research
reports

No. 410 November 1999

Niels Væver Hartvig

A stochastic geometry model for fMRI data

department of
theoretical
statistics

university of
aarhus

A stochastic geometry model for fMRI data

Niels Væver Hartvig*
University of Aarhus

Abstract

Functional magnetic resonance imaging (fMRI) is a principal method for mapping the human brain. fMRI data consist of a sequence of MR scans of the brain acquired during stimulation of specific cortical areas, and the purpose of analysing the data is to detect activated areas, i.e. areas where the intensity changes according to the stimulation paradigm. A common analysis procedure is to estimate the activity pattern non-parametricly by smoothing the data spatially. The focus is then on assessing significance of peaks or clusters in the smoothed activation surface by means of multiple hypothesis testing, rather than assessing the uncertainty of the estimated pattern itself. In this paper we formulate a more structured model for the spatial activation pattern. We achieve this by considering a stochastic geometry model where the activation surface is given by a sum of Gaussian functions, which to some extent can be thought of as individual centres of activation in the brain. The model is formulated in a Bayesian framework, where the prior distribution of the centres is given by a marked point process density. An advantage of this approach is that inference can be carried out by simulation techniques, and hence it is easy, though time consuming, to evaluate the uncertainty of the estimate or to test hypotheses of interest regarding the activation. Furthermore in this framework, we are able to model the temporal pattern of the activation with fewer assumptions than usually imposed. This reveals significant non-stationarities in the analysed data, which violate the common assumption of stationarity of the haemodynamic response.

Keywords: Functional magnetic resonance imaging; Stochastic geometry model; Marked point proces; Markov chain Monte Carlo; State space model; Correlogram

1 Introduction

Functional magnetic resonance imaging (fMRI) is a medical imaging technique where fast MR scanners are used to measure changes in blood oxygenation in the brain. The latter is known as the Blood Oxygen Level Dependent (BOLD) signal. These oxygenation changes

*Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, email: vaever@imf.au.dk

correlate with neural activity in the surrounding tissue, and hence fMRI is an indirect method for measuring activation in the brain. The technique is quite new, one of the first experiments was reported by Kwong *et al.* (1992) and since then the number of publications in the field has grown extremely fast. fMRI is a very attractive modality for imaging the brain, since it is non-invasive and has a good temporal and spatial resolution.

In a typical fMRI experiment a subjects brain is scanned while specific centres are stimulated, for instance the visual cortex can be activated by flashing a light in the eyes. The acquired data consist of a sequence of scans and the aim of the statistical analysis is to identify regions in the images, where the intensity changes according to the stimulus rhythm. A biostatistical introduction to the subject is given in Lange (1996), Lange and Zeger (1997) also contains a good introduction.

The analysis of the data is impeded by the uncertainty of the haemodynamic response to the stimulus. It is well known, that the response is delayed about 6 seconds and dispersed in time compared to the stimulus paradigm, but otherwise there is no general accepted biological model for the response, which can guide us when modelling the signal.

Another problem is the incorporation of spatial structure in the analysis. Of course the spatial activation pattern depends on the type of stimulation, and it is difficult to impose structure on this in a general setting. Instead, a common approach is to marginalize the analysis to a one dimensional time-series problem for each voxel in the scan, see for instance Worsley and Friston (1995), Lange and Zeger (1997) or Bullmore *et al.* (1996). The spatial structure of the data is included in a second step, when the image of activation estimates is convolved with a smoothing kernel to obtain a non-parametric estimate of the activation. In this approach there is no specific model for the spatial pattern of activation. Furthermore the focus is on assessing significance of peaks and clusters in the image by testing thousands of voxel-wise hypotheses simultanously.

In this paper we will focus on the issue of estimating the activation pattern, rather than testing multiple hypotheses. The model for the spatial pattern is based on two fundamental assumptions in the fMRI literature: 1) The activated areas have a spatial extent of several millimetres and 2) the activation pattern is "smooth". Both assumptions are based on the haemodynamic origin of the signal: Even if neural activation is localized to a single voxel, say, the haemodynamic effects will occur in the surrounding venes, and will cover a larger area. We will incorporate these two assumptions in a stochastic geometry model based on marked point processes, see for instance Baddeley and van Lieshout (1993). This is done by modelling the spatial activation surface by a collection of Gaussian functions, which to some extent can be thought of as individual centres in the brain. The model is formulated in a Bayesian setting where the centres *a priori* are distributed as a marked point process; here the points are the locations of the centres and the marks describe the shape and height of the centres. The inference in the model is based on simulation techniques, by which we can estimate the posterior mean of functions of interest, such as the mean activation pattern.

The advantages compared to the common analysis procedure outlined above are many. 1) We don't have to smooth the data spatially, but can retain the detailed resolution of the MR scans. 2) We can assess the uncertainty of the estimated spatial pattern in a Bayesian framework. 3) We can quantify our belief in more specific hypotheses about the activation by

estimating posterior probabilities in the model. 4) Finally we can model the haemodynamic response function, i.e. the temporal pattern of the activation, in a semi-parametric way, which allows for non-stationarities and non-linearities. With the latter approach explicit knowledge of the stimulation paradigm is not required, and we can hence estimate activation which is not time-locked to the stimulation rhythm.

The paper is organized as follows: In Section 2 we formulate the Bayesian model for the spatial activation pattern, and combine this with a simple initial model for the temporal response to obtain a spatio-temporal model. The temporal pattern is assumed to be known and described by a convolution model. This is somewhat restrictive, but it allows us to focus on the spatial pattern for a start, and discuss how we can simulate the latter from the posterior distribution. This is done by an MCMC algorithm, which is described in Section 3. In Section 4 we apply the model to a simulated data set, which is used for estimating prior parameters, and to visual stimulation data. In the next two sections we extend the model in different ways: In Section 5 we describe a state space model for the haemodynamic response function, and demonstrate its ability to model non-stationarities which are indeed present in the data. In Section 6 we extent the covariance structure to account for the spatio-temporal correlation, which is present in the noise. Finally we have a discussion in Section 7 and an appendix where theoretical properties of the MCMC algorithm are studied.

2 The model

2.1 Preprocessing of the data

Suppose the data consist of m scans, acquired with a stimulation paradigm $\pi_1, ..., \pi_m$, where $\pi_t = 1$ indicates stimulation and $\pi_t = 0$ no stimulation at time t. Typically the paradigm is arranged in blocks of, say, 10 scans with stimulation and 10 without. Let V be the set of voxels covering brain tissue, $V \subseteq S$, where S represents a 2 dimensional slice or a 3 dimensional volume of the brain. The dataset is hence given by a set of intensity measurements $Y = \{Y_{it}, i \in V, t = 1, \ldots, m\}$.

The units of the intensities recorded by the MR scanner are arbitrary, and it is common in the literature to report variation of the signal in percent of baseline intensity. In order to consider variation of the intensity in different voxels on the same scale, we have log-transformed the data. Suppose for instance, that the measurement in a given voxel at time t is given by $Y_t = \mu(1 + \varepsilon_t)$, where ε_t is a deviation from the baseline intensity of the voxel. For small deviations we then have

$$\log Y_t = \log \mu + \log(1 + \varepsilon_t) \simeq \log \mu + \varepsilon_t,$$

and hence the magnitude of (structural and random) variations of the log data, can be compared between different timeseries. Furthermore, the unit of the deviations can be thought of as percent of baseline intensity.

We will preprocess the data, such that the images are aligned to correct for subject movement and has been corrected for trends. In our applications we have used a simple

procedure, where each image is aligned to a reference image by minimizing the squared difference between the two images over all translations and rotations. As for the trend correction, we will consider the residuals after subtracting the mean and correcting for a linear trend in each individual time series. The presence of trends and low-frequency fluctuations in fMRI time series is often reported in the literature, though the processes which generate these are not well understood. Modelling these features as linear terms is necessarily an approximation, and more general models such as proposed by Holmes *et al.* (1997) and Petersen *et al.* (1998) may be applied. However, as will be described in Section 5, our aim is to model general temporal response patterns, and hence we are cautious not to remove any fluctuations caused by the haemodynamic response. A linear model is a good compromise in this context.

A basic assumption of the model is that the spatial and temporal patterns of the activation can be modelled separately. Considering an image or a volume of the activation magnitudes $A = \{A_i, i \in V\}$ and a timeseries $\varphi = \{\varphi_t, t = 1, \ldots, m\}$ of the common temporal variation caused by the BOLD effect, we assume that the mean intensity measured in voxel i at time t is given by $EY_{it} = A_i\varphi_t$. We will now describe in detail how the spatial and temporal pattern are modelled.

2.2 A model for the spatial activation pattern

Consider first the case where data only represent a 2 dimensional slice of the brain, that is $V \subseteq S \subseteq \mathbb{R}^2$. The spatial activation pattern will be modelled as a collection of n "activation centres" $X = \{X_1, X_2, \ldots, X_n\}$, each parametrized as $X_j = (\mu_j, a_j, d_j, r_j, \theta_j)$. The global pattern $A(X) = \{A_i(X) \,|\, i \in V\}$ is given by the superposition of n bells,

$$A_i(X) = h(i; X_1) + \cdots + h(i; X_n)$$

where

$$h(i; X_j) = a_j \exp\left\{ -\frac{\pi \log 2}{d_j}\left(\frac{\tilde{i}_1^2}{r_j/(1-r_j)} + \frac{\tilde{i}_2^2}{(1-r_j)/r_j}\right)\right\} \tag{1}$$

and $\tilde{i} = (\tilde{i}_1, \tilde{i}_2) = R(-\theta_j)(i - \mu_j)$. Here $R(\theta)$ is a rotation with angle θ. Hence $h(\cdot; X_j)$ is a Gaussian bell of height a_j centred at $\mu_j \in S$. The parameter $d_j \in R_+$ is the area of the contourellipse at half height, $r_j \in (0, 1)$ is a measure of the eccentricity of the ellipse, more precisely the ratio of the first principal axis and the sum of the two axes, and $\theta_j \in [-\pi/4, \pi/4]$ is the orientation of the ellipse. Notice that the angle is constrained to an interval of length $\pi/2$ to ensure identifiability of the parameters (r, θ).

In order for this specification to be meaningful, we need to restrict heights to be positive and incorporate some regularity in the point pattern. We will achieve this by formulating a prior model for X in the context of marked point processes, see for instance Møller (1999). Each centre $X_j = (\mu_j, a_j, d_j, r_j, \theta_j)$ is a point in $\mathcal{X} = S \times M$, where

$$M = [0, C_a] \times [0, C_d] \times (0, 1) \times [-\pi/4, \pi/4].$$

Here C_a and C_d are natural bounds for the height and area respectively. Though time series with negative activation amplitude is observed, we will initially assume that all bells have positive height. We will discuss later, how negative activation can be accounted for in the model.

Let $\mathcal{X}$ be equipped with the Borel σ-field $\mathcal{S} \times \mathcal{M}$ and the Lebesque measure $\lambda_2 \times \lambda_4$, and let Ω denote the exponential space over $\mathcal{X}$, that is the set of finite sets $\{x_1, \ldots, x_n\}$ where $x_i \in \mathcal{X}$ for all i. The activation profile $X = \{X_1, \ldots, X_n\}$ can then be interpreted as a point process in Ω or equivalently as a marked point process with point space S and mark space M. A priori we will assume that X has density wrt. the unit rate homogenous Poisson process on Ω of the form

$$f(x) \propto \beta^n \left(\prod_{i=1}^{n} \prod_{j=i+1}^{n} \phi(x_i, x_j) \right) \prod_{j=1}^{n} \{p(a_j)p(d_j)p(r_j)\}, \quad x = \{x_1, \ldots, x_n\} \tag{2}$$

where $n = n(x)$ is the number of points in x and β is an intensity parameter. The pairwise interaction function ϕ introduces a regularity in X, discouraging configurations with centres placed "on top" of each other. A popular choice when modelling repulsive point patterns is the Strauss model with interaction radius $\rho > 0$ with respect to a metric $\delta(\cdot, \cdot)$ on $\mathcal{X}$. In this case ϕ is given by

$$\phi(\xi, \eta) = \gamma^{1(\delta(\xi,\eta)<\rho)}, \quad \xi, \eta \in \mathcal{X},$$

with $\gamma \in [0, 1]$ and with the convention that $0^0 = 1$. In our setup, however, we wish to impose a hard-core restriction, which prohibits pairs of centres with distances close to zero. The hard-core model with $\gamma = 0$ is not very suitable in this context, since the posterior distribution will be very sensitive to the choice of ρ. An appropriate alternative is the so-called very-soft-core model of Ogata and Tanemura (1984) with

$$\phi(\xi, \eta) = 1 - \exp\left\{-(\delta(\xi, \eta)/\rho)^p\right\}, \quad \xi, \eta \in \mathcal{X}, \quad p \geq 2. \tag{3}$$

The hard-core model is obtained by setting $p = \infty$, while finite values of p yield a continous interaction function which increases smoothly from 0 to 1 with the distance between two points. A plot of the interaction functions for different values of p can be seen in Figure 1.

