish well between the different sources of signal, i.e. the CAR process and Gaussian noise. The posterior maps of  $z$  and  $z \odot x$  are of high quality in the  $\sigma^2 = 0.5$  case, giving similar posterior map estimates as those obtained in the  $\sigma^2 = 0.01$  case, but with higher residuals. However, when  $\sigma^2 = 1.0$  the maps begin to degrade due to uncertainty as to whether a voxel has a high magnitude value due primarily to true signal or noise.

The MCMC plots of the parameters (not shown) again appear to converge quite

quickly, i.e. within a few hundred iterations. However, the MCMC chains are exploring an incorrect region of the parameter space, but one which is plausible in terms of the data. Since the noise in the data is higher than in the previous cases, the joint marginal distribution of the hyper-parameters will be ‘flatter’. This leads to a higher expectancy for the MCMC procedure to explore incorrect regions of the parameter space and hence become stuck in incorrect modes. In this case, the residual variance  $\sigma^2$  is explaining all of the variance in the response map. This effect can be controlled to some extent by fixing  $\sigma^2$  to be small for the early iterations. This is employed in analyses of real data herein and is explained further in Section 6.

The simulation studies have shown that the procedure is capable of reconstructing the components of the model and estimating the corresponding hyper-parameters at differing levels of SNR.

As is to be expected, the accuracy of the parameter estimates also decreases as the SNR is reduced. Indeed the parameter estimates when  $\sigma^2$  was increased to 1.0 were very poor, although controlling  $\sigma^2$  at the start improves things considerably. It is also to be expected that uncertainty in the reconstruction increases as the SNR increases. However, the quality of reconstruction of the individual fields remains relatively high up to  $\sigma^2 = 0.5$ , even though the parameter estimates are not very precise, further indicating that the model is not very sensitive to perturbation of the hyper-parameter estimates. This means that the main requirement is met, since inference in fMRI is not particularly concerned with the MRF hyper-parameters, but with observed characteristics of the reconstructed fields, e.g. the number, size and location of activated regions.

## 6 Analysis of real data

We considered datasets acquired within a simple periodic on/off experimental paradigm (although the method is readily extended to consider more complex experimental paradigms). Specifically, the data considered here were acquired on a 3T scanner, in an on/off paradigm with 32 cycles. Each cycle contained 7s of speech (on), followed by 28s of silence (off). The subject was requested to attend to the speech but no response was required. A further 2 minutes of imaging in the silent condition were acquired both before and after the experiment, in order to obtain stable baseline estimates. Images were acquired at 584 time points with an inter-scan interval of 2.33 seconds. Each image has 8 parallel coronal slices, consisting of regular lattices with  $256 \times 256$  voxels. Voxel dimensions are  $3 \times 3 \times 8$  mm, the 8 mm being the distance between slices.

We initially adopted the spatial model (2); in Section 7 we give an extension to this model that leads to superior results. The Poisson magnitude parameter estimates for individual axial slices crossing through primary auditory cortex were analyzed for two subjects. We considered one slice from subject 1 for the application of this first model and for the extended model, two neighboring slices from subject 1 and one slice from subject 2.

In order to prevent the possibility of negative values for  $x$ ,  $\log x$  was modeled as a CAR process as outlined in Section 3.1. Other differences for this model compared with that

used in the simulation experiment were as follows:  $\mu$  was the mean of  $\log x$ , started at  $3.5 \approx \log 33$  and a more precise prior variance on  $\mu$  ( $\sigma_1^2 = 0.5$ ) than previous was adopted for reasons discussed below. The prior for  $\beta$  was that used in the informative simulations, i.e.  $Be(400, 1.5)$  (see Section 5). This was required in order to prevent  $\beta$  gradually going to 0, which would lead to  $x$  becoming a constant field with  $\kappa^2$  non-identifiable. (Note that these severe restrictions are required because the model is inadequate for these data as shown later in this section.) As in the  $x \sim \text{CAR}$  restoration,  $\sigma^2$  was again kept low at 0.001 for an initial set of samples, to allow  $\kappa^2$  to settle down. The  $x$ -field was started as a constant field at approximately the maximum of the observed values, i.e. 35, for the reasons stated below.

The parameter start values of the MCMC sampling procedure were as in the simulation experiment, except that  $x$  was started at 20, corresponding to just over half of the maximum estimated magnitude parameter of 35.0, and the  $\sigma^2$  parameter was kept constant at 0.01 for the first 100 iterations and subsequently was sampled in the normal manner via the Gibbs sampler. Note that the prior distributions we use are incomplete in that they do not fully characterize prior knowledge of activation patterns (see Section 3), and therefore regions of the posterior will exist that are feasible in terms of the data but do not give realistic realizations (in terms of our true prior expectations).

The step size  $h$  for the uniform proposal distributions for  $\beta$ ,  $x_s$  were chosen to give an acceptance rate of between 30% and 70%, as advised by the rule-of-thumb suggested by Besag et al. (1995).

In this analysis,  $\sigma^2$  was initially kept constant in the MCMC sampling process in order to let the other parameters settle into a sensible region of the parameter space. If  $\sigma^2$  was not fixed at the start of the chain its MCMC sample value would increase rapidly, so that it could explain all of the variation. The corresponding value for  $\kappa^2$  shrinks (so that virtually no fine scale variability is explained by the CAR process), with  $x$  taking a near constant value approximately equal to the log of the starting value for  $\mu$ . In this situation (i.e. when  $\sigma^2$  explains all of the fine scale spatial variance), the  $z$ -map indicates 1 in strong regions of activity, with  $x$  taking up part of the response value (approximately equal to  $\mu$ ) and with the difference being absorbed through  $\sigma^2$ . The  $\beta$  parameter plays no significant role in the posterior mass and is free to wander. Effectively, the sampler has entered a mode that does not capture the essential structure, from which it is difficult to escape. When the parameters of the model are constrained in the above stated manner, the MCMC traces of the hyper-parameters all appear to behave well.

An alternative strategy for stopping  $\sigma^2$  from exploding would be to put a strong prior on  $\sigma^2$ . This seems undesirable in the present context, since such a prior would bias posterior estimates toward the ‘center’ of the prior distribution (for which currently no evidence exists). The extent of such a bias was tested using the inverse gamma prior for  $\sigma^2$ , which is conjugate in the model. It was found that unrealistically strong prior information was required to keep  $\sigma^2$  under control.

---

Figure 4 about here

---

Figure 4 displays the original Poisson parameter estimates along with the posterior estimates of the fields in the model. The posterior estimate of  $x$  (panel b) provided the required kind of field. It fitted undulation in the activated area in the left hemisphere, but stayed largely flat outside the area of high activity and away from  $x = 0$  (which was guaranteed when strong prior information was given for the location of  $\mu$  and when  $x$  was started well away from  $x = 0$ ). However, the residuals had a spatially correlated structure and had a high magnitude (which can be seen in panel f). Note that the  $v$  map could not simply be an artifact of the spatial smoothing in the data because the smoothing kernel only has significant mass for the width of about 3 or 4 voxels. Therefore the main region of activation was not being smoothed to the extent of influencing the whole brain.

The spatially correlated effect invites an extension of the Bayesian model, which introduces a second CAR MRF that acts across the whole brain. Hitherto, functional neuroimaging has been largely confirmatory in character, with a concentration on the avoidance of false positive errors via adjustments for multiple testing. Activation effects synchronized to the stimulation/task cycles have been confirmed initially in brain areas known from other types of data (e.g. functional deficits from brain lesions), to be involved in functions plausibly related to stimulus analysis or task performance. The issues raised by unexpected forms of ‘activation’ have largely been deferred. The spatially correlated residual structure is stimulus-locked but appears globally (i.e. including regions not thought to be functionally relevant to the task/stimulation).

In the context of the spatial model we have considered so far, the spatially correlated residual structure explains why  $\mu$  needs a relatively strong prior and why  $x$  has to be started far from  $x = 0$ . There are effectively two separate correlation patterns (on different scales) which  $x$  can try to fit, so it was necessary to use strong priors to prevent  $x$  fitting the global, but lower magnitude, correlation structure.

## 7 Adding the intermediate $v$ -field to the Bayesian model

In order to take into account the spatially correlated intermediate response, the spatial model in Equation (2) is extended to that of Equation (13)

$$y = z \odot x + v + \epsilon \tag{13}$$

where a further Markov random field  $v$  has been added to represent this secondary activation. The other components are unchanged. The new  $v$ -field is modeled *a priori* by a CAR process. The hyper-parameters for this CAR process are  $\lambda$  (mean),  $\gamma$  (interaction) and  $\eta^2$  (variance) which all have hyper-priors of the same format as those given to the corresponding hyper-parameters of the  $\log x$  process. The  $x$ ,  $v$  and  $w$  processes are assumed to be mutually independent *a priori*.

The MCMC sampling procedure described in Section 4 is extended to incorporate the  $v$ -field and its hyper-parameters. The sampling procedure for these hyper-parameters is similar to that of the corresponding hyper-parameters in the  $\log x$  process, and indeed the same subroutines are used for the equivalent hyper-parameters of the  $x$ - and  $v$ -fields.

A further change to the sampling code is brought about by the addition of the  $v$ -field terms into the likelihood, i.e. the term in the exponential of the likelihood changes from  $(x \odot z - y)^T (x \odot z - y)$  for the model in Equation (2), to  $(x \odot z + v - y)^T (x \odot z + v - y)$ . Any variables which have the likelihood entering into their conditional distribution require a corresponding adjustment in their full conditionals.

## 7.1 Application of the new model

The starting values and prior distributions are given in table 3. The hyper-prior distributions are more specific than those used in Section 5; the addition of a further MRF into the model has increased the sensitivity to the starting values and choice of hyper-priors. A burn-in period of 20000 iterations was employed with a subsequent sample of 60000 iterations used for inference. The  $\sigma^2$  chain is fixed for a longer period of 500 iterations (at 0.01), as the increase in number of parameters leads to the other hyper-parameters potentially requiring a longer time to become stable within the correct region of their joint sub-space.

---

Table 3 about here

---

In general, for known auditory cortical areas we expect activation at either or both left and right extremes about half way up the coronal slice, together with the less reliable possibility of showing activation at the bottom of the mid-brain. Two subjects' data are presented illustrating this expected activation, one at good SNR and the other at poor.

## Samples from the posterior

### Subject 1 analysis.

In figure 5, the MCMC appears to have converged with the conventional activity assigned to  $z \odot x$ , and the secondary activity to  $v$ . The figure shows the Poisson parameter map and the MMSE estimates for the model maps of slice 1, subject 1. The starting values and hyper-parameters used are listed in table 3.

---

Figure 5 about here

---

After exploring many combinations, it appeared that the structure of posterior estimate maps are largely independent of either the starting values or hyper-priors. In general, strong priors had to be placed on  $\mu$ ,  $\beta$  and  $\kappa^2$  (as well as  $\sigma^2$ ). The  $x$ -field parameters had to be controlled such that the  $x$ -field did not become ‘too flexible’ and have the capability of fitting the  $v$ -response. The hyper-priors for  $\beta$  and  $\sigma^2$  were truncated such that  $\beta$  could not go above 0.2495 and  $\sigma^2$  could not go above 3.0. The first restriction was to prevent  $\mu$  becoming non-identifiable. The problem in that case, is that  $x$  approached  $x = 0$  in non-activated regions and hence false regions of activity were indicated. The second restriction, that on  $\sigma^2$ , prevented the (rare) event of the MCMC jumping to a different mode where  $\sigma^2$  would explain the majority of spatial variation.

The final residual error map in figure 5 (and later maps of residuals in this chapter) appears flat. Of course, the residuals are not totally flat, but the flat image reflects the fact that the residual magnitudes were small compared to the magnitudes of the other fields. When re-scaled as in figure 6, it can be seen that although much of the spatial structure had been removed, we cannot say that it is spatially uncorrelated. This is not surprising because, spatially correlated residuals are not penalized in the decision process for MMSE, which is based on a least squares error loss function.

---

Figure 6 about here

---

Table 4 lists the marginal MCMC output mean, median and standard deviations for the hyper-parameters  $\mu$ ,  $\beta$ ,  $\kappa^2$ ,  $\sigma^2$ ,  $\lambda$ ,  $\gamma$  and  $\eta^2$ , as well as the total RSS for slice 1 of subject 1. The parameter estimates all seem reasonable, but any interpretation has to be considered with care because of the high precision of the hyper-priors.

---

Table 4 about here

---

In order to examine the variability of the MMSE map estimates, samples of each of  $z \odot x + z \odot v$  and  $v$  (not shown) were examined. Some variability in the maps existed, but this was quite small. This appears to reflect the high SNR content and consequent

specificity in the parameter maps.

---

Figure 7 about here

---

Figure 7 shows the MMSE maps for slice 2 of subject 1. The results again separate the parameter estimates into their constituent components ( $z \odot x$ ,  $v$  and  $\epsilon$ ) in a coherent manner and conventional activation can be observed here bilaterally along with slight activation in the lower mid-brain. The starting parameters and hyper-prior distributions were all kept the same as those used for slice 1.

### Subject 2.

Figure 8 shows the MMSE maps for a single slice examined in subject 2, with the starting parameters and hyper-prior distributions as used in subject 1, except that  $\mu$  was started at 2.0 (and is given a hyper-prior with mean 2.0) and the  $x$ -field is started at 24.0. The reason these changes were made was because the magnitude parameter estimates were in general smaller and the reduced start values enabled the  $x$ -field values of the MCMC to reach the correct region more quickly. This was required to prevent the chain entering an incorrect mode; note that fMRI response magnitude is recorded on an arbitrary scale and hence some such calibration of the  $x$ -field is necessary. Despite much lower SNR in the data compared with subject 1, the posterior maps appear coherent, with detected activation in right auditory cortex and the lower mid-brain. Note that the  $v$ -field effect displayed in panel d is less evident than in subject 1.

---

Figure 8 about here

---

Table 5 gives the MMSE parameter estimates for the slice of subject 2. The reduced  $\eta^2$  estimate reflects the lesser spatial structure in the  $v$ -field. For parameter maps such as this one, where there is no strong secondary field it would be better to stick with the more parsimonious model of Section 6. The lower SNR of subject 2 was further reflected in sample maps from the posterior of  $z \odot x + z \odot v$  and  $v$  which had higher variability than those of subject 1. However, the visible variability between sampled maps was still not high.