The metric $\delta(\cdot, \cdot)$ should be defined such that two centres x_1 and x_2 are close, if they are close in space and have similar size and shape. One way of assessing this is by the *J-divergence* (Kullback, 1959) of the corresponding Gaussian functions: By rewriting the expression in (1), we find that the activation intensity $h(\cdot, x_j)$ induced by $x_j = (\mu_j, a_j, d_j, r_j, \theta_j)$ is given by $h(i; x_j) = a_j d_j f_j(i)/\log 2$, where $f_j(\cdot) = f(\cdot; \mu_j, \Sigma_j)$ is a multivariate normal density with mean μ_j and covariance matrix

$$\Sigma = \frac{d_j}{2\pi \log 2} R(\theta) \begin{pmatrix} \frac{r}{1-r} & 0 \\ 0 & \frac{1-r}{r} \end{pmatrix} R(-\theta).$$

The J-divergence between the two densities is now given by

$$\delta(x_1, x_2) = J(f_1, f_2) = \int (f_1(x) - f_2(x)) \log \frac{f_1(x)}{f_2(x)} \lambda_2(dx) \tag{4}$$

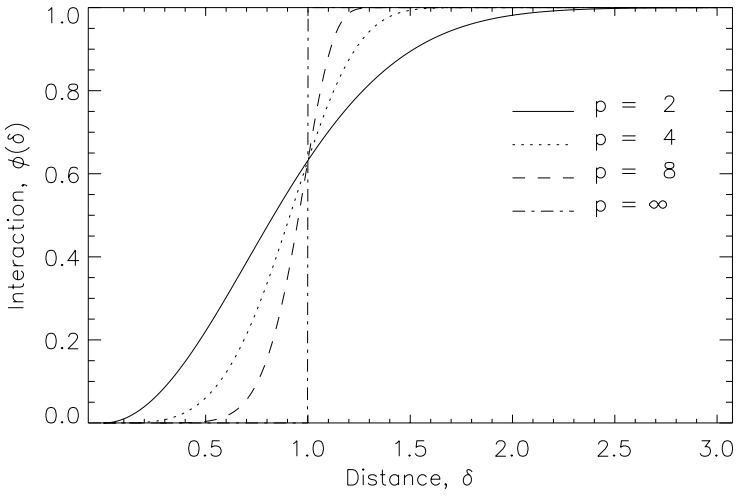

Figure 1: The soft-core interaction function $\phi(\xi, \eta)$ in (3) as a function of the distance $\delta(\xi, \eta)$, $\xi, \eta \in \mathcal{X}$.

and by inserting the means and variances we get

$$\delta(x_1, x_2) = -2 + \frac{1}{2}\left\{(\mu_1 - \mu_2)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\mu_1 - \mu_2) + \text{trace}(\Sigma_2^{-1}\Sigma_1 + \Sigma_1^{-1}\Sigma_2)\right\}. \quad (5)$$

Figure 2 is a plot illustrating distances between pairs of points with this metric. The expression can of course be rewritten in terms of the (d, r, θ)-parametrization of the Gaussian density. The parameter ρ equals 1.

The priors for a and d should be as uniform as possible, yet penalizing values close to zero. The inverted Gamma distribution is a suitable choice in this context, with its light tail near zero and its quite heavy tail for large values. Hence we will assume that $a^{-1} \sim , (2, \beta_a)$ and $d^{-1} \sim , (2, \beta_d)$ with the restrictions that $a \in (0, C_a]$ and $d \in (0, C_d]$. The density of d is

$$p(d) = \exp(\beta_d/C_d)(\beta_d/C_d + 1)^{-1}\beta_d^2 d^{-3}\exp(-\beta_d/d), \quad d \in (0, C_d].$$

In our application we set $\beta_a = 0.05$ and $\beta_d = 200$ mm^2, for comparison the voxels cover an area of 3.61 mm^2 in each slice in our data. As for the axis ratio r we wish to discourage very eccentric ellipses. This can be obtained by a Beta-prior, $r \sim \text{Beta}(5, 5)$. Finally the angle θ is uniformly distributed on $[-\pi/4, \pi/4]$.

With this choice of prior, we assume that the intensity β of the centres is constant over V. An obvious refinement is to include covariate information on the underlying tissue and allow the intensity to depend on the location. Also the experimenter often has good prior knowledge of where the activation is likely to occur, which could be used, when specifying β.

This specification can straightforwardly be generalized to a 3-dimensional setting where $S \subseteq \mathbb{R}^3$. In this case a centre is given by $x = (\mu, a, d, r_1, r_2, \theta_1, \theta_2)$, and the contribution to

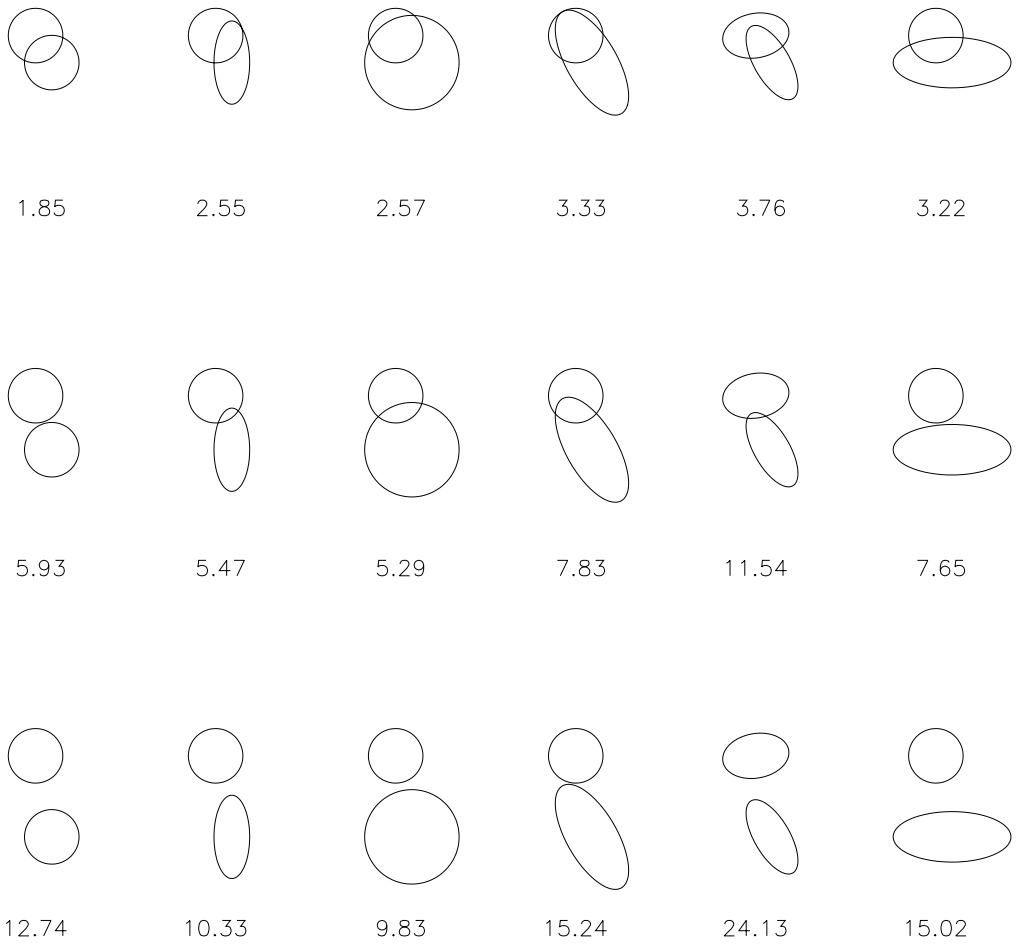

Figure 2: Examples of distances measured by the metric (5) on the product space of points and marks. Illustrated are pairs of countour ellipses at half height of the respective bells, together with their distance.

the activation volume is

$$h(i;x) = a \exp\left\{ -\log 2 \left(\frac{4\pi}{3d}\right)^{2/3} \left(\frac{\tilde{i}_1^2}{(r_1^2/r_2r_3)^{2/3}} + \frac{\tilde{i}_2^2}{(r_2^2/r_1r_3)^{2/3}} + \frac{\tilde{i}_3^2}{(r_3^2/r_1r_2)^{2/3}} \right) \right\}.$$

Here $r_3 = 1 - r_1 - r_2$, $r_i > 0$ for $i = 1, 2, 3$ and

$$\tilde{i} = (\tilde{i}_1, \tilde{i}_2, \tilde{i}_3) = \begin{pmatrix} \cos\theta_1 \cos\theta_2 & -\sin\theta_1 & -\cos\theta_1 \sin\theta_2 \\ \sin\theta_1 \cos\theta_2 & \cos\theta_1 & -\sin\theta_1 \sin\theta_2 \\ \sin\theta_2 & 0 & \cos\theta_2 \end{pmatrix} (i - \mu).$$

With this parametrization d is the volume of the contour ellipsoid at height $a/2$, and r_i is the ratio of the ith main axis and the sum of the three main axis. The angles θ_1 and θ_2 are the rotations in the xy-plane and xz-plane respectively, which are restricted to the interval $[-\pi/4, \pi/4]$. The natural extension of the priors is to assume that $(r_1, r_2) \sim D_2(5, 5)$ where D_2 is the two-dimensional Dirichlet distribution.

2.3 A model for the temporal pattern

In order to obtain a reasonably simple spatio-temporal structure in the model, we will assume that the temporal pattern $\varphi = \{\varphi_t, t = 1, \ldots, m\}$ is approximately the same for all voxels. Hence we assume that any voxel-wise differences in the delay is negligible, and we assume that the shape of the haemodynamic response is the same everywhere. This should be contrasted to, for instance, the approach in Lange and Zeger (1997), where differences from one voxel to another is explicitly accounted for. However, we will discuss later how the model can be extended in order to relax this assumption.

Initially we will follow the approach in Friston *et al.* (1995) and consider φ to be known and given by a convolution between the paradigm π and a Gaussian density of mean 6 seconds and variance 9 seconds, modelling the delay and dispersion of the signal. Hence

$$\varphi_t = \sum_i \pi_{t-i} \frac{T}{\sqrt{2\pi}3} \exp(-\frac{(iT-6)^2}{18}), \tag{6}$$

where T is the repetition time, i.e. the time between two consecutive images. This is a rather simple model, and there is no particular reason for choosing a Gaussian density as the model for the impulse response function, neither is it obvious that the response is stationary. In Section 5 we will describe a more flexible semi-parametric model for the temporal pattern which does not require these assumptions. However the simulation procedure to be presented in the following section simplifies a great deal if we assume a known and fixed response, and we will thus start with this model.

2.4 Combining the spatial and temporal models

Given the centres X and the haemodynamic response function φ, the model for the intensity Y is,

$$Y_{it} = (A_i(x) + \eta_i)\varphi_t + \varepsilon_{it}, \quad \eta_i \sim N(0, \tau^2),\ \varepsilon_{it} \sim N(0, \sigma^2), \quad i \in V, t = 1, \ldots, m \tag{7}$$

where $\{\varepsilon_{it}\}$ and $\{\eta_i\}$ are independent white noise sequences. We assume a simple noise model, with the ε_{it}'s being independent, but more general covariance structures can be incorporated in a theoretically simple way, see Section 6. Also more complicated noise sources may be removed before the analysis, for instance by procedures in Le and Hu (1996) or Petersen *et al.* (1998).

The likelihood function in the model (7) is given by

$$p(Y|x) = (2\pi\sigma^2)^{-\frac{(m-1)|V|}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i\in V}\sum_{t=1}^{m}\left(Y_{it}-\tilde{Y}_i\varphi_t\right)^2\right\} \times$$
$$(2\pi(\sigma^2+\tau^2\mathrm{ss}_\varphi))^{-\frac{V}{2}} \exp\left\{-\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi+\tau^2)}\sum_{i\in V}\left(\tilde{Y}_i - A_i(x)\right)^2\right\}. \quad (8)$$

Here

$$\tilde{Y}_i = \sum_{t=1}^{m} Y_{it}\varphi_t/\mathrm{ss}_\varphi, \quad \mathrm{ss}_\varphi = \sum_{t=1}^{m}\varphi_t^2, \quad (9)$$

is the coefficient of the projection of $\{Y_{it}, t = 1,\ldots,m\}$ on the vectorspace $L = \mathrm{span}\{\varphi\}$. Notice that the likelihood function factorizes into two terms, involving only the projection of Y onto L and onto the orthogonal complement to L, respectively, with X only entering in the latter. Hence we find that $\{\tilde{Y}_i, i \in V\}$ is sufficient for X. The former is a regression image with the voxel-wise estimated activation amplitudes, this is also known as a *Statistical Parametric Map* (SPM) in the fMRI literature (Friston *et al.*, 1994). The estimation of the spatial pattern $A_i(x)$ can hence be viewed as a model based way of smoothing the SPM. This provides a link to more traditional methods, where the SPM is smoothed with a Gaussian filter, and afterwards regarded as a differentiable Gaussian random field for inference purposes. The model in this simplest setting hence provides an alternative estimate for the activation based on the raw SPM and a way of assessing the uncertainty of the estimate. In the more general setting described in Section 5, we can estimate φ semi-parametricly rather than assuming it is known, which is in general not possible in the traditional SPM approach.