## 8 Discussion

### 8.1 Sensitivity of hyper-parameters

The general high sensitivity of the hyper-parameter values is likely to be due to the limited capacity of pairwise structure models to differentiate second-order structure (i.e. the structure of second differences) in the parameter maps. In particular, separation of the  $x$  and  $v$  components of the image would be more robust if their respective levels of smoothness helped determine more directly the deconvolution of the two fields. The conditional specification of the CAR prior models used for  $\log x$  and  $v$  is mis-leading in this respect. When  $\beta$  is close to 0.25, Equation (8) implies that the voxel response has highest conditional probability mass when it is the average of its neighboring values (ignoring the consequentially small effect of the overall mean of the process), i.e. it depends on how well it interpolates the neighborhood. This appears to define smoothness in terms of second-order structure. However, the joint CAR distribution only depends on singleton and pairwise functions and hence the second-order structure is not controlled by the CAR process other than through the mean function.

For example, consider the intrinsic CAR case. The individual conditional distributions seem to imply that configurations have maximum probability of occurring if pixels are the average of their neighbors (i.e. have a smooth second-order structure). However, the posterior joint distribution implies that only the sum of the pairwise differences is relevant to the probability of a configuration, which does not explicitly encourage smooth second-order structure.

For the fMRI parameter maps, large smoothly rising and falling bumps need to be distinguished from undulation of lower magnitude and higher spatial frequency. The current method cannot very well differentiate the two textures, because the effective squared sum of pair-wise differences may be similar.

To obtain further intuitive understanding of why the conditional form can be misleading, consider any path of pixels in the lattice for which the steps in the path are made up of first-order neighbors. If the path eventually returns to its starting point it forms a closed loop. Clearly, there are a large number of different closed loops for lattices of substantial size. The pixel values along the loop depend conditionally only on their direct neighbors, however, the spatial structure imposes further conditions, since all the pairwise differences in closed loops have to sum to 0.

MRF models for continuously distributed magnitude data which aim to directly model second-order spatial structure would incorporate second-order clique functions (at some

increased computational cost), which incorporate second-order structure directly, into the joint distribution. This is a topic for future research.

## 8.2 Alternative model formulations

If the  $v$ - and  $x$ -effects are not additive in the way the model assumes, an alternative formulation to the  $y = z \odot x + v + \epsilon$  model used so far, could be  $y = z \odot x + (1 - z) \odot v + \epsilon$ . In this model the  $v$ -field is ‘switched off’ in regions where conventional activation occurs.

However, when the  $(1 - z) \odot v$  model was implemented via some minor changes to the MCMC code, there was only a negligible difference to the posterior maps. The robustness to the choice between these two models is clearly desirable. The original additive  $y = z \odot x + v + \epsilon$  is retained for reasons of simplicity and practicality, in the absence of any published physiological evidence for the preference of either model.

Adoption of a fully spatio-temporal approach would enable the potential of incorporating an additive  $v$ -field with a delay at each voxel location differing from that of the  $x$ -field. This would lead to a more ‘genuine’ additivity of  $v$  and  $z \odot x$ , though at considerably higher computational cost.

## 8.3 Conclusion

The full Bayesian model as developed in this report incorporates spatial constraint for the estimation of stimulus response patterns in fMRI experiments via the use of Markov random field models for prior distributions. There are two levels at which the MRF models operate to provide a priori spatial smoothness. The first is at the level of the parameter estimates (within the continuous  $x$ -field) which is little different in its force from classical/deterministic smoothing via filters, except that it is placed within the probabilistic framework of MRFs. The hyper-parameters for the  $x$ -field CAR model can be seen as corresponding to the width/shape of smoothing kernels in the classical framework and can similarly be optimized for particular dataset resolutions or types of problem. The second is at the decision level for the existence of activity within voxels, i.e. for the binary  $z$ -field. The MRF prior in this case spreads the decision process to take into account the classification as active or inactive of neighboring voxels. This implements the assumption that a more functionally sensitive analysis will result by imposing a model in which the nature of the hemodynamic response cannot be highly local (true) and is broadly categorical rather than continuous in nature (largely true unless special designs are used to reveal proportional effects). This handles data on the magnitude of activation in a consistent and near-optimal way. This second point would normally be left to the eye of the radiologist, radiographer or neuroscientist and be allowed as “interpretation” in a clinical context. This is a “skilled” reading of the activated area rather than the individual magnitude parameters. In a scientific study this is the aspect of activation that it is difficult to specify and data-reduce in any objective way for the subsequent evaluation of the presence (or absence) of regional activity relating to physiological stimulus.

The potential to fit entirely positive or negative Poisson curves (via the scaling parameter  $y_s$ ), but with variable delays (through the mean parameter  $\lambda_s$ ), has led to the

discovery of a secondary global signal response which is locked to the stimulus paradigm. It is likely that an improved magnitude resolution from using a variable delay, similar to that demonstrated when fitting functions permitting (and generally having) both a positive and negative excursion, is responsible for enabling this finding. Its physiological significance is unclear; it may represent ripple in the “hemodynamic net”, or be an expression of the negative activations that are occasionally found. Incorporation of delay into the Bayesian model may help to give further independent information concerning the probability voxels are activated. The finding of the global secondary field (see forthcoming paper by Kornak, Haggard and Hall for further details) has led to an extension of the Bayesian spatial model, so that this secondary activity can be distinguished from conventional activity. This places greater emphasis on the role of the MRF priors, as they are required to characterize the local structure of the different fields and hence aid the separation of the two types of response.

The model as extended to include the  $v$ -field, can usefully separate parameter estimate maps into the separate components of interest: conventional primary activity (observed in the expected locations for the stimulus used), secondary global activity via the intermediate field and residual noise. This spatial model requires careful implementation for a number of reasons. Firstly, there is the inherent non-identifiability of the CAR model mean and interaction parameters. Secondly, there is the requirement for prior response level calibration and thirdly, care is needed in specifying prior distributions for the hyper-parameters to aid separation of the multiple component fields. However, the second and third problems should be alleviated as improved prior knowledge of the relationships between the model’s hyper-parameters are obtained from future data sets.

The full Bayesian model as developed in this report incorporates spatial constraint via Markov random field models as prior distributions. There are two points at which the Markov random field models operate to provide a posteriori spatial smoothness. The first is at the level of the parameter estimates (within the  $x$ -field which is little different in its force from classical smoothing via filters except that it is placed within the probabilistic framework of MRFs and an implicit model for the spatial covariance structure. The hyper-parameters within the  $x$ -field can be seen as parameters that can be optimized for data or problem area (via posterior estimation) after some preliminary smoothing of the data for reliability of the HDR parameter estimation. The second point is at the decision level for the existence of activity within voxels, i.e. for the  $z$ -field. The MRF prior in this case spreads the decision process to take into account the classification as active or inactive of neighboring voxels. This implements the assumption that a more functionally sensitive analysis will result by imposing an analysis model in which the nature of the hemodynamic response cannot be highly local (true) and is broadly categorical rather than continuous in nature (largely true unless special designs are used to reveal proportional effects). This handles data on the magnitude of activation in a consistent and near-optimal way. This second point would normally be left to the eye of the radiologist, radiographer or neuroscientist and be allowed as “interpretation” in a clinical context. This is a “skilled” reading of the activated area rather than the magnitude parameter of activation. In a scientific study this is the aspect of activation that it is difficult to specify and data-reduce in any objective way for the subsequent probabilistic evaluation of the

presence (or absence) of regional activity relating to a particular physiological stimulus.

## Acknowledgments

We would like to thank the following: Professor Ian L. Dryden for helpful advice and constructive criticism, Dr Bruce Dunham for his support in setting up this project as well as providing useful discussion, Dr Debbie Hall for considerable aid in experimental design and data preparation, Professors Peter J. Green and Peter Morris for their initial examination and criticism of this work. The data were acquired at the Magnetic Resonance Center at the University of Nottingham and the work was supported by the UK Medical Research Council (MRC) as core programme of the MRC Institute of Hearing Research and as training studentship 5145 to John Kornak.

Computation was performed using Fortran 90. NAG library routines were incorporated for the simulation of random variables, as well as for calculating density and inverse density function values. The code is available from the corresponding author upon email request: kornak@itsa.ucsf.edu.

## A Appendix A: Binary map simulations

Simulations from the prior of  $w$  and corresponding  $z$ -fields thresholded to give 30% and 10% activated voxels (figure B), show that the spatial correlation may be too low to correctly model the expected extent of *a priori* connectedness in  $z$ . We can therefore consider the prior distribution to may be too conservative in imposing *a priori* spatial correlation. The simulations of (almost) intrinsic GMRFs were obtained with a conditional variance parameter for  $w_s|w_{-s}$  of 1, interaction parameter as close as numerically possible to the intrinsic limit and no edge correction using `GMRFSim` (Rue, 1999).

---

Figure A.1 about here

---

## B Conditional distribution for $w_s$

We obtain the full conditional distribution for  $w_s$  as follows:

$$\begin{aligned}
\pi(w_s|w_{-s}, y, x, \mu, \beta, \kappa^2, \sigma^2) &\propto \\
&\exp \left\{ -\frac{1}{2\sigma^2} \sum_{s=1}^K (x_s I_{w_s > 0} - y_s)^2 \right\} \exp \left\{ -\frac{1}{2} \sum_{\langle s, t \rangle} (w_s - w_t)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} (x_s z_s - y_s)^2 - \frac{1}{2} \sum_{t \in \partial s} (w_s - w_t)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} (x_s^2 z_s^2 - 2y_s x_s z_s) - \frac{1}{2} \sum_{t \in \partial s} (w_s^2 - 2w_s w_t) \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} z_s (x_s^2 - 2y_s x_s) - 2(w_s - \bar{w}_s)^2 \right\}
\end{aligned}$$

Therefore

$$\begin{aligned}
\pi(w_s|w_{-s}, y, x, \mu, \beta, \kappa^2, \sigma^2) &= \\
&A \exp \left\{ -\frac{1}{2\sigma^2} z_s (x_s^2 - 2y_s x_s) - 2(w_s - \bar{w}_s)^2 \right\} \\
&= \begin{cases} A \exp \{-2(w_s - \bar{w}_s)^2\} & w_s \leq 0 \\ A \exp \left\{ -\frac{1}{2\sigma^2} (x_s^2 - 2y_s x_s) - 2(w_s - \bar{w}_s)^2 \right\} & w_s > 0 \end{cases}
\end{aligned}$$

Since  $\exp \left\{ -\frac{1}{2\sigma^2} (x_s^2 - 2y_s x_s) \right\}$  is a constant factor, i.e. does not depend on  $w_s$ , the conditional distribution for  $w_s$  is proportional to a  $N(\bar{w}_s, \frac{1}{2})$  distribution both above and below zero. However, it is weighted such that the weight for the region where  $w_s > 0$  is  $\exp \left\{ -\frac{1}{2\sigma^2} (x_s^2 - 2y_s x_s) \right\}$  times that of the region where  $w_s \leq 0$ . To sample  $w_s$  it is therefore necessary to find the constant  $A$ . This is achieved by equating the sum of the integrals for the two regions to 1.

$$\begin{aligned}
1 &= \int_{-\infty}^0 A \exp \{-2(w_s - \bar{w}_s)^2\} dw_s + \\
&\int_0^{\infty} A \exp \left\{ -\frac{1}{2\sigma^2} (x_s - 2y_s x_s) - 2(w_s - \bar{w}_s)^2 \right\} dw_s \\
\frac{1}{A} &= \int_{-\infty}^0 \exp \{-2(w_s - \bar{w}_s)^2\} dw_s + \\
&\exp \left\{ -\frac{1}{2\sigma^2} (x_s^2 - 2y_s x_s) \right\} \int_0^{\infty} \exp \{-2(w_s - \bar{w}_s)^2\} dw_s \\
A &= \left\{ \sqrt{2\pi} \Phi \left( -\sqrt{2}\bar{w}_s \right) + \exp \left\{ -\frac{1}{2\sigma^2} (x_s^2 - 2y_s x_s) \right\} \sqrt{2\pi} \left[ 1 - \Phi \left( -\sqrt{2}\bar{w}_s \right) \right] \right\}^{-1}
\end{aligned}$$

where  $\Phi$  indicates the standard normal distribution  $N(0, 1)$ . In order to obtain a Gibbs sample for  $w_s$  the following steps are performed:

1. Calculate  $Pr(w_s \leq 0) = A\sqrt{2\pi} \Phi(-2\bar{w}_s)$
2. Sample a uniform  $U[0, 1]$  random variable  $U$
3. if  $U < Pr(w_s \leq 0)$  set  $w_s = F^{-1}\left(\frac{UF(0)}{Pr(w_s \leq 0)}\right)$   
 else set  $w_s = F^{-1}\left(1 - \left[\frac{(1-U)(1-F(0))}{1-Pr(w_s \leq 0)}\right]\right)$

where  $F$  is the cumulative density function of the  $N(\bar{w}_s, \frac{1}{4})$  distribution;  $I_{w_s > 0}$  gives the corresponding value of  $z_s$ .

## References

- Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), “Bayesian Computation and Stochastic Systems,” *Statistical Science*, 10, 3–66.
- Besag, J. and Kooperberg, C. (1995), “On Conditional and Intrinsic Autoregressions,” *Biometrika*, 82, 733–746.
- Besag, J. E. (1974), “Spatial interaction and the statistical analysis of lattice systems (with discussion),” *Journal of the Royal Statistical Society, series B*, 36, 192–236.
- Besag, J. E., York, J., and Mollié, A. (1991), “Bayesian Image Restoration with Two Applications in Spatial Statistics,” *Ann. Inst. Statist. Math.*, 43, 1–59.
- Cressie, N. A. C. (1993), *Statistics for Spatial Data, Revised Edition*, Wiley.
- De Oliveira, V. (2000), “Bayesian Prediction of Clipped Gaussian Random Fields,” *Computational Statistics and Data Analysis*, 34, 299–314.
- Descombes, X. and Kruggel, F. (1999), “A Markov Pixon Information Approach for Low-Level Image Description,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 482–494.
- Descombes, X., Kruggel, F., and von Cramon, D. Y. (1998), “Spatio-Temporal fMRI Analysis Using Markov Random Fields,” *IEEE Transactions on Medical Imaging*, 17, 1028–1039.
- Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., and Mazziotta, J. C. (1997), *Human Brain Function*, Academic Press.
- Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., and Ashburner, J. (2003a), “Classical and Bayesian Inference in Neuroimaging: Applications,” *NeuroImage*, 16, 484–512.
- Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2003b), “Classical and Bayesian Inference in Neuroimaging: Theory,” *NeuroImage*, 16, 465–483.