The purpose of the random effect term η_i is to regularize the estimate of X. To see why this is necessary, consider the log posterior distribution of X, which up to an additive constant is given by

$$\log p(x|Y) = -\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi+\tau^2)}\sum_{i\in V}\left(\tilde{Y}_i - A_i(x)\right)^2 + \log p(x).$$

Suppose for a moment that $\tau = 0$, corresponding to omitting the random effect η_i above. By inserting sufficiently many small bells, we can obtain a configuration where $A_i(x) = \tilde{Y}_i$ when the latter is positive, and $A_i(x) = 0$ elsewhere. This configuration will minimize the sum of squares above. Even if the prior density of such a pathological point configuration is very small, it will be the maximum aposteriori estimate in the limit as m, and hence

ss_φ, tends to infinity, since the sum of squares will dominate in the limit. By assuming a fixed positive value for τ^2 this undesirable property of the posterior distribution is removed. Intuitively τ^2 is a measure of how well we expect the actual activation surface to be described by a reasonable collection of Gaussian functions, while the purpose of the prior for X is to quantify what we mean by a reasonable collection.

When applying the model, we will insert *ad hoc* estimates of σ^2 and τ^2. An unbiased and consistent estimator for σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{(m-1)|V|} \sum_{i \in V} \sum_{t=1}^{m} \left(Y_{it} - \tilde{Y}_i \varphi_t \right)^2 \sim \sigma^2 \chi^2(f)/f, \;\; f = (m-1)|V|. \tag{10}$$

As for τ^2, we will estimate $\sigma^2/\text{ss}_\varphi + \tau^2$ by considering the regression coefficients $\tilde{Y}_i$. These are distributed as

$$\tilde{Y}_i \sim N(A_i(x), \sigma^2/\text{ss}_\varphi + \tau^2), \quad i \in V,$$

with all $\tilde{Y}_i$'s independent. Letting ∂i denote the 9-voxel neighbourhood of i, we will let

$$\bar{Y}_i = \frac{1}{9} \sum_{j \in \partial i} \tilde{Y}_j \sim N(\bar{A}_i(x), \frac{1}{9}(\sigma^2/\text{ss}_\varphi + \tau^2))$$

for $i \in V^\circ$, where $V^\circ = \{i \in V \,|\, \partial i \subseteq V\}$. By assuming that the activation surface $A_i(x)$ can be approximated by a plane locally around i, we have that $A_i(x) = \bar{A}_i(x)$ and hence that

$$\frac{9}{8|V^\circ|} \sum_{i \in V^\circ} \left(\tilde{Y}_i - \bar{Y}_i \right)^2 \tag{11}$$

is an unbiased and consistent estimator for $\sigma^2/\text{ss}_\varphi + \tau^2$. When the approximation is not exact, we will get a slight positive bias in the estimate for τ^2.

2.5 Modelling negative activation

So far we have only considered areas with increased intensity during stimulation, but in fact in some areas of the brain a parallel decrease in the intensity is observed. This is typically attributed to large veins or other types of non-neural tissue, and as such is not of primary interest in the analysis. However, in order to obtain a realistic model for the data, we need to consider this effect.

A natural way to proceed is to model the activation as $A^+ - A^-$, where A^+ and A^- are positive surfaces describing positive and negative activation respectively, and each has a prior similarly to the surface A described earlier. When doing this, we have to incorporate restrictions in the prior that separates the two surfaces for identifiability reasons. If not, the two surfaces may overlap to an extent where they cannot be individually identified as positive and negative activated areas, but rather positive and negative terms in a general surface, which is not necessarily given by sums of Gaussian functions. As a result of this the

positive and negative centres will become highly correlated, and the interpretation of the activation surface becomes very difficult.

Suppose we let X^+ and X^- be two point processes modelled as described in Section 2.2, determining the postive and negative surfaces respectively. One way of separating the surfaces would be to model the prior as

$$p(X^+, X^-) \propto f(X^+)f(X^-)\exp(-\alpha \sum_{i \in V} A_i(X^+)A_i(X^-)),$$

where $f(\cdot)$ is the density in (2) and $\alpha > 0$. This prior allows some overlap between $A(X^+)$ and $A(X^-)$, but the last term penalizes configuration where $A_i(X^+)$ and $A_i(X^-)$ are both large for some $i \in V$. The parameter α determines the weight of the separation term in the prior. Suppose we let $\alpha^{-1} = \sigma^2/\mathrm{ss}_\varphi + \tau^2$, the variance of $\tilde{Y}_i$, such that $\sqrt{\alpha}A_i$ is given as units of standard deviation of $\tilde{Y}_i$. The posterior obtained with this prior is then

$$\begin{aligned}
p(X^+, X^-|Y) &\propto \exp\left\{-\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi + \tau^2)} \sum_{i \in V} \left(\tilde{Y}_i - [A_i(X^+) - A_i(X^-)]\right)^2\right\} \\
&\quad \times f(X^+)f(X^-)\exp\left\{-\frac{1}{\sigma^2/\mathrm{ss}_\varphi + \tau^2} \sum_{i \in V} A_i(X^+)A_i(X^-)\right\} \\
&\propto \exp\left\{-\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi + \tau^2)} \sum_{i \in V} \left(\tilde{Y}_i - A_i(X^+)\right)^2\right\} f(X^+) \\
&\quad \times \exp\left\{-\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi + \tau^2)} \sum_{i \in V} \left(\tilde{Y}_i + A_i(X^-)\right)^2\right\} f(X^-).
\end{aligned}$$

This shows that X^+ and X^- are independent given the data Y, and the marginal distribution of X^+ is the same as that obtained when ignoring X^- as described in the previous sections. Hence if we choose to separate the surfaces by this choice of prior, we can make inference about X^+ and X^- in their respective marginal distributions, and afterwards combine estimates using the independence of the two point processes. Naturally this prior is only one suitable way of restricting the two activation patterns out of many. As an alternative one might model repulsion between points in the two point processes, or one could consider a hard-core restriction prohibiting the surfaces from overlapping more than a certain amount. In these cases the two point processes would not be independent in the posterior distribution. However, the independence argument above makes it plausible to make separate inference about positive and negative activation, or to only consider positive activation, which is the main parameter of interest. As there are considerable advantages of considering only one type of activation at a time, namely a reduction of the dimensionality of the point processes and improved properties of the simulation algorithm, we will henceforth marginalize the inference in this way.

3 Simulating from the posterior distribution

In order to explore the posterior distribution of the activation centres given the data, we have designed a Metropolis-Hastings algorithm based on the Geyer and Møller (1994) algorithm for general finite point processes. Let x be the current point configuration. We will then propose to 1) insert a new point, 2) remove an existing point or 3) change an existing point, with probabilities p_1, p_2 and p_3 respectively, where $p_1 + p_2 + p_3 = 1$. By "change an existing point" we mean that one of the coordinates of the point is changed, either the position or one of the mark-coordinates.

Let $q_m(x' \mid x)$ denote the proposal density of a new configuration x' based on the current configuration x with move type $m = 1, 2, 3$. The probability of accepting the move is then respectively

$$\alpha_1(x, x') = \min\left\{\frac{p(x' \mid Y)q_2(x|x')p_2}{p(x|Y)q_1(x'|x)p_1}, 1\right\},$$
$$\alpha_2(x, x') = \min\left\{\frac{p(x' \mid Y)q_1(x|x')p_1}{p(x|Y)q_2(x'|x)p_2}, 1\right\},$$
$$\alpha_3(x, x') = \min\left\{\frac{p(x' \mid Y)q_3(x|x')}{p(x|Y)q_3(x'|x)}, 1\right\}.$$

If the move is rejected, the Markov chain stays in x. The proposal distributions are described in detail in the following.

3.1 Insertion of a point

With probability p_1 we propose to add a new point $\xi = (\mu, a, d, r, \theta)$ to the existing point configuration $x = \{x_1, \ldots, x_n\}$. In order to obtain a reasonable acceptance rate for this move, we wish to perform a Gibbs-like update and sample the parameters from a density proportional to the Papangelou conditional intensity $p(x \cup \xi|Y)/p(x|Y)$. However this is a distribution on the 6 dimensional space of points and marks and it is not possible to simulate directly from it. Instead, we will propose the parameters (μ, a, d, r, θ) sequentially, hence the proposal $q_1(x \cup \xi|x)$ is a combination of the terms,

$$q_1(x \cup \xi|x) = q(\mu|x)q(a|\mu, x)q(d|\mu, a, x)q(r|\mu, a, d, x)q(\theta|\mu, a, d, r, x), \tag{12}$$

where we use the generic symbol $q(\cdot|\cdot)$ for a proposal density. We will choose the proposal of a single parameter, a say, such that it resembles the conditional intensity of a point $(\mu, a, d_0, r_0, \theta_0)$ given the current configuration x, where (d_0, r_0, θ_0) are fixed typical values for the remaining parameters and μ is the proposed position of the point. In our applications we have chosen $a_0 = 0.01$, $d_0 = 50$ mm^2 (corresponding to about 14 voxels in our data), $r_0 = 0.5$ and $\theta_0 = 0$. Generally we simulate from discretized approximations to the conditional intensities, the details are given below.

Using (8) we find that when ignoring the priors, the Papangelou intensity of a new point ξ given x is

$$\frac{p(Y|x\cup\xi)}{p(Y|x)} = \exp\left\{-\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi+\tau^2)}\left(\sum_{i\in V} h(i;\xi)^2 - 2\sum_{i\in V} h(i;\xi)(\tilde{Y}_i - A_i(x))\right)\right\}. \qquad (13)$$

By approximating the discrete sum by an integral, we find,

$$\begin{aligned}\sum_{i\in V} h(i;\xi)^2 &\simeq \iint a^2 \exp\left\{-\frac{2\pi\log 2}{d}\left(\frac{x_\theta^2}{r/(1-r)}+\frac{y_\theta^2}{(1-r)/r}\right)\right\} dxdy/(v_xv_y)\\ &= \iint a^2 \exp\left\{-\frac{2\pi\log 2}{d}(x^2+y^2)\right\} dxdy/(v_xv_y)\\ &= a^2 d/(2\log 2 v_x v_y) = a^2\tilde{d}/(2\log 2), \qquad (14)\end{aligned}$$

where v_x and v_y are the length of the voxelsides in mm's and $\tilde{d} = d/(v_xv_y)$ is the area measured in voxels. Above (x_θ, y_θ) represents a translation and rotation of (x,y), and the second equality follows since this transformation together with the coordinate scaling has Jacobian one.

When proposing the position μ we will fix the remaining parameters at $(a_0, d_0, r_0, \theta_0)$ and approximate the Papangelou intensity in (13) with a voxel-wise constant density;

$$q(\mu|x) \propto \exp\left\{\frac{1}{(\sigma^2/\mathrm{ss}_\varphi+\tau^2)}\sum_{i\in V} h(i;\mu,a_0,d_0,r_0,\theta_0)(\tilde{Y}_i - A_i(x))\right\} \quad \text{for } \mu\in V.$$

We will need to calculate a sum over V for all possible values of $\mu\in V$ in order to simulate from this density. Hence an order of $|V|^2$ iterations are required, which can be quite large; in most applications $|V|$ is around 5000. The computational burden can however be reduced, either by only performing the sum over a part of V, where $h(\cdot;\mu,a_0,d_0,r_0,\theta_0)$ is greater than a certain threshold, in which case the number of iterations is $O(|V|)$. Alternatively the convolution can be calculated in the Fourier domain, which requires $O(|V|\log_2|V|)$ iterations when a Fast Fourier Transform algorithm is used, see Press *et al.* (1992).

Considering (13) as a function of the height a, the proposal density is then

$$\begin{aligned}&q(a|\mu,x)\\ &\propto \exp\left\{-\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi+\tau^2)}\left(\frac{a^2 d_0}{2\log 2 v_x v_y} - 2a\sum_{i\in V} h(i;\mu,1,d_0,r_0,\theta_0)(\tilde{Y}_i - A_i(x))\right)\right\}, \qquad (15)\end{aligned}$$

which is a Gaussian distribution,

$$a|\mu,x \sim N\left(\frac{\sum_{i\in V} h(i;\mu,1,d_0,r_0,\theta_0)(\tilde{Y}_i - A_i(x))}{\tilde{d}_0/(2\log 2)}, \frac{\sigma^2/\mathrm{ss}_\varphi+\tau^2}{\tilde{d}_0/(2\log 2)}\right),$$

restricted to the compact interval $(0, C_a]$. As for the three remaining parameters (d, r, θ) we will approximate the conditional intensity with a piecewise log-linear intensity, and sample from the corresponding distribution. When proposing d we will select a grid $(\delta_0, \ldots, \delta_m)$ such that $\delta_0 = 0$, $\delta_m = C_d$ and let

$$q(d|\mu, a, x) \propto \exp\left\{p_{i-1} + \frac{p_i - p_{i-1}}{\delta_i - \delta_{i-1}}(d - \delta_{i-1})\right\} \quad \text{for } d \in (\delta_{i-1}, \delta_i],$$

where

$$p_i = -\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi + \tau^2)}\left(\delta_i \frac{a^2}{2\log 2 v_x v_y} - 2\sum_{i \in V} h(i; \mu, a, \delta_i, r_0, \theta_0)(\tilde{Y}_i - A_i(x))\right) - 3\log\delta_i - \beta_d/\delta_i, \quad (16)$$

for $i = 1, \ldots, m-1$, $p_0 = p_1$ and $p_m = p_{m-1}$. Above the last two terms stem from the prior for d.

The expressions for $q(r|\mu, a, d, x)$ and $q(\theta|\mu, a, d, r, x)$ are derived similarly.

3.2 Removal of a point

With probability p_2 we propose to remove a point. If the current configuration x is empty we do nothing, otherwise we select the candidate between the points in x with equal probability $1/n(x)$.

3.3 Moving a point

With probability p_3 we propose to change a parameter of a randomly selected point. We choose one of the parameters μ, a, d, r or θ with equal probability and a new value is proposed by considering the conditional distribution of the parameter given the other parameters.

Suppose for instance that a point $\xi = (\mu, a, d, r, \theta) \in x$ has been selected and we wish to propose a new position μ' for ξ. Corresponding to the insertion of a new point above, we will then propose the position by simulating from a distribution which has voxel-wise constant density

$$q(\mu'|x) \propto \exp\left\{\frac{1}{(\sigma^2/\mathrm{ss}_\varphi + \tau^2)}\sum_{i \in V} h(i; \mu', a, d, r, \theta)(\tilde{Y}_i - A_i(x\backslash\xi))\right\}, \quad \mu' \in V.$$

For the parameters r, d and θ we consider a neighbourhood of the current value, and approximate the conditional density as in (16) above. In our application, we have chosen a neighbourhood of 100 mm^2 for d, 0.3 for r and 0.35 for θ.

Finally, the height a is simulated from a normal distribution as when proposing a new point,

$$a|x \sim N\left(\frac{\sum_{i \in V} h(i; \mu, 1, d, r, \theta)(\tilde{Y}_i - A_i(x\backslash\xi))}{\tilde{d}/(2\log 2)}, \frac{\sigma^2/\mathrm{ss}_\varphi + \tau^2}{\tilde{d}/(2\log 2)}\right).$$

4 Simulation study

Estimation of an activation surface can be carried out by simulating from the posterior distribution of the surface given the data. However, we are left with the problem of determining sensible values for the parameters of the prior for X, sensible in the sense that the estimated activation surface corresponds well with the underlying true surface. To this end we have generated a training data set by simulating from the model (7), with a known underlying activation pattern A. An image of the latter can be seen in Figure 3. The image was generated to mimic a "true" activation image, with coherent regions of activation of both small and moderate sizes. In each region the activation level in individual voxels were simulated from a normal distribution with a common mean. Finally the image was smoothed with a Gaussian kernel to obtain a smooth activation image. Naturally an image obtained in this way cannot be reproduced exactly by a single realization of the activation pattern of model, in this sense the training data is not different from real fMRI data sets. However by using the posterior mean of the activation pattern as an estimate of the latter, we can reproduce more general patterns than those represented by the prior model.

We will fix the parameters of the priors for d, a and r at the values given earlier. Hence we are left with the intensity β, the scaling parameter ρ and the order of the soft-core prior p. Since we need to perform an entire run of the MCMC algorithm for each combination of parameter values, it is only possible to perform a crude estimation where a few different values of each parameter are tried. For each set of parameters we produced 400000 samples from the MCMC algorithm and stored every 100'th sample. After an initial burn-in of 500 subsamples, the chain was judged to be stationary from plots of diagnostics of the simulated point patterns (not shown). We will measure the goodness-of-fit of the model by the posterior mean of the L^2 distance between the activation surface and the true surface. We will estimate this quantity by

$$\widehat{GOF} = \frac{1}{N}\sum_{j=1}^{N}\left\{\sum_{i\in V}(A_i(x^{(j)}) - A_i)^2\right\}^{1/2},$$

where $x^{(1)}, \ldots, x^{(N)}$ is a sequence of simulations from the MCMC algorithm. The variance of this estimate was estimated by the method of batch means with batch sizes of 25. The variance estimate gives an idea of the level of uncertainty, but as it depends on the chosen batch size, it should be interpreted with care.

The true values for the standard deviations were respectively $\sigma = 0.03$ and $\tau = 0.005$. The estimates obtained by (10) and (11) were $\hat{\sigma} = 0.02996$ and $\hat{\tau} = 0.005074$. Table 4 shows the goodness-of-fit of the model with different parameters values. The model which yields the best fit is the one with $\beta = 0.01$, $\rho = 5$ and $p = 10$. Though the largest changes in the GOF measure occur when varying β, the four rows of $\beta = 10^{-4}$ indicates that some degree of regularity ($\rho > 0$) improves the goodness-of-fit. In the last column of the table is an estimate of the mean integrated activation, that is the integral of the activation surface. This can be considered as a summary statistic of the total level of activation.

Evidently the choice of prior parameters affects the final result to some extent. Usually, we would prefer to estimate the parameters in an empirical Bayes fashion, however as the maximum likelihood estimates can in general correspond to a prior that favours meaningless point configurations, cf. the discussion in Section 2.4, this is not an advisable strategy. The fully Bayesian approach, with hyperpriors on the parameters, is an alternative. However, it is not obvious how one should simulate the posterior distribution of the parameters, since the unknown normalization constant of the point process density would enter in the Metropolis-Hastings ratio. Instead we will fix the prior parameters at the values which yield the best fit in the simulation study. Though the result will to some extent depend on this choice, we note from the table that statistics of interst, such as the mean integrated activation, varies only litte, and in no systematic way, with the parameters.

β	ρ	p	$\widehat{GOF}$ (s.e.$\times 10^4$)	Int. act. (s.e.)
10^{-4}	5	2	0.1342 (2.22)	4.89 (0.029)
10^{-4}	20	2	0.1345 (2.59)	4.43 (0.019)
10^{-4}	5	10	0.1357 (2.16)	4.63 (0.022)
10^{-4}	20	10	0.1328 (1.97)	4.57 (0.018)
10^{-4}	0	-	0.1355 (3.01)	4.68 (0.022)
10^{-2}	5	10	0.1305 (2.87)	4.86 (0.026)
10^{-6}	5	10	0.1428 (2.00)	4.83 (0.025)

Table 1: Estimates of the goodness-of-fit of the model with different parameter values. Standard errors due to the simulation are given in parentheses. In the row with $\rho = 0$ no interaction between the points was included in the model. In the last column is an estimate of the mean integrated activation.

In Figure 3 is the estimate of the posterior mean activation image under the best model above. For comparison, the smoothed SPM estimate of the activation is also displayed in the figure. This is obtained by smoothing the regression image $\{\tilde{Y}_i\}$ with a Gaussian kernel of FWHM 3 voxels. The latter denotes the full width of half maximum of the smoothing kernel, this is the typical measure for the width of a smoothing kernel in the medical imaging literature. Both estimates tend to oversmooth the true image, due to the smoothness assumptions underlying them both, but the bias is largest for the SPM. This is more clearly seen in the plot in Figure 4 which shows the number of voxels with an activation level higher than a given threshold for the true image, the posterior mean activation image and the smoothed SPM. One cannot reduce the oversmoothing of the SPM by reducing the width of the smoothing kernel, since the theory of Gaussian random fields, used for making inference in the SPM, requires the discrete image to be a reasonable approximation to a differentiable spatial process.

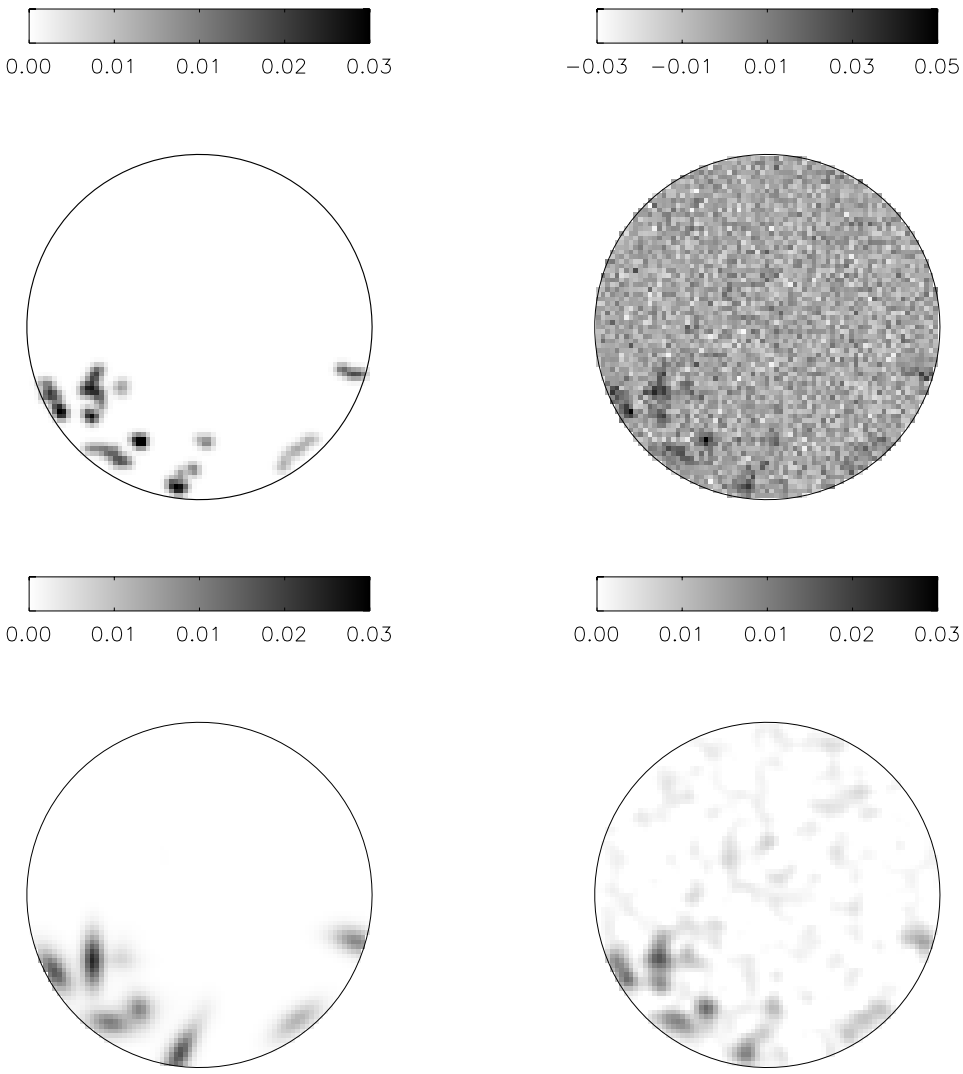

Figure 3: Top Left: The artificial activation image used for generating training data. Intensity values range from 0.0 to 0.04, but the image is clipped at 0.03 for display purposes. Top Right: The regression image $\{\tilde{Y}_i\}$ obtained from the training data. Bottom left: The estimate of the mean posterior activation with $\beta = 0.01$, $\rho = 5$, and $p = 10$. Bottom right: The regression image smoothed by a kernel of FWHM 3 voxels. Note that the color scale differs in the upper right image compared to the three others.