- Friston, K. J., Worsley, K. J., Frackowiack, R. S. J., Mazziotta, J. C., and Evans, A. C. (1994), “Assessing the Significance of Focal Activations Using their Spatial Extent,” *Human Brain Mapping*, 1, 210–220.
- Genovese, C. R. (2000), “A Bayesian Time-Course Model for Functional Magnetic Resonance Imaging Data (with comments),” *Journal of the American Statistical Association*, 95, 691–719.
- Gössl, C., Auer, D., and Fahrmeir, L. (1999), “A Bayesian approach for spatial connectivity in fMRI, Poster No. 483,” *Biometrics*, 57, 554–562.
- Green, P. J. (1996), “MCMC in image analysis,” in Gilks, W. R. et al., eds., *Markov chain Monte Carlo in practice*, chap. 21, Chapman & Hall, pp. 381–399.
- Guyon, X. (1995), *Random Fields on a Network — Modeling, Statistics, and Applications*, Springer-Verlag.
- Hartvig, N. V. (2002), “A Stochastic Geometry Model for Functional Magnetic Resonance Images,” *Scandinavian Journal of Statistics*, 29, 333–354.
- Kaiser, M. S. and Cressie, N. (2000), “The Construction of Multivariate Distributions from Markov Random Fields,” *Journal of Multivariate Analysis*, 73, 199–220.
- Kornak, J., Haggard, M. P., and Hall, D. (2000), “fMRI functional mapping with alternative parameters of haemodynamic response in auditory cortex,” *British Journal of Audiology*, 34, 2, 96–97.
- Kornak, J., Haggard, M. P., and O’Hagan, A. (1999), “Parameterisation of the BOLD Haemodynamic Response in fMRI Incorporated within a Bayesian Multiplicative Markov Random Field Model,” in Mardia, K. V. et al., eds., *Spatial Temporal Modelling and its Applications*, Leeds University Press. The 18th Leeds Annual Statistical Research (L.A.S.R.) Workshop, pp. 27–30.
- Melas, D. E. and Wilson, S. P. (1997), “Texture Based Image Segmentation using the Double MRF Model,” in Mardia, K. V. et al., eds., *The Art and Science of Bayesian Image Analysis*, Leeds University Press. The 17th Leeds Annual Statistical Research (L.A.S.R.) Workshop, pp. 169–174.
- Rajapakse, J. C. and Piyaratna, J. (2001), “Bayesian Approach to Segmentation of Statistical Parametric Maps,” *IEEE Transactions on Biomedical Engineering*, 48, 1186–1194.
- Ripley, B. D. (1981), *Spatial Statistics*, Wiley.
- (1988), *Statistical Inference for Spatial Processes*, Cambridge University Press.

- Taskinen, I. (1999), “A Bayesian Smoothing Method for Detecting Brain Activations,” in Mardia, K. V. et al., eds., *Spatial Temporal Modelling and its Applications*, Leeds University Press. The 18th Leeds Annual Statistical Research (L.A.S.R.) Workshop, pp. 25–26.
- Tjelmeland, H. and Besag, J. E. (1998), “Markov Random Fields with Higher-order Interactions,” *Scandinavian Journal of Statistics*, 25, 415–433.
- Weir, I. S. (1997), “Fully Bayesian Reconstructions from Single-photon Emission Computed Tomography Data,” *Journal of the American Statistical Association*, 92, 49–60.
- Weir, I. S. and Pettitt, A. N. (1999), “Spatial Modelling for Binary Data using a Hidden Conditional Autoregressive Gaussian Process: a Multivariate Extension of the Probit model,” *Statistics and Computing*, 9, 77–86.
- Winkler, G. (1995), *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer.
- Worsley, K. J. (1994), “Local Maxima and the Expected Euler Characteristic of Excursion Sets of  $\chi^2$ ,  $F$  and  $t$  Fields,” *Advances in Applied Probability*, 26, 13–42.

Table 1:

Posterior parameter estimates for the different beta priors,  $\sigma^2 = 0.01$ .

Parameter	$Be(0, 0)$		$Be(1, 1)$		$Be(400, 1.5)$	
	mean	std dev	mean	std dev	mean	std dev
$\mu$	4.77	53.49	5.31	0.689	5.27	1.47
$\beta$	0.2499252	0.00018	0.242676	0.00060	0.24911	0.00071
$\kappa^2$	1.07	0.118	1.11	0.124	1.08	0.118
$\sigma^2$	0.00982	0.000942	0.00981	0.000941	0.00982	0.000942
RSS	2.393		2.236		2.238	

Table 2:

Posterior parameter estimates for the different levels of  $\sigma^2$ ,  $4\beta \sim Be(1, 1)$ .

Parameter	$\sigma^2 = 0.01$		$\sigma^2 = 0.5$		$\sigma^2 = 1.0$	
	mean	std dev	mean	std dev	mean	std dev
$\mu$	5.31	0.689	3.91	0.395	4.54	$8.8 \times 10^{-5}$
$\beta$	0.242676	0.00060	0.23615	0.0112	0.13803	0.0746
$\kappa^2$	1.11	0.124	0.76	0.123	0.0	0.0
$\sigma^2$	0.00981	0.000941	0.250	0.0256	2.38	0.178
RSS	2.236		60.5		871.2	

Table 3:

Start values and hyper-prior distributions used in MCMC.

Parameter	Starting value	Prior distribution
$\mu$	3.5	$N(3.5, 0.1)$
$\beta$	0.249	$Be(400, 1.5)$
$\kappa^2$	0.1	$Ga(500, 10)$
$\sigma^2$	0.01	$Ga(50, 5)$
$\lambda$	0.0	$N(0.0, 5.0)$
$\gamma$	0.2495	$Be(1, 1)$
$\eta^2$	0.5	$Ga(1, 1)$
$x$	25.0	
$v$	0.0	
$z$	0	

Table 4:

Posterior parameter estimates for subject 1 –slice 1.

Parameter	mean	median	std dev
$\mu$	3.39	3.39	0.09
$\beta$	0.2491	0.2493	0.0004
$\kappa^2$	0.120	0.118	0.015
$\sigma^2$	0.030	0.030	0.0015
$\lambda$	-1.10	-1.19	1.52
$\gamma$	0.2497	0.2498	0.0003
$\eta^2$	9.30	9.29	0.25
RSS		5.09	

Table 5:

Posterior parameter estimates for the slice of subject 2.