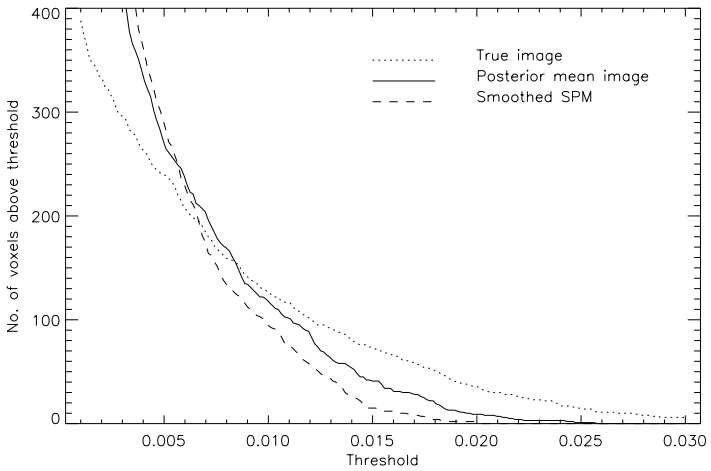

Figure 4: The number of voxels with activation level exceeding a given threshold. Shown is respectively the true image, the posterior mean activation image and the smoothed SPM.

4.1 An analysis of a visual stimulation dataset

We will apply the method to an fMRI dataset consisting of 90 scans, acquired while a light was periodicly flashed in the eye of the subject. The scans, which was obtained by a method denoted Echo-Planar Imaging (EPI), was recorded every 2 seconds during a 3 minute period. The stimulation was arranged in blocks of 20 seconds off, 20 second on, 20 seconds off etc., with 4 complete on-off cycles during the session. Each scans consists of 128 by 128 voxels each covering an area of 1.875×1.875 mm in a slice of thickness 5 mm.

In general the magnetization of the tissue will be highest in the initial scans, causing an increased intensity in the beginning of the time series. After a couple of scans an equilibrium is obtained, and the intensity stabilizes at a steady level, and we will hence discard the first 5 scans and only consider the remaining 85 in the analysis.

The variance estimates were $\hat{\sigma} = 0.0294$ and $\hat{\tau} = 0.00421$. We generated 1 million simulations from the MCMC algorithm and subsampled every 100'th sample. In Figure 5 is a plot of two diagnostics of the simulated point process, namely the number of points and the L^2 norm of the residual image, $\{\tilde{Y}_i - A_i(x)\}$. As can be seen from the plots there is an initial burn-in period of about 2000 subsamples, after which the chain stabilizes to a stationary level. However, as the auto-correlation plot for the number of points shows, the samples are somewhat correlated, and it would be worthwhile to improve the mixing properties of the algorithm to speed up convergence. The acceptance probabilities for the different movetypes are listed in Table 2.

In Figure 6 is a plot of the posterior mean activation image and the posterior standard deviation of the activation, calculated voxelwise. The back of the head is in the top of the images. It is evident from the images, that there are large areas of activation in the back of the brain, which corresponds to the location of the visual cortex, that processes visual

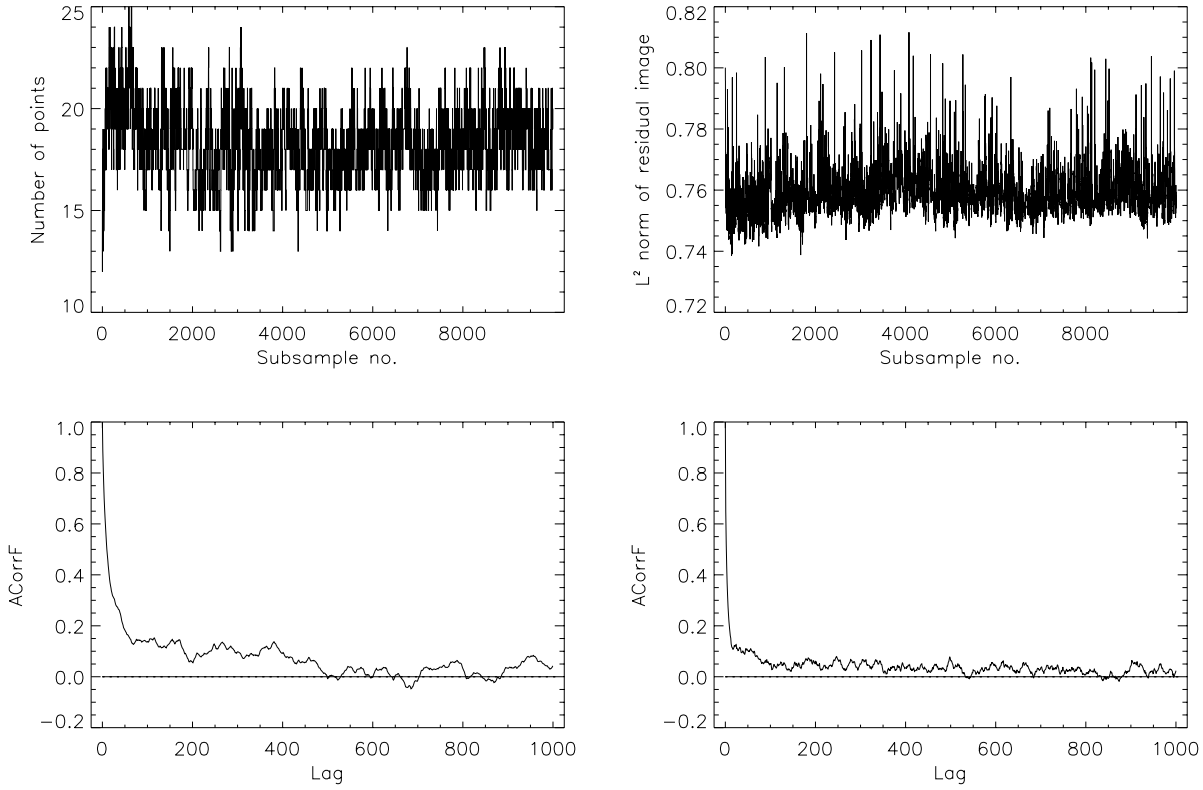

Figure 5: Diagnostics plots of the simulations obtained by subsampling every 100'th iteration of the MCMC algorithm. Shown is to the left the number of points in X and to the right the L^2 norm of the residual image, $\{Y_i - A_i(X)\}$. Below are auto-correlation plots of the two timeseries.

Move type	Acceptance (%) Independent noise	Acceptance (%) Correlated noise
Insert point	5.51	4.72
Delete point	5.53	4.75
Update position	15.78	11.35
Update height	43.92	25.98
Update area	35.63	23.76
Update angle	66.86	52.10
Update ratio	59.84	45.43

Table 2: Acceptance probabilities for the different move types in the MCMC algorithm. The correlated noise model will be described in Section 6.

impressions. For comparison is also the smoothed SPM, which appears to oversmooth the image, also in this example.

The posterior variance image gives some idea of the uncertainty of the activation estimate. The area which has large posterior variance turned out to posses some time series which were more noisy than the remaining ones, this lack of fit of the model is hence reflected in a larger uncertainty of the estimate in this area. Alternatively the uncertainty can be quantified by the posterior probabilities of individual voxels having activation level greater than a certain threshold, 0.009 say. The latter corresponds to the standard error of usual voxel-wise regression estimates for the activation level. The posterior probabilities are displayed as an image in the figure.

Often the interest is on a particular summary statistic, such as the activation area, measured in terms of number of activated voxels. For the current data set the estimate of the mean activation area is 491.0 voxels, and the standard deviation of the area is estimated to 26.0. More specific hypotheses about the activation pattern may be evaluated by estimating posterior probabilities of events of interest.

Finally we will estimate the shape of the response function, given the estimated activation surface. Recall that $Y_{it} = A_i\varphi_t + \varepsilon_{it}$, hence for known A the m.l.e. of φ_t is given by $\hat{\varphi}_t = \sum_i A_i Y_{it} / \sum_i A_i^2$ for $t = 1, \ldots, m$. By inserting the estimate of the posterior mean activation surface displayed in Figure 6, we get the estimate of φ plotted in Figure 7. Overlaid on the plot is the model response function given in (6). As can be seen from this plot, there are substantial differences between the model, and the observered response. The observed response does not appear to be stationary, for instance the last peak is higher than the 3 first, and the dip below baseline is more prominent after the first and third cycle than after the second. In the next section we will describe a method for modelling such non-stationarities in a semi-parametric setting.

5 A semi-parametric model for the haemodynamic response

Though the model for the haemodynamic response (6) is verified empirically to give a reasonable fit to the observed response, its limitations were demonstrated in the previous section. Several hypothesis have been proposed to explain the complex interplay between the local blood flow and oxygenation changes and the BOLD signal, yet there is still not concensus of the quantitative relationship between these. Clearly the choice of a convolution model with a Gaussian impulse response function is somewhat *ad hoc* in this context. Alternative models for the impulse response has been proposed, such as Gamma densities by Lange and Zeger (1997), Friston *et al.* (1998) and FIR filters by Nielsen *et al.* (1997). The question of whether a convolution model is appropriate is however not clear; in some circumstances the response is approximately linear (Dale and Buckner, 1997) while in others it is highly non-linear (Vazquez and Noll, 1998). Also a relevant question is whether the response is stationary over time, or if the response changes with general alertness and learning as suggested

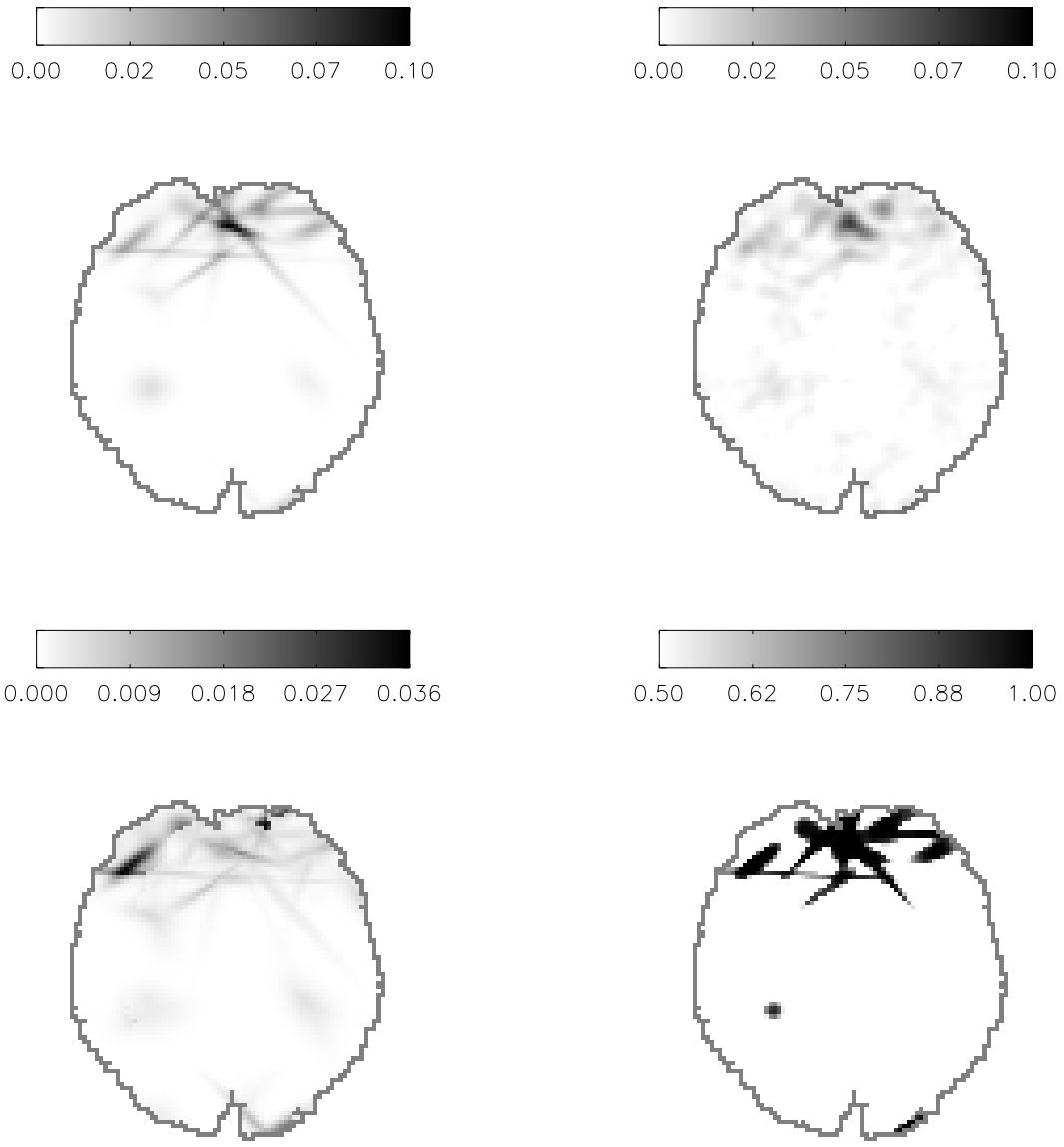

Figure 6: Top left: Monte Carlo estimate of posterior mean activation surface. Top right: Smoothed SPM estimate of activation surface. Bottom left: Estimate of voxelwise posterior standard deviation. Bottom right: Voxelwise posterior probability of activation level greater than 0.009. The images represent a slice of the brain, the upper part of the images correspond to the back. The grey line represents the surface of the cortex.