Parameter	mean	median	std dev
$\mu$	1.94	1.93	0.13
$\beta$	0.2430	0.2432	0.0007
$\kappa^2$	0.103	0.102	0.0014
$\sigma^2$	0.026	0.026	0.0014
$\lambda$	0.0087	0.0095	0.13
$\gamma$	0.2435	0.2435	0.0001
$\eta^2$	1.67	1.57	0.04
RSS		5.22	

Figure 1. Schematic of the BOLD HDR function. There is a short and small initial dip followed by the main rise and (slower) fall. The response subsequently dips below the baseline before settling to the baseline state.

Figure 2. Simulation study data and results for the  $\sigma^2 = 0.01$  and  $4\beta \sim Be(0,0)$  prior case. Panel (a) displays the simulated spatial magnitudes (response data). The central disk consists of the CAR process (representing a region of activity) plus Gaussian noise  $\sigma^2 = 0.01$  (elsewhere consists of Gaussian noise only). The other maps show the corresponding posterior MMSE estimates of the other fields in the model; (b)  $x$ , (c)  $z$ , (d)  $w$ , (e)  $z \odot x$  and (f)  $\epsilon$ .

Figure 3. MCMC plots for the parameters  $\mu$  and  $\beta$  ( $Be(0,0)$  prior,  $\sigma^2 = 0.01$ ). For periods where the trace for  $\beta$  approaches 0.25, the parameter  $\mu$  becomes non-identifiable and the variability of its trace increases.

Figure 4. Map (a) is that of the Poisson parameter estimates for slice 1 of subject 1. Maps (b) – (f) are the posterior map estimates: (b)  $x$ , (c)  $z$ , (d)  $w$ , (e)  $z \odot x$ , (f)  $\epsilon$ . Dark regions correspond to high values. The residual map has high spatial correlation.

Figure 5. [subject 1 – slice 1] Map (a) is that of the Poisson parameter estimates for slice 1 of subject 1. Maps (b) – (f) are the posterior map estimates: (b)  $x$ , (c)  $z$ , (d)  $v$ , (e)  $z \odot x + z \odot v$ , (f)  $\epsilon$ . The magnitude of the residuals is negligible compared those of the previous model displayed in Figure 4.

Figure 6. [subject 1 – slice 1] The map shows the spatial distribution of the re-scaled residuals from Fig. 5. Spatial structure continues to exist in the residuals but its magnitude is much reduced over that in Figure 4.

Figure 7. [subject 1 – slice 2] Map (a) is that of the Poisson parameter estimates for slice 2 of subject 1. Maps (b) – (f) are the posterior map estimates: (b)  $x$ , (c)  $z$ , (d)  $v$ , (e)  $z \odot x + z \odot v$ , (f)  $\epsilon$ . Bi-lateral activation is observed in this slice along with activation in the lower mid-brain.

Figure 8. [subject 2] Map (a) is that of the Poisson parameter estimates for the considered slice subject 2. Maps (b) – (f) are the posterior map estimates: (b)  $x$ , (c)  $z$ , (d)  $v$ , (e)  $z \odot x + z \odot v$ , (f)  $\epsilon$ . Notice that the secondary activation map displayed in panel (d) has less structure than those displayed in panel (d) of Figures 5 and 7 for subject 1.

Figure A.1.  $100 \times 100$  simulations of (almost) intrinsic first order GMRFs (a), (b) are simulations of the GMRFs, (c), (d) are the thresholded versions with 30% of voxels active (black = 1), (d), (e) are thresholded for 10% active voxels.

Figure 1:

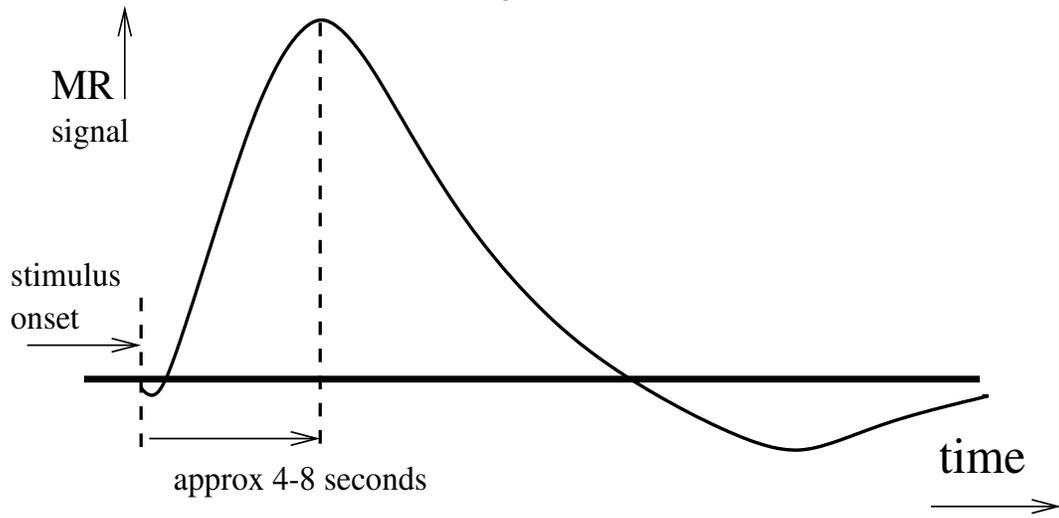


Figure 2:

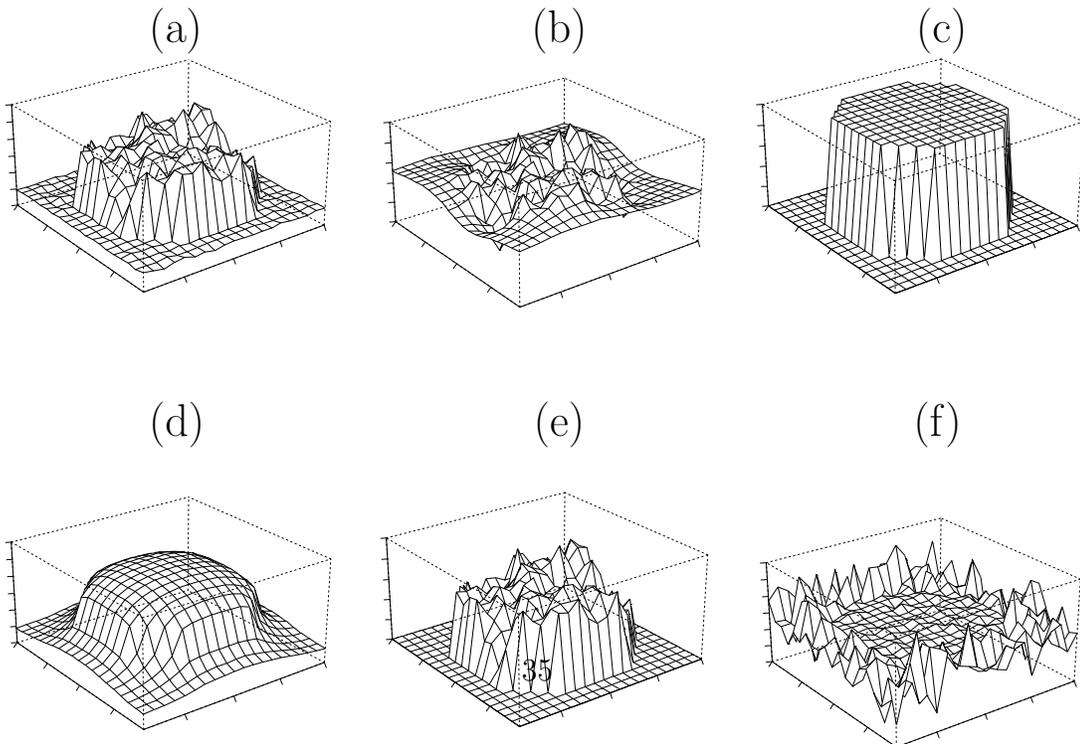


Figure 3:

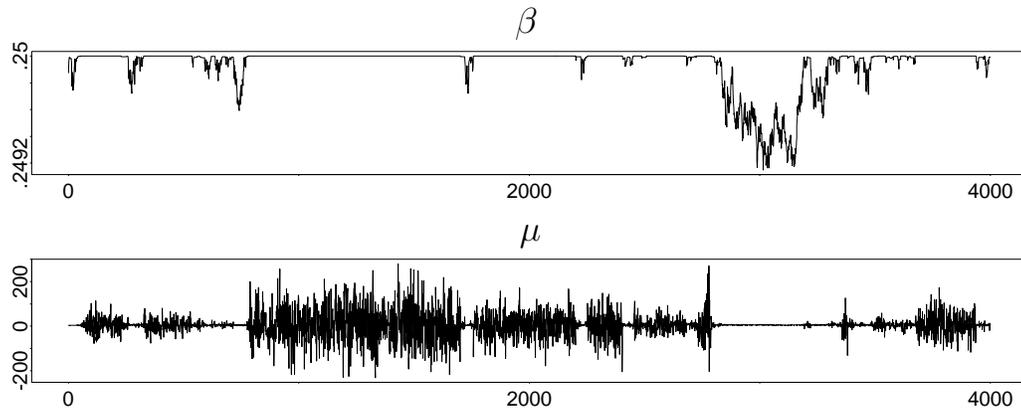


Figure 4:

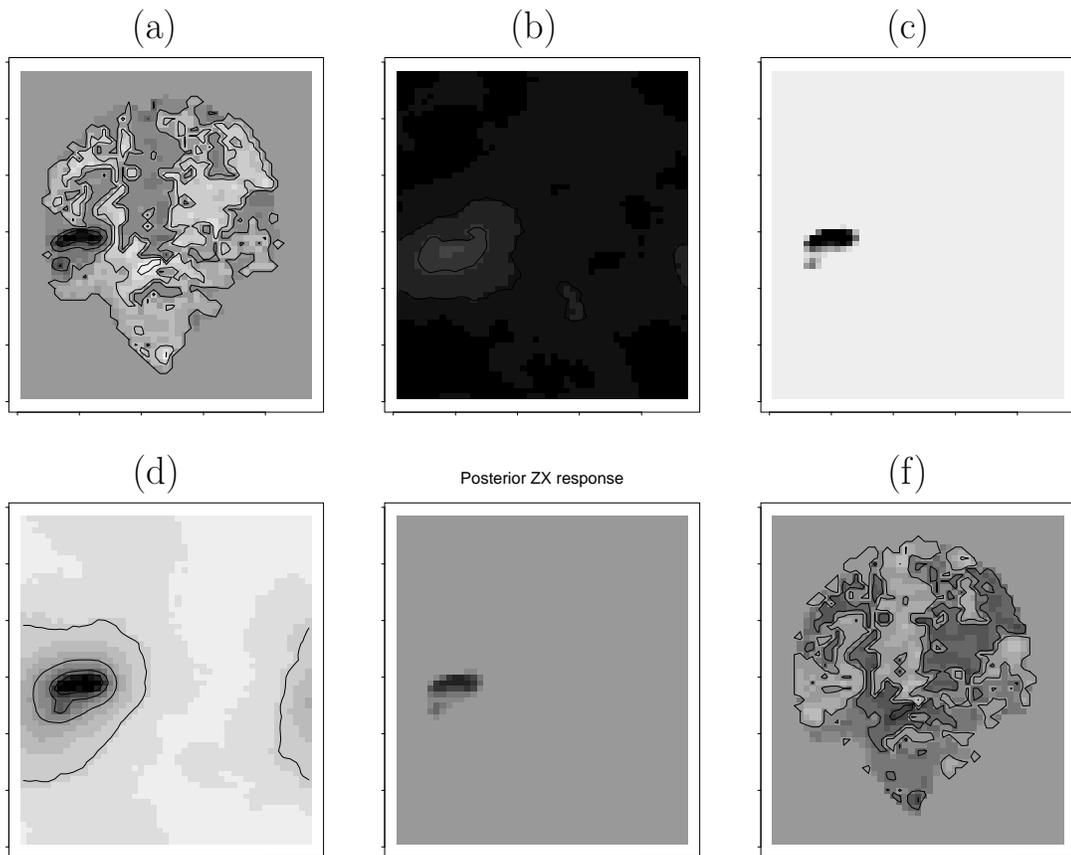


Figure 5:

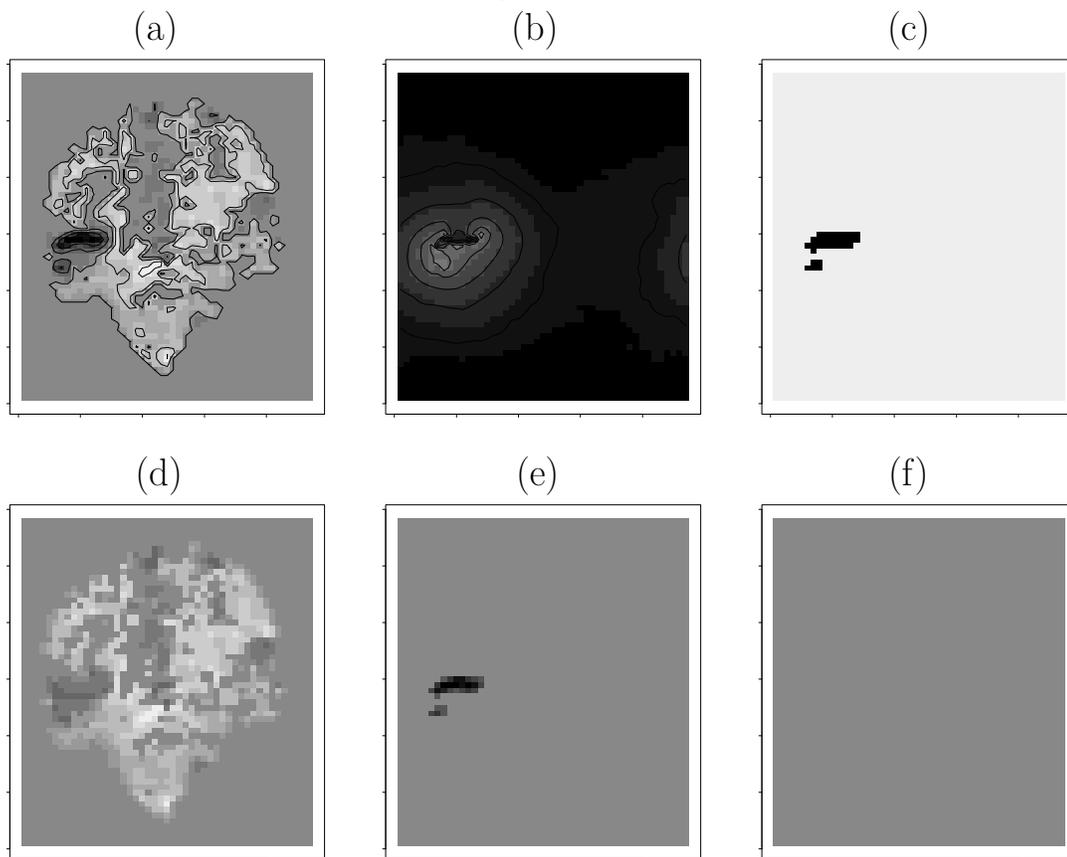


Figure 6:

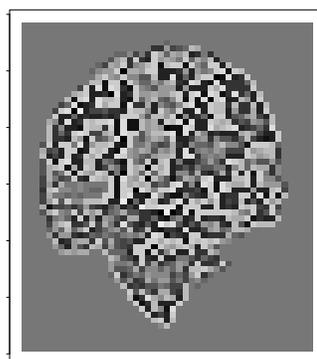


Figure 7:

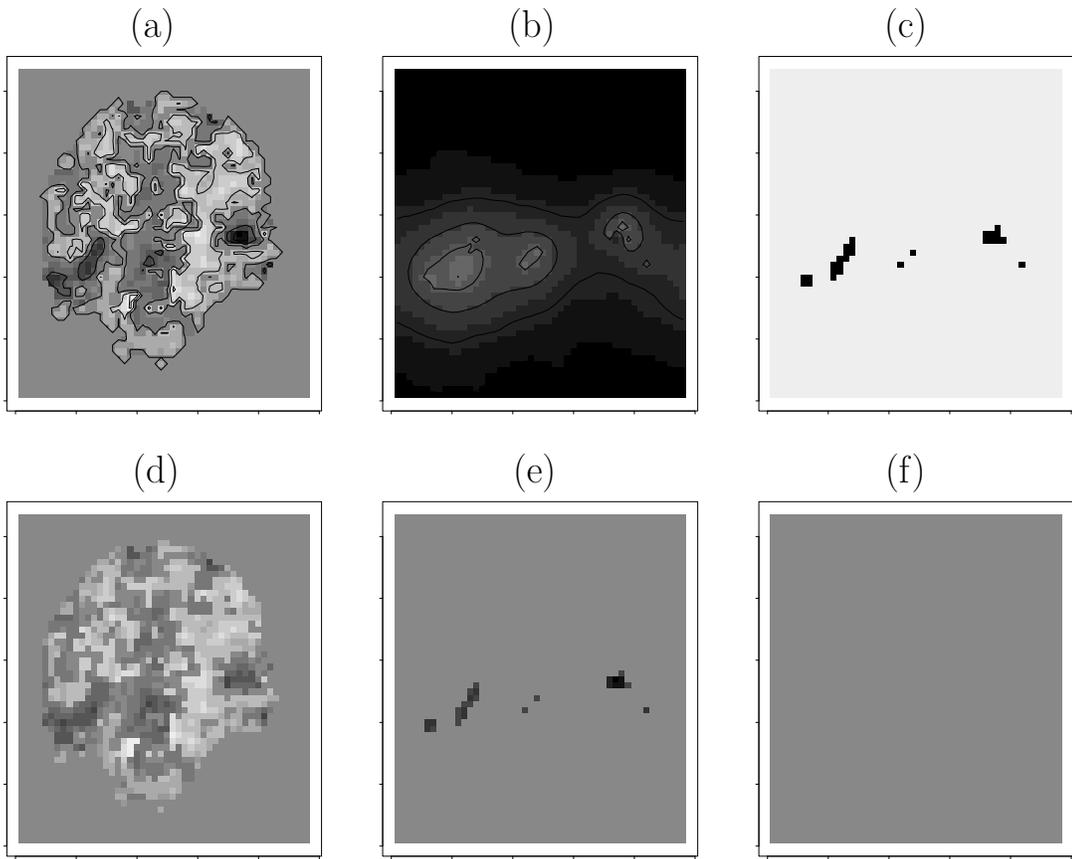


Figure 8:

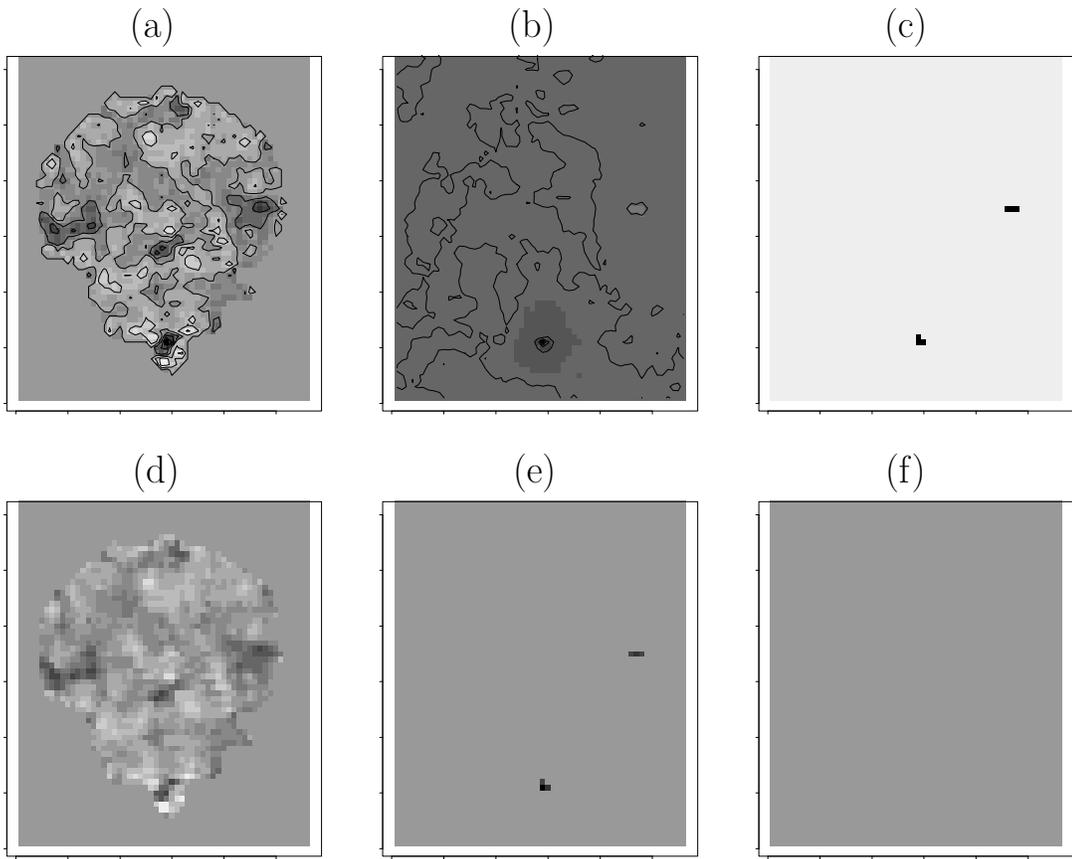


Figure A.1