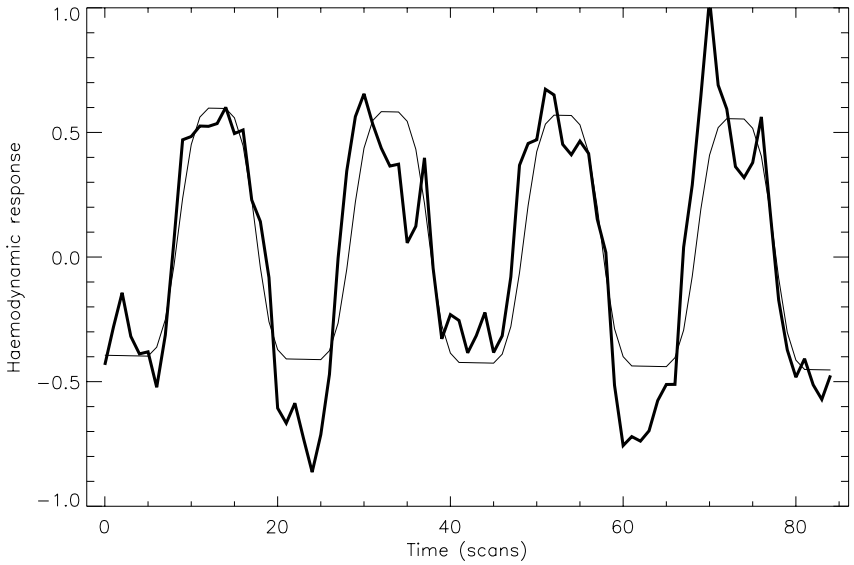

Figure 7: Thick line: Maximum likelihood estimate of the haemodynamic response function under the assumption that the spatial activation pattern is known and given by the estimate in Figure 6. Thin line: The model for the haemodynamic response in (6).

by Gaschler-Markefski *et al.* (1997).

Considering the complexity of temporal response, it seems very appealing to model the latter in a semi-parametric setting. In this framework we do not have to assume stationarity over time or additivity of the response. Instead we will assume a prior of the form

$$\varphi_t = \lambda_t + \nu_t, \quad \nu_t - \nu_{t-1} \sim N(0, \kappa^2), \quad t = 1, 2, \ldots, m. \tag{17}$$

where $\{\nu_t - \nu_{t-1}\}$ are independent and $\nu_0 = 0$. Here the mean λ_t is a simple model, such as that in (6) used in the previous section, reflecting the overall structure of the response. However φ is allowed to deviate a lot from the mean, via the random walk structure of the noise terms ν_t. The variance κ^2 governs the smoothness of $\varphi_t - \lambda_t$.

Combining this prior for φ with the prior for spatial activation pattern (2) we can make inference about (X, φ) through the simultaneous posterior distribution $P(X, \varphi|Y)$. For computational reasons we will in fact consider the posterior distribution of (X, Y, η) where $\eta = \{\eta_i, i \in V\}$ are the random intercepts in the model (7). This posterior is given by

$$p(X, \varphi, \eta|Y) \propto P(Y|X, \varphi, \eta)P(X)P(\varphi)P(\eta),$$

where the likelihood term is obtained by conditioning on η in (7). The variable η can be considered as an auxiliary variable in the simulation algorithm, since it is not of interest in itself, but simulation of the two other variables X and φ becomes much easier, when we condition on η.

We can generate a Markov chain which has the posterior as invariant distribution by a variable-at-a-time Metropolis-Hastings algorithm, where we iteratively update one parameter given the two others. When updating X, the proposals are as described earlier, though with the modification that we replace $A_i(x)$ with $A_i(x) + \eta_i$ and set $\tau^2 = 0$ in the formulas in Section 3, in order to account for the fact that we condition on η. A similar modification applies to the likelihood function in (8), when calculating the acceptance ratio.

When updating η, we will simulate directly from the conditional distribution given (Y, X, φ). This is hence a Gibbs update, which will always yield acceptance rates of 1. It can easily be verified that

$$\eta_i | Y, X, \varphi \sim N\left(\frac{\tau^2}{\tau^2 + \sigma^2/\mathrm{ss}_\varphi}(\tilde{Y}_i - A_i(X)), \tau^2(1 - \frac{\tau^2}{\tau^2 + \sigma^2/\mathrm{ss}_\varphi})\right), \tag{18}$$

with all η_i's conditionally independent.

We will simulate directly from the conditional distribution of φ given (Y, X, η), as well, since this is also normal, where mean and variance can be calculated as follows. Let $Y_{\star t} = (Y_{it})_{i \in V}$ denote the image recorded at time t, regarded as a $|V|$-dimensional vector, and let $\varepsilon_{\star t}$ be defined correspondingly. In the following, all distributions are conditionally on $A = A(X)$ and η. Then the model (7) states that

$$Y_{\star t} = (A + \eta)\varphi_t + \varepsilon_{\star t}, \quad \varepsilon_{\star t} \sim N(0, \sigma^2 I_{|V|}), \quad t = 1, \ldots, m,$$

which combined with the prior (17) is a linear Gaussian state space model. Hence it is not difficult to see that if we condition on Y, φ has a Gaussian distribution. The literature on state space models is extensive, hence we will just give the formulas for the conditional mean and variance of φ as given by the Kalman smoother, and refer to West and Harrison (1989), for instance, for the proofs.

Let $D_t = \sigma\{Y_{\star 1}, \ldots, Y_{\star t}\}$ denote the information up to time t and suppose that $\varphi_{t-1} | D_{t-1} \sim N(\mu_{t-1}, C_{t-1})$. This is true for $t = 1$ when we consider the initial distribution of φ_0 as a degenerate normal distribution concentrated at 0. The Kalman filter then gives that φ_t given D_t is also normal, $\varphi_t | D_t \sim N(\mu_t, C_t)$, where

$$\begin{aligned} C_t^{-1} &= \|A + \eta\|^2/\sigma^2 + (C_{t-1} + \kappa^2)^{-1}, \\ \mu_t &= \mu_{t-1} + \lambda_t - \lambda_{t-1} + \frac{C_t}{\sigma^2}(A + \eta)'(Y_{\star t} - (A + \eta)(\mu_{t-1} + \lambda_t - \lambda_{t-1})). \end{aligned}$$

Here prime denotes the transpose matrix.

In order to simulate from the distribution of φ_t given $Y = D_m$, we will also consider the Kalman smoother. Suppose that $\varphi_{t+1} | \varphi_{t+2}, D_m \sim N(\bar{\mu}_{t+1}, \bar{C}_{t+1})$. By the recursion above, this is true for $t + 1 = m$ with $\bar{\mu}_m = \mu_m$ and $\bar{C}_m = C_m$. Then we have that $\varphi_t | \varphi_{t+1}, D_m \sim N(\bar{\mu}_t, \bar{C}_t)$, where

$$\bar{C}_t = C_t - \frac{C_t^2}{C_t + \kappa^2}, \quad \bar{\mu}_t = \mu_t + \frac{C_t}{C_t + \kappa^2}(\varphi_{t+1} - \mu_t - \lambda_{t+1} + \lambda_t). \tag{19}$$

In fact this is the conditional distribution of φ_t given $\varphi_{t+1}, \ldots, \varphi_m, D_m$, by conditional independence of φ_t and $\varphi_{t+2}, \ldots, \varphi_m$, and hence we can use this recursion to simulate $\varphi|D_m$: We simply simulate the φ_t's one at a time, starting from the back with φ_m.

We note here, that a collection of response functions could be modelled by allowing a multidimensional φ. We could assign different functions to different groups of centres, and in this way account for regional differences in the response. The formulas above would be slightly more complicated, but at least for moderate dimensions of φ the recursive simulation routine would still be very efficient.

In Figure 8 is a plot based 1000000 simulations of this Markov chain. We considered the visual stimulation data of the previous section, though preprocessed in a slightly different way, as we removed some low-frequency trends with very large magnitude from the data, in order to stabilize the algorithm. The plot illustrates the estimated posterior mean $E(\varphi|Y)$ with confidence limits for φ based on the posterior variance. For comparison is an overlay of the initial model (6). The plot shows the same deviations from stationarity as was indicated by Figure 7.

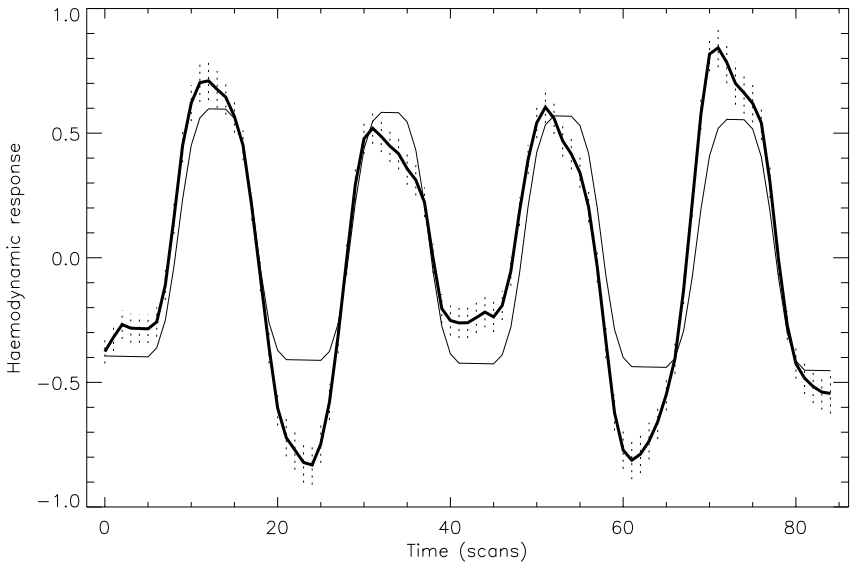

Figure 8: Monte Carlo estimate of the posterior mean of the haemodynamic response function based on 10000 subsamples of 1000000 simulations. Overlaid is pointwise 95%-confidence regions based on the estimated posterior variance. The thin line is the prior mean of the response given by the convolution model in (6).

One consequence of modelling φ in this way, is that the paradigm is only vaguely included in the model, in the sense that the response function is not time-locked to the paradigm, but is allowed to drift by the random walk structure. It might seem unwise to ignore a relevant covariate like this, however, in some experiments the *actual* paradigm is not directly controllable by the experimenter and hence precise information of this is not available. For

instance in memory processing or other mental stimulation experiments, it is not possible to end the stimulation at an exact time point. Furthermore with this formulation, we may detect subtle activation patterns, which depends on the paradigm in more complex ways. An example of the latter is the XOR signal of Lange *et al.* (1999).

6 Accounting for correlated noise

The initial model in (7) assumed that the noise was uncorrelated both temporally and spatially. This is necessarily a somewhat optimistic assumption. The noise sources in fMRI data are both of physiological and physical origin. The pixel values are constructed by inverse Fourier transforms of a sequence of mesurements of currents in a coil over a short time period. Hence there is no physical separation of the pixels, which could justify independence. The temporal correlation is likely to arise from physiological sources, but also intrinsicly in the MR scanner.

In order to investigate the correlation of the noise, we will consider the residuals in the model (7), $r_{it} = Y_{it} - \hat{Y}_i \varphi_t$, where we assume the response function φ is known. The empirical temporal and spatial correlograms are respectively,

$$\hat{\gamma}_i(l) = \frac{\sum_{t=1}^{m-l} r_{i,t} r_{i,t+l}}{\sum_{t=1}^{m} r_{i,t}^2} \quad l = 1, 2, \ldots, m-1,\ i \in V,$$

$$\hat{\lambda}_t(k) = \frac{\sum_{i \in V_k} r_{i,t} r_{i+k,t}}{\sum_{i \in V} r_{i,t}^2}, \quad k \in \mathbb{Z}^2, t = 1, \ldots, m,$$

where $V_k = \{i \in V | i + k \in V\}$. We estimate the correlograms voxel-by-voxel respectively scan-by-scan in order to assess whether the correlation is stationary over voxels and scans. In Figure 9 is a plot of $\hat{\gamma}_i(1)$ as a function of $i \in V$ with estimated 95%-confidence bounds based on a global AR(1) model $\gamma_i(1) = \gamma(1)$ and $\gamma_i(l) = 0$ for $l > 1$ for all i. Displayed is also a plot of $\lambda_t((1,0))$ as a function of t with estimated 95%-confidence bounds based on the spatial model described below. About 10% of the points fall outside the confidence bounds in each plot, which indicates that the temporal (spatial) correlation structure is not the same in all voxels (scans). This is not really surprising, as it merely reflects the inhomogeneity of the underlying tissue. The observation suggests that a non-separable covariance model, which allows for the different temporal structures, should be fitted to the data. One such model is that proposed by Lange and Zeger (1997), where the voxel time series are considered in the frequency domain, and different spatial covariance models are fitted to different frequencies. However, since we need to invert (or Cholesky decompose) the large spatio-temporal covariance matrix in order to calculate the likelihood function, we will have to restrict ourselves to reasonably simple covariance structures. For this reason we will only consider a separable model in the following sense. Let $\varepsilon = \{\varepsilon_{it}, i \in V, t = 1, \ldots, m\}$ be the noise terms in (7) regarded as a $|V| \times m$ matrix, then

$$\varepsilon \sim N_{|V| \times m}(0, \sigma^2\, , \ \otimes \Lambda).$$

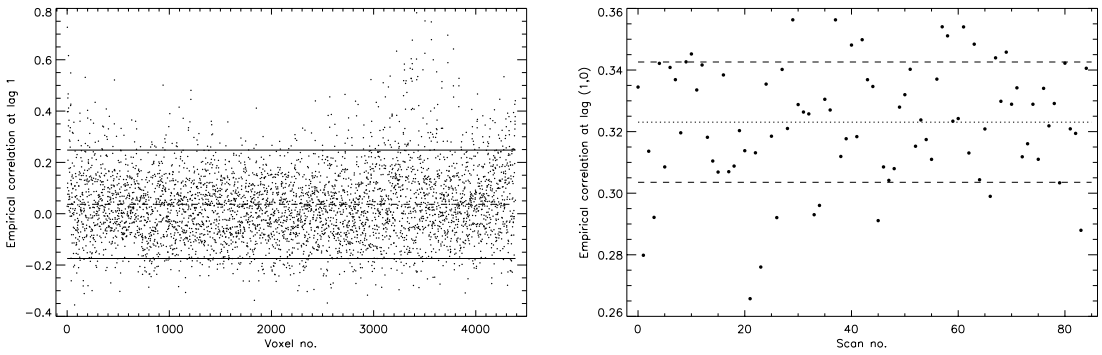

Figure 9: Left: The empirical temporal correlations at lag 1, $\hat{\gamma}_i(1)$, as a function of voxel number. The 95%-confidence bounds are based on a common AR(1) model for all voxels. Right: The empirical spatial correlation at lag (1,0), $\hat{\lambda}_t((1,0))$, as a function of scan number. The 95%-confidence bounds are based on a common spatial model for all scans (see text.)

where $\otimes$ denotes the Kronecker product and where , and Λ are $|V| \times |V|$ and $m \times m$ correlation matrices.

Global correlation estimates are obtained by the averages $\hat{\gamma}(l) = |V|^{-1} \sum_{i \in V} \hat{\gamma}_i(l)$ and $\hat{\lambda}(k) = m^{-1} \sum_{t=1}^{m} \hat{\lambda}_t(k)$. A plot of the empirical temporal correlations can be seen in Figure 10. The fitted AR(1) model with $\hat{\gamma}(1) = 0.0367$ gives a reasonable fit to the observed correlations.

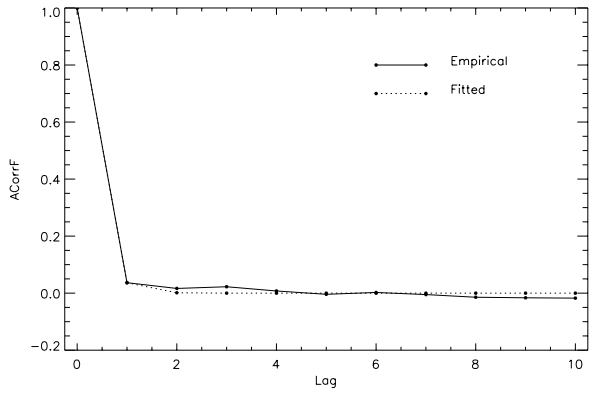

Figure 10: The empirical temporal correlogram for all voxels with the fitted AR(1) correlogram.

A plot of the empirical spatial correlation can be seen in Figure 11. The correlation is clearly non-isotropic and furthermore there is evidence of negative correlation at lag 2 voxels. The plot indicates that observations further than a distance of 2 voxels apart are

almost uncorrelated, which suggests a moving average type model,

$$\varepsilon_i = \sum_{j \in D} g_j U_{i+j}, \quad i \in V,$$

where $\{U_j, j \in \mathbb{Z}^2\}$ is white noise and D is some neighbourhood of the origin. Here and in the following we will only consider the spatial covariance structure of a single scan and hence ignore the temporal index t in the notation. The parameters $g_j, j \in D$ can be estimated by fitting the model to the empirical covariances and an estimate of the spatial correlation matrix , may be calculated. The problem with this approach, however, is that one needs to invert , , or at least compute the Cholesky decomposition , $= LL'$ where L is lower triangular, in order to calculate the likelihood function. In our data set there are more than 4000 voxels constituting V which makes it very demanding to decompose the correlation matrix. As a practical alternative to the above model, we propose to parametrize the Cholesky square root L rather than , itself, and hence consider the model $\varepsilon \sim N_{|V|}(0, \sigma^2 LL')$ where L is parametrized as follows. Let $\tilde{L} = \{\tilde{l}_{ij}\}$ be a lower triangular matrix, such that

$$\tilde{l}_{ij} = \begin{cases} \tilde{l}_{i-j} & \text{if } i > j,\ i - j \in D, \\ 1 & \text{if } i = j, \\ 0 & \text{else,} \end{cases}$$

where D is a neighbourhood of the origin, not necessarily equal to the neighbourhood in the moving average model above. Then we let $L = \{l_{ij}\}$ be given by

$$l_{ij} = \frac{\tilde{l}_{ij}}{\left(\sum_{j \le i} \tilde{l}_{ij}^2\right)^{1/2}} \quad i, j \in V$$

The normalization above ensures that LL' is correlation matrix. Notice that when ignoring edge-effects, the model will be stationary since $L_{ij} = L_{i-j}$.

In this formulation , can be calculated by a matrix product, and expressions such as z', ^{-1}z for $z \in \mathbb{R}^{|V|}$, which enters in the likelihood function, can be calculated by

$$z',\ ^{-1}z = \|L^{-1}z\|^2 = \|v\|^2, \tag{20}$$

where v is the solution to $Lv = z$ which can be obtained easily due to the lower-triangularity of L. Notice that we have to order the indices in $|V|$ when expressing the spatial correlation as the matrix , . The correlation between ε_i and ε_j is given by

$$\operatorname{corr}(\varepsilon_i, \varepsilon_j) = \sum_{k \le \min(i,j)} l_{i-k} l_{j-k}. \tag{21}$$

Hence the model has the peculiar property, that the covariance structure depends on the ordering of the indices. From a theoretical point of view this is difficult to accept, since the

ordering is arbitrarily chosen. From a practical point of view, however, the ordering is chosen in any natural way, and the model is judged by how well it fits data. We will demonstrate in a moment that the model fits data well, and since the computational advantages by working with L rather than , or , $^{-1}$ are considerable, we favour this method.

The parameters may be estimated by fitting the implied correlation (21) to the empirical covariance,

$$\hat{l} = \text{argmin} \sum_{\{j \in \mathbb{Z}^2 : D \cap (D-j) \neq \emptyset\}} \left(\sum_{k \leq \min(0,j)} l_{-k} l_{j-k} - \hat{\gamma}(j) \right)^2 .$$

We have chosen the natural lexicographic ordering of the voxel indices (x, y) and have parametrized the model by letting D include 3. order neighbours, which gives 6 free parameters. In Figure 11 is a plot of the empirical correlation along the 4 equiangular directions $(0, \pi/4, \pi/2, \pi)$, together with the fitted correlation. Clearly the model fits the data quite well.

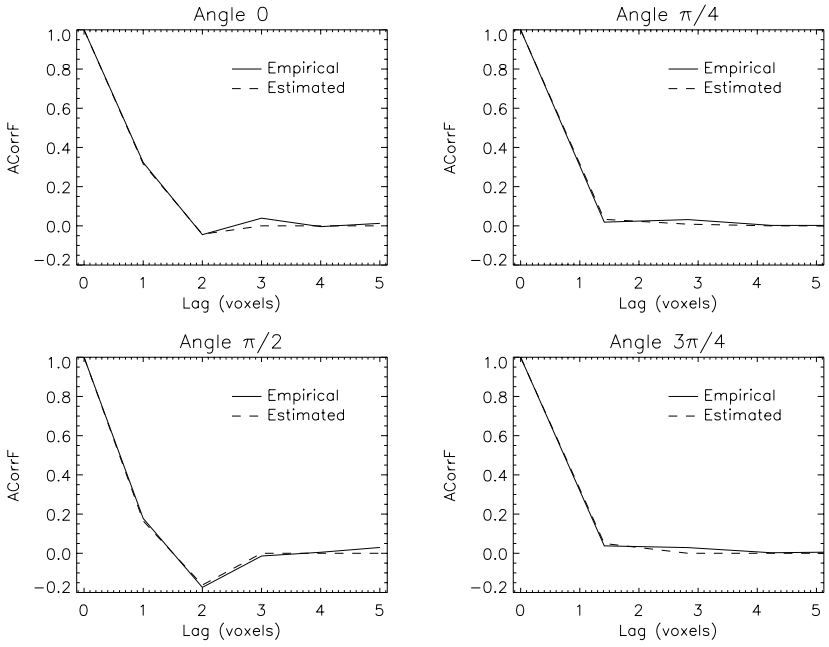

Figure 11: Empirical and fitted spatial correlations along the four equiangular directions $(0, \pi/4, \pi/2, 3\pi/4)$

Incorporating estimates of , and Λ in the model (7) is quite straightforward if we assume that these are reasonable precise estimates, and hence can be regarded as fixed. Consider the data Y as a $|V| \times m$ matrix, and let $Y^\circ = Y(M^{-1})'$, and $\varphi^\circ = M^{-1}\varphi$, where M is the lower triangular Cholesky square root of Λ. Then the conditional likelihood function, where

we condition on the values of η, is given by

$$p(Y|x,\varphi,\eta) = (2\pi\sigma^2)^{-\frac{m|V|}{2}}|,\,|^{-\frac{m}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{m}\|L^{-1}(Y^\circ_{\star t} - \tilde{Y}^\circ\varphi^\circ_t)\|^2\right\}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2/\mathrm{ss}^\circ_\varphi}\|L^{-1}(\tilde{Y}^\circ - A - \eta)\|^2\right\}, \quad (22)$$

where

$$\tilde{Y}^\circ_i = \sum_{t=1}^{m} Y^\circ_{it}\varphi^\circ_t/\mathrm{ss}^\circ_\varphi, \quad \mathrm{ss}^\circ_\varphi = \sum_{t=1}^{m}\varphi^{\circ 2}_t,$$

is defined equivalently to $\tilde{Y}$ in (9). Recall from Section 3 and 5 that the proposal distributions for updating X and φ in the MCMC algorithm were based on the model with independent noise. A big advantage from an application point of view is that these need not be changed when we incorporate correlated noise terms, if we are willing to accept slightly worse mixing properties. By simply substituting the expressions for the likelihood ratio in the Metropolis-Hastings ratio, we ensure that the chain converges to the correct posterior distribution. Inference on X can hence be made by simulating (X,η) iteratively, where the distribution of $\eta|X,Y,\varphi$ is as in (18).

We considered the visual stimulation data again, and simulated 1000000 samples of X using the same MCMC algorithm as described in Section 3, but with the modified likelihood function. As expected, the acceptance rates decreased a bit, see Table 2. The effect of accounting for correlation in the noise shows both in the activity estimate itself and in an increased uncertainty of the latter. The mean activation image is illustrated in Figure 12. The largest difference, compared to the activation image obtained with the uncorrelated noise model in Figure 6, is the circular region in the back of the brain, which is much larger in the current image. As an example of how the variance of the estimate increases, we may consider the number of activated voxels. The mean and standard deviation of this are estimated to 543.1 respectively 31.4, the corresponding figures from the uncorrelated model in Section 4.1 were 491.0 and 26.0.

7 Discussion

We have proposed a spatio-temporal model for fMRI data which explicitly accounts for the fact that the signal changes are locally coherent in both space and time. This assumption is often implicitly included in the analysis of fMRI data, when spatial and temporal filtering are applied prior to the analysis, but rarely included explicitly in a model. The relation (8) shows that in the simplest setting the procedure is effectively fitting ellipsoids of different sizes and orientations to a regression image, and assessing the significance of these. The random field theory has counterparts to this procedure, namely the search for local maxima in both scale and space, Siegmund and Worsley (1995), and in the space of ellipses with different orientation and shape, Shafie *et al.* (1998). The method is, however, fundamentally

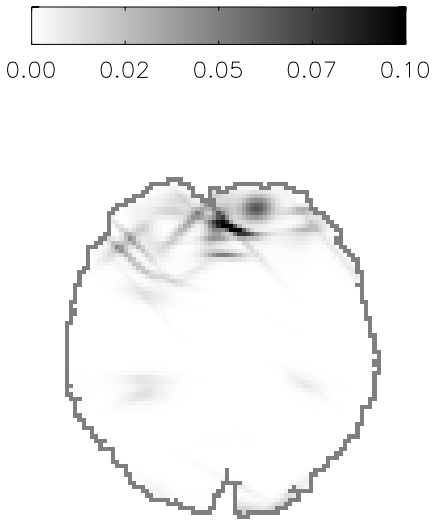

Figure 12: The estimate of the posterior mean activation image in the correlated noise model.

different from the random field approach. The latter provides a framework for testing the null-hypothesis of no activation in each individual voxel with correction for the large number of tests performed. As was pointed out by Keith Worsley in the discussion of Lange and Zeger (1997), what is really an estimation problem is hence answered by a large number of statistical tests, with corresponding conceptual and mathematical problems. With the proposed method the focus is shifted towards estimating the activation pattern by use of standard Bayesian methods, rather than testing for activation in individual voxels.

Assessing the uncertaincy of the estimated activated pattern is theoretically easy by considering posterior variances. This allows us to evaluate the significance of hypothesis of interest within single subjects. Alternatively we may record estimates and standard errors of relevant features of the activation in different experiments, and use this when comparing different groups of subjects. This might for instance be the mean activity level in a certain region of the brain, where the latter could be identified individually in each subject from high-resolution anatomical scans acquired simultaneously with the fMRI scans.

The approach of modelling the temporal response in a non-parametric setting with few assumptions seems appealing to us, given the uncertainty about the nature of the haemodynamic effects in different stimulation types. Also the fact that the modelled response depends only vaquely on the specified paradigm is an advantage when analysing data where the actual paradigm is difficult to determine. Naturally the method has it's limitations. Firstly the prior we have formulated, restricts the response to be sufficiently smooth, and one could imagine that this is not the case in the recent event-related pardigms Buckner (1998), where several stimulation types are rapidly interchanged. To analyse these data

with our method, an alternative prior should be formulated, possibly by incorporating the paradigm and assuming some sort of stationarity.

Another limitation is the assumption, that the response is the same everywhere in the brain. Authors such as Lee *et al.* (1995) and Kornak *et al.* (1999) have found, by fitting simple parametric response functions to fMRI time series, that the delay can vary with a few seconds over the activated regions. Though the semi-parametric model is limited by the assumption of constant delays, it is advantageous in the sense that it can capture more general response patterns than those proposed by these authors. An obvious way of relaxing the assumption of constant delay and shape is by working with a collection of response functions, and assigning different functions to different centres, hence we would only assume that the response is locally similar. In this formulation we would in fact search for any spatial regions of similar temporal pattern, and not only paradigm related patterns.

Acknowledgements

The author is grateful for inspiring and constructive discussions with Jens Ledet Jensen and Rasmus Waagepetersen. Thanks also to Hans Stødkilde-Jørgensen from the MR-ResearchCentre at Skejby Sygehus for kindly providing the data.

A Ergodicity properties of the Markov chain

Theoretical results for the algorithm described in section 3 are studied in Geyer and Møller (1994), Møller (1999) and Geyer (1999). Geyer has a stability condition, namely that the point process has bounded conditional intensities, which implies Harris recurrence and geometric ergodicity of the algorithm, that he studies. This situation is, however, slightly different, since our proposal density for inserting a new point $q_2(\xi \cup x|x)$ is not constant, as in Geyer's algorithm, but has a rather complicated structure. However, when restricting the support of the prior in a natural way $q_2(\xi \cup x|x)$ is bounded below, which turns out to be sufficient to apply Geyer's method.

PROPOSITION 1 *There exists a constant M such that*

$$p(x \cup \xi|Y) \leq Mp(x|Y) \quad \forall x \in \Omega, \xi \in \mathcal{X}.$$

PROOF Recall that $\phi(x_i, x_j)$ given by (3) is less than 1, hence we have,

$$\begin{aligned}
\frac{p(x \cup \xi|Y)}{p(x|Y)} = \frac{p(x \cup \xi)}{p(x)} \frac{p(Y|x \cup \xi)}{p(Y|x)} &= \beta \prod_{\eta \in x} \phi(\xi, \eta) p(a, d, r) \\
&\times \exp\left\{ -\frac{1}{2(\sigma^2/\mathrm{ss}_\varphi + \tau^2)} \left(\sum_{i \in V} h(i;\xi)^2 - 2 \sum_{i \in V} h(i;\xi)(\tilde{Y}_i - A_i(x)) \right) \right\} \\
&\leq c \exp\left\{ \frac{1}{(\sigma^2/\mathrm{ss}_\varphi + \tau^2)} \sum_{i \in V} h(i;\xi) \tilde{Y}_i \right\} \\
&\leq c \exp\left\{ \frac{1}{(\sigma^2/\mathrm{ss}_\varphi + \tau^2)} \left(C_d C_a^2 / (2 \log 2 \, v_x v_y) \sum_{i \in V} \tilde{Y}_i^2 \right)^{1/2} \right\}.
\end{aligned}$$

Here the last inequality follows from (14). Let M denote the expression in the last line, this does not depend on x og ξ and the proof is complete. □

In the rest of this section we will restrict the support of the prior to the region D given by

$$D = \{x \in \Omega | \sum_{j=1}^{n(x)} h(i; x_j) < C_a \ \forall i \in V\}.$$

This assumption says that not only is C_a a natural upper bound for the height of individual activation bells, but also for the image obtained by combining all bells. As C_a was chosen arbitrarily large, this is not a restriction in practice.

PROPOSITION 2 *There exists a $\delta > 0$ such that*

$$q_1(\xi \cup x|x) \geq \delta \quad \forall \, \xi \in \mathcal{X}, x \in \Omega \text{ such that } x \cup \xi \in D.$$

PROOF We will show that each of the factors in (12) is bounded below. Considering $q(\mu|x)$ we have

$$q(\mu|x) = Z^{-1} \exp\left\{ \frac{1}{\sigma^2/\mathrm{ss}_\varphi + \tau^2} \sum_{i \in V} h(i; \mu, a_0, d_0, r_0, \theta_0)(\tilde{Y}_i - A_i(x)) \right\},$$

where Z is the normalizing constant, that is the sum over $\mu \in V$ of the last term. By an evaluation such as that in the proof of Proposition 1 we have that this term is bounded above by a finite constant c_1, hence $Z^{-1} \geq (|V|c_1)^{-1}$. By the assumption that $x \cup \xi \in D$ we have that $A_i(x) < C_a$ for all $i \in V$ such that the last term is bounded below by a positive constant c_2. Hence we have $q(\mu|x) \geq c_2/(|V|c_1)$. A similar evaluation of $q(a|\mu, x)$, combined with the fact that the proposal is restricted to a bounded interval, shows that $q(a|\mu, x) \geq c_3$ for a positive constant c_3 and all x, μ and a. Likewise p_i defined in (16) is bounded below and above for all i, and hence so is $q(d|a, \mu, x)$. The proposals for r and θ are equivalent to the one for d. □

PROPOSITION 3 *The algorithm simulates a Markov chain that is Harris recurrent and geometrically ergodic.*

These properties are desirable, since they ensure that the chain will converge to the correct stationary distribution geometrically fast, such that a central limit theorem holds.

PROOF By Propositions 1 and 2 the probability of accepting an upstep can be dominated as follows,

$$\min\left\{1, \frac{p(x \cup \xi|Y)}{p(x|Y)} \frac{1}{(n+1)q_1(x \cup \xi|x)}\right\} \leq \frac{M}{\delta(n+1)}.$$

As the number of points n tends to infinity, the expression on the right hand side tends to zero. The probability of accepting a downstep is

$$\min\left\{1, \frac{p(x|Y)}{p(x \cup \xi|Y)}(n+1)q_1(x \cup \xi|x)\right\} \geq 1,$$

for n large enough. Hence if the number of points gets very large, the propability of accepting a further upsted is almost zero while we will allways accept a downstep. This guarantees a drift towards a smaller number of points which again implies geometrically ergodicity. We refer to the arguments in the proofs of Propositions 2 and 3 in Geyer (1999) for details. □

References

Baddeley, A.J. and van Lieshout, M.N.M. (1993) Stochastic geometry models in high-level vision. In K.V. Mardia and G.K. Kanji (eds.), *Statistics and Images*, vol. 1, chap. 11, pp. 231–256, Appl. Statist.

Barndorff-Nielsen, O., Kendall, W. and Lieshout, M. (eds.) (1999) *Stochastic Geomety. Likelihood and Computation.* Chapman & Hall/CRC.

Buckner, R.L. (1998) Event-related fMRI and the hemodynamic response. *Human Brain Mapping*, **6**, 373–377.

Bullmore, E., Brammer, M., Williams, S.C.R., Rabe-Hesketh, S. *et al.* (1996) Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.*, **35**, 261–277.

Dale, A.M. and Buckner, R.L. (1997) Selective averaging of individual trials using fMRI. *Neuroimage*, **5**, S47.

Friston, K., Josephs, O., Rees, G. and Turner, R. (1998) Nonlinear event-related responses in fMRI. *Magn. Reson. Med.*, **39**, 41–52.

Friston, K.J., Jezzard, P. and Turner, R. (1994) The analysis of functional MRI time-series. *Human Brain Mapping*, **1**, 153–171.

Friston, K.J., Holmes, A.P., Poline, J.B. *et al.* (1995) Analysis of fMRI time-series revisited. *Neuroimage*, **2**, 45–53.

Gaschler-Markefski, B., Baumgart, F., Tempelmann, C. *et al.* (1997) Statistical methods in functional magnetic resonance imaging with respect to nonstationary time-series auditory cortex activity. *Magn. Reson. Med.*, **38**, 811–820.

Geyer, C. (1999) Likelihood inference for spatial point processes. In Barndorff-Nielsen *et al.* (1999), chap. 3.

Geyer, C.J. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.*, **21**, 359–373.

Holmes, A.P., Josephs, O., Büchel, C. *et al.* (1997) Statistical modelling of low-frequency confounds in fMRI. *Neuroimage*, **5**, S480.

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 for efficient spatial inference. In K.V. Mardia, R.G. Aykroyd and I.L. Dryden (eds.), *Spatial Temporal Modelling and its Applications*. Leeds University Press.

Kullback, S. (1959) *Information Theory and Statistics*. John Wiley & Sons, Inc.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E. *et al.* (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA*, **89**, 5675–5679.

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

Lange, N. and Zeger, S.L. (1997) Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion). *Appl. Statist.*, **46**, 1–29.

Lange, N., Strother, S., Anderson, J. *et al.* (1999) Plurality and resemblance in fMRI data analysis. *NeuroImage*, **10**, 282–303.

Le, T.H. and Hu, X. (1996) Retrospective estimation and correction of physiological artifacts in fMRI by direct extraction of physiological activity from MR data. *Magn. Reson. Med.*, **35**, 290–298.

Lee, A.T., Glover, G.H. and Meyer, C.H. (1995) Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. *Magn. Reson. Med.*, **33**, 745–754.

Møller, J. (1999) Markov chain Monte Carlo and spatial point processes. In Barndorff-Nielsen *et al.* (1999), chap. 4.

Nielsen, F.A., Hansen, L.K., Toft, P. *et al.* (1997) Comparison of two convolution models for fMRI time series. *Neuroimage*, **5**, S473.

Ogata, Y. and Tanemura, M. (1984) Likelihood analysis of spatial point patterns. *J. R. Statist. Soc. B*, **46**, 496–518.

Petersen, N.V., Jensen, J.L., Burchhardt, J. and Stødkilde-Jørgensen, H. (1998) State space models for physiological noise in fMRI time series. *Neuroimage*, **7**, S592.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) *Numerical Recipes in C*. Cambridge University Press, second edn.

Shafie, K., Worsley, K.J., Wolforth, M. and Evans, A.C. (1998) Rotation space: Detecting functional activation by searching over rotated and scaled filters. *Neuroimage*, **7**, S755.

Siegmund, D.O. and Worsley, K.J. (1995) Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Stat.*, **23**, 608–639.

Vazquez, A.L. and Noll, D.C. (1998) Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage*, **7**, 108–118.

West, M. and Harrison, J. (1989) *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.

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